

## ARTICLE OPEN



# The Sherpa hypothesis: Phenotype-Preserving Disordered Proteins stabilize the phenotypes of neurons and oligodendrocytes

Vic Norris <sup>1</sup>✉, Judit Oláh <sup>2</sup>, Sergey N. Krylov<sup>3</sup>, Vladimir N. Uversky <sup>4</sup> and Judit Ovádi<sup>2</sup>

Intrinsically disordered proteins (IDPs), which can interact with many partner proteins, are central to many physiological functions and to various pathologies that include neurodegeneration. Here, we introduce the *Sherpa* hypothesis, according to which a subset of stable IDPs that we term Phenotype-Preserving Disordered Proteins (PPDP) play a central role in protecting cell phenotypes from perturbations. To illustrate and test this hypothesis, we computer-simulate some salient features of how cells evolve and differentiate in the presence of either a single PPDP or two incompatible PPDPs. We relate this virtual experiment to the pathological interactions between two PPDPs,  $\alpha$ -synuclein and Tubulin Polymerization Promoting Protein/p25, in neurodegenerative disorders. Finally, we discuss the implications of the *Sherpa* hypothesis for aptamer-based therapies of such disorders.

npj Systems Biology and Applications (2023)9:31 ; <https://doi.org/10.1038/s41540-023-00291-8>

## INTRODUCTION

Intrinsically disordered proteins (IDPs) transiently adopt many different conformations that range from the compact to the extended forms unhindered by energetic constraints<sup>1</sup>. IDPs are common, with the eukaryotic proteome estimated to contain around ~20% disordered proteins<sup>2,3</sup>. Disordered regulatory domains can constitute switches as their structures can be changed significantly by post-translational modifications<sup>4</sup>. IDPs and proteins that contain intrinsically disordered regions, which for convenience we class here as IDPs, are implicated in liquid-liquid phase separation and the formation of intracellular biomolecular condensates<sup>5,6</sup>. IDPs are frequently involved in cell signalling networks because they can bind to many partner proteins with both high specificity and low affinity (hence transiently)<sup>7,8</sup>; these disorder-based interactions can take several forms<sup>1,9,10</sup>. Some IDPs display a tendency to self-associate and to form stable aggregates<sup>11,12</sup>.

IDPs are frequently implicated in various maladies, such as cancer and neurodegenerative diseases<sup>13</sup>. For example, mutations in the disordered protein p53 have been identified in colon, lung, breast, and brain cancers<sup>14</sup>. Protein aggregation in the brain is central to neurodegenerative diseases, with  $\alpha$ -synuclein (SYN) and Tubulin Polymerization Promoting Protein/p25 (TPPP) aggregation being characteristic of Parkinson's disease and multiple system atrophy<sup>15–17</sup>, and  $\beta$ -amyloid and tau aggregation being characteristic of Alzheimer's disease<sup>18,19</sup>. Aggregation of these unfolded/misfolded proteins generally leads to the formation of characteristic proteinaceous deposits in neurodegenerative diseases<sup>20</sup>.

Despite the well-established importance of many IDPs for pathogenesis, the normal function of these proteins is not entirely clear. Indeed, understanding their physiological functions might allow the development of therapies without side-effects. It has been argued that one advantage of disorder may be that it makes an economical use of resources because the interface between an IDP

and a partner protein requires fewer residues from the IDP than a similar interface between ordered proteins<sup>21</sup>. It is also believed that the flexibility of IDPs enables them to perform functions that are complementary to those of ordered proteins<sup>22</sup>. The structural heterogeneity and conformational plasticity of IDPs constitute the foundation of the structure-function continuum model, in which a protein exists as a dynamic conformational ensemble containing multiple proteoforms characterized by a broad spectrum of structural features and possessing various functional potentials<sup>23–25</sup>, with SYN being considered as one of the illustrative examples of this protein structure-function continuum model<sup>26</sup>.

Fully understanding the function of a protein may well require understanding the living system within which this protein operates. One approach to understanding living systems is to model them as networks of connections between elements or nodes. In a pioneering investigation of such networks, Kauffman highlighted a fundamental problem that confronts cells containing many interacting constituents, namely, how such cells manage to obtain the reproducible phenotypes needed for natural selection to be effective when a 'hyper-astronomical' number of combinations of these constituents is apparently available<sup>27</sup>. The solution we favour is to consider the cell as a set of elements or macromolecules, out of which only a few – rather than a number limited only by the total number of constituents – are selected to determine the cell's behaviour or phenotype at any given time<sup>28,29</sup>. This subset of macromolecules, which determines the phenotype, is selected on the basis that: i) the subset must be coherent (i.e., the macromolecules in this subset must work together and not against one another); ii) the behaviour at one time must be coherent with the behaviour at the previous time (i.e., any behaviour-determining subset of macromolecules must continue the work of the previous subset).

Here, we propose a hypothesis for the function of stable IDPs inspired by Himalayan mountain guides or Sherpas. A Sherpa does

<sup>1</sup>Laboratory of Microbiology Signals and Microenvironment, University of Rouen, 76821 Mont Saint Aignan, France. <sup>2</sup>Institute of Enzymology, Research Centre for Natural Sciences, Budapest H-1117, Hungary. <sup>3</sup>Centre for Research on Biomolecular Interactions, York University, Toronto ON M3J1P3, Canada. <sup>4</sup>Department of Molecular Medicine, Morsani College of Medicine, University of South Florida, Tampa, FL 33612, USA. ✉email: victor.norris@univ-rouen.fr

more than helping to carry the material needed for an expedition; he can act as an indispensable guide for those climbing a particular mountain, choosing the route and keeping the climbers on it despite perturbations. Climbing different mountains requires different Sherpas. In our hypothesis, one class of the IDPs, Phenotype-Preserving Disordered Proteins (PPDPs), limits the phenotypes available to cells so they can differentiate. The PPDP achieves this restriction of phenotype space by its interactions with its partner proteins and by its own stability, which allows it to act as a kind of memory. To illustrate and validate the hypothesis, we introduce the equivalent of an PPDP into a version of a learning program, *Coco*, which simulates how a cell evolves into a differentiated state via a competition between selection for a state that is internally coherent (i.e., that the constituents work together) and selection for a historical coherence (i.e., that the relationship between successive states is meaningful). We find that even in this very simple, easy-to-analyse, simulation of cell evolution, a PPDP-like element can have a selectable function that protects the phenotype. We also find biological evidence in the neurodegenerative pathways of SYN and TPPP consistent with the Sherpa hypothesis and, finally, we mention the implications for therapy.

## RESULTS

### The Sherpa hypothesis

We propose that a PPDP can play a central role in the determination of phenotypes by guiding the cell through successive patterns or trajectories of gene and protein activity and by helping maintain these patterns despite perturbations (Fig. 1). The phenotype-stabilising actions of the PPDPs are achieved because these 'chameleon' proteins undergo a succession of conformational changes via their binding to different partners (not necessarily in a particular order) as determined by the intracellular conditions. In this scenario, the first partner to bind the PPDP reduces the conformational space available to the PPDP so that only a subset of the original conformations become available (and perhaps some new conformations that were not in the original space), then the second partner binds and reduces that space still further. This sequence of conformational changes may sometimes be accompanied by a sequence of post-translational modifications to the PPDP and its partners that determines the functioning of the ensemble of these proteins. The result is a phenotype-determining, multi-partner assembly. The binding of a different first partner would send the PPDP off into a different region of conformational space where the PPDP's function would be altered. If the cell were subjected to a major perturbation, the PPDP would help maintain the cell in its differentiated state or trajectory because it binds, brings together, and can protect (or inactivate) its partners, which are necessary for the differentiation. In the Sherpa hypothesis, different PPDPs can be associated with different, incompatible, differentiated states; the downside of this is that the introduction of a second PPDP into a cell could result in a serious perturbation of its phenotype; if this second PPDP were then removed, the first PPDP would then help restore the phenotype.

### The *Coco* program: Learning vs. phenotype

To investigate the possible roles of proteins that are stable and that can interact with many partners in preserving the phenotype, we model the role of the interactions of a subset of macromolecules in determining the phenotype of the cell at any given time. In this modelling, we use a previously published, machine-learning program, *Coco*<sup>28,29</sup>, that simulates, at a very simple level, how a cell evolves into a differentiated state (Fig. 2). The idea here is to show that, during a learning task mimicking natural selection, conditions can exist that select a PPDP-like element. This program

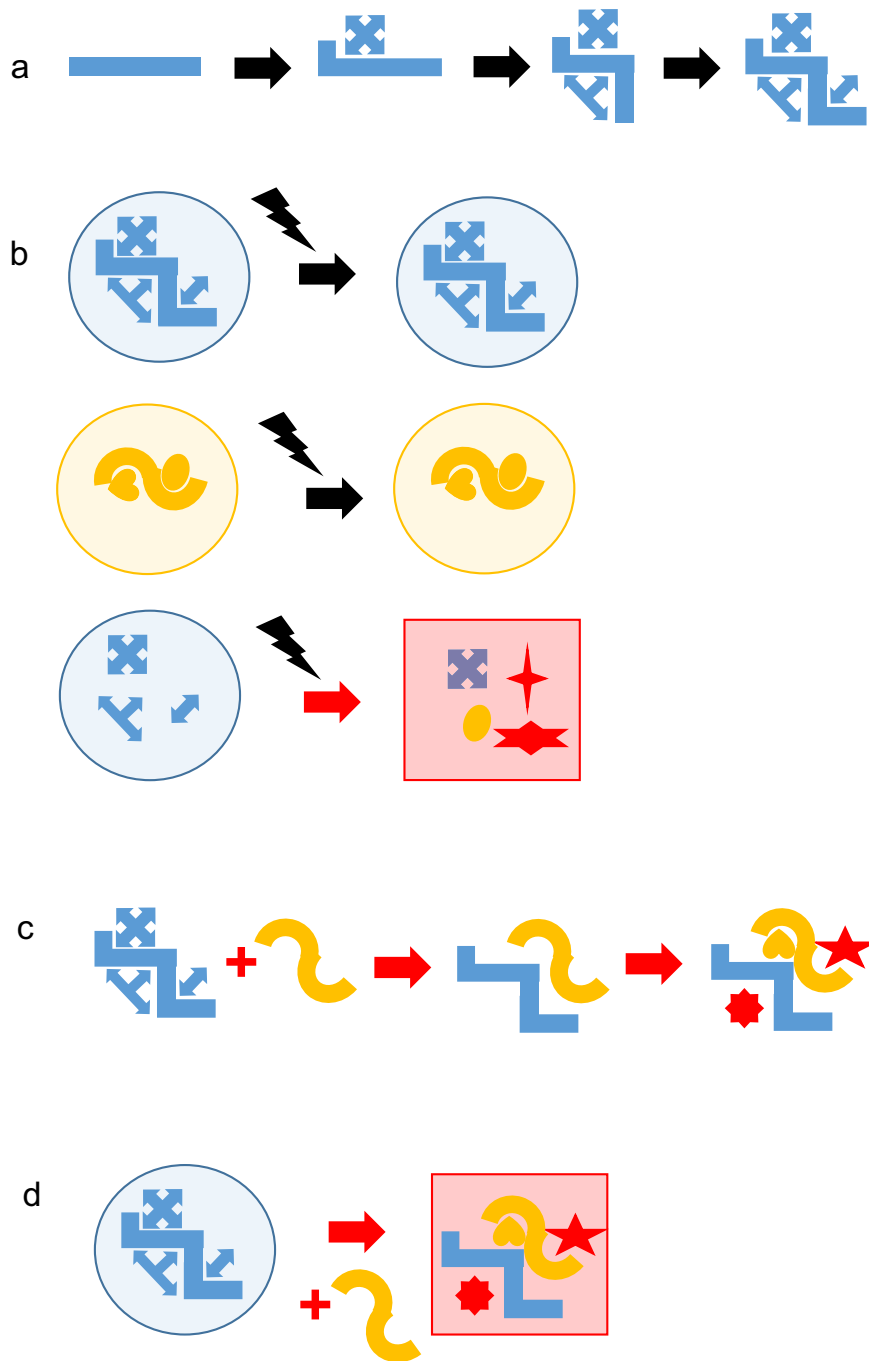
can have thousands of elements that can mimic macromolecules such as proteins; each element can be either *active* or *inactive* in determining the behaviour of the system in a time-step; an element is *active* if it is a member of the current state of the *Active* subset of elements (and inactive if it is not). Although *Coco* can learn with a state of the *Active* subset containing from three to over a hundred elements (not shown), for convenience, in this version, the state of the *Active* subset contains only six elements selected from a total set of a thousand possible elements.

The composition of a state of the *Active* subset in terms of elements changes at each time step. Membership of a state of the *Active* subset is made based on connectivity; each element has two fields, a *Now* field and a *Next* field, that contain the addresses of other elements; the presence of these addresses in an element's fields correspond to the connections that this element makes with other elements. The strength of the *Next* connections that Element-X makes to Element-Y is calculated by counting the number of times the address of Element-Y occurs in the *Next* field of Element-X (and similarly for its *Now* field). The membership of the *Active* subset is calculated for each time step by ranking the strengths of the connections of the *Next* fields of the elements in the present *Active* subset. This entails summing the number of times the address of a particular element occurs in the *Next* fields of all the elements in the present *Active* subset, then ranking these sums, then selecting the element with the highest *Next* rank and comparing it with the element with the highest *Now* rank. This latter element is chosen via the connections in the *Now* fields of those elements already selected as members of the new developing *Active* subset. Note that the same address can occur in both the *Next* and *Now* fields, but this is kept separate in the ranking. To put it differently, the score of the element at the top of the *Next* ranking is compared with the score of the element at the top of the *Now* ranking and the element with the higher score is then elected for membership of the *Active* subset. This coherence competition between the *Next* and *Now* rankings results in the system going through a sequence of states of the *Active* subset. By appropriately rewarding and punishing the contents of the states of the *Active* subsets (i.e., by changing the addresses in the fields of the *Active* elements), the system can learn to respond to its environment in an evolutionary process that leads to a sequence of states of *Active* subsets equivalent to differentiation.

### Simulation: A PPDP constrains phenotype space

The acquisition of a subset of phenotypes or phenotype trajectory by a cell via natural selection during its evolution was simulated by *Coco*. This program learns to select its constituent elements (alias 'macromolecules') to be in the subset of elements that determines the system's behaviour in successive time-steps; this selection is based on the connections between elements; these connections result from reward and punishment strategies that strengthen or weaken these connections (cf. Methods). *Coco* was given the task of learning to respond correctly to the input sequence of (1, 2, 3, 2, 3) from the environment with an output sequence of (1000, 999, 999, 1000, 998) to that environment; in these input-output sequences, *Coco* has to learn to respond to the input of Element-1 with an output of Element-1000, this is followed by an input of Element-2 and a learnt output of Element-999 etc. In the absence of a PPDP-like element, the *active* elements (i.e., those with their addresses in the *Active* subset) were drawn from the entire set of 1000 elements (Fig. 3).

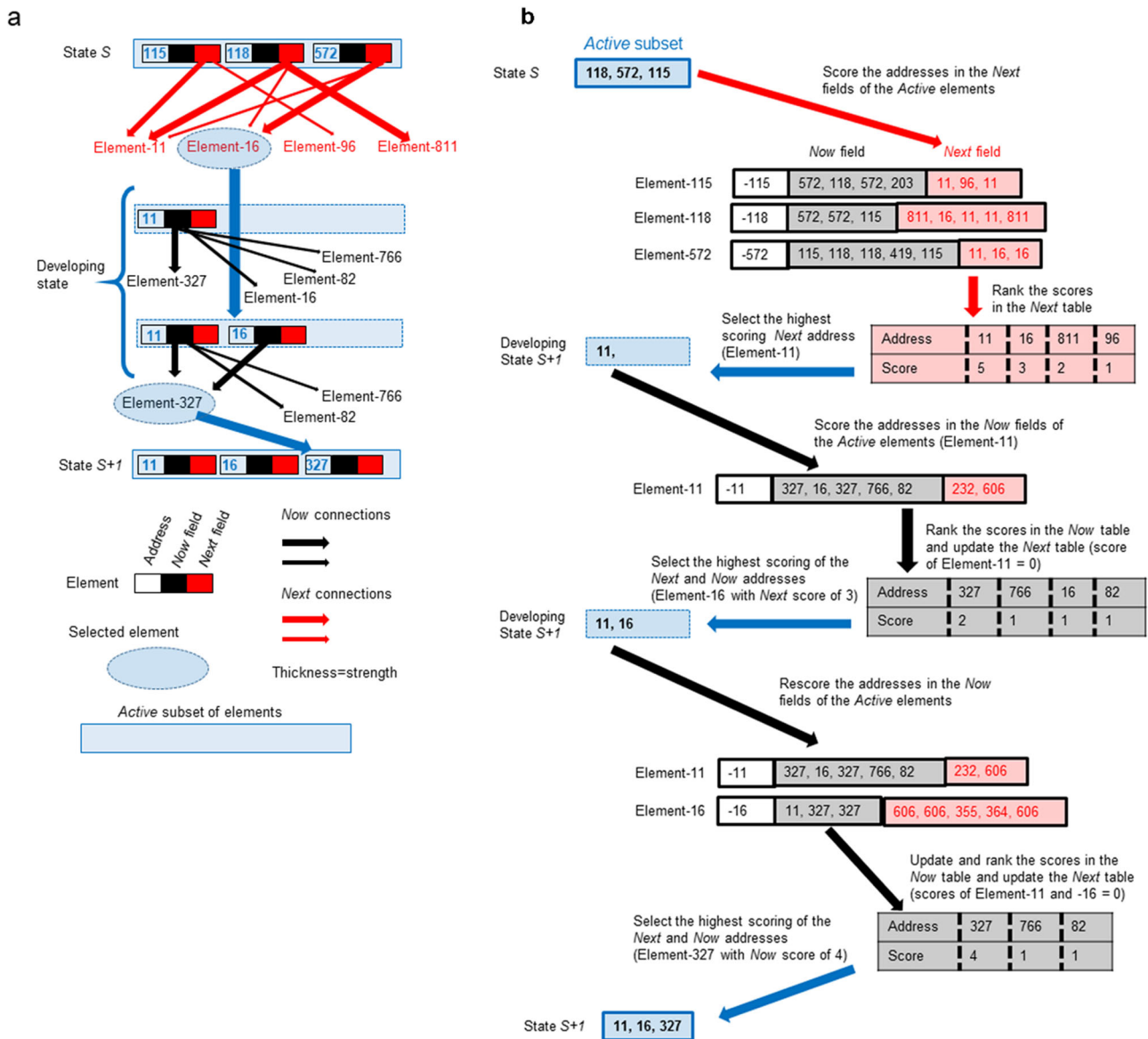
However, in the presence of a PPDP-like element (Element-10), all the *active* elements were drawn from the elements connected to the PPDP-like element, that is, the addresses of all these elements were in the *Now* and *Next* fields of Element-10 (fields that only contain addresses 111 to 170) and no other elements were *active* (Fig. 3). This shows how the presence of a stable, highly connected element in *Coco* can steer its evolution so its



**Fig. 1 The Sherpa hypothesis.** A PPDP plays its physiological role when it interacts with its physiological partners to activate or inactivate them and to co-locate them via phenotype-determining interactions. **a** In physiological conditions, the structure of a PPDP changes progressively (initially a blue rectangle and finally a blue zigzag shape) as its partner proteins (blue rectilinear shapes) are associated with it. **b** Top panel, the PPDP maintains the differentiated state of the cell despite a perturbation (black lightning); middle panel, a different PPDP (thick, dark yellow, curved line) with different partner proteins (dark yellow curved shapes) maintains a different differentiated state despite a perturbation; bottom panel, in the absence of a PPDP, the differentiated state (circular cell) is lost (square red cell corresponding to the absence of blue partner proteins and presence of red stars and of an inappropriate, dark yellow, partner). **c** If two different PPDP bind to one another, the appropriate partner proteins (blue and dark yellow) no longer bind whilst other, inappropriate, proteins (red) may bind. **d** In pathological conditions, the introduction of a second PPDP (dark yellow) causes a loss of the phenotype (red square). Black arrow, physiological path; red arrow, pathological path; red stars, proteins that should not be present in either of the physiological conditions.

states involve only a very limited subset of the elements available. One implication of this result is that the presence of a stable, potentially highly connected, PPDP in an evolving cell or protocell limits the range of regulatory molecules responsible for its phenotype. A related implication is that a set of different, incompatible PPDPs with different connections could favour the

selection of different phenotypes (or more exactly phenotype trajectories) during evolution; this is because this situation would help avoid confusion due to the same molecules being involved in incompatible phenotypes. In other words, there appears to be a good reason why PPDPs should have been selected during evolution.



**Fig. 2 The principle of the program.** **a** The phenotypic state of the cell is represented by the contents of the *Active* subset, which here contains only three elements (for simplicity). The composition of the *Active* subset is determined by a competition between the *Now* and *Next* connectivities. Initially, the *Next* connectivity (red arrows), which represents the coherence of phenotypic states over time, is used to select the first element (e.g., macromolecule) for membership of the developing state of the *Active* subset. Then, the *Now* connectivity (black arrows), which represents the coherence of the phenotypic state at a particular time, is also used to select the elements. The composition of the *Active* subset after learning (alias evolution by natural selection) is a solution to the problem of satisfying the two types of coherence. **b** Details of how consecutive states of the *Active* subset are calculated. Note that the sizes of the *Now* and *Next* fields varies. The *Now* table of element addresses and their scores is grey and those of the *Next* table are pink.

### A PPDP helps preserve a phenotype

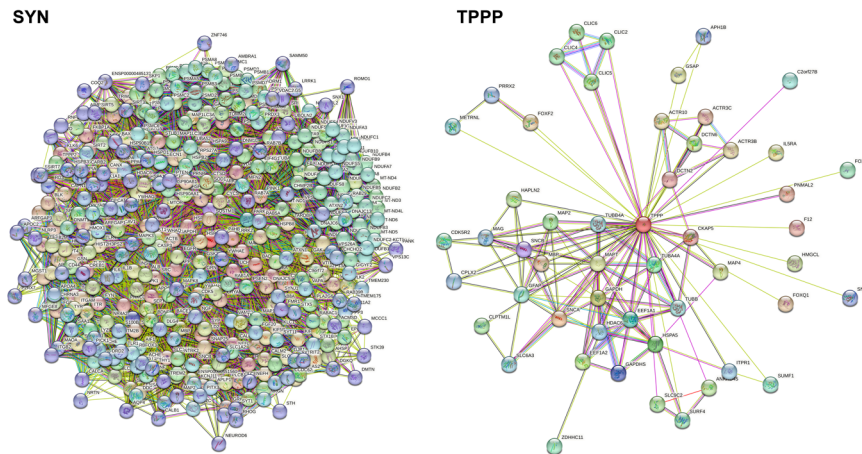
Once a PPDP has been selected during the evolution, we propose that it is maintained because it helps preserve the phenotype (or trajectory of phenotypes). Such preservation would entail resisting perturbations that could otherwise result in the loss of the phenotype. To illustrate this, we first explored the effects of a major perturbation on *Coco* once it had learnt the standard sequence (corresponding to a trajectory of phenotypes). This perturbation took two forms. Firstly, we reasoned that the number of factors determining the phenotype is important and that it varies with the environment; indeed, in the case of an exponentially growing bacterial population, a change in the environment results in major changes in the number of factors<sup>30</sup>. We therefore simulated this number as the size of the state of the

*Active* subset, which we doubled by going abruptly from six elements to twelve. Secondly, in the next time step where there would otherwise be an input of 1 in the *Active* subset, we overwrote this state by creating an entirely new state in which we force into it the addresses of inactive elements. These elements could be considered as corresponding to the regulatory factors appropriate for a different phenotype.

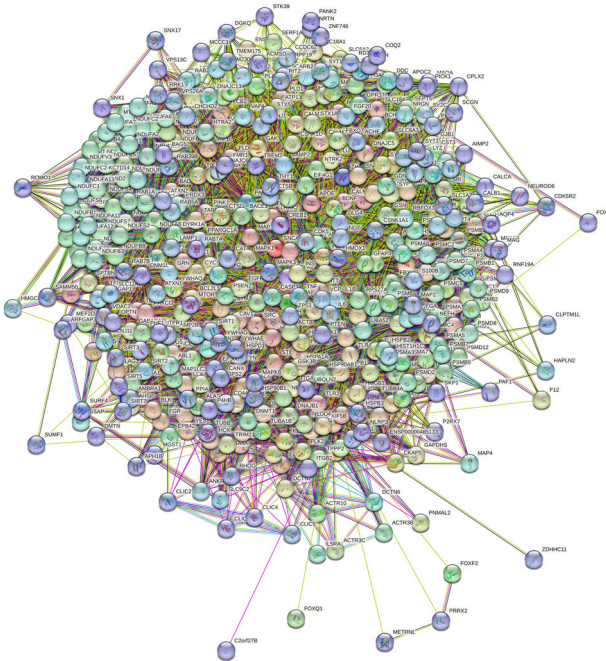
In the absence of a PPDP-like element, the perturbation resulted in a complete loss of the learnt states (Supplementary Fig. 1a). Hardly any of the elements that were present in these states were present after the perturbation; Elements-22 and -122 are exceptions in occurring both before and after the perturbation. In contrast, in the presence of the PPDP-like Element-10, the learnt states are preserved (Supplementary Fig. 1b); for example, all the elements in one of the *Active* states shown (153, 114, 1000, 145, 10







**Fig. 4 Interaction networks of SYN and TPPP.** Left panel: SYN-centred protein-protein interaction network generated by the online database resource Search Tool for the Retrieval of Interacting Genes (STRING)<sup>73</sup> using confidence of 0.45 for the minimum required interaction score (Table S1). This network contains 401 nodes connected by 8,400 edges. The average node degree of the network is 41.9, and its average clustering coefficient (see Methods) is 0.605. Right panel: TPPP-centred protein-protein interaction network generated by STRING<sup>73</sup> using confidence of 0.45 for the minimum required interaction score (Table S2). This network contains 52 nodes connected by 147 edges. The average node degree of the network is 5.65, and its average clustering coefficient is 0.826. SYN is denoted by SNCA in the STRING.



**Fig. 5 Joint SYN-TPPP protein-protein interaction network.** Joint SYN-TPPP-centred protein-protein interaction network generated by STRING<sup>73</sup> using confidence of 0.45 for the minimum required interaction score (Table S3). This network contains 442 nodes connected by 8866 edges. The average node degree of the network is 40.1, and its average clustering coefficient is 0.5796.

SYN and TPPP proteins display multifunctional characteristics due to their multiple interactions (Fig. 4 and Supplementary Fig. 4).

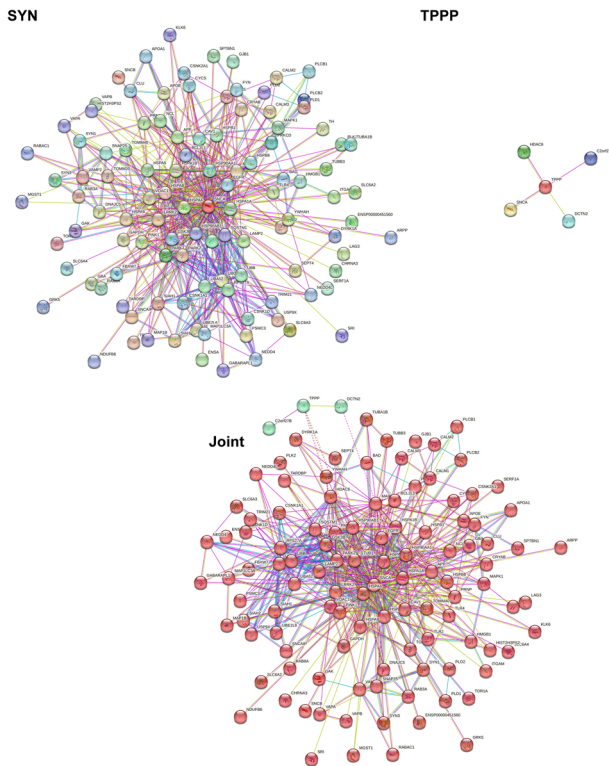
The SYN- and TPPP-centred STRING networks were generated using protein search by name option of STRING using the corresponding UniProt IDs (P37840 and O94811 for SYN and TPPP, respectively) (Fig. 4), whereas the SYN-TPPP network was generated using the Multiple Proteins search mode (Fig. 5). In all cases, the confidence of 0.45 was used for the minimum required interaction score (Excel files with the detailed information are presented in the Supplementary materials as Supplementary Tables 1–3). Figure 4 illustrates the interactions of SYN as well as

those of TPPP in both physiological and pathological conditions; however, most of the data refer to pathological circumstances such as the association between SYN and TPPP. The TPPP-derived aggregation of SYN can disrupt the phenotype of the SYN and TPPP networks in both neurons and OLGs. Although the TPPP-centred, protein-protein interaction (PPI) network includes SYN, and although the SYN-centred, PPI network includes TPPP, there are just 9 common proteins in these two networks: glyceraldehyde-3-phosphate dehydrogenase (GAPDH), glial fibrillary acidic protein (GFAP), 78 kDa glucose-regulated protein (HSPA5), HDAC6, microtubule-associated protein tau (MAPT), microtubule-associated protein 2 (MAP2), myelin basic protein (MBP), beta-synuclein (SNCB), and solute carrier family 6 (neurotransmitter transporter, dopamine) member 3 (SLC6A3).

It might be argued that when the PPI networks of two PPDPs are identical (which is not the case for the SYN and TPPP networks) the simultaneous presence of the two PPDPs does not perturb the phenotype whilst when the PPIs are very different such simultaneous presence *does* perturb the phenotype. Simulation of the simultaneous presence of two PPDP-like elements with different networks, such as SYN and TPPP, does indeed result in a loss of the learnt state (Supplementary Fig. 2). The implication of this is that since the PPI networks generated for SYN and TPPP have little overlap, the probability that the introduction of TPPP and the TPPP network into a neuron will disrupt the phenotype of the SYN network (and vice versa for an oligodendrocyte) is greater than the probability that the two networks will reinforce one another's phenotypes.

As mentioned above, partner proteins specific to SYN include lipoprotein-associated, phospholipase A2, vesicle-associated membrane protein 2 (VAMP2) and synaptosomal-associated protein 25 (SNAP25) (in the SNARE complex) and the dopamine transporter<sup>37–39</sup>, whilst partner proteins specific to TPPP include LIM kinase 1<sup>36,40</sup>. In the Sherpa hypothesis, these proteins could be considered as candidates represented by Elements-111 to –170, the addresses of which are in the fields of the PPDP-element. Partners common to SYN and TPPP include tubulin/microtubule, HDAC6/SIRT2, GAPDH and DJ-1<sup>36,41–44</sup>. Since these proteins may undergo post-translational modifications, it is conceivable that differences in modification would determine differences in their roles in the SYN and TPPP networks.

In order to obtain additional information pertaining to the actual physical interactions of SYN, TPPP, and SYN-TPPP with their



**Fig. 6** STRING-generated physical subnetworks (the edges indicate that the proteins are part of a physical complex) for SYN, TPPP and SYN-TPPP. Statistical parameters for these three physical subnetworks: SYN: 110 nodes, 522 edges, average node degree 9.49, average local clustering coefficient 0.707; TPPP: 5 nodes, 5 edges, average node degree 1.6, average local clustering coefficient 0.8; SYN-TPPP: 113 nodes, 548 edges, average node degree 9.7, average local clustering coefficient 0.689. In the physical subnetwork for SYN-TPPP, nodes are coloured based on the clustering using the Markov Cluster Algorithm (MCL) option of STRING<sup>73</sup>. The network was clustered to the MCL inflation parameter of 2. “Red” cluster includes 110 nodes that correspond to the proteins physically interacting with SYN, whereas three members of the “green” cluster corresponds to the members of the TPPP physical complex.

partners (Fig. 6), a network type “physical subnetwork” was selected from the STRING basic settings in all three networks (SYN-, TPPP-, and SYN-TPPP-centred). The joint network of these two PPDPs (SYN and TPPP) show two clusters with clear separation of a very large SYN-centred cluster from the small TPPP-based cluster containing just three proteins. It is important to remember that this analysis does not permit differentiation between the “date” (where interactions between different partners take place at different locations and times) and “party” (where most of the interactions occur simultaneously) type of protein-protein interactions. Analysis of the PPI network shown in Fig. 6 indicated that in addition to binding to SYN, TPPP also interacts with another member of the SYN-based physical subnetwork, HDAC6. Furthermore, dynactin subunit 2 (DCTN2) from the TPPP physical subnetwork interacts with MAPT that also belongs to the SYN-based physical subnetwork. Note that both HDAC6 and MAPT are highly connected hubs, interacting with 23 and 31 proteins, respectively. Finally, according to the Human Protein Atlas<sup>45</sup>, the physical subnetwork of SYN in neurons alone contains 78 nodes (as against 110 in all cells); these nodes include both proteins expressed in all cells (such as Heat shock protein family A member 1 A (HSPA1A) and GAPDH) and proteins enriched in neurons (such as SNAP25, Synapsin-1 (SYN1) and VAMP2). Similarly, the physical network of TPPP in OLGs alone contains 3 nodes (as against 5 nodes in all cells), namely, TPPP, HDAC6 and DCTN2.

## DISCUSSION

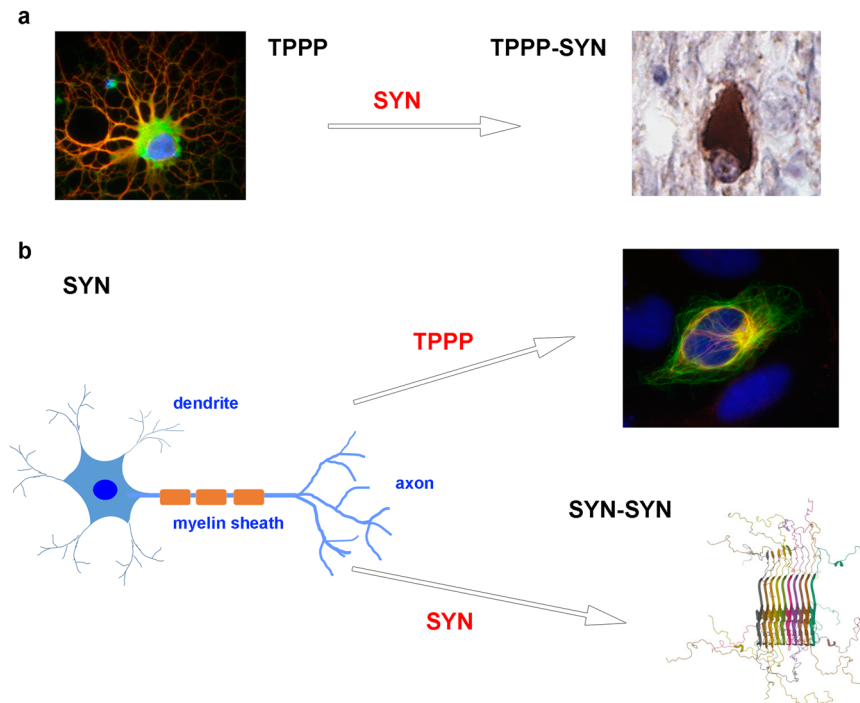
The basic idea of the Sherpa hypothesis is that the IDPs corresponding to PPDPs can contribute to the preservation of the cellular phenotype by counteracting perturbations and influencing evolutionary processes. Potential perturbations include the presence or absence of proteins and ligands that interact with the PPDP, along with post-translational modifications and mutations; these perturbations can affect the properties of the PPDP, such as its association with itself and its partners, and hence the preservation of the phenotype. For example, the presence of SYN perturbs the physiological association of TPPP with the microtubule network as does N- and C-terminal truncation of TPPP;<sup>46</sup> in the case of SYN, post-translational modifications and mutations of SYN can affect its aggregation<sup>47,48</sup> whilst certain mutations (A53T, A53E, and G51D) can result in an aggressive pathological phenotype<sup>49</sup>.

A significant fraction of proteins do not have a well-defined 3D structure, and these IDPs with their extensive interaction potency display multiple physiological and pathological functions; they often serve as hub proteins<sup>50</sup>. IDPs such as SYN and TPPP characteristically display a remarkable conformational flexibility and structural plasticity. These chameleon proteins are considered to be unstable and are targets of proteolytic degradation, at least, in their nascent forms. However, they readily develop multiple interactions with ligands and macromolecules that can prevent their proteolytic degradation<sup>51</sup>. Their associations with their partner proteins, which may have special functions, are predominantly determined by the intracellular milieu (which itself depends on the cell type).

Molecular chaperones can buffer the destabilizing mutations in individual proteins by providing robustness during proteome evolution. According to Pechmann and Frydman<sup>52</sup>, the IDPs, including the hub proteins (which interact with multiple partners in networks), exhibit more non-conservative substitutions at the expense of enhanced chaperone assistance, thereby highlighting an intricate interplay of molecular chaperones and protein disorder in the course of the network evolution. Mutations were ranked according to their predicted effect on protein stability, and those that were highly destabilizing in more than 20% of their occurrences were considered non-conservative. This interplay balances the possible cost of mutations on protein folding/stability and the benefit of new interactions/functions<sup>52</sup>. In fact, there are molecular chaperones that specialize in folding unfolded/misfolded proteins in order to counteract destabilization and so prevent the disruption of multiple cellular processes. While unique biological functions of proteins often require unique 3D-structures, the IDPs and proteins that contain both IDRs and ordered regions are also functional, being able to engage in biological activities and perform seemingly impossible tasks such as the phenotype stabilization proposed in the Sherpa hypothesis.

The idea that one function of the IDPs, or at least of the PPDPs, is to preserve the phenotype via their stabilizing properties is the key element of the Sherpa hypothesis as illustrated in Fig. 1. The importance of the unfolded/misfolded proteins with non-conservative mutations in the evolution of protein networks has been demonstrated at the proteome level<sup>52</sup>. Therefore, the opportunity for the preservation of the phenotype of the cells by PPDPs is reasonable, especially if they are associated with partners at the very moment of their expression. The ongoing complexation of these proteins by the PPDPs is modelled in the *Coco* program by the many interactions possible for the PPDP elements and by the continued presence of the PPDPs (via the assumption of a zero Downtime i.e., complete stability) for the PPDP elements themselves. The results show that, in this sort of learning program, the associations of PPDPs with their specific partners could constrain the space of new phenotypes explored during evolution. It should be noted that the Sherpa hypothesis





**Fig. 7 A phenotypic alteration caused by IDPs leads to Parkinsonism.** **a** TPPP-derived differentiated OLG proceeds to aggregation by SYN leading to the formation of intracellular inclusion found in multiple system atrophy<sup>33,74</sup>. **b** TPPP and/or SYN-induced aggregation of SYN in neuron; SYN-TPPP assembly-promoted Lewy body characteristic of Parkinson's disease<sup>75</sup>. Atomic-resolution structure of alpha-synuclein fibrils (PDB ID: 2N0A)<sup>76</sup>.

and its simulation by *Coco* are not at the level of molecular mechanisms where compartmentalisation, phase separation, mutations, specific concentrations, post-translational modifications etc. should be taken into account individually.

Different basins of the energy landscapes correspond to different phenotypes such as neurons and OLGs or their progenitor and differentiated forms, which depend on the presence of TPPP and/or SYN<sup>33,53–55</sup>. The “depth” and number of the basins determine the outcome of perturbations. If the local minimum is too deep, the phenotype is stable and no transformation occurs into other phenotypes in the course of the evolution; however, at a modest local minimum, the stability can be reduced. The local minima characteristic of neurons and differentiated OLGs may, of course, differ. One can argue that the depth of basins for these phenotypes is strongly determined by the expression of – or the transformation by – the two PPDP proteins, that is, by their intracellular concentrations. In particular, it is reasonable to suppose that the initial concentration of a PPDP is an important factor in these basins (i.e., in the differentiation of a cell and in the maintenance of its differentiated state). In physiological conditions, where the total concentrations of SYN and TPPP in the brain and cerebrospinal fluid are similar<sup>56,57</sup>, the TPPP in neurons and the SYN in OLGs are below the detection level whilst, in pathological conditions, SYN and TPPP are co-enriched and co-localized in both neurons and OLGs. In other words, the differentiated state is stable if the second PPDP is undetectable.

A model system has been established with SYN and TPPP, which are expressed endogenously in neurons and OLGs, respectively, in normal human brain (Fig. 7)<sup>33,53–55</sup>. The *Coco* program simulates certain related experimental results, namely, the promotion of phenotype alterations due to the hetero-association of the disordered TPPP with another disordered protein, SYN, which occur because of the transmission of SYN into OLGs or of TPPP into neuronal cells<sup>17,58–62</sup>. In addition, the pathological overexpression of SYN could be considered as acting as a second PPDP resulting in

SYN assembly with phenotypic alteration as illustrated by the *Coco* program. Within the aggregates that ultimately result from SYN/TPPP interactions, the partner proteins lose their functions<sup>63</sup>, a situation that corresponds to the perturbations predicted by the Sherpa hypothesis and to the results of our simulation. These perturbations lead to neurodegenerative disorders or other diseases<sup>17,47,64–67</sup>.

In the case of these PPDPs, the therapeutic approaches opened up by the Sherpa hypothesis include those that limit perturbations, which would help maintain the functioning of the PPDP and its partners. Another approach is the construction/selection of novel, highly specific, synthetic or designed PPDP helpers (that could be activators or inhibitors) out of a set of oligonucleotide or peptide aptamers; each individual aptamer would target the PPDP or, possibly, the partners of the PPDP. Such aptamers could be linked to one another to form a single construct for the modulation of both the homologous and the heterologous associations of PPDPs within the cells. These constructs would activate or inactivate the physiological and pathological proteins as appropriate, which might well suffice to avoid the pathological consequences of the association of SYN-SYN and SYN-TPPP; consequently, such synthetic PPDP helpers could act potentially as therapeutic agents. Encouragingly, recent studies have revealed that both oligonucleotide- and peptide-based aptamers, modulators of protein-protein interactions, can effectively counteract the pathological aggregations of these PPDPs<sup>46,68–72</sup>.

Finally, the associations of a PPDP with its partner proteins may help other types of differentiated cells resist perturbations. In the case of TPPP, its stabilization of the microtubule network is central to differentiation; this stabilization is disrupted by the pathological presence of SYN. The hypothesis may therefore even prove helpful in understanding and treating diseases like Parkinson's disease and multiple system atrophy in which the interactions between SYN and TPPP are implicated<sup>15–17,36,62</sup>.



## METHODS

### Coco program

In this version of *Coco*, a thousand elements are connected initially with one another in a pattern that has some structure. In essence, connections were made at random between the 36 elements in a group of elements (e.g., between elements with addresses 50 to 86) for each of nine groups (with addresses 150 to 186, 250 to 286, 350 to 386 etc.). Three of the thousand elements are chosen to be input elements (with addresses 1, 2 and 3), and three are chosen to be output elements (998, 999 and 1000). *Coco* learns to respond to repeated cycles of an input sequence of (1, 2, 3, 2, 3) with an output sequence of (1000, 999, 999, 1000, 998). Most elements have a *Now* and a *Next* field that can vary in size from containing the address of just one element to containing the addresses of up to 8 different elements per field (N.B., the address of a single element can occur many times in a field). The size of an element's fields increases if this element is part of a successful response (i.e., is part of the state of an *Active* subset that is rewarded) and decreases if it is part of an unsuccessful one (i.e., is part of a state that is punished). The contents of these fields also change with learning, such that the number of times an element's address is present in the fields of another element increases if the elements are both part of a rewarded state or sequence of states of the *Active* subset and, reciprocally, decreases in the case of punishment. In the rewarding and punishing process, an input (which corresponds to an address of elements-1 or -2 or -3) being inserted into a state of the *Active* subset must be followed by an output (corresponding to 998, 999 or 1000) appearing in the same of in one of the following two states of the *Active* subset. If an output does not appear, *Coco* forces a randomly chosen address of an output into the last of the three states of the *Active* subset, thereby allowing an appropriate rewarding or punishing. Downtime is an important characteristic of the elements in *Coco*. It corresponds to the time needed for the resynthesis of a regulatory molecule following the degradation of this molecule when its function is no longer needed. Once an element has been part of a state of the *Active* subset, it cannot be part of a new state until there have been 12 more states of the *Active* subset.

To model a stable, highly connected PPDP like SYN and TPPP in *Coco*, we choose the element with address 10 to represent the PPDP; in other words, 10 identifies the element. We give Element-10 a Downtime of zero and a high strength via its *Next* connections so that it can be present in consecutive states of the *Active* subset (corresponding to a stable protein). We also give Element-10 large *Now* and *Next* fields of fixed sizes that can each contain 60 addresses (corresponding to the many proteins with which an IDP might interact). These addresses run from 111 to 140 in Element-10's *Now* field and from 141 to 170 in its *Next* field; each of these addresses occurs twice. It is important to note that the addresses in Element-10's fields are fixed at the start of the program and cannot be changed – in other words, they are exempt from the rewarding and punishing routines.

To model the effects of a perturbation, we first let *Coco* learn and, when this learning is stable, disable the rewarding and punishing routines along with those responsible for forced inputs from the environment (1, 2 and 3) and for forced outputs. *Coco* then continues with an unchanging cycle of states of the *Active* subset, which we take to represent a cell that has undergone natural selection then remaining in a differentiated state or phenotype (more accurately, a cycle of phenotypes) despite the absence of selective pressure. At this stage, we subject *Coco* to a major perturbation. Firstly, in the time step 999, we double the size of the *Active* subset so that from then on a state of the *Active* subset contains 12 addresses. Secondly, in the next time step where there would otherwise be an input of 1, we create a new state of the *Active* subset; the addresses in this new state are those

of elements that do not belong to those that are in the learnt sequence (i.e., they are inactive elements).

To model the effects of the introduction of a second IDP, we use a slightly different version of *Coco*. After the program has learnt, we create a second IDP using the inactive elements. The pattern of connections between the elements that are *active* in one of the last twenty-four states of the *Active* subset is recorded (which more than covers an entire learnt sequence of states). This pattern is then mimicked using only those elements that never have addresses in the *Active* subset (they therefore have Downtimes of zero); one of these elements has a similar pattern of connections to the first IDP (Element-10) and is designated as the second IDP. This IDP is given the same properties as the first IDP with a Downtime of zero and a *Next* connectivity score that ensures it is always in a state of the *Active* subset. The second IDP is introduced into a state of the *Active* subset along with some of the elements to which this second IDP is connected. *Coco* then continues as before.

Finally, we use the modified version of *Coco* to investigate whether the removal of the second IDP by putting its *Next* connectivity score to zero can lead to the restoration of the original, learnt sequence of the states of the *Active* subset.

Availability of computer code and algorithm: The code used to generate some of the results and an overview of the programs are given in Supplementary Information.

### Analysis of interactability of SYN and TPPP

Interactability of human SYN (UniProt ID: P37840) and TPPP (UniProt ID: O94811) was analysed by the online database resource Search Tool for the Retrieval of Interacting Genes (STRING)<sup>73</sup>. STRING generates a network of predicted associations based on predicted and experimentally-validated information on the interaction partners of a protein of interest<sup>73</sup>. In the corresponding network, the nodes correspond to proteins, whereas the edges show predicted or known functional associations. Seven types of evidence are used to build the corresponding network, and are indicated by the differently coloured lines: a green line represents neighbourhood evidence; a red line – the presence of fusion evidence; a purple line – experimental evidence; a blue line – co-occurrence evidence; a light blue line – database evidence; a yellow line – text mining evidence; and a black line – co-expression evidence<sup>73</sup>.

In this study, STRING was used in two different modes: to create PPI networks centred on human SYN or TPPP and to produce a joint SYN-TPPP PPI network. Individual protein-centred PPI networks and a joint SYN-TPPP-centred network were generated by STRING (<https://string-db.org/>) using a custom value of 500 first-shell interactions and custom confidence level of 0.45. The resulting PPI networks were further analysed using STRING-embedded routines in order to retrieve the network-related statistics, such as: the number of nodes (proteins); the number of edges (interactions); average node degree (average number of interactions per protein); average local clustering coefficient (which defines how close the neighbours are to being a complete clique – if a local clustering coefficient is equal to 1, then every neighbour connected to a given node  $N_i$  is also connected to every other node within the neighbourhood, and if it is equal to 0, then no node that is connected to a given node  $N_i$  connects to any other node that is connected to  $N_i$ ); expected number of edges (which is a number of interactions among the proteins in a random set of proteins of similar size); and a PPI enrichment p-value (which is a reflection of the fact that query proteins in the analysed PPI network have more interactions among themselves than what would be expected for a random set of proteins of similar size, drawn from the genome. It was pointed out that such an enrichment indicates that the proteins are at least partially

biologically connected, as a group). We also looked for the actual physical interactions of SYN, TPPP, and SYN-TPPP with their partners. To this end, a network type “physical subnetwork” was selected from the STRING basic settings in all three networks (SYN-, TPPP-, and SYN-TPPP-centred). Furthermore, we clustered the physical SYN-TPPP subnetwork using the Markov Cluster Algorithm (MCL) option of STRING, with the network being clustered to the MCL inflation parameter of 2.

### Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

### DATA AVAILABILITY

All data generated or analysed during this study are included in this published article (and its Supplementary Information files).

Received: 11 March 2023; Accepted: 19 June 2023;

Published online: 11 July 2023

### REFERENCES

- Wright, P. E. & Dyson, H. J. Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev.* **16**, 18–29 (2015).
- Oldfield, C. J. et al. Comparing and combining predictors of mostly disordered proteins. *Biochemistry* **44**, 1989–2000 (2005).
- Peng, Y. et al. A Metastable Contact and Structural Disorder in the Estrogen Receptor Transactivation Domain. *Structure* **27**, 229–240.e224 (2019).
- Suryadinata, R., Sadowski, M. & Sarcevic, B. Control of cell cycle progression by phosphorylation of cyclin-dependent kinase (CDK) substrates. *Biosci. Rep.* **30**, 243–255 (2010).
- Uversky, V. N., Kuznetsova, I. M., Turoverov, K. K. & Zaslavsky, B. Intrinsically disordered proteins as crucial constituents of cellular aqueous two phase systems and coacervates. *FEBS Lett.* **589**, 15–22 (2015).
- Banani, S. F., Lee, H. O., Hyman, A. A. & Rosen, M. K. Biomolecular condensates: organizers of cellular biochemistry. *Nat. Rev.* **18**, 285–298 (2017).
- Dunker, A. K., Cortese, M. S., Romero, P., Iakoucheva, L. M. & Uversky, V. N. Flexible nets. The roles of intrinsic disorder in protein interaction networks. *FEBS J.* **272**, 5129–5148 (2005).
- Uversky, V. N., Oldfield, C. J. & Dunker, A. K. Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. *J. Mol. Recognit.* **18**, 343–384 (2005).
- Sharma, R., Raduly, Z., Miskei, M. & Fuxreiter, M. Fuzzy complexes: Specific binding without complete folding. *FEBS Lett.* **589**, 2533–2542 (2015).
- Uversky, V. N. Multitude of binding modes attainable by intrinsically disordered proteins: a portrait gallery of disorder-based complexes. *Chem. Soc. Rev.* **40**, 1623–1634 (2011).
- Uversky, V. N. Amyloidogenesis of natively unfolded proteins. *Curr. Alzheimer Res.* **5**, 260–287 (2008).
- Uversky, V. N. Targeting intrinsically disordered proteins in neurodegenerative and protein dysfunction diseases: another illustration of the D(2) concept. *Expert Rev. Proteom.* **7**, 543–564 (2010).
- Uversky, V. N., Oldfield, C. J. & Dunker, A. K. Intrinsically disordered proteins in human diseases: introducing the D2 concept. *Annu. Rev. Biophys.* **37**, 215–246 (2008).
- Hollstein, M., Sidransky, D., Vogelstein, B. & Harris, C. C. p53 mutations in human cancers. *Science (New York, N.Y.)* **253**, 49–53 (1991).
- Spillantini, M. G. et al. Alpha-synuclein in Lewy bodies. *Nature* **388**, 839–840 (1997).
- Baba, M. et al. Aggregation of alpha-synuclein in Lewy bodies of sporadic Parkinson's disease and dementia with Lewy bodies. *Am. J. Pathol.* **152**, 879–884 (1998).
- Kovacs, G. G. et al. Natively unfolded tubulin polymerization promoting protein TPPP/p25 is a common marker of alpha-synucleinopathies. *Neurobiol. Dis.* **17**, 155–162 (2004).
- Irvine, G. B., El-Agnaf, O. M., Shankar, G. M. & Walsh, D. M. Protein aggregation in the brain: the molecular basis for Alzheimer's and Parkinson's diseases. *Mol. Med. (Cambridge, Mass)* **14**, 451–464 (2008).
- Trejo-Lopez, J. A., Yachnis, A. T. & Prokop, S. Neuropathology of Alzheimer's Disease. *Neurotherapeutics* **19**, 173–185 (2022).
- Mochizuki, H., Choong, C. J. & Masliah, E. A refined concept: alpha-synuclein dysregulation disease. *Neurochem. Int.* **119**, 84–96 (2018).
- Liu, Z. & Huang, Y. Advantages of proteins being disordered. *Protein Sci.* **23**, 539–550 (2014).
- Beveridge, R. & Calabrese, A. N. Structural Proteomics Methods to Interrogate the Conformations and Dynamics of Intrinsically Disordered Proteins. *Front. Chem.* **9**, 603639 (2021).
- Uversky, V. N. Dancing Protein Clouds: The Strange Biology and Chaotic Physics of Intrinsically Disordered Proteins. *J. Biol. Chem.* **291**, 6681–6688 (2016).
- Uversky, V. N. p53 Proteoforms and Intrinsic Disorder: An Illustration of the Protein Structure-Function Continuum Concept. *Int. J. Mol. Sci.* **17**, 1874 (2016).
- Uversky, V. N. Protein intrinsic disorder and structure-function continuum. *Prog. Mol. Biol. Transl. Sci.* **166**, 1–17 (2019).
- Uversky, V. N. Looking at the recent advances in understanding alpha-synuclein and its aggregation through the proteoform prism. *F1000Research* **6**, 525 (2017).
- Kauffman, S. *At home in the Universe, the search for the laws of complexity.* (Penguin, London); (1996).
- Norris, V., Engel, M. & Demarty, M. Modelling biological systems with competitive coherence. *Adv. Artif. Neural Syst.* **2012**, 1–20 (2012).
- Norris, V. Competitive Coherence Generates Qualia in Bacteria and Other Living Systems. *Biology* **10**, 1034 (2021).
- Vohradsky, J. & Ramsden, J. J. Genome resource utilization during prokaryotic development. *FASEB J.* **15**, 2054–2056 (2001).
- Cheng, F., Vivacqua, G. & Yu, S. The role of alpha-synuclein in neurotransmission and synaptic plasticity. *J. Chem. Neuroanat.* **42**, 242–248 (2011).
- Longhena, F., Faustini, G., Spillantini, M. G. & Bellucci, A. Living in Promiscuity: The Multiple Partners of Alpha-Synuclein at the Synapse in Physiology and Pathology. *Int. J. Mol. Sci.* **20**, 141 (2019).
- Lehotzky, A. et al. Tubulin polymerization-promoting protein (TPPP/p25) is critical for oligodendrocyte differentiation. *Glia* **58**, 157–168 (2010).
- Fu, M. M. et al. The Golgi Outpost Protein TPPP Nucleates Microtubules and Is Critical for Myelination. *Cell* **179**, 132–146.e114 (2019).
- Hisahara, S. et al. SIRT1 decelerates morphological processing of oligodendrocyte cell lines and regulates the expression of cytoskeleton-related oligodendrocyte proteins. *Biochem Biophys. Res Commun.* **546**, 7–14 (2021).
- Olah, J. et al. Microtubule-Associated Proteins with Regulatory Functions by Day and Pathological Potency at Night. *Cells* **9**, 357 (2020).
- Sun, J. et al. Functional cooperation of alpha-synuclein and VAMP2 in synaptic vesicle recycling. *Proc. Natl Acad. Sci. USA* **116**, 11113–11115 (2019).
- Gao, V., Briano, J. A., Komer, L. E. & Burre, J. Functional and Pathological Effects of alpha-Synuclein on Synaptic SNARE Complexes. *J. Mol. Biol.* **435**, 167714 (2023).
- Pingale, T. D. & Gupta, G. L. Novel therapeutic approaches for Parkinson's disease by targeting brain cholesterol homeostasis. *J. Pharm. Pharmacol.* **73**, 862–873 (2021).
- Acevedo, K. et al. The phosphorylation of p25/TPPP by LIM kinase 1 inhibits its ability to assemble microtubules. *Exp. cell Res.* **313**, 4091–4106 (2007).
- Cartelli, D. et al. alpha-Synuclein is a Novel Microtubule Dynamase. *Sci. Rep.* **6**, 33289 (2016).
- de Oliveira, R. M. et al. The mechanism of sirtuin 2-mediated exacerbation of alpha-synuclein toxicity in models of Parkinson disease. *PLoS Biol.* **15**, e2000374 (2017).
- Kumar, R. et al. Partially oxidized DJ-1 inhibits alpha-synuclein nucleation and remodels mature alpha-synuclein fibrils in vitro. *Commun. Biol.* **2**, 395 (2019).
- Olah, J., Lehotzky, A., Szenasi, T. & Ovadi, J. Anti-Aggregative Effect of the Anti-oxidant DJ-1 on the TPPP/p25-Derived Pathological Associations of Alpha-Synuclein. *Cells* **10**, 2909 (2021).
- The Human Protein Atlas. (2023). <https://www.proteinatlas.org/>.
- Tokesi, N. et al. Identification of motives mediating alternative functions of the neomorphic moonlighting TPPP/p25. *Biochimica et. biophysica acta* **1842**, 547–557 (2014).
- Kalia, L. V. & Lang, A. E. Parkinson's disease. *Lancet (Lond., Engl.)* **386**, 896–912 (2015).
- Schmid, A. W., Fauvet, B., Moniatte, M. & Lashuel, H. A. Alpha-synuclein post-translational modifications as potential biomarkers for Parkinson disease and other synucleinopathies. *Mol. Cell Proteom.* **12**, 3543–3558 (2013).
- Whittaker, H. T., Qui, Y., Bettencourt, C. & Houlden, H. Multiple system atrophy: genetic risks and alpha-synuclein mutations. *F1000Research* **6**, 2072 (2017).
- Hu, G., Wu, Z., Uversky, V. N. & Kurgan, L. Functional Analysis of Human Hub Proteins and Their Interactors Involved in the Intrinsic Disorder-Enriched Interactions. *Int. J. Mol. Sci.* **18**, 2761 (2017).
- Lehotzky, A. et al. Co-Transmission of Alpha-Synuclein and TPPP/p25 Inhibits Their Proteolytic Degradation in Human Cell Models. *Front. Mol. Biosci.* **8**, 666026 (2021).
- Pechmann, S. & Frydman, J. Interplay between chaperones and protein disorder promotes the evolution of protein networks. *PLoS Comput. Biol.* **10**, e1003674 (2014).

53. Maroteaux, L., Campanelli, J. T. & Scheller, R. H. Synuclein: a neuron-specific protein localized to the nucleus and presynaptic nerve terminal. *J. Neurosci.* **8**, 2804–2815 (1988).
54. Takahashi, M. et al. A brain-specific protein p25 is localized and associated with oligodendrocytes, neuropil, and fiber-like structures of the CA3 hippocampal region in the rat brain. *J. Neurochem.* **60**, 228–235 (1993).
55. Bates, C. A. & Zheng, W. Brain disposition of alpha-Synuclein: roles of brain barrier systems and implications for Parkinson's disease. *Fluids barriers CNS* **11**, 17 (2014).
56. Vincze, O. et al. A new myelin protein, TPPP/p25, reduced in demyelinated lesions is enriched in cerebrospinal fluid of multiple sclerosis. *Biochem Biophys. Res Commun.* **409**, 137–141 (2011).
57. Parnetti, L. et al. Parkinson's and Lewy body dementia CSF biomarkers. *Clin. Chim. acta; Int. J. Clin. Chem.* **495**, 318–325 (2019).
58. Wakabayashi, K., Yoshimoto, M., Tsuji, S. & Takahashi, H. Alpha-synuclein immunoreactivity in glial cytoplasmic inclusions in multiple system atrophy. *Neurosci. Lett.* **249**, 180–182 (1998).
59. Ota, K. et al. Relocation of p25alpha/tubulin polymerization promoting protein from the nucleus to the perinuclear cytoplasm in the oligodendroglia of sporadic and COQ2 mutant multiple system atrophy. *Acta neuropathologica Commun.* **2**, 136 (2014).
60. Mavroei, P. et al. Endogenous oligodendroglial alpha-synuclein and TPPP/p25alpha orchestrate alpha-synuclein pathology in experimental multiple system atrophy models. *Acta neuropathologica* **138**, 415–441 (2019).
61. Nishimura, Y. et al. Early and extensive alterations of glial connexins, distal oligodendroglial pathology type demyelination, and nodal/paranodal pathology are characteristic of multiple system atrophy. *Brain Pathol. (Zur., Switz.)* **33**, e13131 (2022).
62. Surguchov, A. & Surguchev, A. Synucleins: New Data on Misfolding, Aggregation and Role in Diseases. *Biomedicines* **10**, 3241 (2022).
63. Wakabayashi, K. et al. The Lewy body in Parkinson's disease and related neurodegenerative disorders. *Mol. Neurobiol.* **47**, 495–508 (2013).
64. Preusser, M., Lehotzky, A., Budka, H., Ovadi, J. & Kovacs, G. G. TPPP/p25 in brain tumours: expression in non-neoplastic oligodendrocytes but not in oligodendroglial cells. *Acta neuropathologica* **113**, 213–215 (2007).
65. Feng, D. D., Cai, W. & Chen, X. The associations between Parkinson's disease and cancer: the plot thickens. *Transl. neurodegeneration* **4**, 20 (2015).
66. Olah, J., Bertrand, P. & Ovadi, J. Role of the microtubule-associated TPPP/p25 in Parkinson's and related diseases and its therapeutic potential. *Expert Rev. Proteom.* **14**, 301–309 (2017).
67. Surguchev, A. A., Emamzadeh, F. N. & Surguchov, A. Cell Responses to Extracellular alpha-Synuclein. *Molecules (Basel, Switz.)* **24**, 305 (2019).
68. Szunyogh, S., Olah, J., Szenasi, T., Szabo, A. & Ovadi, J. Targeting the interface of the pathological complex of alpha-synuclein and TPPP/p25. *Biochimica et. biophysica acta* **1852**, 2653–2661 (2015).
69. Szenasi, T. et al. Challenging drug target for Parkinson's disease: Pathological complex of the chameleon TPPP/p25 and alpha-synuclein proteins. *Biochim Biophys. Acta Mol. Basis Dis.* **1863**, 310–323 (2017).
70. Zheng, Y. et al. Novel DNA Aptamers for Parkinson's Disease Treatment Inhibit alpha-Synuclein Aggregation and Facilitate its Degradation. *Mol. Ther.* **11**, 228–242 (2018).
71. Ren, X. et al. Exosomal DNA Aptamer Targeting alpha-Synuclein Aggregates Reduced Neuropathological Deficits in a Mouse Parkinson's Disease Model. *Mol. Ther.* **17**, 726–740 (2019).
72. Khan, I., Preeti, K., Fernandes, V., Khatri, D. K. & Singh, S. B. Role of MicroRNAs, Aptamers in Neuroinflammation and Neurodegenerative Disorders. *Cell. Mol. Neurobiol.* **42**, 2075–2095 (2021).
73. Szklarczyk, D. et al. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic acids Res.* **51**, D638–D646 (2023).
74. Orosz, F. et al. TPPP/p25: from unfolded protein to misfolding disease: prediction and experiments. *Biol. Cell* **96**, 701–711 (2004).
75. Tokes, N. et al. TPPP/p25 promotes tubulin acetylation by inhibiting histone deacetylase 6. *J. Biol. Chem.* **285**, 17896–17906 (2010).
76. Tuttle, M. D. et al. Solid-state NMR structure of a pathogenic fibril of full-length human alpha-synuclein. *Nat. Struct. Mol. Biol.* **23**, 409–415 (2016).

## ACKNOWLEDGEMENTS

This work was funded in part by the Natural Sciences and Engineering Research Council of Canada (Grant SPG-P 521331–2018 to S.K.).

## AUTHOR CONTRIBUTIONS

V.N. wrote the program; V.U. provided the String data; J.Ov. and J.OI. provided the information on experiments with the disordered SYN and TPPP proteins on which the hypothesis is based; S.K. provided aptamer information; all authors contributed to the hypothesis, analysed the data and wrote the paper.

## COMPETING INTERESTS

The authors declare no competing interests.

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41540-023-00291-8>.

**Correspondence** and requests for materials should be addressed to Vic Norris.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023