

PCCP

Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.



PCCP

ARTICLE

Protein structural robustness to mutations: an *in silico* investigation

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

Mounia Achoch^a, Rodrigo Dorantes-Gilardi^b, Chris Wymant^c, Giovanni Feverati^d, Kave Salamatian^a, Laurent Vuillon^b and Claire Lesieur^{e†}

Proteins possess qualities of robustness and adaptability to perturbations such as mutations, but occasionally fail to withstand them, resulting in loss of function. Here the structural impact of mutations is investigated independently of the functional impact. Primarily, we aim at understanding the mechanisms of structural robustness, pre-requisite for functional integrity. The structural changes due to mutations propagate from the site of mutation to residues much more distant than typical scales of chemical interactions, following a cascade mechanism. This can trigger dramatic changes or subtle ones, consistent with a loss of function and disease, or the emergence of new functions. Robustness is enhanced by changes producing alternative structures, in good agreement with the view that proteins are dynamic objects fulfilling their functions from a set of conformations. This result, robust alternative structures, is also coherent with epistasis or rescue mutations, more generally with non-additive mutational effects and compensatory mutations. To achieve this study, we have developed the first algorithm, referred to as Amino Acid Rank (AAR), which follows the structural changes associated with mutations from the site of the mutation to the entire protein structure and quantifies the changes so mutations can be ranked accordingly. Assessing the paths of changes opens the possibility to assume secondary mutations for compensatory mechanisms.

Introduction

How proteins sustain and adapt their biological functions, or fail to do so, is a complex question. The structure and function of proteins are defined by amino acid sequences which naturally vary upon genetic mutations. The robustness of proteins against mutations depends on the impact on the protein function of the structural changes arising from the mutations, changes which are not much investigated¹. Proteins are strongly resistant to single amino acid mutations: most amino acids can be mutated without loss of function², i.e. such mutations are functionally neutral. Less frequently, with a frequency about 10^{-9} per site, mutations lead to the emergence of new functions (innovation)³. Alternatively, there are pathological mutations which lead to a loss of function. The present view of neutral mutations is that some are adaptive because their combination with other mutations drives functional evolution through non-additive effects (e.g. functional promiscuity or epistasis)³. Non-additive effects are also involved in rescue mechanisms, where the negative effect of pathological mutation is

neutralized by a mutation at a second site^{2, 4-6}. Generally, protein robustness, protein innovation and protein adaptation refer to the impact of mutations on the biological function of proteins.

On the other hand, the structural changes which are tolerated by a protein without jeopardizing the protein functionality (functional robustness or emergence of a new function) or those who on the contrary lead to loss of functions, are rarely looked into. Yet, even little understanding of the underlying structural changes would be instructive to address pathological mutations or help designing new enzymes. The gap between the studies on functional and structural robustness is due to several issues. To investigate functional robustness, a protein prototype is chosen, every individual amino acid is mutated and the function of the mutants is tested experimentally⁷. Likewise, studying structural robustness, namely maintenance of the structural integrity necessary for a biological function, implies to choose a protein prototype, mutate every individual amino acid, crystallize each mutant, solve each structure and compare the ones which share the same function. First, this is technically and financially challenging as well as time consuming. Second, the goal is to understand if a protein structure is built to bear mutational changes and if so, to investigate by what mechanisms. Thus an experimental approach is not appropriate because some mutations would fail to produce a structure but for reasons not necessarily related to structural robustness. A mutation might prevent folding and acquisition of a stable structure, but have no impact on the structural robustness. For instance, the B subunits of the pentamers of the cholera toxin and the heat labile enterotoxin maintain a pentamer at pH 5.0 but do not reassemble at this pH⁸⁻¹¹. Also a mutation leading to a new structure and a new function might not easily be identified as such, experimentally. On the other hand, *in silico* mutations produce structural changes in

^a Laboratoire d'informatique systèmes, traitement de l'information et de la connaissance (LISTIC), Université de Savoie, Annecy le Vieux, France

^b Laboratoire de mathématiques (LAMA UMR 5127), Université Savoie Mont Blanc, CNRS, Le Bourget du Lac, France

^c Medical Research Council Centre for Outbreak Analysis and Modelling, Department of Infectious Disease Epidemiology, Imperial College London, London, United Kingdom

^d Federation de recherche Fr3405, Modelisation, Simulations, Interactions Fondamentales, Annecy-le-Vieux, France

^e CNRS-UCBL, IXXI-ENS-Lyon, Laboratoire AMPERE, Lyon, France

† Corresponding author: claire.lesieur@ens-lyon.fr
See DOI: 10.1039/x0xx00000x

order to generate a stable structure. *In silico* methods cannot create a new structure or destroy a structure from a mutation, they produce a set of conformations close to the wild-type structure. This is a relevant framework to investigate the structural changes which underlie structural robustness as a general issue rather than having to restrict the study on specific mutations. The third issue is the lack of tools to measure and compare the effects of mutations on a structure, comparison needed to understand the mechanisms by which the protein structure bears the changes. There exist programs to compare global structure features (e.g. RMSD) and visualize structural differences¹²⁻¹⁵. But here it is about following changes from a local perturbation, the site of the mutation, to the entire protein structure.

To circumvent these difficulties, we have adopted the following strategy. We have worked on the atomic structure of the pentamer of the cholera toxin B subunit (CtxB₅) because it is a stable protein with an OB-fold, structure common to many other proteins with different sequences. We can therefore assume that the structure is naturally robust to mutational changes. We have generated a set of *in silico* mutations using Fold X, which produces structural changes maintaining a reliable structure¹⁶. Let us recall that the goal of the study is not to predict the effects of experimental mutations on a structure, but to have a set of mutations appropriate to explore structural robustness. The dataset is the individual mutation of all the amino acids which compose the toxin interface. To analyse the structural changes due to mutations, we have modelled the toxin interfaces as networks of amino acids in interaction such that the structural properties are compared through network comparison. The analysis of the networks helped us to build an ad hoc algorithm, called Amino Acid Rank (AAR) which takes into account all structural changes observed in the dataset, quantifies them and ranks the mutations accordingly.

Finally, we have analysed the results of AAR in terms of structural robustness. The results indicate that mutations generate structural changes at different scales (local or long range) in a cascade mechanism and independently of the local changes on the mutation site and of the nature of the mutation. Structural robustness relies not only on mutations producing no or little changes but also on mutations producing significant structural changes but generating redundant conformations, in good agreement with the recent definition of protein as an ensemble of conformations fulfilling one function. Thus, the redundancy produces alternative structures necessary for having conformations functionally distinct upon secondary mutations, consistently with "adaptive neutral mutations". An example of non-additive mutations is provided not in the context of emerging functions but as a correction mechanism of a cancer-related mutation reported in the tetrameric domain of the tumour suppressor p53. This error-correction mechanism is not conceivable if structural robustness is based only on a lack of structural changes upon mutation. The identification of a second site mutation capable of correcting default is possible because of the new algorithm AAR.

Methods

AminoAcidRank (AAR) algorithm. *Function SpetralPro.* The goal is to model a protein interface by a hotspots network. A protein interface is made of the amino acids of one chain which interact with the amino acids of adjacent chains. These amino acids are referred to as hotspots. To construct a hot-spot network, we first define its atomic network. Using the atomic coordinates from a

PDB, all distances between atoms of one chain and atoms of adjacent chains are computed. Two atoms share a link if they are within 5 Å distance. Two hotspots share a link if they have at least one of their respective atoms within 5 Å distance from one another. It is convenient to represent the hot-spot network as its adjacency matrix A. If N is the number of hotspots in the protein, then A is the N × N matrix with value a_{i,j} in row i and column j if i and j are connected by a link, and 0 otherwise. The weighted adjacency matrix W is defined by w_{i,j}, the weight of the link connecting i and j, that is the number of atomic links between amino acid i and amino acid j. The adjacency matrix A is defined by a_{i,j} equals 1 if w_{i,j} > 0, otherwise a_{i,j} equals to 0.

Function Arank. A mutated PDB is generated with Fold X introducing a single hotspot mutation of a residue at position r. The function SpectralPro is then applied on the mutated PDB. To compute the quantity of structural changes produced by the mutation, a N × N « difference matrix » D is defined as follows: d_{i,j} = w_{i,j}^{mut} - w_{i,j}^{wt} where d_{i,j} is the entry value of D at row i and column j, w_{i,j}^{mut} is the weight of the mutated network at row i and column j and w_{i,j}^{wt} is the weight of the wild type (WT) network at row i column j.

The structural changes produced by the mutation on the entire structure (Global changes, arank_r) are computed as the sum of the absolute value of all the entries of D (that is $\sum_{i,j} |d_{i,j}|$). The structural changes at the position of the mutation (local changes, local_r) are computed as the sum of the absolute values of all entries of D at row j (that is $\sum_j |d_{i,j}|$). The arank_r values are used to rank mutations according to the amount of structural changes they produce.

Function Backup. This is to compute the redundancy of every link of the WT hotspot network. The backup links are sought within the local secondary structure around every hotspot link based on the known hydrogen bonding of secondary structure. That is any (i,j) links located within a distance of 4 residues along the sequence on both chains of the considered hotspot link is computed as its backup link. Details are provided in the AAR pseudocode.

The AAR pseudo code is provided in the electronic supplementary information (ESI).

Fold X. Mutations were computed using the protein design tool of Fold X (version 3 beta)^{16,17}. Only the protein design function was used for mutagenesis using the PDB 1EEI as the wild-type (WT) structure. Details and run parameters are in the electronic supplementary information (ESI). Essentially the run parameters are chosen to minimize their impact on the network construction, to be applicable broadly on X-ray structures, and not to depend too strongly on a high quality structure. Here the qualities of the structures need to be at ~ 2.5 Å or above resolution.

Results and discussion

The aim is to investigate the structural changes that a protein may go through from individual mutations of its amino acids, still maintaining a stable structure. As a model of study, we use CtxB₅,

focusing on the amino acids that compose the toxin interface, so-called hotspots. A protein structure is built on atomic interactions between its amino acids, likewise for a protein interface. Thus to analyze the structural changes that take place in the toxin interface upon mutation, first intermolecular atomic interactions need to be established. The exact atomic interactions are intractable due to the large size of the system. Atomic interactions rely on chemical nature of atoms, distances between atoms and the atom environment (atomic neighbors). In order to take these parameters into account, the following procedure is undertaken (Methods). The distances between all atoms of one chain and all atoms of an adjacent chain, referred to as interatomic distances are calculated from the X-ray coordinates of CtxB₅ provided by the RCSB Protein Data Bank (PDB code 1EEI). All interatomic distances within 5 Å are considered as chemical interactions, without distinguishing the nature of the atoms (methods). This approximation is reasonable because every type of chemical interactions (van der Waals, electrostatic, hydrogen bonds, etc) between the atoms of amino acids carbon, oxygen, nitrogen, sulfur and hydrogen fall within a distance of less than 5 Å¹⁸. The chemical nature of atoms is not considered also because it is assumed that two atoms in the X-ray structure would not be close if they ought to chemically clash. They are either necessarily chemically compatible or their neighbors' shielding prevent them from clashing.

To each hotspot is associated a weight w_i , equals to $\sum_j w_{i,j}$, which is the total number of its links (intermolecular atomic distances within 5 Å, see methods). The pairs of atoms which are within 5 Å distance are coarse-grained to their respective amino acids in order to associate to each hotspot a number of amino acids in physical contacts (degree), number a_i , equals to $\sum_j a_{i,j}$. Because all the distances within 5 Å of every atom are considered, the algorithm intrinsically accounts for the neighbor atoms. The weight and the degree can be considered as proxy of the probability of interactions of the amino acid, the higher the degree the more likely the amino acid is to have an interaction.

Survey of the structural changes

Our algorithm Amino Acid Rank (AAR) after establishing the amino acids and the interactions that composed the toxin interface with the above procedure, models the interface as a network of amino acids in intermolecular interactions (Methods, Function SpectralPro). The amino acids that have at least one intermolecular atomic distance within 5 Å are linked and referred to as hotspots. The CtxB₅ interface has 58 hotspots forming the nodes of the network, these are also recognized as hotspots by other programs available¹⁹. There are no histidine or cysteine hotspots.

We systematically mutate every hotspot one by one. In the current work we restrict ourselves to mutations to asparagine residue for simplicity, asparagine having average chemical and geometrical properties. For example it has a residue that is polar rather than hydrophobic or charged, and has an average number of atoms as compared to other amino acids. Mutations to other amino acids will be considered in future work.

In silico mutations are performed using Fold X (Methods)¹⁶ to generate a mutated structure, from which a mutated toxin interface and a mutated network are produced by the AAR algorithm. To capture the structural changes associated with a mutation, AAR compares the networks after and before mutation and extracts all modified amino acid links (Methods, Function Arank). Mutations change the positions of atoms which modify the intermolecular atomic distances and so the nodes, degrees and weights of the

network. To quantify the structural changes produced by a mutation at position r within the entire structure ($arank_r$), AAR sums the absolute values of the differences between the weights after and before mutation of all the nodes of the networks, the higher the $arank_r$, the larger the structural changes (Table 1). A change in weight means some atoms have become closer or further away, implying atomic interaction rearrangements. Depletion of an amino acid link means that the two hotspots have no more atoms within 5 Å distance. Addition of a new link means that the two hotspots have moved closer so they have atoms within 5 Å distance. These are amino acid link rearrangements. To qualitatively describe the mutations, a sphere of influence is defined as the number of modified amino acids by the mutation and by the distances between the site of the mutation and the modified residue the furthest from it (Table 1). Two distances are measured, geodesic and Euclidian. The geodesic distance is measured by the number of chemical links to be crossed to go from the site of mutation to the modified residue the furthest from it by the shortest path and the Euclidian distance is measured between the two residues in Ångström (Fig. 1). The spheres of influence of the fifty eight mutations are shown on their respective X-ray structures in Fig. S1 (see the electronic supplementary information, ESI), highlighting the broad diversity of structural changes in quantity and quality. The $arank_r$ values vary from 182 to 2, ten mutations have an $arank_r$ below the first quartile while fifteen have an $arank_r$ above the third quartile, and thus most mutations generate significant changes (Table 1). The changes involve side chain atoms only since the RMSD is zero for all mutations. No more than 10 % of the native interfacial contacts are lost upon mutations. On average the mutations modified eight hotspots; a quarter modifies only up to five hotspots and a quarter modifies more than eleven. Thirteen mutations out of fifty-eight produce only local perturbations, namely structural changes of residues in physical contact with the site of the mutation and so located within the chemical reach of the mutated residue (Euclidian and geodesic distances within 5 Å and 1, respectively). Forty-five mutations produce global changes, namely changes beyond physical contact and chemical reach of the mutated residue. Eighteen modify residues located at distances above 10 Å. The maximum long range modification is 17 Å. The mechanism of the long range modifications is chemically sound since the changes are going from hotspots chemically linked to hotspots chemically linked in a step-by-step manner as determined from the geodesic distances (Fig. 1). This cascade mechanism seems related to the secondary structure of the mutated residue since out of eighteen residues belonging to α -helices, seventeen produce a cascade (long range changes) upon mutation (95 %). Out of twenty-six which belong to a β -structure, thirteen produce a cascade (50 %) while out of fourteen which belong to a loop, twelve produce a cascade (86 %). This relation would need to be verified and further explored on a dataset. There are twelve mutations for which the changes do not go from hotspots to hotspots but go from the mutated residue to its intramolecular contacts, which subsequently modify their hotspots (Table 1, column Intra). It is still a step-by-step mechanism, but through intramolecular and intermolecular links. Thus the results highlight paths of changes between amino acids of the interface and amino acids outside it. Likewise, mutations of amino acids outside the interface are capable of modifying hotspots' degrees (work in progress). This is consistent with the mechanisms of protein assembly combining folding and association steps in a coordinated manner (for review see²⁰). A step-by-step mechanism is described in other real networks as Peer-to-Peer mechanisms (P2P)²¹.

As selected examples, the mutations K69N, A64N, L31N and I39N are considered in details because they allow covering the chemical and geometrical properties of amino acids (small, medium and large side chain, hydrophobic, charged and polar chemical nature). Their spheres of influence are shown in the X-ray structures of the respective mutants (Figure 2A). Large modifications are seen for the K69N and A64N mutants while fewer modifications take place for the mutants I39N and L31N. The mutations K69N and A64N are among the top disruptive ones with $arank_r$ values equal to 182 (first rank) and 112 (fifth rank), respectively (Table 1). This highlights that the extent of the structural changes cannot be inferred by the difference between the nature of the original and mutated residue since lysine is bigger and has more atoms than asparagine while alanine is smaller and has less atoms. This is further supported by the fact that the mutations of other lysine or alanine such as K34N and A102N have different AAR values (Table 1). To consolidate this point the spheres of influences shown in figure S1 are sorted by amino acid type, subsequently sorted by decreasing values of $arank_r$.

Now if the mutation K69N is compared to the mutation L31N, the latter has an $arank_r$ value ten times lower than 182. Yet the residue L31 has a degree 9 and a weight 74, significantly higher than the degree and weight of the residue K69, 2 and 29, respectively. Like the nature of the residue, the degree or the weight does not condition the extent of the structural changes. This is further evidenced by plotting the $arank_r$ values against the weight of the original residue before mutation for the fifty eight mutations (Fig. 2B). The linear correlation is weak (Fig. 2B, $R^2 = 0.27$), indicating that mutation of an amino acid with a high weight does not systematically lead to large structural changes, and likewise mutation of an amino acid with a low weight does not necessarily lead to few structural changes.

The $arank_r$ values are then plotted against the local weight changes (local, weight differences on the mutated residue after and before mutation, see methods), and again a rather weak linear correlation is observed (Fig. 2C, $R^2 = 0.44$). This indicates that global changes are not proportional to local changes. Moreover, only some mutations have $arank_r$ values which fall on the straight-line of slope two implying local changes (Fig. 2C, red line). Most mutations have $arank_r$ values outside this line and so they produce global changes and involve cascades. If there are only local changes, that is weight changes on the mutated residue and nowhere else, then the global changes are twice the local changes because the global changes count the weight changes on the mutated node and on its endpoint nodes. This confirms that mutations produce changes at different scales as shown by the spheres of influence (Fig. S1). The absence of correlations between the $arank_r$ values and the local weight before mutation or the local weight changes remains true even if the networks are built with cut offs 4 and 6 Å instead of 5 Å. Thus these properties are invariant within the experimental error of X-ray structures (~ 1 Å). It is interesting to discuss the two AAR outliers, the mutation R67N and the mutation K69N because they have similar local and global changes (Table 1). What is different however is their fraction of local changes: R67N has lost 24 % of its interactions (24/101, ratio local weight difference to weight before mutation) while K69N has lost 77 % (23/30). The fraction of local changes does not correlate either with the global changes measured by AAR (not shown).

Structural robustness, fragility and adaptation

To assess whether the structure of a protein is built to bear mutational effects, we propose to consider the structural changes produced in the CtxB interface by the mutations and see if they are consistent with all known mutational effects: robustness, innovation, adaptation/rescue and pathology.

The first key point is that the mutations yield structural impact at different scales (Table 1, Fig. 2, Fig. S1). This means there is no *a priori* specific scale (e.g. 5 Å) at which structural changes can be detectable and it is necessary to measure them locally as well as globally. This is in good agreement with other studies showing both direct and indirect physical interactions in co-evolving residues¹. Local structural changes, namely modification within the chemical reach of the site of the mutation is consistent with enzymatic innovation or adaptation which does not lead to a full reorganization of the global structure. Global structural changes are consistent with pathologies where a single mutation is enough to jeopardize a structure and consequently a function. Of course, this does not imply that enzymatic innovation and pathology occurs only via local and global changes, respectively. This all depends on the scale at which the function is regulated by the structure.

The scaling does not explain adaptation through epistasis, rescue mechanism, or compensatory mutations (non-additive effects). Let us consider the pre-requisite for such effects: a mutation at a site 1 with an effect 1 (Mutant 1) and a mutation at a site 2 with an effect 2 (Mutant 2). Non-additive effects mean the consequences of the combination of mutations 1 and 2 are different from the consequences of mutation 2 (or of mutation 1) individually. This implies that the structures of the mutant 1 (or of mutant 2) and of the wild-type are different, otherwise they would react similarly upon the secondary mutation (Fig. 3). In other words, a robust mutation that leads to a rescue mechanism or a compensatory effect upon a second site mutation necessarily has a structure distinct from the WT one. This suggests that functional robustness is built on mutations with no structural impact (neutral mutation) as well as on mutations producing distinct structural solutions functionally equivalent to the WT one (adaptive mutations). If true, this means among networks different from the WT one (i.e. $Arank_r \neq 0$), some should be WT-alternative and other should be dissimilar. To investigate this possibility, the four mutations K69N, A64N, L31N and I39N are considered again. The structural changes due to these mutations are schematized by networks before and after mutation on Fig. 4. Let us first consider the mutations K69N and A64N which both have significant structural changes, namely high $arank_r$ (Fig. 4A). The K69N mutation modifies the layout of the WT network substantially, since it reduces the atomic interactions between the region of interface composed of residues 63 to 67 of one chain and residues 73 and 65 of the adjacent chain, and simultaneously increases the atomic interactions between the residue 67 of one chain and the residues 27 to 37 on the adjacent chain. This is well-illustrated on the X-ray structures (Fig. 4A). Moreover, the mutation also depletes the only two weak ties of the WT network, namely the links (31, 50) and (63, 53) which connect two regions of interface otherwise unconnected.

On the contrary, the networks A64 and N64 have a similar layout (Fig. 4A). In fact, the N64 network appears like a WT alternative network with more amino acid links, but the same regions are connected. The K69N and A64N mutations well-illustrate the distinction between structural changes and alternative structural solutions. The mutations I39N and L31N have low $arank_r$ (14 and 18, respectively) but a similar result can be observed (Fig. 4B). Only the link (39, 8) is depleted in the I39N mutation, not modifying the network significantly since there are other linked residues in the

vicinity of the link (39, 8) (Fig. 4B). In contrast, the L31N, even though it also yields a single link depletion (31, 50), the mutated network is not equivalent to the WT one because it lacks the only link that was connecting the regions 50, 64-68, 88 and 96-98 through the intermolecular link (31, 50) (Fig. 4B). It is therefore important to acknowledge that structural changes large or small yield alternative networks or not. So, the quality of structural changes must also somehow be incorporated in order to anticipate the impact of a mutation. Because of the scaling issue and the cascade mechanism, establishing the appropriate measure for alternative networks to sort out robust (neutral and adaptive) and fragile mutations is complex and beyond the scope of the present work.

The obvious difference between the A64N and I39N alternative networks and the altered K69N and L31N networks is the redundancy of amino acid and atomic links in the formers. This is reminiscent of peer-to-peer networks, which are robust to perturbation because they have more links than necessary -'back up' links- such that depletion or addition of links is tolerated by generating several alternative networks²². To see if alternative structures and networks exist in proteins, we have measured backup amino acid links in the interface of CtxB₅. Two amino acid links (i, j) and (i', j'), which belong to the same secondary structural element, defined as the residues $-i-$ and $-i'-$ are four amino acids apart along the sequence and likewise for their respective $-j-$ and $-j'-$ residues, are considered to backup each other. This is because the integrity of the secondary structure relies on at least the amino acid links which participate to the hydrogen bonding. The maximum distance of four amino acids apart along the sequence corresponds to a helix turn ($i + j - 4$) so backup links are counted within this range of distance along the backbone. Based on this definition of backup, AAR calculates the number of backup links for each link of the WT network (Methods, Function backup). Out of 92 links of amino acids, only the two weak ties have no backup. Eleven links have 1 to 3 backups, fifty-two have 4 to 13 backups and twenty-seven have more than fourteen backups. A backup network of the WT toxin interface is shown in Fig. 5, with the number of backups of each link described by a colour code. The network shows a non-uniform distribution of the number of backup per hotspots within the structure that may indicate fragile areas. This result supports the possibility of having neutral structural changes through addition and/or depletion of links producing alternative networks and structural robustness (Fig. 5). The backup for the residues K69, A64, L31 and I39 are 16, 41, 80 and 26, respectively. The mutations A64N and I39N which have a redundant network also have a higher backup than K69N. The L31N has a highest backup but the amino acid link (31, 50) has none. This illustrates the complexity in assessing robustness due to the scaling problem (robustness of a node, of a link or of a region/community). Nevertheless the results are encouraging to further explore the concept of backup as a measure of robustness and fragility.

WT alternative networks lay the ground for non-additive mutational effects because different atomic interactions would cope differently with secondary mutations. A mutation not tolerated in a WT network/structure might be tolerated in a mutated WT alternative network. We tested this possibility to further support a mechanism of robustness via alternative WT networks. The cancer-related mutation G334V reported for the tetrameric domain of the tumour suppressor p53, is used as a default mutation case²³. The goal is to find a second site mutation which alone produces neutral structural

changes and a WT alternative network but coupled with the G334V mutation prevents its structural damages, corroborating non-additive effects through alternative networks. The impact of the G334V mutation on the protein conformation is such that X-ray crystallography is inapplicable and there is no fiber structure available yet. The mutation G334V is generated *in silico* from the WT atomic structure (PDB 1SAK) using Fold X instead. The interface between chains D and B is analysed. The G334V mutation leads to a large amount of structural changes as the AAR is 286, there are side chain and backbone atom rearrangements since the RMSD is 0.03 Å. The sphere of influence reveals long range changes up to residues at geodesic distances five and Euclidian distance 15 Å from the residue 334 (Fig. 6). The structural changes go from the residue 334 up to the residue 324 on the N-terminal end and up to the residue 352 on the C-terminal end (Fig. 6A). The mutation does not change the degree of the residue 334 but it changes the degree of its intramolecular amino acid neighbours, residues 333 and 337, in a cascade mechanism (Fig 6). As a result, the residue 337 loses its pairing with the residues 345, 349 and 352, keeps its pairing only with the residue 348, reducing the connectivity within the interface region composed of the residues 345 to 352 and 337 to 341 (Fig. 6B). Moreover, the residue 333 also loses pairing with the residue 345 removing a link between the interface region composed of residues 330-334 and 325-328 and the interface region composed of the residues 337-341 and 345-352 (Fig. 6B). It is possible that the rigidity between these two regions loosen up after depletion of the link 345-333. The residue N345 is at the cross-road of the structural changes produced by the mutation G334V. We tested if a mutation at this position could reinforce the atomic interactions of the network such that it becomes robust to the G334V mutation. Again *in silico* mutations are performed using Fold X. The network of the single mutant N345D is similar to the WT network except for an increase of the weights (number of atomic interactions) of the links (345, 333), (345, 341), (337, 348) and (337, 349) and a decrease of the weight of the link (337, 345) (Figure 6B). The double mutant N345D+G334V has structural changes on half as many residues as the mutant G334V, it maintains both links (345, 333) and (345, 337) and its network looks like the WT one, apart from an additional link between the residue 333 and 352 found as well in the single mutant G334V (Fig. 6B). The small changes in the atomic interactions produced by the N345D prevent the residue 337 from moving away after the mutation of the residue 334 and prevent the loss of the link (333, 345). This is a non-additive mutational effect since the effects of the individual mutations differ from the effects of combined mutations; the effects of the G334V are lost when combined with the N345D mutation. This suggests that a second site mutation producing a compensatory effect is to be found among the residues modified by the first site mutation, namely it is on the sphere of influence of the first site mutation. This hypothesis is supported by the observation that on average, in the interface of CtxB₅, eight amino acids are modified by mutation and on average deleterious mutations can be compensated by nine mutations^{1, 24}.

Conclusions

The work investigates the mechanisms proteins use to resist structural changes upon mutations, as a groundwork to understand functional robustness. Assuming that all proteins bear mutations by similar mechanisms, a case of study is a good model of investigation. The first challenge is to elaborate a set of mutations producing structural perturbations still maintaining a viable structure to look at. The solution proposed is to mutate *in silico*

every amino acid of the interface of the B subunit pentamer of the cholera toxin and to monitor structural changes via a network model of the interface. A network representation is interesting because it allows measuring local to global changes and to investigate the capacity of proteins to cope with perturbation²⁵. The relevance of network models in the study of structures for protein dynamics is now well established²⁶⁻³². The second achievement is the AAR algorithm which quantifies all structural changes between wild-type and mutant structures by simply counting the changes in their number of atomic interactions. AAR is fast (less than one second for a protein of 103 amino acids), thorough and applicable on the Cartesian coordinates of any atomic structures.

One novel finding is that structural changes follow a cascade mechanism where the local reorganization of the atoms at the site of the mutation disturbs the chemical neighbors of the mutated residue which in turn disturb their chemical neighbors, etc as in a domino effect. What triggers the cascade is not yet identified but it is neither the degree nor the weight of the original residues nor the fraction of local changes. This differs from networks where perturbations propagate through hubs (highly connected nodes)³³. Instead, the changes propagate stepwise from hotspot to hotspot, from the site of the mutation to its neighbors (local change) to the rest of the protein (global change). This cascade mechanism results in major changes in interactions stretching out to large distances, or to more subtle changes. As mentioned already, the formers are consistent with pathological mutations while the latter accommodate adaptability and emergence of new functions through structural rearrangements which do not completely modify the protein conformation⁷. A cascade mechanism is also consistent with allostery, although multiple perturbations -as found in binding- are not tested here³⁴. The cascade mechanism is more reliable than propagation of changes through hubs in a network with a power law distribution (few hubs, many low degree) because it tallies with experimental evidences on the functional impact of mutations. In a hub-regulated network, the mutation of hubs would lead to massive change, and pathologies; the mutation of residues with low degree would lead to local changes and explain robustness^{35,36}. Yet, it would be difficult to account for the emergence of new function through few subtle changes as well as for adaptive mutations (non-additive mutation effects), since there would be little or large changes. Moreover, proteins do not have hubs in terms of having nodes with a significantly higher degree than other nodes, they have nodes with average degree²⁵.

The second novelty is the mechanism of robustness through alternative structures, rather than just unchanged structures. This fits the updated definition of protein function: an ensemble of conformations³⁷. This also lays the ground for adaptability because it allows for non-additive effects, error corrections or epistasis^{6,38}. The presence of backup links in the WT network, which allows addition and depletion of links without altering substantially the network layout, might be a clue for identifying what triggers the cascade. Backup and alternative solutions are a current mechanism of robustness, reported for other real networks such as peer-to-peer networks or other biological networks^{39,40}.

In summary, the extent of structural changes produced by mutations does not depend on the degree of the mutated residue, and it does not condition the impact of a mutation on the structure. The impact of mutation involves more complex mechanisms which remain to be deciphered⁴¹. Altogether the mechanisms of structural changes observed through an *in silico* approach are consistent with all known functional effects of mutations (robustness, innovation,

adaptation and pathology) supporting the approach as well as the hypothesis that structural robustness is embedded in the structure of the protein.

Acknowledgements

We are very thankful to Paul Sorba and Sylvie Ricard-Blum for critical reading of the manuscript. Financial support of the Federation of Research MSIF (Modélisation, Simulations and Interaction Fundamentals) is gratefully acknowledged.

References

1. D. N. Ivankov, A. V. Finkelstein and F. A. Kondrashov, *Curr Opin Struct Biol*, 2014, **26**, 104-112.
2. A. Toth-Petroczy and D. S. Tawfik, *Curr Opin Struct Biol*, 2014, **26C**, 131-138.
3. G. Amitai, R. D. Gupta and D. S. Tawfik, *HFSP J*, 2007, **1**, 67-78.
4. E. A. Ortlund, J. T. Bridgham, M. R. Redinbo and J. W. Thornton, *Science*, 2007, **317**, 1544-1548.
5. M. L. Salverda, E. Dellus, F. A. Gorter, A. J. Debets, J. van der Oost, R. F. Hoekstra, D. S. Tawfik and J. A. de Visser, *PLoS genetics*, 2011, **7**, e1001321.
6. O. Demir, R. Baronio, F. Salehi, C. D. Wassman, L. Hall, G. W. Hatfield, R. Chamberlin, P. Kaiser, R. H. Lathrop and R. E. Amaro, *PLoS Comput Biol*, 2011, **7**, e1002238.
7. R. N. McLaughlin, Jr., F. J. Poelwijk, A. Raman, W. S. Gosal and R. Ranganathan, *Nature*, 2012, **491**, 138-142.
8. L. W. Ruddock, J. J. Coen, C. Cheesman, R. B. Freedman and T. R. Hirst, *J Biol Chem*, 1996b, **271**, 19118-19123.
9. L. W. Ruddock, S. P. Ruston, S. M. Kelly, N. C. Price, R. B. Freedman and T. R. Hirst, *J Biol Chem*, 1995, **270**, 29953-29958.
10. M. J. De Wolf, G. A. Van Dessel, A. R. Lagrou, H. J. Hilderson and W. S. Dierick, *Biochemistry*, 1987, **26**, 3799-3806.
11. J. Zrimi, A. Ng Ling, E. Giri-Rachman Arifin, G. Feverati and C. Lesieur, *PLoS One*, 2010, **5**, e15347.
12. J. Hsin, A. Arkhipov, Y. Yin, J. E. Stone and K. Schulten, *Current protocols in bioinformatics / editorial board, Andreas D. Baxeavanis ... [et al.]*, 2008, **Chapter 5**, Unit 5 7.
13. W. Humphrey, A. Dalke and K. Schulten, *Journal of molecular graphics*, 1996, **14**, 33-38, 27-38.
14. O. Bachar, D. Fischer, R. Nussinov and H. Wolfson, *Protein Eng*, 1993, **6**, 279-288.
15. M. Shatsky, R. Nussinov and H. J. Wolfson, *Methods Mol Biol*, 2008, **413**, 125-146.
16. J. Schymkowitz, J. Borg, F. Stricher, R. Nys, F. Rousseau and L. Serrano, *Nucleic acids research*, 2005, **33**, W382-W388.
17. R. Guerois, J. E. Nielsen and L. Serrano, *Journal of molecular biology*, 2002, **320**, 369-387.
18. G. Gronau, S. T. Krishnaji, M. E. Kinahan, T. Giesa, J. Y. Wong, D. L. Kaplan and M. J. Buehler, *Biomaterials*, 2012, **33**, 8240-8255.
19. M. Achoch, G. Feverati, L. Vuillon, K. Salamatian and C. Lesieur, Belgrade, Serbia, 2013.
20. C. Lesieur, *Oligomerization of Chemical and Biological Compounds*, 2014, DOI: 10.5772/58576
21. V. N. Padmanabhan, H. J. Wang and P. A. Chou, *11th IEEE International Conference on Network Protocols*,

- Proceedings*, 2003, DOI: Doi 10.1109/Icnp.2003.1249753, 16-27.
22. S. Boyd, A. Ghosh, B. Prabhakar and D. Shah, 2005.
 23. Y. Higashimoto, Y. Asanomi, S. Takakusagi, M. S. Lewis, K. Uosaki, S. R. Durell, C. W. Anderson, E. Appella and K. Sakaguchi, *Biochemistry*, 2006, **45**, 1608-1619.
 24. A. F. Poon and L. Chao, *Evolution; international journal of organic evolution*, 2006, **60**, 2032-2043.
 25. L. Vuillon and C. Lesieur, *Curr Opin Struct Biol*, 2015, **31**, 1-8.
 26. V. A. Feher, J. D. Durrant, A. T. Van Wart and R. E. Amaro, *Curr Opin Struct Biol*, 2014, **25**, 98-103.
 27. L. Di Paola and A. Giuliani, *Current opinion in structural biology*, 2015, **31**, 43-48.
 28. B. Barz, D. J. Wales and B. Strodel, *J Phys Chem B*, 2014, **118**, 1003-1011.
 29. K. V. Brinda and S. Vishveshwara, *Biophys J*, 2005, **89**, 4159-4170.
 30. G. Feverati, M. Achoch, L. Vuillon and C. Lesieur, *PLoS One*, 2014, **9**, e94745.
 31. D. M. Leitner, S. Buchenberg, P. Brettel and G. Stock, *The Journal of chemical physics*, 2015, **142**, 075101.
 32. D. M. Leitner, *The Journal of chemical physics*, 2009, **130**, 195101.
 33. A. L. Barabasi and Z. N. Oltvai, *Nature reviews. Genetics*, 2004, **5**, 101-113.
 34. G. M. Suel, S. W. Lockless, M. A. Wall and R. Ranganathan, *Nat Struct Biol*, 2003, **10**, 59-69.
 35. Y. Y. Liu, J. J. Slotine and A. L. Barabasi, *Nature*, 2011, **473**, 167-173.
 36. R. Albert, H. Jeong and A. L. Barabasi, *Nature*, 2000, **406**, 378-382.
 37. G. Parisi, D. J. Zea, A. M. Monzon and C. Marino-Buslje, *Curr Opin Struct Biol*, 2015, **32C**, 58-65.
 38. E. Dellus-Gur, M. Elias, E. Caselli, F. Prati, M. L. Salverda, J. A. de Visser, J. S. Fraser and D. S. Tawfik, *J Mol Biol*, 2015, DOI: 10.1016/j.jmb.2015.05.011.
 39. A. Wagner, *Biophys J*, 2014, **106**, 955-965.
 40. J. L. Payne and A. Wagner, *Science*, 2014, **343**, 875-877.
 41. T. Liu, S. T. Whitten and V. J. Hilser, *Proc Natl Acad Sci U S A*, 2007, **104**, 4347-4352.

Figure legend.

Figure 1. Schematic of the cascade mechanism underlying the structural changes associated with mutations. As the most disruptive mutation, K69N is chosen to illustrate the paths of the structural changes going from the site of mutation to elsewhere in the interface. The K69N mutation modified the atomic interactions of twenty-two hotspots of the interface covering a distance of fifteen Ångströms. The paths of changes are schematically described by arrows going from hotspots (nodes, black circles) chemically linked to hotspots chemically linked, the chemical distances (5 Å) are illustrated by dotted semi-circles. However, because the structure is a three dimensional object, the Euclidian distance between the site of mutation and the residue modified the further from it cannot be calculated from the schematic. The geodesic distances are the number of chemical links crossed to go from one hotspot to another. The structural changes of K69N cover three chemical links.

Figure 2. Local degrees and global changes. A. Spheres of influence. Only two adjacent chains D and E of CtxB₅ are represented in pale and dark grey strands, respectively (PDB 1EEI).

The toxin interface is in ribbon. The residues modified by mutations are spacefilled and the mutated residues are red. The left panel shows the location of the four mutated hotspots K69, A64, L31 and I39 on the WT structure. The other panels on the right are their respective spheres of influence as shown on their respective X-ray structures. **B. Weak correlation between the original weighted degree of the mutated residue and the amount of structural changes after mutation measured by AAR.** Arank_r values are plotted against the weights of each hotspot *-i-* before mutation w_{iWT} . The dotted line is the linear correlation. **C. Global vs local changes.** Arank_r values are plotted against local_r values (methods, local weighted degree differences ($|w_{imut} - w_{iWT}|$)). The dotted line is the linear correlation and the red line is for $\gamma = 2x$.

Figure 3. Schematics of additive and non-additive mutational effects. A WT network maintaining two segments together through four links of amino acids is drawn. Two sites of mutations M₁ and M₂ are considered. **Non compensatory mutations (Upper schematic).** If M₁ implies no structural and network reorganisation, then M₂ has the same effect on the WT and M1 mutated network. **Compensatory mutations (Lower schematic).** If M₂ does not have the same effect of the WT and M₁ mutated networks, then the M₁ and WT structures and networks are different.

Figure 4. Structural robustness. A. Networks of K69 and A64 residues, before and after mutation. Networks of the sphere of influence with hotspots nodes and links of hotspots as links. Zoom on a subset of interfacial residues in the X-ray structures of K69 and N69 (balls and stick representation). The numbers are the sequence position of the residues. The residue 69 of chain E and the residue 67 of chain D are shown in CPK and yellow, respectively. The residues of the chain E are otherwise colored in green. The backbone shows that both structures are in the same position. **B. Networks of the spheres of influence of the residue I39 and L31, before and after mutation.** Legend as in 4A.

Figure 5. Backup network of the WT interface. Structural robustness is based on the presence of backup links that allow bearing addition and depletion of links without structural impact. The nodes of the backup networks represent the hotspots, the size of the nodes represents their degree. The links represent pairs of hotspots and the colors of the links represent the number of backup for each link within a range indicated by the color scale on the right. The reddest link, the least backup interactions the pair of amino acid has. The arrows indicate the positions of the two nodes with weak ties (50, 31) and (53, 63). The letters on the network are the chains on which the hotspots are located.

Figure 6. Non-additive in silico mutations G334V and N345D in the p53 tetrameric domain. A. p53 WT. Left panel. The chains B (light grey) and D (dark grey) of the WT p53 are shown in backbone representation (PDB 1SAK) except for the residues of the sphere of influence of the mutation G334V, spacefilled. Right panel. As on left, but with a strand representation but in strands except for the residues indicated in balls and sticks. The cascade of changes is illustrated by arrows. **B. Networks of the WT, G334V, N345D and G334+N345D spheres of influence.** Legend as in figure 4. The mutated residues are in red. The open circles are the residues whose degrees are modified by the mutation. Arrows illustrated the path of structural changes going from the residue 334 to the residue 352. The red lines are for added (continuous) and depleted

links (dotted) of amino acids. Black thick and thin lines are for increased and decreased weights, respectively.

Table 1. Mutations features

Mutations	arank _i	Global Changes				Local changes			
		# modified hotspots	Geodesic	Euclidian	Intra	$\alpha_{i\text{WT}}$	$W_{i\text{WT}}$	$\Delta\alpha_{i(\text{Mut-WT})}$	$ \Delta W_{i(\text{Mut-WT})} $
K69N	182	22	3	15	0	2	30	-1	23
R67N	178	16	4	9	0	9	101	-5	24
Y76N	143	8	2	6	0	4	40	-3	37
Q3N	120	4	1	5	0	4	43	-1	26
A64N	112	18	3	10	0	4	20	5	41
Y12N	102	10	4	9	0	4	41	-4	44
T78N	97	5	3	8	0	1	2	0	1
A32N	94	11	2	5	0	5	35	2	44
E29N	90	11	3	10	0	6	77	0	30
R73N	88	18	3	15	0	4	42	-1	32
Y27N	86	16	3	13	0	5	41	-2	8
E66N	82	15	3	13	1	2	33	0	8
A98N	80	8	2	11	0	3	23	1	38
M101N	76	11	2	13	0	6	60	0	5
F25N	72	6	2	5	0	3	39	0	28
N103K	70	7	2	5	0	4	44	-2	33
A80N	66	9	3	9	0	1	1	1	24
K23N	66	6	3	11	0	1	7	-1	7
G33N	60	7	2	6	0	3	25	1	29
T71N	56	12	10	10	0	3	31	0	4
K81N	54	5	3	9	0	1	1	0	0
D70N	53	16	5	16	1	2	28	-1	10
L77N	51	14	3	10	0	4	8	1	3
S26N	48	5	1	5	0	2	15	2	25
P2N	48	7	2	7	0	4	19	0	14
V50N	46	14	4	17	1	1	1	0	1
R35N	46	9	1	6	1	5	47	-1	9
E36N	42	15	2	14	1	5	36	-2	1
Q61N	38	10	3	8	0	4	39	0	6
A97N	38	8	2	5	1	3	34	0	15
T28N	36	8	1	5	0	4	35	3	16
E11N	36	4	2	5	0	1	15	0	12
S100N	34	5	1	5	0	2	22	1	17
T1N	33	7	1	5	0	5	32	0	1
I99N	32	8	2	10	0	3	36	1	15
P93N	32	6	2	5	0	3	31	0	3
S30N	30	7	2	5	0	5	31	2	14
I58N	26	6	2	5	0	3	10	-3	10
I74N	20	10	4	9	0	3	7	-2	5
K34N	20	4	1	5	1	3	11	0	4
L31N	18	8	1	5	0	9	74	-1	1
S60N	16	8	3	7	0	2	18	0	3
L8N	16	11	2	5	1	5	19	-2	3
K63N	16	9	3	12	0	4	19	-2	6
W88N	16	7	2	11	1	3	11	-2	7
I65N	16	8	2	11	0	1	7	0	3
M68N	16	7	2	5	0	3	31	-1	8
Q49N	16	4	1	7	1	1	7	-1	7
N4K	15	5	2	5	0	1	4	3	11
I39N	14	7	1	5	0	4	16	-1	5
P53N	12	5	2	5	0	1	3	2	3
M37N	12	4	1	5	0	3	8	-2	4
I24N	12	4	2	8	1	1	1	0	0
A102N	12	3	2	5	0	3	26	0	6
T92N	8	2	1	5	0	2	16	0	4
I96N	2	3	2	5	0	1	6	0	0
I5N	2	3	2	6	1	1	7	0	0
T47N	2	2	1	5	0	1	10	0	1

i is a hotspot, k_i its degree; W_i its weighted degree; the Euclidian distances are Ångström.

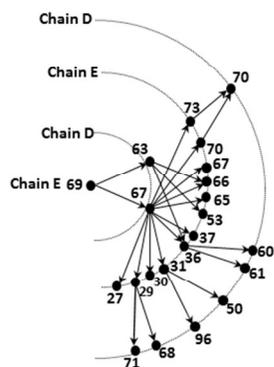
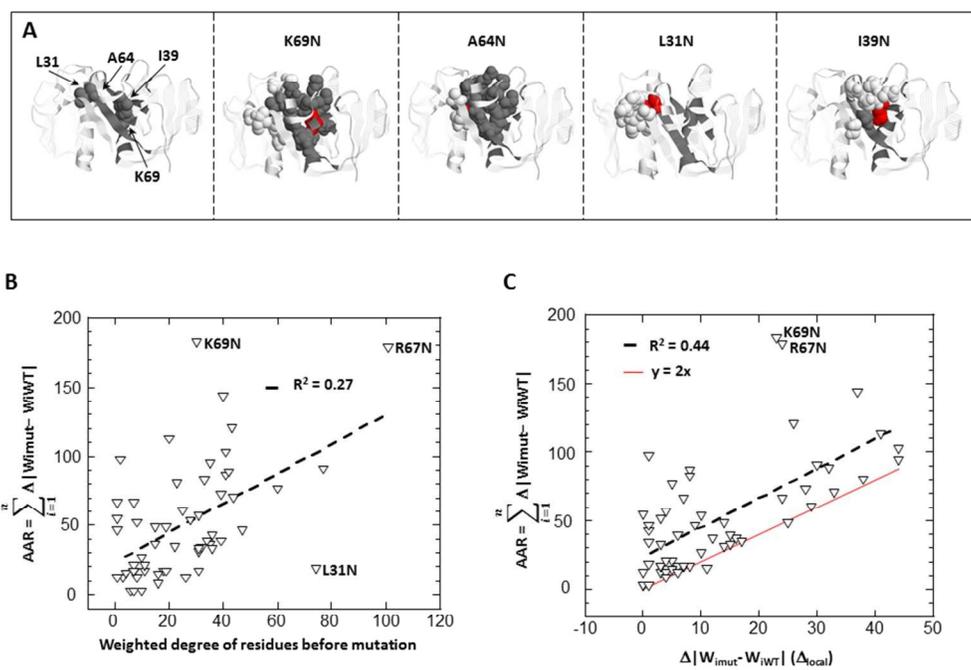


Figure 1

254x190mm (96 x 96 DPI)



254x190mm (96 x 96 DPI)

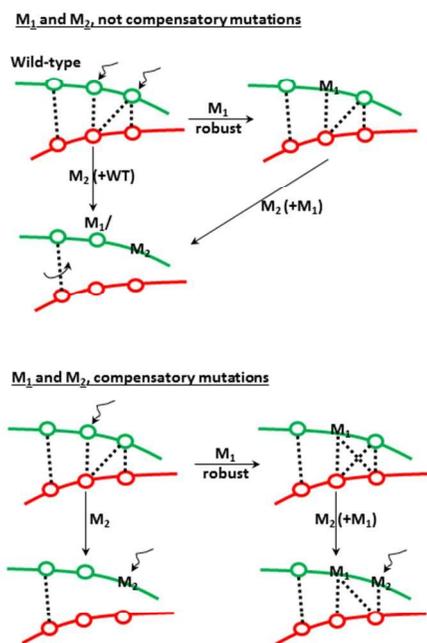
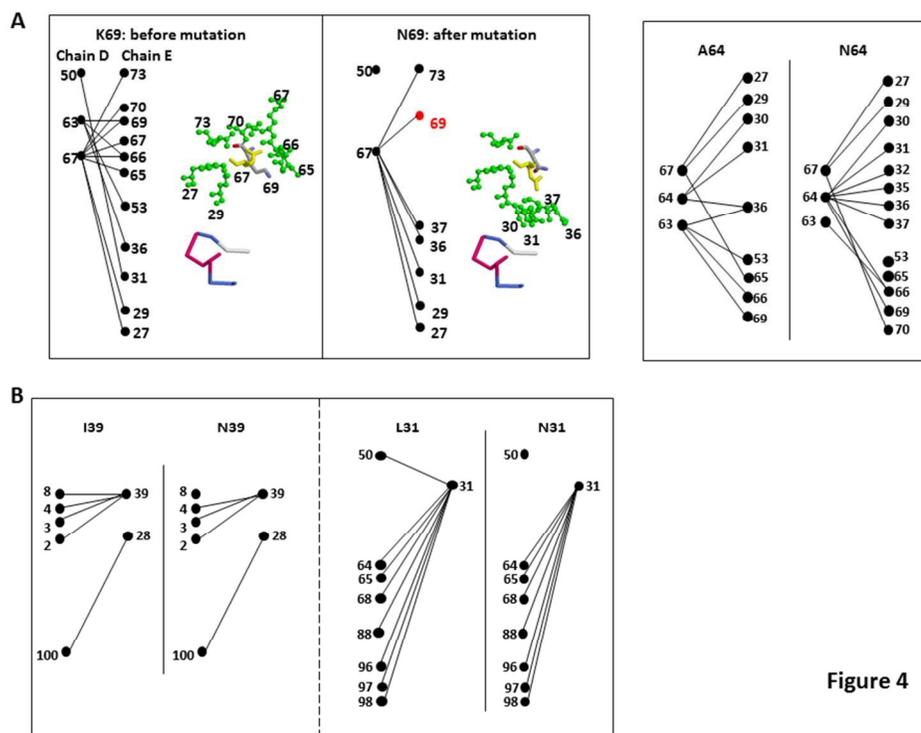


Figure 3

254x190mm (96 x 96 DPI)



254x190mm (96 x 96 DPI)

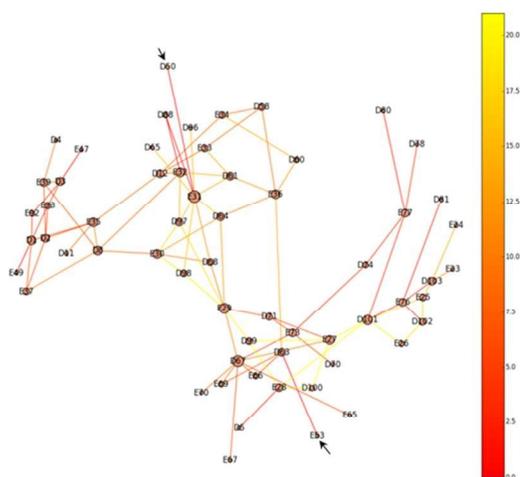


Figure 5

254x190mm (96 x 96 DPI)

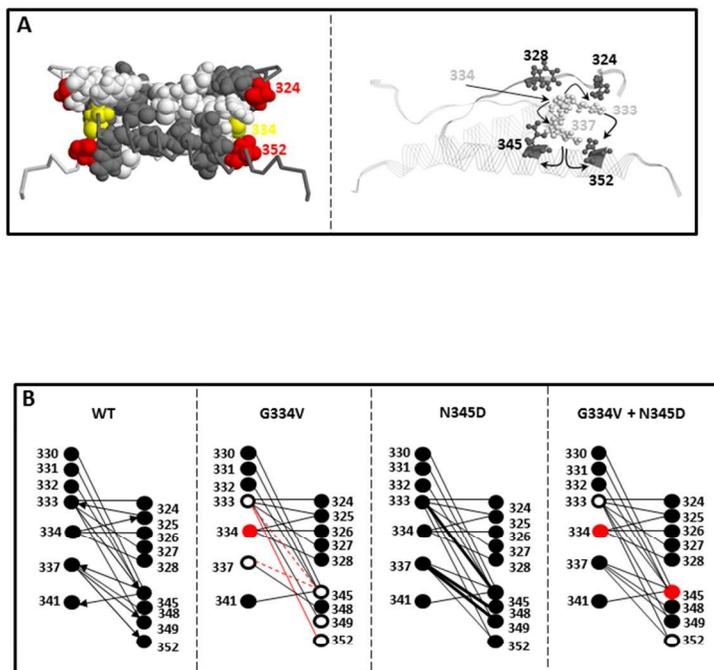


Figure 6

254x190mm (96 x 96 DPI)