 doi:10.20944/preprints201812.0273.v1

# Structural perspective on revealing and altering molecular mechanisms of genetic variants linked with diseases

**Yunhui Peng, Emil Alexov, Sankar Basu***

**Department of Physics and Astronomy, Clemson University, SC, USA**

*) corresponding author: nemo8130@gmail.com

**Key words:** mutations, disease-causing mutations, polymorphism, folding free energy change, binding free energy change, drug discovery, in silico screening

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 December 2018 doi:10.20944/preprints201812.0273.v1

**Abstract**

*Structural information of biological macromolecules is crucial and necessary to deliver predictions about the effects of mutations – whether polymorphic or deleterious (i.e., disease causing), wherein, thermodynamic parameters, namely, folding and binding free energies potentially serve as effective biomarkers. It is to be emphasized that the effect of a mutation depends on various factors, including the type of protein (globular, membrane or intrinsically disordered protein) and the structural context to which it occurs. Furthermore, due to the intrinsic plasticity of proteins, even mutations involving radical change of the structural and physico-chemical properties of the amino acids (native vs. mutant) can still have minimal effects of protein thermodynamics. However, if a mutation causes significant perturbation of either folding or binding free energies, it is quite likely to be deleterious. Mitigating such effects is a promising alternative to the traditional approaches of designing inhibitors. This can be done by structure-based in silico screening of small molecules for which binding to the dysfunctional protein restores its wild type thermodynamics. In this review we emphasize on the effects of mutations on two important biophysical characteristics, stability and binding affinity and how structures can be used for structure-based drug design to mitigate the effects of disease-causing variants on the above biophysical characteristics.*

## Effect of mutations on stability and binding

Study of the effect of amino acid mutations within proteins has been a traditional chapter in protein science. Earlier studies were applying site-directed mutagenesis to assess the importance of an amino acid for stability and function of the corresponding protein. Nowadays, the focus has shifted to understanding the effects caused by genetic variants, non-synonymous single nucleoside polymorphism (nsSNP), with respect to disease predisposition. The phage-T4 lysozyme, for example, has served as one of the most well studied systems with regard to mutations [1]. These studies were facilitated by the availability of X-ray structures (native and mutants) allowing for structural investigations of the effects on protein packing, stability and activity. At the same time, lysozyme mutants in human have also been studied to characterize the molecular mechanism of diseases like hereditary systemic amyloidosis [2]. Similarly, the barnase-barstar protein-inhibitor complex was subjected to extensive mutagenesis to reveal the role of various residues on binding affinity [3-5]. This high-resolution complex has served as a model system to study protein-protein recognition by single and double mutant cycles [4, 5]. Such studies have also served to rationalize optimization theories in the binding energetics [5, 6] generally applicable to protein-protein recognition. Recent advances in the study of genetic (DNA) variants in the same system have also revealed their influence in the manifestation of differential immunogenicity and this very property has then been applied in bio-therapeutics, for example, by constructing heterodimeric barnase-barstar DNA vaccine molecules [7], path-breaking in the development of novel DNA vaccines with increased potency. Effectively, numerous works in molecular biophysics were and are focusing on understanding the effects of mutations on protein stability and binding. Below we review the relevant concepts and works associated with the two most fundamental biophysical events in protein science, folding and binding.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 December 2018 doi:10.20944/preprints201812.0273.v1

To begin with effects of amino acid substitutions on protein folding, we emphasize that the same substitution may have different effect occurring in globular, membrane or intrinsically disordered proteins. It is perhaps good to reiterate that globular proteins are characterized by the presence of densely packed interiors (hydrophobic core) with packing densities (0.7 to 0.8) resembling to that of crystalline solids [8] and, that, interior packing is known to be one of the most dominant forces in protein folding [9], also related to the stability, dynamics and the *de novo* design of the foldable globules. The dense interior packing within globular proteins is known to be achieved by a nucleation-condensation of 'packing motifs'[10], concomitant to the rapid collapse of hydrophobic residues in an aqueous environment. On the other hand, helix packing in integral membrane proteins[11] inserted within the lipid bilayer does not involve the 'hydrophobic effect' and yet scales to an equivalent magnitude of packing to that of the globular proteins [12]. The polar vs. hydrophobic environment presented in the two cases demand differential amino acid compositions in the two types of proteins to achieve an equivalent degree of packing in both. Interestingly, small hydrophobic (Gly, Ala)[11, 12] and small hydroxyl-containing (Ser, Thr)[12] amino acids have been found to be contributing the most in tight packing of helices in membrane proteins as opposed to large hydrophobic and aromatic residues[10, 13] in globular protein interiors. Apart from the tight packing of helices, membrane proteins are also known to involve distinct pattern of charges[14, 15] embedded in their sequence to remain stable and active within the amphiphilic lipid bilayer. In dramatic contrast, in case of intrinsically disordered proteins (IDPs), the interior packing is practically negligible [16], since, (unlike globular proteins) the few hydrophobic residues in them are so placed that forbids the possibility of a hydrophobic collapse to attain a stable fold with a well packed core. This, in fact, enables them to retain their characteristic disorder or dynamic flexibility by means of existing as conformational ensembles rather than a single stably folded global minima structure likewise to either globular[9] or membrane proteins[17]. The major component in retaining this dynamic flexibility in IDPs is electrostatic interactions[18] involving hydrogen bonds, salt-bridges, charge-dipole, and dipole-dipole interactions. Hence, when subjected to mutational studies, the sites to perform the mutations are chosen based on the knowledge-based prediction of the anticipated differential effect in folding, stability and dynamics for the three major class of proteins. Studies of mutation(s) performed on these three major class of proteins have thus naturally been chosen based on their anticipated effect in folding, stability and dynamics. For example, salt-bridge mutations have served to constitute one of the prime chapters in understanding the *modus operanddi* in IDPs[19, 20] while the study of hydrophobic core mutations have been traditionally used to probe interior packing within globular proteins [21] to be discussed in more details in the next section. Electrostatics also serves as an indispensable component in the folding and stability of globular proteins [22]. For membrane proteins, mutations have been chosen mostly based on structure-function relationships[23] like oligomerization[24], thermostability[25] etc. involving both packing and electrostatics. There have also been instances of strategic point mutations (e.g., involving proline, the well-known helix-breaker) introducing kinks **(Fig. 1)** and wedging on transmembrane helix-helix interfaces [26]. Apart from the specific emphasis on individual structure-functional attributes of these different classes of proteins, mutational studies have also been attempted as a mean to trace their evolutionary origin (or common ancestor), particularly relevant in the context of the 'globular-disordered interface' [27, 28] in proteins.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 December 2018 doi:10.20944/preprints201812.0273.v1

**Mutation and its Compensation: Structural Plasticity and Conformational Relaxation**

Structural studies have shown how conformational relaxation of both main- and side-chain atoms could compensate the deleterious effects of mutations, thereby preserving the overall fold [29,30]. The random mutation of the 12 out of 13 core residues of ribonuclease barnase is an example where 23% of the mutants retained their enzymatic activity *in-vivo* [31]. Other similar studies followed and the idea of introducing strategic multiple mutations was eventually extended into the realm of *de novo* design of proteins [32]. Parallel $(\alpha/\beta)_8$ – TIM barrel [33, 34] served as an exemplary early model system, wherein, the specificity in side-chain packing as well as the pattern of hydrophobicities, both, were revealed to play their part. However, from all such studies, it was unmistakable that conformational plasticity is an inherent feature in proteins, resulting in structural relaxation to reduce the effect of mutations – which is particularly applicable in the context of multiple core mutations **(Fig. 2)** in foldable globules [35], also relevant for IDP's [36, 37].

**Mutations in IDP's as compared to globular and membrane proteins**

Although one of the hallmarks of IDP's is to harbor a high degree of structural plasticity, this may not always guarantee compensation of the deleterious effects caused by certain missense mutations. We should recall that many human diseases like cancer, diabetes, neurodegenerative and cardiovascular disorders are associated with IDPs. Interestingly, similar mutational prototypes have generally been found to be more damaging in IDPs than in globular proteins [38-40] – which is somewhat paradoxical, given the fact that IDPs have a substantially greater degree of structural plasticity and therefore is expected to have a corresponding greater potential to compensate for the mutational damage than that of globular proteins. However, the stepwise macromolecular and cellular consequences of a certain mutational prototype is hierarchical, multi-factored and complex. For example, amyloidosis is the macromolecular end-effect responsible for many neurodegenerative disorders (e.g., Parkison's, Alzheimer's) which, in-turn, is a consequence of β-aggregation. Again, β-aggregation is more complex than to be accounted for by hydrophobicity and / or β-sheet propensity of a protein region alone[38]. Interestingly, IDPs have been found to be more prone to amyloidosis in spite of having a much lower aggregation propensity (having only 1/3$^{rd}$ of aggregation nucleating regions) compared to globular (and membrane) proteins [38]. This high aggregation propensity also explains the considerable amount of structural frustration in globular proteins [38]. As a matter of fact, the 'disease-associated missense mutations' in IDPs are also found in a higher prevalence with greater functional impact [41] than the 'neutral polymorphisms'[42]. More importantly, the IDP-disease-mutations are found to be associated with the 'disorder-to-order transitions'[36] at a far greater frequency than the polymorphic ones [41]. The cancerous mutations in p53[42, 43] in its DNA-binding domain are classic examples of IDP-disease-mutations, wherein, dramatic destabilization of the domain renders it disordered at physiological conditions [44]. Overall, there are many investigations associated with mutational studies on IDPs revealing their molecular evolution [28] and pathological features [20]. Traditionally, the mutations can be viewed as mostly 'hereditary' [20], chosen on the basis of geographic and ethnic variations, pedigree of individual families with a history of a certain (say, the Alzheimer's[59]) disease. At the molecular level, one of the major insights revealed by these mutational studies have been the influential role of salt-bridges in mediating the 'mutation-

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 December 2018 doi:10.20944/preprints201812.0273.v1

induced rigidity' associated with enhanced aggregation of the candidate IDP – which has also found support by recent molecular dynamic studies exploring the nitty-gritty and transient nature of salt-bridge dynamics **(Fig. 3)** in IDPs [19].

**The activity-stability trade off within proteins: emergence of temperature sensitive mutations[1]**

Proteins manifest their presence in the cell by performing their function(s), frequently referred to as activity. It is well understood that evolution does not optimize stability (say, thermal stability) or activity *per se* within proteins (particularly enzymes), rather optimizes a fine balance between the two [45,46]. This optimization is intrinsic to the continuous endless evolutionary selection pressure consistent with the evolution of new protein function(s) [47]. The coupling of an enzyme's resistance to thermal unfolding and its ability to perform the conformational changes required for its function constitutes is central to this so-called 'activity – stability trade-off'[48, 49] - which has served to rationalize the fundamentals of enzyme-bioenergetics for organisms growing across a wide array of temperature. As a matter of fact, disease causing mutations have also been found to be stabilizing sometimes (although rarely), in intracellular $Ca^{2+}$ release channels [50], [51]. A recent critical appraisal on the subject, however, has categorically pointed out that such a trade-off is not imposed by any physico-chemical constraints [48]. This observation was used to carry appropriate thermodynamic modifications to overcome the *natural* limits of enzyme efficiency [47] in biotechnological applications while designing chemical and genetic enzyme modifications [52]. One other key application emerging from the basic studies in enzyme-bioenergetics has certainly been the art of designing temperature sensitive (Ts) mutations [53, 54], favorably aiding research in protein design, bio-therapeutics, structural genomics and genetic engineering. Web-servers [55] and other *in-silico* methodologies [56] predicting high-accuracy Ts mutational hotspots in proteins have been instrumental, for example, in probing molecular defects and identifying drug targets in deadly viral genomes [57], in characterizing evolutionarily conserved essential (housekeeping) genes and pathways across organisms [58, 59] and also in formulating and validating novel concepts like neighborhood properties[60] contributing even further to the methodology development in the design of appropriate mutants.

**Probing the role of mutations in diseases: tracking changes in thermodynamic parameters**

Changes in folding and binding free energies ( ΔG) are the standard thermodynamic measures of the effect of mutations on protein stability and binding [61]. It has been demonstrated that for assessing the effect, one needs to take into account the relative change in ΔΔG with respect to the $\Delta G_{WT}$ rather than considering ΔΔG alone [62]. Changes in G were used to characterize sequence and structural patterns on human disease-causing amino acid variants [63]. Particular attention was paid on mutations involving reversal of biophysical characteristics of the wild type residue(s). Thus, salt-bridge mutations are typically disease-causing as demonstrated in the case of hyper-aldosteronism, wherein, mere removal of the charge (while keeping intact the side chain geometry) on a single strategic amino acid site (Glu → Gln) [64] effectively dismantling a critical salt-bridge was found to be critical. Salt-bridge mutations have also been revealed in IDPs to be

[1] Mutations that manifest deleterious altered function(s) only at elevated (or lowered) temperatures and remain silent otherwise at normal (physiological) temperature.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 December 2018 doi:10.20944/preprints201812.0273.v1

disease-associated with enhanced aggregation (in Alzheimer's and Parkinson's Diseases) of the proteins [20]. Similar observations have also been revealed by molecular dynamic simulations on IDPs, wherein, considerable reduction in the conformational variation was found in $A\beta_{42}$ upon dismantling both high persistence as well as transient salt-bridges [19]. Many cancer related mutations are associated with altered protein-protein interactions [65] and stability [66].

Because of the importance of thermodynamic parameters ( G) in revealing the pathogenicity of mutations, many computational approaches were developed to predict folding and binding free energy changes. These approaches vary from sequence based [67, 68], to structure based [69-71], depending on the input [61, 63]. Methodologies vary from empirical approaches [2, 73], first-principle approaches combination of knowledge-based terms and physics [5-77] to machine learning approaches [8, 79]. It should be emphasized that for effective drug discovery, one needs to know not only the thermodynamic effects of mutation, but also the 3D structure of the target biomolecule, as outlined in the next sections.

**Statistical classification of mutations based on their degree of harmfulness**

The effect of some mutations are more pronounced [80] than others. To that end, statistical studies [3,81] have broadly classified the nsSNPs into two major categories: (i) polymorphic (or harmless) and (ii) disease variants. The influential factors considered in such statistical studies are genetic variations, frequency of occurrence and statistical measure(s) of the degree of harmfulness (Pd, also known as disease index) [63,82]. The object of the exercise was to find any empirical correlations between the variation type and the degree of harmfulness, if any. To that end, the entire combinatorial space of 380 possible amino acid mutations[2] (that can occur from a set of 20 naturally occurring amino acids) were explored and the frequency of each mutation in the corresponding database was recorded. To overcome any possible database-bias (which is trivial to all knowledge-based statistics), the calculations were repeated as a mean to cross-validate the results on updated database(s). Major observations were that in the HumVar dataset [83], 108 out of 380 possible mutations were never found, and, contrastingly, the top 26 most frequent variants made up as much as 46% of the whole dataset. As a matter of fact, only about $1/4^{th}$ (only 87 out of 380) of the variants were found to belong to the "harmless" category [63] in the same database. As a cross-validation, when the analysis was repeated in an expanded dataset of more than 3-fold increased size, the results had very good correlations (R=0.90). In spite of this apparent similarity, the later study [63] pointed out the subtle differences between the two cases. More precisely, a jump was observed in the 'polymorphic-to-disease variant' ratio from 0.74 to 1.54 in the new compared to the older dataset. Such observations are indicative of the fact that many statistical studies are biased towards database availability. Such limitations of this knowledge-based approach speak in favor of using G as a more reliable and preferred biomarker on a case-to-case basis to make predictions about the effect of a particular mutation and then conclude on its pathogenicity.

[2] 20 amino acids each can be replaced by one out of the other 19 makes it 380

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 December 2018 doi:10.20944/preprints201812.0273.v1

**Computational approaches to mitigate the effects of disease-causing genetic variants:**
With the rapid development of computer techniques, computer-aided approaches have currently been widely applied in aiding early-stage drug discovery in both industrial as well as academic projects [84-87]. By discovering the potential compounds that target and affect the function of the specific proteins, the biological process can be modulated to mitigate or eliminate the disease-causing effects [85, 87]. Advances in human genome projects have provided a large plethora of target proteins for drug discovery projects [88, 89]. Meanwhile, breakthroughs in structural biology have offered in-depth structural information of more and more targets and elucidated the disease mechanisms at the molecular level [90-93]. Such advances have further stimulated the application of computational approaches to integrate the available structural information, functional mechanism and physico-chemical properties to drug discovery [86, 94]. Discovery of compounds to mitigate or eliminate the disease-causing effects induced by a specific amino acid mutation is the main goal of Personalized Medicines [95]. Benefiting from individual's genomic information (by means of comparing to the sequence consensus of the standard human genome), such compounds can be developed into potential drugs, targeting specific disease-associated mutations on the individuals to provide precise treatment.

**Importance of elucidating and clustering the effects of mutations' in drug design:**
In terms of a drug-design process, targeting specific disease-causing mutations, elucidation of the effects of mutations is of great importance, especially for the approaches requiring information about target protein structure. Alone with aforementioned computational approaches, this can further be integrated into the drug design pipeline. For example, free energy calculation methods are used to determine the dominant effects of mutations, whether affecting protein stability, protein binding or both. With the in-depth analysis of the effect of mutations at molecular-level, the disease-causing mutations in the target proteins can further be clustered by their major effects such as destabilizing mutation, catalytic mutations, mutations affecting dimerization or protein conformations [90, 96-98]. Such type of classification can help designing drugs for certain groups of mutations with similar effects and is thus applicable to a broader spectrum of diagnosis and therapy.

**Structure-based approach in drug design:**
Structure-based drug design (SBDD) is the computational approach that rely on knowledge of the 3D structure of the biological targets to identify or design the potential chemical structure suitable for clinical tests [94, 99]. With the explosion of genomic, functional and structural information in recent decades, majority of biological targets with 3D structure have been identified and stimulated the applications of structure-based approaches in the current design pipeline. SSDB is popular for virtual screening to filter the drug-like compounds from a large library of small molecules, including widely applied approaches such as docking and structure-based pharcorphore [87]. While the established high-throughput screening (HTS) allows for automatic testing of a wide range of compounds (up to millions), the low success rate and high cost limit its applications. Alternatively, one can use computational approaches to reduce the numbers of compounds subjected to testing [85, 100]. Below, we discuss the applications of two of the most popular approaches, namely, docking and structure-based pharmacophore design in the context of SBDD, targeting disease-causing mutations.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 December 2018 doi:10.20944/preprints201812.0273.v1

**Docking:**

Docking is one of the most common approaches for compound screening and the basic idea is to use scoring functions to evaluate the fitness of the target protein in complex with the docked compound [87]. Currently, vast docking programs have been developed to perform fast docking calculations with a wide array of protocols and scoring functions, such as Dock6 [101], Autodock Vina [102], Glide [103], Surflex [104] and many others. Such approaches require the structure of the target protein, either experimentally solved or modeled structure. As mentioned above, one should do intensive modeling to generate the best representative structure or set of structures to be subjected to docking [86, 105]. In the past, SBDD has been widely applied in mitigating the effects of mutations related to many common diseases. Examples include the p53 protein, which is the so-called "guardian protein" in cancer, functioning as a tumor-suppressor [106]. It is to be noted that p53 is a tumor suppressor gene (i.e., mutations are of the 'loss of function' type) and is therefore heterozygous recessive; i.e., both copies of the gene needs to be mutated to manifest the mutant phenotype. Again, only, some mutations in p53 result in the malfunctioning of the protein and increases the risk of cancers [107]. In cancer patients, mutations destabilizing the DNA binding to p53 are frequently observed and rescuing the functions of the 'mutant p53 protein' is of one central objective in current cancer research [108, 109]. In the past, it has been shown that binding of small molecules can stabilize the DNA binding domain and rescue mutant functions [109]. Recent work modeled the wild-type and several mutants [110] to elucidate the mechanism of p53 reactivation [110]. A novel transiently open L1/L3 pocket was identified and indicated the exposure of Cys-124 in the formation of such cavity [110]. Such finding is crucial as Cys-124 has been suggested to be the covalent docking site for known alkylating p53 stabilizers [111] while, compounds can be docked into this pocket to search for other potential stabilizers. As a matter of fact, 1,324 compounds from the NCI/DTP Open Chemical Repository Diversity Set II were docked onto the generated ensemble structures of R273H cancer mutant out of which 45 compounds were selected for biological assay [110]. Finally, one compound, stictic acid (NSC-87511) was experimentally validated to be an efficient reactivation compound for mutant p53 [110].

Besides cancer research, docking based screening is also used in rare diseases. Snyder-Robinson Syndrome (SRS) is a rare X-linked mental disease, caused by the malfunction of an important human enzyme, the spermine synthase [112]. Spermine synthase functions as homo-dimer and mutations affecting the dimerization such as G56S are shown to abolish the enzyme activity to result in the disease [62, 112]. Recent work has targeted identification of dimer stabilizers by binding to the mutant homo-dimer interface [113]. So to speak, integrated large commercial compound libraries were used for docking-based virtual screening with the representative structures [113]. The best-ranked 51 compounds were then subjected to experimental screening out of which three top-ranked compounds (also known as 'leads') have been shown to enhance the catalytic activity up to 30% [113, 114].

**Structure-based Pharmacophore design:**

Pharmacophore models can be used to make an ensemble of abstract steric and electronic features representing macromolecular (target protein) interactions with drug-like small molecules [115, 116]. In other words, three-dimensional arrangement of these features such as hydrophobic

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 December 2018 doi:10.20944/preprints201812.0273.v1

centroids, aromatic rings and hydrogen bonds are representation of the binding mode between the ligand and the target [116, 117]. Pharmacophores are generated from common features of active ligands, which are identified by aligning or superimposing the conformers of either ligand-target complexes or known active molecules [117]. Multiple degenerate atomic models can potentially be output from pharmacophore modeling programs requiring further optimization and validation to select the best one. Pharmacophore models are commonly used for virtual screening of active small molecules from large compound database[115-117]. Such approaches can be more efficient than docking for certain targets, especially when a large number of existing known active compounds are available [118].

Pharmacophore models have also been used for identifying active molecules to mitigate the effects of mutations in many diseases [119-122]. For the cases where a sufficient number of active molecules are previously known for generating high-quality pharmacophore models, pharmacophore proves to be a powerful tool for drug 'lead' identification [116]. Recent work has applied structure-based pharmacophore analysis to identify the novel ROS-1 inhibitors to curb the drug resistance problem caused by mutations [121]. Proto-oncogene receptor tyrosine kinase ROS-1 is ectopicly and oncogenicly expressed in many cancers, mainly in non-small cell lung cancer (NSCLC) [123]. ROS-1 is highly homologous with the kinase domain of anaplastic lymphoma kinase (ALK) and FDA approved ALK inhibitors such as Crizotinib are experimentally validated as therapeutics against ROS-1 driven tumors [121]. However, these commercial ROS-1 inhibitors lack a broad spectrum of activity due to the growing resistance from ROS-1 mutations, primarily G2032R [124]. Following on, a pharmacophore model was built using the complex structure of both wildtype and mutant ROS-1 with previously known inhibitors to identify more general inhibitors against both WT and mutant [121]. Pharmacophore-based virtual screening was then performed to selected candidates from commercial databases with further filtering and scoring analysis. Five hits were eventually identified with good binding affinities to both WT and mutant [121].

**Ligand-based approaches in drug design:**

In the lack of structural information of the target protein(s), the aforementioned structure-based approaches may not be suitable for drug design. As an alternative, ligand-based drug design (LBDD) can be applied for such cases [125]. Ligand-based methods only focus on the analysis of physico-chemical properties of known ligands that interact with the target of interests. Most popular approaches, however, are the quantitative structure activity relationship (QSAR) models and the ligand-based pharmacophore modeling [125]. In terms of drug design, targeting the mutant proteins, LBDD could be efficient for novel discovered mutations whose effects have not yet been investigated **(Fig. 4)**.

The basic assumption in ligand-based drug design is that small molecules with similar shape and biophysical properties will likewise interact with the same target receptor [117, 125]. By identifying the fingerprints of known active ligands and constructing LBDD models, large databases can be screened to retrieve the novel compounds as potential leads for the target of interest [125]. QSAR is a widely applied LBDD approach, which utilizes mathematical models to correlate the physiochemical properties of compounds to their experimentally measured activity in context of

 doi:10.20944/preprints201812.0273.v1

the corresponding biological process. Generally, QSAR methodology identifies the molecular descriptors associated with properties of the ligands and further uses mathematical models to discover correlations between molecular descriptors and the biological activity. Finally, these QSAR models are tested and validated for the prediction of biological activity of compounds. As it stands, the current *state-of-the-art* is to apply the QSAR models widely in computer-aided drug design, targeting the mutant protein(s). One major success is the discovery of the potential corrector for cystic fibrosis (CF) mutations, namely, F508del in cystic fibrosis transmembrane conductance regulator gene (CFTR) [126]. F508del is the most frequent CF causing mutation, which leads to the improper folding of the protein and its degradation [127]. Subsequent to the identification, QSAR analysis has further been applied to guide the synthesis of novel compounds to treat CF by improving the trafficking of the mutant CFTR (the CF corrector) [128]. Recent works have collected all compounds known to improve the F508del trafficking and then applied QSAR analysis to indicate the critical chemical descriptors for the potential F508del correctors [128]. A novel predictive model was then constructed with these identified descriptors to provide further guiding to the design and optimization of novel correctors [128].

As previously described, pharmacophore defines the interaction framework among the active ligands and their specific targets [116, 117]. Not necessarily relying on the structure of target protein(s), pharmacophore models can be generated with libraries of active ligands alone and trained for discrimination between active and inactive molecules [116]. Thus, pharmacophore models are widely used for virtual screening especially when lacking the target structure. In addition, as the pharmacophore model represent the binding (or interaction map) of 'active compounds-target interaction', it provides a plausible relationship between the structure and the ligand activity and could help to elucidate the underlying mechanism of biological modulator and further guide the design of the novel active compounds [116]. For example, by exploring the different pharmacological properties, recent studies have seem to improve the potency of existing pharmacophore and designed novel epidermal growth factor receptor (EGFR) inhibitor potentially inhibited by primary mutants (L858R, del9) and drug-resistant mutants, such as, L858R/T790M [122].

With the advances in understanding of structural and functional characteristics of biological targets, structure-based approaches have gained widespread popularity. However, it should be noted that the combination of ligand and structure-based approaches is expected to add significantly more to the current *'state-of-the-art'* [129, 130]. Such combinations can either be sequential, parallel or hybrid, integrated contextually into a drug discovery pipeline, and, have already shown much promise [130].

**Conclusion:**

Macromolecular structural analyses may potentially be used to aid probing of genetic variants linked with disease. Such studies are usually complemented by a wide range of biophysical solution assays and computational modeling. Research along these directions have also opened up avenues towards developing diagnostic tools and plausible therapeutics. In such a context, it is of foremost importance to conceptualize (i) how traditionally mutational effects on protein stability and binding have been probed and (ii) the basis of the differential affects of mutations to

 doi:10.20944/preprints201812.0273.v1

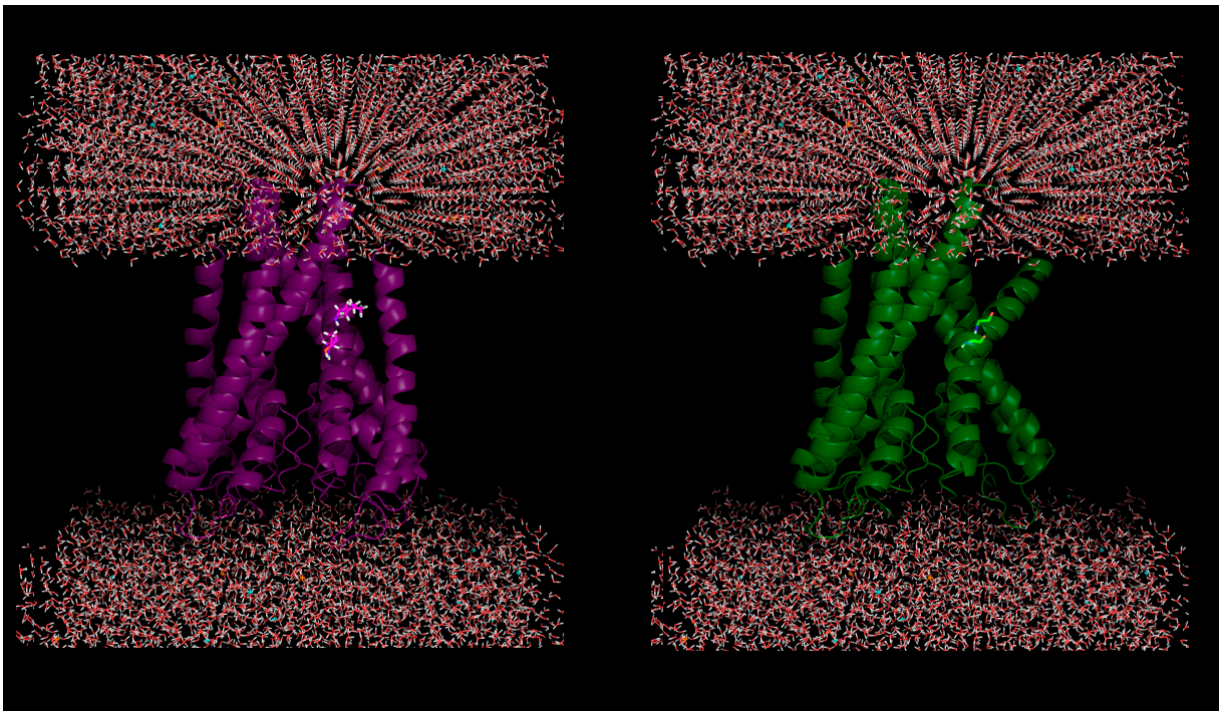
different class of proteins (globular, membrane and disordered proteins) based on conformational relaxation, structural plasticity, compensation and other physico-chemical factors. The activity-stability trade off in proteins is one of the key evolutionary concepts in such a context. As a probing technique, we particularly highlight the importance of tracking changes in thermodynamic parameters ($\Delta G_{WT}$) and also take the opportunity to discuss the limitations of knowledge based approaches as is the statistical classification of mutations based on their degree of harmfulness. The review particularly highlights the emergence of the ever-so-promising recent approach to computationally mitigate the effects of disease-causing genetic variants, alternative to the traditional approaches in designing inhibitors. A wide array of structure-based approaches in drug design including docking, structure-based pharmacophore design and ligand-based approaches have been vividly discussed along with their proper context of applicability, as to whether they are to be aided in presence or absence of the experimental coordinates of the target protein. A wide array of examples have been presented, anecdotic at instances, and, yet, with a sincere attempt to reveal generalized trends or rule of thumbs wherever applicable.

**Acknowledgements:** The work was supported by a grant from NIH, grant number 1R01GM125639.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 December 2018 doi:10.20944/preprints201812.0273.v1

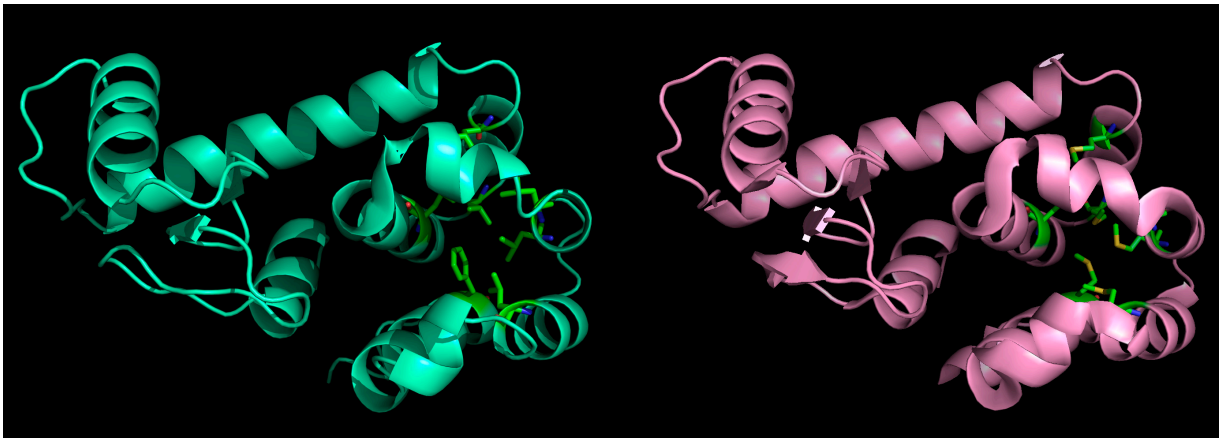
## Figures

**Figure 1. Plausible effect of mutations in membrane proteins:** Helical kink is introduced due to mutations of two successive are residues to glycine (helix breaker). The mutant generated computationally - due to the lack of any such experimental structures: as would be extremely difficult to crystallize !



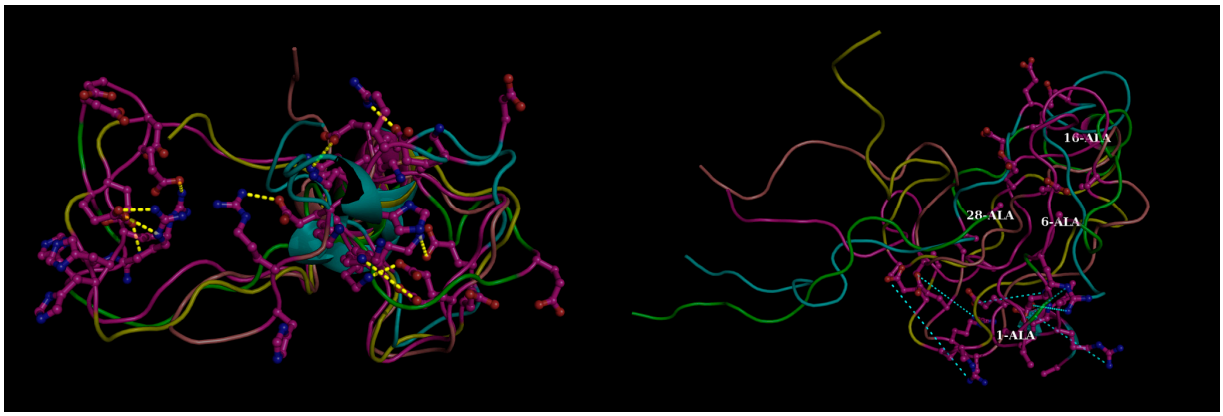
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 December 2018 doi:10.20944/preprints201812.0273.v1

**Figure 2**. **Effect of mutations in globular proteins:** 7 core (hydrophobic) residues mutated to methionine (left panel: native, right: mutant) in phage T4-lysozyme and yet the fold is preserved without almost any marked distortions. This happens because of 'structural relaxation' in proteins due to their inherent conformational plasticity (adaptability to changes).



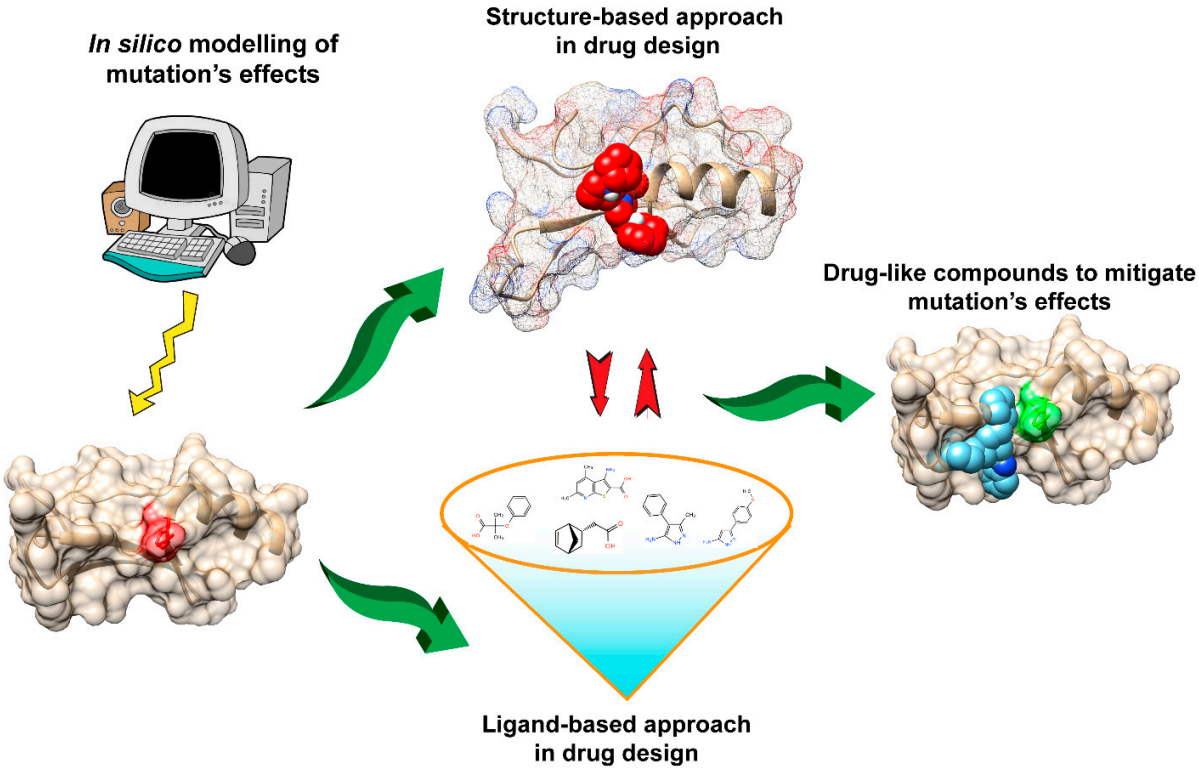
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 December 2018 doi:10.20944/preprints201812.0273.v1

**Figure 3. Effect of mutations in disordered proteins.** 4 transient (flitting) salt bridge forming charged residues (1-Asp, 6-His, 16-Lys, 28-Lys) mutated to Alanine in beta amyloid (Aβ42) resulting in the dismantling in salt-bridges globally throughout the structural ensemble (Left Panel: Mutant compared to the Right: Native). The yellow dashed lines in the left panel (native) shows the salt-bridges found individually in the five randomly chosen conformers (within 4 Å) while the same connections are shown by thinner cyan dashed lines in the right panel to portray the absence of these ionic interactions (far far greater than 4 Å). Molecular Dynamic simulation trajectories collected from Basu and Biswas, BBA Proteomics, 2018, 1866, 624-641 (Ref. 19) and figure reconstructed in Pymol.



*Preprints* (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 December 2018 doi:10.20944/preprints201812.0273.v1

**Figure 4.** Schematic presentation of drug discovery process to mitigate the effects of disease-causing mutations.



 doi:10.20944/preprints201812.0273.v1

**References**


1. Shoichet, B. K.; Baase, W. A.; Kuroki, R.; Matthews, B. W. A relationship between protein stability and protein function. *Proceedings of the National Academy of Sciences* **1995,** 92, 452-456.
2. Pepys, M. B.; Hawkins, P. N.; Booth, D. R.; Vigushin, D. M.; Tennent, G. A.; Soutar, A. K.; Totty, N.; Nguyen, O.; Blake, C. C. F.; Terry, C. J.; Feest, T. G.; Zalin, A. M.; Hsuan, J. J. Human lysozyme gene mutations cause hereditary systemic amyloidosis. *Nature* **1993,** 362, 553-557.
3. Hartley, R. W. Directed mutagenesis and barnase-barstar recognition. *Biochemistry* **1993,** 32, 5978-5984.
4. Buckle, A. M.; Schreiber, G.; Fersht, A. R. Protein-Protein Recognition - Crystal Structural-Analysis of a Barnase Barstar Complex at 2.0-Angstrom Resolution. *Biochemistry* **1994,** 33, 8878-8889.
5. Schreiber, G.; Fersht, A. R. Energetics of protein-protein interactions: analysis of the barnase-barstar interface by single mutations and double mutant cycles. *J Mol Biol* **1995,** 248, 478-86.
6. Wang, T.; Tomic, S.; Gabdoulline, R. R.; Wade, R. C. How optimal are the binding energetics of barnase and barstar? *Biophysical Journal* **2004,** 87, 1618-1630.
7. Spang, H. C. L.; Braathen, R.; Bogen, B. Heterodimeric Barnase-Barstar Vaccine Molecules: Influence of One versus Two Targeting Units Specific for Antigen Presenting Cells. *Plos One* **2012,** 7.
8. Richards, F. M. The interpretation of protein structures: Total volume, group volume distributions and packing density. *Journal of Molecular Biology* **1974,** 82, 1-14.
9. Dill, K. A. Dominant forces in protein folding. *Biochemistry* **1990,** 29, 7133-7155.
10. Basu, S.; Bhattacharyya, D.; Banerjee, R. Mapping the distribution of packing topologies within protein interiors shows predominant preference for specific packing motifs. *BMC bioinformatics* **2011,** 12, 195.
11. Javadpour, M. M.; Eilers, M.; Groesbeek, M.; Smith, S. O. Helix packing in polytopic membrane proteins: Role of glycine in transmembrane helix association. *Biophysical Journal* **1999,** 77, 1609-1618.
12. Eilers, M.; Shekar, S. C.; Shieh, T.; Smith, S. O.; Fleming, P. J. Internal packing of helical membrane proteins. *Proceedings of the National Academy of Sciences of the United States of America* **2000,** 97, 5796-5801.
13. Banerjee, R.; Sen, M.; Bhattacharya, D.; Saha, P. The jigsaw puzzle model: search for conformational specificity in protein interiors. *Journal of Molecular Biology* **2003,** 333, 211-226.
14. Charneski, C. A.; Hurst, L. D. Positive Charge Loading at Protein Termini Is Due to Membrane Protein Topology, Not a Translational Ramp. *Molecular Biology and Evolution* **2014,** 31, 70-84.
15. Harley, C. A.; Tipper, D. J. The role of charged residues in determining transmembrane protein insertion orientation in yeast. *Journal of Biological Chemistry* **1996,** 271, 24625-24633.
16. Uversky, V. N. Unusual biophysics of intrinsically disordered proteins. *Biochimica Et Biophysica Acta* **2013,** 1834, 932-951.

17. Skach, W. R. Cellular mechanisms of membrane protein folding. *Nature Structural & Molecular Biology* **2009,** 16, 606-612.
18. Nakamura, H. Roles of electrostatic interaction in proteins. *Quarterly Reviews of Biophysics* **1996,** 29, 1-90.
19. Basu, S.; Biswas, P. Salt-bridge dynamics in intrinsically disordered proteins: A trade-off between electrostatic interactions and structural flexibility. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2018,** 1866, 624-641.
20. Coskuner-Weber, O.; Uversky, V. N. Insights into the Molecular Mechanisms of Alzheimer's and Parkinson's Diseases with Molecular Simulations: Understanding the Roles of Artificial and Pathological Missense Mutations in Intrinsically Disordered Proteins Related to Pathology. *International Journal of Molecular Sciences* **2018,** 19.
21. Gassner, N. C.; Baase, W. A.; Matthews, B. W. A test of the "jigsaw puzzle" model for protein folding by multiple methionine substitutions within the core of T4 lysozyme. *Proceedings of the National Academy of Sciences of the United States of America* **1996,** 93, 12155-12158.
22. Basu, S.; Bhattacharyya, D.; Banerjee, R. Applications of complementarity plot in error detection and structure validation of proteins. *Indian Journal of Biochemistry & Biophysics* **2014,** 51, 188-200.
23. Liang, J.; Naveed, H.; Jimenez-Morales, D.; Adamian, L.; Lin, M. S. Computational studies of membrane proteins: Models and predictions for biological understanding. *Biochimica Et Biophysica Acta-Biomembranes* **2012,** 1818, 927-941.
24. Taylor, M. S.; Fung, H. K.; Rajgaria, R.; Filizola, M.; Weinstein, H.; Floudas, C. A. Mutations affecting the oligomerization interface of G-protein-coupled receptors revealed by a novel de novo protein design framework. *Biophysical Journal* **2008,** 94, 2470-2481.
25. Zhou, Y. F.; Bowie, J. U. Building a thermostable membrane protein. *Journal of Biological Chemistry* **2000,** 275, 6975-6979.
26. Schmidt, T.; Situ, A. J.; Ulmer, T. S. Structural and thermodynamic basis of proline-induced transmembrane complex stabilization. *Scientific Reports* **2016,** 6.
27. Baruah, A.; Biswas, P. Globular–disorder transition in proteins: a compromise between hydrophobic and electrostatic interactions? **2016,** 18, 23207-23214.
28. Zhu, H.; Sepulveda, E.; Hartmann, M. D.; Kogenaru, M.; Ursinus, A.; Sulz, E.; Albrecht, R.; Coles, M.; Martin, J.; Lupas, A. N. Origin of a folded repeat protein from an intrinsically disordered ancestor. *eLife* 5.
29. Eriksson, A. E.; Baase, W. A.; Zhang, X. J.; Heinz, D. W.; Blaber, M.; Baldwin, E. P.; Matthews, B. W. Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. *Science (New York, N.Y.)* **1992,** 255, 178-183.
30. Buckle, A. M.; Cramer, P.; Fersht, A. R. Structural and Energetic Responses to Cavity-Creating Mutations in Hydrophobic Cores: Observation of a Buried Water Molecule and the Hydrophilic Nature of Such Hydrophobic Cavities. *Biochemistry* **1996,** 35, 4298-4305.
31. Axe, D. D.; Foster, N. W.; Fersht, A. R. Active barnase variants with completely random hydrophobic cores. *Proceedings of the National Academy of Sciences of the United States of America* **1996,** 93, 5590-5594.
32. Dahiyat, B. I.; Sarisky, C. A.; Mayo, S. L. De novo protein design: towards fully automated sequence selection. *Journal of Molecular Biology* **1997,** 273, 789-796.

33. Goraj, K.; Renard, A.; Martial, J. A. Synthesis, purification and initial structural characterization of octarellin, a de novo polypeptide modelled on the alpha/beta-barrel proteins. *Protein Engineering* **1990,** 3, 259-266.
34. Offredi, F.; Dubail, F.; Kischel, P.; Sarinski, K.; Stern, A. S.; Van de Weerdt, C.; Hoch, J. C.; Prosperi, C.; François, J. M.; Mayo, S. L.; Martial, J. A. De novo backbone and sequence design of an idealized alpha/beta-barrel protein: evidence of stable tertiary structure. *Journal of Molecular Biology* **2003,** 325, 163-174.
35. Teng, S.; Madej, T.; Panchenko, A.; Alexov, E. Modeling effects of human single nucleotide polymorphisms on protein-protein interactions. *Biophys J* **2009,** 96, 2178-88.
36. Proteus: A Random Forest Classifier to Predict disorder-to-order Transitioning Binding Regions in Intrinsically Disordered Proteins. | bioRxiv.
37. Theillet, F.-X.; Kalmar, L.; Tompa, P.; Han, K.-H.; Selenko, P.; Dunker, A. K.; Daughdrill, G. W.; Uversky, V. N. The alphabet of intrinsic disorder. *Intrinsically Disordered Proteins* **2013,** 1, e24360.
38. Linding, R.; Schymkowitz, J.; Rousseau, F.; Diella, F.; Serrano, L. A comparative study of the relationship between protein structure and beta-aggregation in globular and intrinsically disordered proteins. *Journal of Molecular Biology* **2004,** 342, 345-353.
39. Yoneda, J. S.; Miles, A. J.; Araujo, A. P. U.; Wallace, B. A. Differential dehydration effects on globular proteins and intrinsically disordered proteins during film formation. *Protein Science* **2017,** 26, 718-726.
40. Teilum, K.; Olsen, J. G.; Kragelund, B. B. Globular and disordered-the non-identical twins in protein-protein interactions. *Front Mol Biosci* **2015,** 2, 40.
41. Vacic, V.; Iakoucheva, L. M. Disease mutations in disordered regions-exception to the rule? *Molecular Biosystems* **2012,** 8, 27-32.
42. p53 polymorphisms and mutation in human lung cancer. *Toxicologic Pathology* **2004,** 32, 756-757.
43. Joerger, A. C.; Fersht, A. R. Structural biology of the tumor suppressor p53 and cancer-associated mutants. *Advances in Cancer Research, Vol 97* **2007,** 97, 1-+.
44. Bullock, A. N.; Henckel, J.; DeDecker, B. S.; Johnson, C. M.; Nikolova, P. V.; Proctor, M. R.; Lane, D. P.; Fersht, A. R. Thermodynamic stability of wild-type and mutant p53 core domain. *Proceedings of the National Academy of Sciences of the United States of America* **1997,** 94, 14338-14342.
45. Fields, P. A. Review: Protein function at thermal extremes: balancing stability and flexibility. *Comparative Biochemistry and Physiology. Part A, Molecular & Integrative Physiology* **2001,** 129, 417-431.
46. Viña, J. Biochemical adaptation: Mechanism and process in physiological evolution. *Biochemistry and Molecular Biology Education* 30, 215-216.
47. Tokuriki, N.; Stricher, F.; Serrano, L.; Tawfik, D. S. How Protein Stability and New Functions Trade Off. *PLOS Comput Biol* **2008,** 4, e1000002.
48. Miller, S. R. An appraisal of the enzyme stability-activity trade-off. *Evolution* 71, 1876-1887.
49. Boonyuen, U.; Chamchoy, K.; Swangsri, T.; Junkree, T.; Day, N. P. J.; White, N. J.; Imwong, M. A trade off between catalytic activity and protein stability determines the clinical

manifestations of glucose-6-phosphate dehydrogenase (G6PD) deficiency. *International Journal of Biological Macromolecules* **2017,** 104, 145-156.
50. Takano, K.; Liu, D.; Tarpey, P.; Gallant, E.; Lam, A.; Witham, S.; Alexov, E.; Chaubey, A.; Stevenson, R. E.; Schwartz, C. E.; Board, P. G.; Dulhunty, A. F. An X-linked channelopathy with cardiomegaly due to a CLIC2 mutation enhancing ryanodine receptor channel activity. *Human Molecular Genetics* **2012,** 21, 4497-4507.
51. Witham, S.; Takano, K.; Schwartz, C.; Alexov, E. A missense mutation in CLIC2 associated with intellectual disability is predicted by in silico modeling to affect protein stability and dynamics. *Proteins-Structure Function and Bioinformatics* **2011,** 79, 2444-2454.
52. Siddiqui, K. S. Defying the activity-stability trade-off in enzymes: taking advantage of entropy to enhance activity and thermostability. *Critical Reviews in Biotechnology* **2017,** 37, 309-322.
53. Chakshusmathi, G.; Mondal, K.; Lakshmi, G. S.; Singh, G.; Roy, A.; Ch, R. B.; Madhusudhanan, S.; Varadarajan, R. Design of temperature-sensitive mutants solely from amino acid sequence. *Proceedings of the National Academy of Sciences* **2004,** 101, 7925-7930.
54. Tan, G.; Chen, M.; Foote, C.; Tan, C. Temperature-Sensitive Mutations Made Easy: Generating Conditional Mutations by Using Temperature-Sensitive Inteins That Function Within Different Temperature Ranges. *Genetics* **2009,** 183, 13-22.
55. Tan, K. P.; Khare, S.; Varadarajan, R.; Madhusudhan, M. S. TSpred: a web server for the rational design of temperature-sensitive mutants. *Nucleic Acids Research* **2014,** 42, W277-W284.
56. Poultney, C. S.; Butterfoss, G. L.; Gutwein, M. R.; Drew, K.; Gresham, D.; Gunsalus, K. C.; Shasha, D. E.; Bonneau, R. Rational Design of Temperature-Sensitive Alleles Using Computational Structure Prediction. *PLOS ONE* **2011,** 6, e23947.
57. Lulla, V.; Sawicki, D. L.; Sawicki, S. G.; Lulla, A.; Merits, A.; Ahola, T. Molecular defects caused by temperature-sensitive mutations in Semliki Forest virus nsP1. *Journal of Virology* **2008,** 82, 9236-9244.
58. Tian, J.; Yam, C.; Balasundaram, G.; Wang, H.; Gore, A.; Sampath, K. A temperature-sensitive mutation in the nodal-related gene cyclops reveals that the floor plate is induced during gastrulation in zebrafish. *Development* **2003,** 130, 3331-3342.
59. Li, Z. J.; Vizeacoumar, F. J.; Bahr, S.; Li, J. J.; Warringer, J.; Vizeacoumar, F. S.; Min, R. Q.; VanderSluis, B.; Bellay, J.; DeVit, M.; Fleming, J. A.; Stephens, A.; Haase, J.; Lin, Z. Y.; Baryshnikova, A.; Lu, H.; Yan, Z.; Jin, K.; Barker, S.; Datti, A.; Giaever, G.; Nislow, C.; Bulawa, C.; Myers, C. L.; Costanzo, M.; Gingras, A. C.; Zhang, Z. L.; Blomberg, A.; Bloom, K.; Andrews, B.; Boone, C. Systematic exploration of essential yeast gene function with temperature-sensitive mutants. *Nature Biotechnology* **2011,** 29, 361-U105.
60. Lockwood, S.; Krishnamoorthy, B.; Ye, P. Neighborhood Properties Are Important Determinants of Temperature Sensitive Mutations. *Plos One* **2011,** 6.
61. Kucukkal, T. G.; Petukh, M.; Li, L.; Alexov, E. Structural and physico-chemical effects of disease and non-disease nsSNPs on proteins. *Current Opinion in Structural Biology* **2015,** 32, 18-24.
62. Peng, Y.; Alexov, E. Investigating the linkage between disease-causing amino acid variants and their effect on protein stability and binding. *Proteins* **2016,** 84, 232-9.
63. Petukh, M.; Kucukkal, T. G.; Alexov, E. On human disease-causing amino acid variants: statistical study of sequence and structural patterns. *Human Mutation* **2015,** 36, 524-534.

64. Monticone, S.; Bandulik, S.; Stindl, J.; Zilbermint, M.; Dedov, I.; Mulatero, P.; Allgaeuer, M.; Lee, C.-C. R.; Stratakis, C. A.; Williams, T. A.; Tiulpakov, A. A case of severe hyperaldosteronism caused by a de novo mutation affecting a critical salt bridge Kir3.4 residue. *The Journal of Clinical Endocrinology and Metabolism* **2015,** 100, E114-118.
65. Nishi, H.; Tyagi, M.; Teng, S.; Shoemaker, B. A.; Hashimoto, K.; Alexov, E.; Wuchty, S.; Panchenko, A. R. Cancer missense mutations alter binding properties of proteins and their interaction networks. *PLoS One* **2013,** 8, e66273.
66. Li, M.; Kales, S. C.; Ma, K.; Shoemaker, B. A.; Crespo-Barreto, J.; Cangelosi, A. L.; Lipkowitz, S.; Panchenko, A. R. Balancing Protein Stability and Activity in Cancer: A New Approach for Identifying Driver Mutations Affecting CBL Ubiquitin Ligase Activation. *Cancer Res* **2016,** 76, 561-71.
67. Capriotti, E.; Fariselli, P.; Casadio, R. A neural-network-based method for predicting protein stability changes upon single point mutations. *Bioinformatics* **2004,** 20 Suppl 1, i63-8.
68. Capriotti, E.; Fariselli, P.; Casadio, R. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Res* **2005,** 33, W306-10.
69. Pires, D. E.; Ascher, D. B.; Blundell, T. L. DUET: a server for predicting effects of mutations on protein stability using an integrated computational approach. *Nucleic Acids Res* **2014,** 42, W314-9.
70. Pires, D. E.; Ascher, D. B.; Blundell, T. L. mCSM: predicting the effects of mutations in proteins using graph-based signatures. *Bioinformatics* **2014,** 30, 335-42.
71. Worth, C. L.; Preissner, R.; Blundell, T. L. SDM--a server for predicting effects of mutations on protein stability and malfunction. *Nucleic Acids Res* **2011,** 39, W215-22.
72. Blanco, J. D.; Radusky, L.; Climente-Gonzalez, H.; Serrano, L. FoldX accurate structural protein-DNA binding prediction using PADA1 (Protein Assisted DNA Assembly 1). *Nucleic Acids Res* **2018,** 46, 3852-3863.
73. Schymkowitz, J.; Borg, J.; Stricher, F.; Nys, R.; Rousseau, F.; Serrano, L. The FoldX web server: an online force field. *Nucleic Acids Res* **2005,** 33, W382-8.
74. Zhang, Z.; Wang, L.; Gao, Y.; Zhang, J.; Zhenirovskyy, M.; Alexov, E. Predicting folding free energy changes upon single point mutations. *Bioinformatics* **2012,** 28, 664-71.
75. Getov, I.; Petukh, M.; Alexov, E. SAAFEC: Predicting the Effect of Single Point Mutations on Protein Folding Free Energy Using a Knowledge-Modified MM/PBSA Approach. *Int J Mol Sci* **2016,** 17, 512.
76. Li, M.; Petukh, M.; Alexov, E.; Panchenko, A. R. Predicting the Impact of Missense Mutations on Protein-Protein Binding Affinity. *J Chem Theory Comput* **2014,** 10, 1770-1780.
77. Petukh, M.; Dai, L.; Alexov, E. SAAMBE: Webserver to Predict the Charge of Binding Free Energy Caused by Amino Acids Mutations. *Int J Mol Sci* **2016,** 17, 547.
78. Cang, Z.; Wei, G. W. Analysis and prediction of protein folding energy changes upon mutation by element specific persistent homology. *Bioinformatics* **2017,** 33, 3549-3557.
79. Cang, Z.; Wei, G. W. TopologyNet: Topology based deep convolutional and multi-task neural networks for biomolecular property predictions. *PLoS Comput Biol* **2017,** 13, e1005690.
80. Lodish, H.; Berk, A.; Zipursky, S. L.; Matsudaira, P.; Baltimore, D.; Darnell, J. Mutations: Types and Causes. *Molecular Cell Biology. 4th edition* **2000**.

81. Gao, M.; Zhou, H.; Skolnick, J. Insights into disease-associated mutations in the human proteome through protein structural analysis. *Structure (London, England : 1993)* **2015,** 23, 1362-1369.
82. Casadio, R.; Vassura, M.; Tiwari, S.; Fariselli, P.; Luigi Martelli, P. Correlating disease-related mutations to their effect on protein stability: a large-scale analysis of the human proteome. *Human Mutation* **2011,** 32, 1161-1170.
83. Adzhubei, I. A.; Schmidt, S.; Peshkin, L.; Ramensky, V. E.; Gerasimova, A.; Bork, P.; Kondrashov, A. S.; Sunyaev, S. R. A method and server for predicting damaging missense mutations. *Nature methods* **2010,** 7, 248-249.
84. Lounnas, V.; Ritschel, T.; Kelder, J.; McGuire, R.; Bywater, R. P.; Foloppe, N. Current progress in Structure-Based Rational Drug Design marks a new mindset in drug discovery. *Comput Struct Biotechnol J* **2013,** 5, e201302011.
85. Hughes, J. P.; Rees, S.; Kalindjian, S. B.; Philpott, K. L. Principles of early drug discovery. *Br J Pharmacol* **2011,** 162, 1239-49.
86. Michel, J. Current and emerging opportunities for molecular simulations in structure-based drug design. *Phys Chem Chem Phys* **2014,** 16, 4465-77.
87. Hung, C. L.; Chen, C. C. Computational approaches for drug discovery. *Drug Dev Res* **2014,** 75, 412-8.
88. Sawicki, M. P.; Samara, G.; Hurwitz, M.; Passaro, E. Human Genome Project. *The American Journal of Surgery* **1993,** 165, 258-264.
89. Genomes Project, C.; Abecasis, G. R.; Altshuler, D.; Auton, A.; Brooks, L. D.; Durbin, R. M.; Gibbs, R. A.; Hurles, M. E.; McVean, G. A. A map of human genome variation from population-scale sequencing. *Nature* **2010,** 467, 1061-73.
90. Peng, Y.; Norris, J.; Schwartz, C.; Alexov, E. Revealing the Effects of Missense Mutations Causing Snyder-Robinson Syndrome on the Stability and Dimerization of Spermine Synthase. *Int J Mol Sci* **2016,** 17.
91. Li, L.; Jia, Z.; Peng, Y.; Godar, S.; Getov, I.; Teng, S.; Alper, J.; Alexov, E. Forces and Disease: Electrostatic force differences caused by mutations in kinesin motor domains can distinguish between disease-causing and non-disease-causing mutations. *Sci Rep* **2017,** 7, 8237.
92. Spellicy, C. J.; Norris, J.; Bend, R.; Bupp, C.; Mester, P.; Reynolds, T.; Dean, J.; Peng, Y.; Alexov, E.; Schwartz, C. E.; Stevenson, R. S.; M, J. F. Key apoptotic genes APAF1 and CASP9 implicated in recurrent folate-resistant neural tube defects. *Eur J Hum Genet* **2018,** 26, 420-427.
93. Vaidyanathan, K.; Niranjan, T.; Selvan, N.; Teo, C. F.; May, M.; Patel, S.; Weatherly, B.; Skinner, C.; Opitz, J.; Carey, J.; Viskochil, D.; Gecz, J.; Shaw, M.; Peng, Y.; Alexov, E.; Wang, T.; Schwartz, C.; Wells, L. Identification and characterization of a missense mutation in the O-linked beta-N-acetylglucosamine (O-GlcNAc) transferase gene that segregates with X-linked intellectual disability. *J Biol Chem* **2017,** 292, 8948-8963.
94. Chen, W.-T.; Hong, C.-J.; Lin, Y.-T.; Chang, W.-H.; Huang, H.-T.; Liao, J.-Y.; Chang, Y.-J.; Hsieh, Y.-F.; Cheng, C.-Y.; Liu, H.-C.; Chen, Y.-R.; Cheng, I. H. Amyloid-beta (Aβ) D7H mutation increases oligomeric Aβ42 and alters properties of Aβ-zinc/copper assemblies. *PloS One* **2012,** 7, e35807.
95. Alexov, E. Advances in Human Biology: Combining Genetics and Molecular Biophysics to Pave the Way for Personalized Diagnostics and Medicine. *Advances in Biology* **2014,** 2014, 1-16.

96. Yang, Y.; Kucukkal, T. G.; Li, J.; Alexov, E.; Cao, W. Binding Analysis of Methyl-CpG Binding Domain of MeCP2 and Rett Syndrome Mutations. *ACS Chem Biol* **2016,** 11, 2706-2715.
97. Peng, Y.; Myers, R.; Zhang, W.; Alexov, E. Computational Investigation of the Missense Mutations in DHCR7 Gene Associated with Smith-Lemli-Opitz Syndrome. *International Journal of Molecular Sciences* **2018,** 19.
98. Peng, Y.; Suryadi, J.; Yang, Y.; Kucukkal, T. G.; Cao, W.; Alexov, E. Mutations in the KDM5C ARID Domain and Their Plausible Association with Syndromic Claes-Jensen-Type Disease. *Int J Mol Sci* **2015,** 16, 27270-87.
99. Ferreira, L. G.; Dos Santos, R. N.; Oliva, G.; Andricopulo, A. D. Molecular docking and structure-based drug design strategies. *Molecules* **2015,** 20, 13384-421.
100. Bleicher, K. H.; Bohm, H. J.; Muller, K.; Alanine, A. I. Hit and lead generation: beyond high-throughput screening. *Nat Rev Drug Discov* **2003,** 2, 369-78.
101. Lang, P. T.; Brozell, S. R.; Mukherjee, S.; Pettersen, E. F.; Meng, E. C.; Thomas, V.; Rizzo, R. C.; Case, D. A.; James, T. L.; Kuntz, I. D. DOCK 6: combining techniques to model RNA-small molecule complexes. *RNA* **2009,** 15, 1219-30.
102. Trott, O.; Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* **2010,** 31, 455-61.
103. Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem* **2004,** 47, 1739-49.
104. Jain, A. N. Surflex: fully automatic flexible molecular docking using a molecular similarity-based search engine. *J Med Chem* **2003,** 46, 499-511.
105. Nair, P. C.; Miners, J. O. Molecular dynamics simulations: from structure function relationships to drug discovery. *In Silico Pharmacol* **2014,** 2, 4.
106. Vogelstein, B.; Lane, D.; Levine, A. J. Surfing the p53 network. *Nature* **2000,** 408, 307-10.
107. Hollstein, M.; Sidransky, D.; Vogelstein, B.; Harris, C. C. P53 Mutations in Human Cancers. *Science* **1991,** 253, 49-53.
108. Muller, P. A.; Vousden, K. H. Mutant p53 in cancer: new functions and therapeutic opportunities. *Cancer Cell* **2014,** 25, 304-17.
109. Bullock, A. N.; Fersht, A. R. Rescuing the function of mutant p53. *Nat Rev Cancer* **2001,** 1, 68-76.
110. Wassman, C. D.; Baronio, R.; Demir, O.; Wallentine, B. D.; Chen, C. K.; Hall, L. V.; Salehi, F.; Lin, D. W.; Chung, B. P.; Hatfield, G. W.; Richard Chamberlin, A.; Luecke, H.; Lathrop, R. H.; Kaiser, P.; Amaro, R. E. Computational identification of a transiently open L1/S3 pocket for reactivation of mutant p53. *Nat Commun* **2013,** 4, 1407.
111. Kaar, J. L.; Basse, N.; Joerger, A. C.; Stephens, E.; Rutherford, T. J.; Fersht, A. R. Stabilization of mutant p53 via alkylation of cysteines and effects on DNA binding. *Protein Sci* **2010,** 19, 2267-78.
112. Pegg, A. E.; Michael, A. J. Spermine synthase. *Cell Mol Life Sci* **2010,** 67, 113-21.

 doi:10.20944/preprints201812.0273.v1

113. Zhang, Z.; Martiny, V.; Lagorce, D.; Ikeguchi, Y.; Alexov, E.; Miteva, M. A. Rational design of small-molecule stabilizers of spermine synthase dimer by virtual screening and free energy-based approach. *PLoS One* **2014,** 9, e110884.
114. Zhang, Z.; Witham, S.; Petukh, M.; Moroy, G.; Miteva, M.; Ikeguchi, Y.; Alexov, E. A rational free energy-based approach to understanding and targeting disease-causing missense mutations. *J Am Med Inform Assoc* **2013,** 20, 643-51.
115. Dror, O.; Schneidman-Duhovny, D.; Inbar, Y.; Nussinov, R.; Wolfson, H. J. Novel approach for efficient pharmacophore-based virtual screening: method and applications. *J Chem Inf Model* **2009,** 49, 2333-43.
116. Kaserer, T.; Beck, K. R.; Akram, M.; Odermatt, A.; Schuster, D. Pharmacophore Models and Pharmacophore-Based Virtual Screening: Concepts and Applications Exemplified on Hydroxysteroid Dehydrogenases. *Molecules* **2015,** 20, 22799-832.
117. Lee, C. H.; Huang, H. C.; Juan, H. F. Reviewing ligand-based rational drug design: the search for an ATP synthase inhibitor. *Int J Mol Sci* **2011,** 12, 5304-18.
118. Chen, Z.; Li, H. L.; Zhang, Q. J.; Bao, X. G.; Yu, K. Q.; Luo, X. M.; Zhu, W. L.; Jiang, H. L. Pharmacophore-based virtual screening versus docking-based virtual screening: a benchmark comparison against eight targets. *Acta Pharmacol Sin* **2009,** 30, 1694-708.
119. Singh, P. K.; Silakari, O. Molecular dynamics guided development of indole based dual inhibitors of EGFR (T790M) and c-MET. *Bioorg Chem* **2018,** 79, 163-170.
120. Springsteel, M. F.; Galietta, L. J. V.; Ma, T.; By, K.; Berger, G. O.; Yang, H.; Dicus, C. W.; Choung, W.; Quan, C.; Shelat, A. A.; Guy, R. K.; Verkman, A. S.; Kurth, M. J.; Nantz, M. H. Benzoflavone activators of the cystic fibrosis transmembrane conductance regulator: towards a pharmacophore model for the nucleotide-binding domain. *Bioorganic & Medicinal Chemistry* **2003,** 11, 4113-4120.
121. Pathak, D.; Chadha, N.; Silakari, O. Identification of non-resistant ROS-1 inhibitors using structure based pharmacophore analysis. *J Mol Graph Model* **2016,** 70, 85-93.
122. Wang, A.; Li, X.; Wu, H.; Zou, F.; Yan, X. E.; Chen, C.; Hu, C.; Yu, K.; Wang, W.; Zhao, P.; Wu, J.; Qi, Z.; Wang, W.; Wang, B.; Wang, L.; Ren, T.; Zhang, S.; Yun, C. H.; Liu, J.; Liu, Q. Discovery of (R)-1-(3-(4-Amino-3-(3-chloro-4-(pyridin-2-ylmethoxy)phenyl)-1H-pyrazolo[3,4-d]py rimidin-1-yl)piperidin-1-yl)prop-2-en-1-one (CHMFL-EGFR-202) as a Novel Irreversible EGFR Mutant Kinase Inhibitor with a Distinct Binding Mode. *J Med Chem* **2017,** 60, 2944-2962.
123. Goldstraw, P.; Ball, D.; Jett, J. R.; Le Chevalier, T.; Lim, E.; Nicholson, A. G.; Shepherd, F. A. Non-small-cell lung cancer. *The Lancet* **2011,** 378, 1727-1740.
124. Awad, M. M.; Katayama, R.; McTigue, M.; Liu, W.; Deng, Y. L.; Brooun, A.; Friboulet, L.; Huang, D.; Falk, M. D.; Timofeevski, S.; Wilner, K. D.; Lockerman, E. L.; Khan, T. M.; Mahmood, S.; Gainor, J. F.; Digumarthy, S. R.; Stone, J. R.; Mino-Kenudson, M.; Christensen, J. G.; Iafrate, A. J.; Engelman, J. A.; Shaw, A. T. Acquired resistance to crizotinib from a mutation in CD74-ROS1. *N Engl J Med* **2013,** 368, 2395-401.
125. Acharya, C.; Coop, A.; E. Polli, J.; D. MacKerell, A. Recent Advances in Ligand-Based Drug Design: Relevance and Utility of the Conformationally Sampled Pharmacophore Approach. *Current Computer Aided-Drug Design* **2011,** 7, 10-22.

 doi:10.20944/preprints201812.0273.v1

126. Kerem, B.; Rommens, J.; Buchanan, J.; Markiewicz, D.; Cox, T.; Chakravarti, A.; Buchwald, M.; Tsui, L. Identification of the cystic fibrosis gene: genetic analysis. *Science* **1989,** 245, 1073-1080.
127. Noy, E.; Senderowitz, H. Combating cystic fibrosis: in search for CF transmembrane conductance regulator (CFTR) modulators. *ChemMedChem* **2011,** 6, 243-51.
128. Liessi, N.; Cichero, E.; Pesce, E.; Arkel, M.; Salis, A.; Tomati, V.; Paccagnella, M.; Damonte, G.; Tasso, B.; Galietta, L. J. V.; Pedemonte, N.; Fossa, P.; Millo, E. Synthesis and biological evaluation of novel thiazole- VX-809 hybrid derivatives as F508del correctors by QSAR-based filtering tools. *Eur J Med Chem* **2018,** 144, 179-200.
129. Wilson, G. L.; Lill, M. A. Integrating structure-based and ligand-based approaches for computational drug design. *Future Med Chem* **2011,** 3, 735-50.
130. Drwal, M. N.; Griffith, R. Combination of ligand- and structure-based methods in virtual screening. *Drug Discov Today Technol* **2013,** 10, e395-401.