Journal of Mathematical Chemistry
https://doi.org/10.1007/s10910-023-01545-w

ORIGINAL PAPER

In silico induction of missense mutation in NNRTI protein: computational modelling and stability study of modelled proteins

Laxmi Sule[1] · Swagata Gupta[2] · Nilanjana Jain[3] · Nitin S. Sapre[4]

Received: 21 June 2023 / Accepted: 26 October 2023

Abstract

The paper presents in silico mutational research on the energetics of the most resolved protein structure of HIV-1 NNRTIs, the 4G1Q HIV-1 reverse transcriptase protein. Twenty nearby residues that are less than 6 Å from the center of the embedded ligand are subjected to in silico alterations. For the current study, 380 unique proteins are generated in silico using a set of 380 surrounding residues that have been altered. Analyses and comparison with the parent protein have been done to determine the impact of mutation on the change in folding-unfolding free energy (G), protein stability, and solvation energy. To evaluate the impact of mutation (i) by and (ii) on a particular amino acid residue, a two-fold investigation is conducted. The findings imply that in 12 designed proteins (G -3.0), folding-unfolding is significantly favored, resulting in the creation of a highly stable conformation. The 11 designed proteins are unstable because the positive values of $G > 0.5$ in these proteins point to unfavorable mutations. However, the G value in 171 designed proteins is -1.0, which suggests that mutations cause designed proteins to adopt a stable conformation. According to the findings, out of the 380 created proteins, 11 had extremely unfavorable, 69 less-favorable, and 270 favorable folding-unfolding changes.

Keywords Modelled Proteins · In Silico · Mutation · NNRTIs · AIDS

1 Introduction

Proteins are engaged in highly selective interactions in micro to macro living systems. Variation (Mutation) in the sequence causes significant perturbations or complete abolishment of function, potentially leading to diseases. There is an important need to understand the impacts of variation in the protein structure. The stability of

The work was presented at the 8th Indo-US Workshop on Mathematical Chemistry.

Extended author information available on the last page of the article

Published online: 28 November 2023

proteins plays an important role in characterizing their functions, activity and regulation [1].

One of the possible ways to assess the effect of a mutation on protein binding affinity/stability is to experimentally measure it. However, these methods can be time-consuming and costly. With the advancements and amalgamation of computing technology with chemistry, physics, and biology, it has become convenient to estimate the impact of mutations on protein stability/energy theoretically with near accuracy to the experimental results [2].

The current era of genome sequencing has unraveled a large number of human genetic variations, many of which may affect protein binding and function [3].

Protein stability refers to the ability of a protein to maintain its native three-dimensional structure under a given set of conditions. The Gibbs free energy (ΔG) is a thermodynamic parameter that describes the tendency of a system to change spontaneously from one state to another. In the context of protein stability, ΔG is a measure of the free energy difference between the folded (native) and unfolded (denatured) states of the protein [4].

A negative ΔG value indicates that the protein is stable in its folded state, while a positive ΔG value indicates that the protein is unstable and has a tendency to unfold. The magnitude of ΔG reflects the strength of the interactions that stabilize the folded protein, such as hydrogen bonds, hydrophobic interactions, and electrostatic interactions [5].

Experimental techniques such as protein folding assays, circular dichroism spectroscopy, and differential scanning calorimetry can be used to measure protein stability and ΔG values under various conditions, such as changes in temperature, pH, and ionic strength. Computational methods such as molecular dynamics simulations and free energy calculations can also be used to predict protein stability and ΔG values based on the protein's structure and environmental conditions [6].

AIDS pandemic, caused by the retrovirus HIV-1, has claimed more than 30 million lives over the past four decades. Antiretroviral (ART), which is required for the whole life, has transformed the disease into a little manageable one. The CD + T lymphocyte is the main target cell through which HIV-1 enters, by binding to its receptor CD4 and to the co-receptors i.e., CC-chemokine receptor-5 (CCR5). The fusion of the viral and human cell membranes, prompted by this binding, initiates a complex intracellular life cycle, producing new viruses [7].

The stability of mutants in the context of HIV proteins, especially in relation to their binding to anti-AIDS drugs like rilpivirine, as well as their impact on pathogenicity and virulence, can vary significantly. It's important to note that HIV is known for its high mutation rate, which can lead to the emergence of drug-resistant variants and altered pathogenicity. Here is some information, along with references, on these aspects:

Rilpivirine Resistance: Rilpivirine is a non-nucleoside reverse transcriptase inhibitor (NNRTI) used in the treatment of HIV. Resistance to rilpivirine can develop due to mutations in the HIV reverse transcriptase gene, specifically in the NNRTI-binding pocket. Common Mutations: Common mutations associated with rilpivirine resistance include K103N, Y181C, and E138A. These mutations can reduce the binding affinity of the drug to the reverse transcriptase enzyme, leading to reduced drug efficacy [8]. Essential amino acids, such as tryptophan (W) such as W229 and W234, which contribute to hydrophobic interactions in the NNRTI binding pocket, are involved in rivilpivirine's binding to the reverse transcriptase enzyme. In order for Rilpivirine and the enzyme to create hydrogen bonds and engage in hydrophobic interactions, tyrosine (Y), as demonstrated by Y181, is essential. Furthermore, phenylalanine (F) residues, such as F-227, contribute to the hydrophobic pocket that Rilpivirine binds to, increasing the affinity of its binding [8].

The basic mechanism of action of Rilpivirine as an NNRTI is derived from this combination of certain amino acid interactions inside the reverse transcriptase enzyme.

Mutations in the reverse transcriptase gene can affect the binding of rilpivirine to the enzyme's active site. The loss of drug binding affinity can result in reduced inhibition of reverse transcription, allowing the virus to replicate. The specific mutations determine the degree of resistance, with some mutants showing higher resistance levels than others. The effect on drug binding can be assessed through in vitro studies and molecular modeling [9].

Mutations in HIV can influence viral pathogenicity and virulence. Some mutations may lead to changes in viral proteins that affect the virus's ability to infect and replicate within host cells. Mutations that enhance viral fitness, replication, and immune evasion can contribute to increased pathogenicity. Conversely, some mutations may reduce viral fitness and replication. Studies on the impact of specific mutations on viral pathogenicity are ongoing, and the results can vary depending on the viral strain and host factors [10]. Due to its capacity to target the immune system, particularly CD4 + T cells, which are essential for the body's defence against diseases, HIV is a very dangerous virus. These cells are the main target of the virus, which causes their depletion, impairs immunity, and increases susceptibility to a variety of opportunistic infections and cancers. The ability of HIV to elude immune response, develop persistent infection, and gradually weaken immune system activities is largely responsible for its pathogenicity [10].

HIV's virulence varies from person to person and is influenced by the virus strain, the immunological system of the individual, and the accessibility of treatment. A rapid course of the disease is caused by certain strains of HIV that are more virulent than others. Treatment becomes much more difficult since the virus can mutate quickly, resulting in the creation of drug-resistant forms [10].

HIV's capacity to incorporate its genetic material into the host's DNA is another factor contributing to its virulence; this allows the virus to create a latent reservoir of infected cells that can reawaken and release virus particles even after years of successful antiretroviral therapy. Finding a treatment for HIV is significantly hampered by this viral reservoir.

The emergence of drug-resistant mutants, including those resistant to rilpivirine, poses a clinical challenge in the management of HIV infection. Alternative antiretroviral regimens may be required for individuals with drug-resistant strains. Monitoring for drug resistance through genotypic and phenotypic testing is essential in HIV clinical care to guide treatment decisions [10].

Computational Chemistry is a multidisciplinary field that combines principles of chemistry, physics, and computer science to investigate and understand chemical phenomena using computational methods. It involves the development and application of theoretical models, algorithms, and software tools to study various aspects of molecular systems, such as their structures, properties, and reactivity. Computational chemistry is a highly sophisticated branch of chemistry that uses computer simulations and mathematical models to study chemical systems. It involves the use of theoretical methods, algorithms, and computer programs to estimate the properties and behaviour of molecules, materials, chemical reactions etc.

The use of computational methods in chemistry has revolutionized the way researchers approach the study of molecules and materials. It enables the exploration of complex chemical systems that are often difficult or even impossible to study experimentally. Computational chemistry techniques provide insights into molecular interactions, reaction mechanisms, and properties of compounds, helping researchers to design new drugs, catalysts, and materials.

Computational chemistry has many applications, including drug discovery, materials science, catalysis, and environmental chemistry. By using computational methods, the properties of molecules and materials can be predicted to near accuracy without the need for expensive and time-consuming experiments. This helps in saving time thereby faster and more efficient development of new drugs, materials, and technologies.

Computational chemistry is a broader field that encompasses a wide range of computational methods and techniques used to study chemical systems. In addition to MD simulations and protein modelling, computational chemistry also includes techniques such as quantum chemistry, molecular mechanics, and molecular docking, among others [11].

Some of the commonly used computational chemistry methods include computer aided drug design (CADD) including, molecular mechanics, quantum mechanics, density functional theory, and molecular dynamics simulations. These methods vary in their level of accuracy and computational cost and are chosen based on the specific research question and available computational resources.

Overall, computational chemistry plays an important role in advancing our understanding of chemical systems and developing new technologies that can improve our lives.

Computer-aided drug design (CADD) is a computational approach that involves the use of computer algorithms and software to assist in the drug discovery process. This approach uses various computational tools to identify potential drug candidates and optimize their properties before they are tested in the laboratory [12].

CADD has become an essential tool in drug discovery, allowing researchers to rapidly screen large numbers of compounds and optimize their properties before investing time and resources in expensive experimental studies.

Virtual screening is a computational technique used to predict the potential activity of small molecules (ligands) against a specific target protein. It involves the use of computer software to analyse large databases of molecules and predict their affinity and activity for a specific target. It can be used in drug discovery to identify potential drug candidates that can bind to the target protein and modulate its activity [13]. It is a powerful tool in drug discovery as it can significantly reduce the time and cost involved in the drug discovery process by identifying potential drug candidates with high affinity and specificity for the target protein.

Molecular Dynamics (MD) simulation is a computational technique used in computational chemistry to study the behaviour of atoms and molecules over time [14]. In an MD simulation, the system of interest is described by a set of equations of motion that define the behavior of each atom or molecule in the system. The equations of motion take into account the interactions between atoms or molecules, which are described by a potential energy function. MD simulations can be used to study a wide range of chemical and biochemical systems, including proteins, DNA, and small molecules. They can provide insights into the dynamics and thermodynamics of these systems, such as the conformational changes that occur in proteins and the binding of ligands to enzymes. The simulation proceeds by solving the equations of motion numerically, typically using a numerical integration method such as the Verlet algorithm or the leapfrog algorithm [15]. The simulation calculates the position, velocity, and acceleration of each atom or molecule at each time step, and the positions of the atoms or molecules are updated based on these calculations.

Molecular dynamics (MD) simulations are one common type of simulation used in this field. MD simulations involve the use of computational models to simulate the motion of atoms and molecules over time. In the context of protein modelling, MD simulations can be used to study the structural and dynamic properties of proteins, including their folding and unfolding processes, interactions with ligands, and conformational changes [16].

Protein modelling is the process of predicting the three-dimensional structure of a protein from its amino acid sequence. The three-dimensional structure of a protein is essential to understanding its function, interactions, and biochemical properties. There are several methods used to model protein structures, including homology modelling, ab initio modelling, and molecular dynamics simulations.

Homology modelling assumes that the amino acid sequence of a protein is similar to that of a known protein with a similar function and structure [17]. In homology modelling, the known protein structure is used as a template to predict the structure of the target protein. The accuracy of homology modelling depends on the similarity between the amino acid sequences of the target protein and the template protein.

Ab initio modelling, also known as *de novo* modelling, is a method that predicts the structure of a protein without using a template structure. Ab initio modelling is based on physical principles such as energy minimization and can be computationally expensive. This method is more challenging than homology modelling but can be used for proteins that do not have a close homolog with a known structure [18].

Protein modelling is an essential tool for understanding protein function and structure. It has applications in drug design, protein engineering, and understanding the mechanisms of protein-protein interactions. A protein could have multiple structures available, and if another structure of the same protein is used, the predicted change in stability for structure-based methods might be different. The mutation causes a change in the stability of a protein.

DUET online server is used for these computations. DUET consolidates two reciprocal approaches (mCSM and SDM) in a agreement vaticination, attained by combining the results of the separate styles in an optimized predictor using

Support Vector Machines (SVM) [16]. The system improves the overall delicacy of the prognostications in comparison with either system collectively and performs as well as or better than analogous styles. DUET is a bioinformatics web garçon created for gaining sapience into the goods of nsSNPs on protein stability. It integrates two reciprocal styles into an agreement/ optimized vaticination, as a way to work the stylish of SDM, a statistical implicit energy function that relies on negotiation tables deduced from homologous protein families which incorporates constraints on residue surroundings during elaboration, and mCSM, a machine literacy algorithm that takes into account the residue 3D physicochemical terrain epitomized as a graph- grounded structural hand [19].

Mutations can be classified into three categories (a) “Good” which increases fitness, (b) “Indifferent or Neutral”, as the effects are too small and, (c) “Bad” which decreases fitness [19].

$\Delta\Delta$G results will fall into three categories:

A. $\Delta\Delta G > 0.5$: Positive results suggest that a mutation would be destabilizing. These mutations are residues that are usually avoided during design and can be classified as “Bad”.
B. $0.5 > \Delta\Delta G > -0.5$: Things that are near 0 are within the noise range so should be considered indifferent or neutral. These can be included in the design to allow more neutral changes in the protein that may compensate for changes in the protein. These can be classified as “Neutral” or “Indifferent”.
C. $\Delta\Delta G < -0.5$: Negative results suggest that the mutation would lead to a more stable protein and can be classified as “Good”.

Protein modelling of missense mutations involves predicting the structural and functional consequences of amino acid substitutions that alter the protein sequence. Missense mutations are single-nucleotide variations that change a single amino acid residue in a protein sequence, potentially affecting protein stability, interactions, or enzymatic activity.

There are several computational tools and methods available for protein modelling of missense mutations, including homology modelling, molecular dynamics simulations, and machine learning-based approaches. These methods use various algorithms to predict the effect of a missense mutation on protein structure and function, such as changes in protein stability, folding, dynamics, and interactions [20–24].

One common approach is to compare the predicted structure and stability of the wild-type protein with that of the mutated protein. If the mutation destabilizes the protein or alters its structural integrity, it may affect the protein’s function or interactions with other molecules.

Overall, protein modelling of missense mutations can provide valuable insights into the potential effects of genetic variations on protein structure and function, which can help in understanding the molecular basis of genetic diseases and designing therapeutic interventions.

The present study is undertaken to asses the impact of *in silico* mutations on the basis of ΔΔG as a measure of stability.

2 Materials and methods

This is an attempt to study the impact of the mutation "on" and "by" specific amino acid residues. An in-silico introduction of missenses investigation has been undertaken to test the effect of mutation on the stability of the newly designed proteins.

In the present study HIV-1 NNRTI protein, namely 4G1Q [25], downloaded from protein data bank (http://www.rcsb.org), was used to perform mutation and assess and compare relative stability of designed proteins with the parent protein [26]. DUET server was used for performing mutations in 4G1Q on twenty neighbouring residues, surrounding the active ligand, within the vicinity of 6 Å from the centre of the ligand [19]. A dataset of 380 designed proteins is created. Further, ΔΔG was estimated for all the 380 designed proteins for comparing their relative stability with the parent protein, 4G1Q. The snapshot of protein 4g1q is presented in Fig. 1.

The FASTA sequence of the protein 4g1q is given herewith.

>4G1Q_1|Chain A|Reverse transcriptase/ribonuclease H|Human immunodeficiency virus type 1 (11,678).

MVPISPIETVPVKLKPGMDGPKVKQWPLTEEKIKALVEICTEME-KEGKISKIGPENPYNTPVFAIKKKDSTKWRKLVDFRELNKRTQDFW-EVQLGIPHPAGLKKKKSVTVLDVGDAYFSVPLDEDFRKYTAFTIP-SINNETPGIRYQYNVLPQGWKGSPAIFQSSMTKILEPFAAQNPDIVI-YQYMDDLYVGSDLEIGQHRTKIEELRQHLLRWGLTTPDKKHQKEPP-

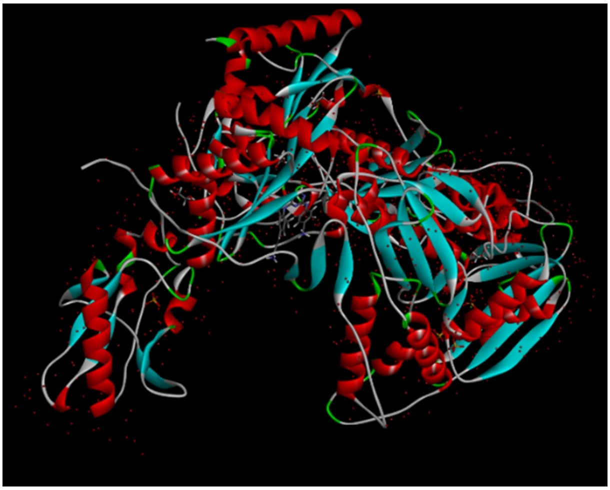

Fig. 1 Snapshot of 4g1q.pdb

FLWMGYELHPDKWTVQPIVLPEKDSWTVNDIQKLVGKLNWASQIYPGIKVRQLSKLLRGTKALTEVIPLTEEAELELAENREILKEPVHGVYYDPSKDLIAEIQKQGQGQWTYQIYQEPFKNLKTGKYARMRGAHTNDVKQLTEAVQKITTESIVIWGKTPKFKLPIQKETWETWWTEYWQATWIPEWEFVNTPPLVKLWYQLEKEPIVGAETFYVDGAANRETKLGKAGYVTNKGRQKVVPLTNTTNQKTELQAIYLALQDSGLEVNIVTDSQYALGIIQAQPDKSESELVNQIIEQLIKKEKVYLAWVPAHKGIGGNEQVDKLVSAG.

> 4G1Q_2|Chain B|p51 RT|Human immunodeficiency virus type 1 (11,678).

PISPIETVPVKLKPGMDGPKVKQWPLTEEKIKALVEICTEMEKEGKISKIGPENPYNTPVFAIKKKDSTKWRKLVDFRELNKRTQDFWEVQLGIPHPAGLKKKKSVTVLDVGDAYFSVPLDEDFRKYTAFTIPSINNETPGIRYQYNVLPQGWKGSPAIFQSSMTKILEPFKKQNPDIVIYQYMDDLYVGSDLEIGQHRTKIEELRQHLLRWGLTTPDKKHQKEPPFLWMGYELHPDKWTVQPIVLPEKDSWTVNDIQKLVGKLNWASQIYPGIKVRQLSKLLRGTKALTEVIPLTEEAELELAENREILKEPVHGVYYDPSKDLIAEIQKQGQGQWTYQIYQEPFKNLKTGKYARMRGAHTNDVKQLTEAVQKITTESIVIWGKTPKFKLPIQKETWETWWTEYWQATWIPEWEFVNTPPLVKLWYQ.

The major focus of the work was to understand the impact of single point mutation on the stability of protein using ΔΔG as a measure of stability. In order to understand the impacts of non-synonymous single nucleotide polymorphisms (nsSNPs) on the structure and function of the proteome, as well as to guide protein engineering, accurate in silico methodologies are needed to study and prognosticate their goods on protein stability. The change in folding free energy upon mutation (ΔΔG in kcal/ mol) is used as the measure to understand the impact of the mutation. DUET, a web server for an intertwined computational approach to study missense mutations in proteins is.

In order to do so, complementary information regarding the mutation, analogous as secondary structure (used by SDM) and a pharmacophore vector that accounts for the changes between wild- type and mutant residue (used by mCSM) are also calculated and used by DUET. As described previously, the pharmacophore vector is attained by comparing the frequency of eight possible grain characteristics between wild- type and mutant remainders (positive, negative, hydrophobic, hydrogen patron, hydrogen acceptor, sulphur and neutral [19].

The DUET (Distance-Dependent, United, Enhanced Sampling) server is a computational tool used for estimating changes in binding free energy upon mutation or interaction in molecular systems, such as protein-protein or protein-ligand complexes. DUET utilizes molecular dynamics simulations and the end-point method to calculate these energy changes. Here is an overview of how calculations are done to determine changes in free energy using the DUET server.

2.1 Preparation of input structures

The user typically provides two input structures for DUET analysis: the wild-type (WT) and mutant (MT) structures. These structures can represent a protein-protein complex, protein-ligand complex, or any other molecular system of interest. The WT structure represents the original or reference state, while the MT structure represents the mutant or perturbed state.

2.2 Molecular dynamics (MD) simulations

DUET performs molecular dynamics simulations for both the WT and MT structures. MD simulations involve the numerical integration of Newton's equations of motion to simulate the behavior of atoms and molecules over time. During the simulations, the system's potential energy is continuously sampled. DUET uses the CHARMM force field to calculate energy terms, including van der Waals interactions, electrostatic interactions, and solvation energies.

2.3 Energy calculations

At various time points during the MD simulations, DUET calculates the potential energy of the system for both the WT and MT states. The energy terms include intra-molecular energies (energies within the molecules), inter-molecular energies (energies between molecules, e.g., protein-protein or protein-ligand interactions), and solvation energies. Calculation of Binding Free Energy: DUET uses the energy values obtained from the MD simulations to calculate the binding free energy difference (ΔΔG) between the WT and MT states. The binding free energy is typically calculated using the following equation:

$$\Delta\Delta G = \Delta G\_MT - \Delta G\_WT$$

where ΔG_MT is the free energy of the mutant (MT) state, and ΔG_WT is the free energy of the wild-type (WT) state.

DUET performs statistical analysis on the energy data obtained from multiple MD trajectories to improve accuracy and reliability.

3 Result and discussion on duet results

The results of missenses caused by inducing mutations in a protein (4g1q.pdb) molecule and their effects on the stability of designed proteins are detailed in this section.

Missenses were introduced in a total of 20 AARs in silico and mutated *de novo* design of *380* proteins is carried out. The stability of the designed proteins is carried out by comparing their ΔΔG values, which is a metric for comparing how a single point mutation affects protein stability, with the parent protein 4G1Q.

The impact of the mutations on protein stability based on ΔΔG are assessed in two ways:

A. Impact on stability of designed protein on mutation of a specific surrounding amino acid residue.
B. Impact on stability of designed protein by mutation of a specific amino acid residue.

The ΔΔG values of all the 380 designed proteins, on a mutation of surrounding amino acid residues, are presented in Table 1, in which ΔΔG for 4G1Q is taken as zero and comparisons are made.

A bar graph showing the comparative ΔΔG values of all the 380 designed proteins is presented in Fig. 2. All values above the x-axis indicate the ΔΔG values of proteins which are unstable than parent 4g1q while those below the x-axis (negative) indicate the ΔΔG values of protein which are stable than the parent 4g1q.

The results thus obtained from the estimation of ΔΔG using the DUET server table presented in 1, it is observed that of 380 designed (mutated) proteins a total of 41 exhibit positive while 339 exhibit negative ΔΔG values. This suggests 339 stable proteins while 41 unstable proteins are obtained, indicating stabilization effect of mutation in nearly 90% cases.

3.1 Effect of mutation on stability of a specific surrounding amino acid residue

Table 2 presents the order of stability of newly designed proteins formed on mutations of a specific AAR. This also gives a detailed insight into the effect of mutation of a specific AAR.

The subsequent data, which show that 380 new designed proteins were produced on a single point mutation, are taken from Table 2. A single point mutation yields 339 stable proteins out of the 380 designed proteins, while 41 designed proteins that are less stable than parent 4G1Q are obtained. All the designed proteins that are obtained by mutating F227, P225, P236, V106, W229, Y183 and Y318 are observed to be more stable than parent 4G1Q, suggesting no effect of mutation on these AARs positions. While mutating P226, Y181, and Y188 mutation produces a total of Fifty-Four (out of Fifty-Seven i.e., Eighteen each) proteins, more stable than 4G1Q are obtained, suggesting mutations of these AARs also stabilizes the designed (mutated) protein but to a lesser extent. 21 out of 41 the unstable proteins were obtained when lysine (K) amino acid residues namely K101, K102 and K103 are mutated. The highest number (08) of unstable designed proteins are obtained when K101 is mutated, while mutation of K102 and K103 yielded 7 and 6 unstable designed proteins, respectively. This suggests mutation of lysine might be highly important in deciding the stability of a protein. This further suggest that introduction

Table 1 ΔΔG values for the designed proteins* as obtained from DUET server

		Amino acid residues causing mutation																				Unsta-ble proteins
S. No.	Sur-rounding AA Residues number	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	
1.	E138_B	−0.325	−0.126	−0.826	0	−0.412	−0.897	−0.767	0.567	−0.224	0.547	0.797	−0.604	−0.166	−0.597	−0.235	−1.089	−0.652	0.304	−0.368	−0.349	4
2.	F227	−3.032	−1.913	−3.156	−2.858	0	−3.255	−1.484	−1.063	−2.386	−1.501	−0.957	−2.905	−2.031	−2.887	−0.193	−3.593	−3.236	−1.859	−0.364	−0.545	0
3.	G190	−0.645	−1.255	−2.179	−2.129	−1.067	0	−1.655	0.242	−1.494	0.081	−0.451	−1.344	−0.979	−1.633	−1.132	−1.739	−1.441	0.16	−1.402	−1.233	3
4.	H235	−1.548	−0.597	−1.515	−1.34	0.388	−1.895	0	−0.323	−1.141	−0.411	−0.649	−1.606	−1.086	−1.223	−1.636	−1.65	−1.197	−0.665	0.476	0.495	3
5.	K101	−0.603	−0.573	−0.048	0.236	−0.133	−0.91	−0.82	0.756	0	0.66	0.044	−0.607	−0.157	−0.536	−0.183	−1.428	−0.815	0.418	−0.146	0.149	6
6.	K102	−0.219	−0.471	0.388	0.547	−0.1	−0.502	−0.982	0.747	0	0.704	0.263	−0.129	0.601	−0.264	−0.256	−0.899	−0.484	0.421	−0.038	0.143	8
7.	K103	−0.364	−0.731	0.03	0.255	−0.306	−0.896	−1.118	0.655	0	0.53	0.162	−0.43	−0.442	−0.626	−0.296	−1.387	−0.813	0.361	−0.143	0.143	7
8.	L100	−2.46	−1.552	−2.81	−2.612	−1.619	−3.065	−2.402	−0.918	−2.074	0	−0.863	−2.222	−1.521	−2.327	−1.218	−2.985	−2.632	−1.739	−1.872	0.168	1
9.	L228	−0.94	−0.547	−0.15	0.033	−0.701	−1.113	−0.374	−0.043	−0.295	0	−0.347	−0.248	−0.843	−0.402	0.2	−0.905	−0.589	−0.16	−0.737	−0.569	2
10.	P095	−1.563	−1.154	−2.626	−2.36	−1.183	−2.651	−2.175	0.08	−1.607	0.143	−0.105	−1.812	0	−1.739	−1.305	−2.548	−2.035	−0.352	−1.655	−1.39	2
11.	P225	−0.841	−0.385	−0.981	−1.037	−0.936	−0.401	−0.769	−0.196	−0.77	−0.194	−0.174	−0.557	0	−0.664	−0.289	−0.962	−0.834	−0.378	−0.936	−0.861	0
12.	P226	−0.815	−0.352	−1.018	−1.25	−1.068	−1.594	−1.045	−0.35	−0.777	0.026	−0.11	−0.583	0	−0.891	−0.422	−1.081	−0.918	−0.57	−0.773	−1.039	1
13.	P236	−1.362	−0.368	−1.455	−1.369	−0.754	−1.78	−1.233	−0.256	−1.014	−0.212	−0.051	−0.859	0	−0.972	−0.697	−1.558	−1.156	−0.509	−1.004	−0.951	0
14.	V106	−1.991	−1.202	−1.651	−1.254	−1.267	−2.355	−1.627	−0.247	−1.268	−0.425	−0.616	−1.34	−1.487	−1.31	−0.737	−2.194	−1.721	0	−1.435	−1.179	0
15.	V179	−0.759	−0.695	−0.197	−0.129	−0.729	−0.894	−0.373	0.092	−0.495	−0.054	−0.522	−0.496	−0.455	−0.681	0.008	−1.107	−0.82	0	−0.758	−0.513	2
16.	W229	−1.652	−1.5	−2.502	−2.346	−0.677	−1.547	−1.616	−0.519	−1.936	−0.811	−0.853	−2.614	−1.091	−2.32	−2.044	−2.647	−2.392	−1.026	0	−0.767	0
17.	Y181	−2.272	−1.355	−1.62	−1.359	−0.361	−2.523	−0.866	−0.715	−1.657	−0.844	−1.065	−2.419	−1.852	−2.166	−1.426	−2.968	−2.537	−1.184	0.086	0	1
18.	Y183	−2.226	−1.134	−1.823	−1.6	−0.878	−2.378	−0.8	−1.067	−1.425	−1.165	−1.15	−2.088	−1.846	−1.948	−1.171	−2.532	−2.12	−1.262	−0.466	0	0
19.	Y188	−2.335	−1.106	−2.639	−2.3	−0.055	−2.675	−1.36	−0.514	−1.583	−0.772	−0.669	−2.587	−1.778	−2.271	−1.578	−3.021	−2.508	−0.943	0.405	0	1
20.	Y318	−3.512	−1.64	−3.437	−3.151	−1.222	−3.955	−2.341	−2.159	−2.06	−2.115	−1.925	−2.784	−2.946	−2.628	−1.624	−3.215	−2.863	−2.711	−0.6	0	0

Table 1 (continued)

S. No.	Surrounding AA Residues number	Amino acid residues causing mutation																				Unstable proteins
		A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	
	Unstable Proteins	0	0	2	4	1	0	0	7	0	7	4	0	1	0	2	0	0	5	3	5	41

*The value of 0 (Zero) is for parent protein (4gq1)

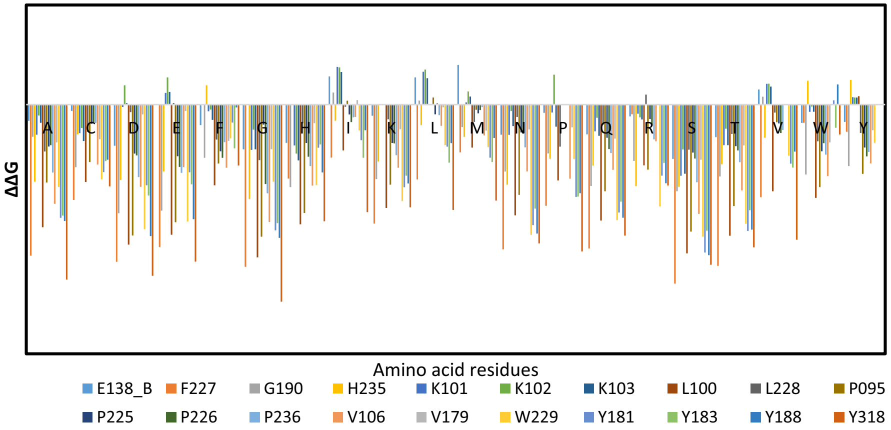

Fig. 2 Comparative ΔΔG of designed proteins on mutation of 20 AARs

Table 2 Order of stability of designed protein of mutation of a specific SAAR*

S. No.	Surrounding AA residues	Order of stability on mutation*
1.	E138_B	S>G>D>H>T>N>Q>F>W>Y>A>R>K>P>C>*E*>V>L>I>M
2.	F227	S>G>T>D>A>N>Q>E>K>P>C>V>L>H>I>M>Y>W>R>*F*
3.	G190	D>E>S>H>Q>K>T>W>N>C>Y>R>F>P>A>M>*G*>L>V>I
4.	H235	G>S>R>N>A>D>E>Q>T>K>P>V>M>C>L>I>*H*>F>W>Y
5.	K101	S>G>H>T>N>A>C>Q>R>P>W>F>D>*K*>M>Y>E>V>L>I
6.	K102	H>S>G>T>C>Q>R>A>N>F>W>*K*>Y>M>D>V>E>P>L>I
7.	K103	S>H>G>T>C>Q>P>N>A>F>R>W>*K*>D>Y>M>E>V>L>I
8.	L100	G>S>D>T>E>A>H>Q>N>K>W>V>F>C>P>R>I>M>*L*>Y
9.	L228	G>A>S>P>W>F>T>Y>C>Q>H>M>K>N>V>D>I>*L*>E>R
10.	P095	G>D>S>E>H>T>N>Q>W>K>A>Y>R>F>C>V>M>*P*>I>L
11.	P225	E>D>S>W>F>Y>A>T>K>H>Q>N>G>C>V>R>I>L>M>*P*
12.	P226	G>E>S>F>H>Y>D>T>Q>A>K>W>N>V>R>C>I>M>*P*>L
13.	P236	G>S>D>E>A>H>T>K>W>Q>Y>N>F>R>V>C>I>L>M>*P*
14.	V106	G>S>A>T>D>H>P>W>N>Q>K>F>E>C>Y>R>M>L>I>*V*
15.	V179	S>G>T>A>W>F>C>Q>M>Y>N>K>P>H>D>E>L>*V*>R>I
16.	W229	S>N>D>T>E>Q>R>K>A>H>G>C>P>V>M>L>Y>F>I>*W*
17.	Y181	S>T>G>N>A>Q>P>K>D>R>E>C>V>M>H>L>I>F>*Y*>W
18.	Y183	S>G>A>T>N>Q>P>D>E>K>V>R>L>M>C>I>F>H>W>*Y*
19.	Y188	S>G>D>N>T>A>E>Q>P>K>R>H>C>V>L>M>I>F>*Y*>W
20.	Y318	G>A>D>S>E>P>T>N>V>Q>H>I>L>K>M>C>R>F>W>*Y*

*The protein depicted in italic is parent 4G1Q (unmutated)

of instability might affect the process of denaturation and in all probabilities enhance it, i.e. when lysine is mutated the stability of a protein decreases.

3.2 Effect of mutation on stability by a specific amino acid residue

Table 3 presents the impact of mutation by a specific mutation on the stability of designed proteins.

Table 3 Effect of mutation by specific AAR*

S. No.		Order of stability on mutation*
1.	A	Y318>F227>L100>Y188>Y181>Y183>V106>W229>P095>H235>P236>L228>P225>P226>V179>G190>K101>K103>E138_B>K102
2.	C	F227>Y318>L100>W229>Y181>G190>V106>P095>Y183>Y188>K103>V179>H235>K101>L228>K102>P225>P236>P226>E138_B
3.	D	Y318>F227>L100>Y188>P095>W229>G190>Y183>V106>Y181>H235>P236>P226>P225>E138_B>V179>L228>K101>K103>K102
4.	E	Y318>F227>L100>P095>W229>Y188>G190>Y183>P236>Y181>H235>V106>P226>P225>V179>*E138_B*>L228>K101>K103>K102
5.	F	L100>V106>Y318>P095>P226>G190>P225>Y183>P236>V179>L228>W229>E138_B>Y181>K103>K101>K102>Y188>*F227*>H235
6.	G	Y318>F227>L100>Y188>P095>Y181>Y183>V106>H235>P236>P226>W229>L228>K101>E138_B>K103>V179>K102>P225>*G190*
7.	H	L100>Y318>P095>G190>V106>W229>F227>Y188>P236>K103>P226>K102>Y181>K101>Y183>P225>E138_B>L228>V179>*H235*
8.	I	Y318>Y183>F227>L100>Y181>W229>Y188>P226>H235>P236>V106>P225>L228>P095>V179>G190>E138_B>*K103>K102>K101*
9.	K	F227>L100>Y318>W229>Y181>P095>Y188>G190>Y183>V106>H235>P236>P226>P225>V179>L228>E138_B>*K103=K102=K101*
10.	L	Y318>F227>Y183>Y181>W229>Y188>V106>H235>P236>P225>V179>*L100=L228*>P226>G190>P095>K103>E138_B>K101>K102
11.	M	Y318>Y183>Y181>F227>L100>W229>Y188>H235>V106>V179>G190>L228>P225>P226>P095>P236>K101>K103>K102>E138_B
12.	N	F227>Y318>W229>Y188>Y181>L100>Y183>P095>H235>G190>V106>P236>K101>E138_B>P226>P225>V179>K103>L228>K102
13.	P	Y318>F227>Y181>Y183>Y188>L100>V106>W229>H235>G190>L228>V179>K103>E138_B>K101>*P095=P236=P226=P225*>K102
14.	Q	F227>Y318>L100>W229>Y188>Y181>Y183>P095>G190>V106>H235>P236>P226>V179>P225>K103>E138_B>K101>L228>K102
15.	R	W229>H235>Y318>Y188>Y181>P095>L100>Y183>G190>V106>P236>P226>K103>P225>K102>E138_B>F227>K101>V179>L228
16.	S	F227>Y318>Y188>L100>Y181>W229>P095>Y183>V106>G190>H235>P236>K101>K103>V179>E138_B>P226>P225>L228>K102
17.	T	F227>Y318>L100>Y181>Y188>W229>Y183>P095>V106>G190>H235>P236>P226>P225>V179>K101>K103>E138_B>L228>K102
18.	V	Y318>F227>L100>Y183>Y181>W229>Y188>H235>P226>P236>P225>P095>L228>*V106=V179*>G190>E138_B>K103>K101>K102
19.	W	L100>P095>V106>G190>P236>P225>P226>V179>L228>Y318>Y183>E138_B>F227>K101>K103>K102>*W229*>Y181>Y188>H235
20.	Y	P095>G190>V106>P226>P236>P225>W229>L228>F227>V179>E138_B>*Y318=Y183=Y181=Y188*>K103>K102>K101>L100>H235

*The proteins depicted in italic are 4G1Q (unmutated)

The following conclusions are drawn from Table 3 on the influence of mutation caused by a particular AAR. The AARs G, H, and K impact the stability of 4G1Q the most and on mutation by these AARs all the *de novo* designed proteins are observed to be more stable than parent 4G1Q. A little lesser Impact is observed when mutations is performed by F and P, wherein only one designed protein, less stable than parent 4G1Q is obtained for each mutation. The Lysine (K) AAR produces the highest number (07) of unstable designed proteins. Of the various impacts of mutation, in 10 cases where K102 is mutated, most unstable designed proteins are obtained. Surprisingly, mutation by and mutation of lysine is creating instability in the designed protein suggesting that neither lysine should be mutated nor it should be used for mutation.

The designed proteins have been classified on the basis of mutation of a specific AAR and their stability (ΔΔG) range. Table 4 shows the details of these mutations and stability (ΔΔG range) of designed proteins.

Table 4 provides the following observations: as previously mentioned, 339 stable and 41 unstable designer proteins are obtained. Of the 339 stable designer protein, 12 highly stable designed proteins are obtained on the mutation of F227, L100, Y188 and Y318. Their ΔΔG values thus obtained are between −4.0 and −3.0. Of these 12 designed proteins it is observed that the maximum number (05) of most stable proteins are obtained when Y318 is mutated. These 12 stable proteins are obtained on mutation of hydrophobic AARs. 58 proteins having ΔΔG values between −3.0 and −2.0 are obtained. Of these highest number (09 each) of designed proteins, within this stability range, is obtained when L100 and Y318 are mutated. 113 proteins having ΔΔG values between −2.0 and −1.0 are obtained. These can be classified as moderately stable.87 proteins having ΔΔG values between −2.0 and −1.0 are obtained and these can be classified relatively less stable. 99 designer protein having ΔΔG values between −0.5 and 0.5 are obtained, and the stability of these cannot be justified as the ΔΔG values are within the noise range so should be considered indifferent or neutral. A total of 11 highly unstable designed proteins are obtained on the mutation of E138_B, K101, K102, and K103. Their ΔΔG values thus obtained are greater than 0.5. The unstable designed proteins are obtained when the charged AARs (E and K) are mutated.

Table 5 shows that all of the designer proteins produced by the mutations of F227, P225, P236, V106, W229, Y183, and Y318 are more stable than the parent 4G1Q. These results of the present study are contrary to the belief that mutation induces instability in the protein and the naturally occurring proteins acquire most stable form. The Lysine residues (101, 102 and 103) are the most affected AARs and they produce least number of stable designer proteins. Though, the missenses are induced in silico, the results need to be verified practically.

Another way in which the designed proteins have been classified is on the basis of mutation by a specific AAR and their stability range (ΔΔG). Table 6 shows the details of these mutations and stability (ΔΔG range) of designed proteins.

Table 4 Classification of designed proteins

S. No.	Mutated AA residue	ΔΔG	Stable than 4G1Q					4G1Q	Unstable than 4G1Q	
			-4.0–≤ 3.0	-3.0–≤ 2.0	-2.0–≤ -1.0	-1.0–≤ -0.5	$-0.5 < 0.0$	0	> 0.0–≤ 0.5	> 0.5
1.	E138_B		0	0	1	6	8	1	1	3
2.	F227		5	5	5	2	2	1	0	0
3.	G190		0	2	11	2	1	1	3	0
4.	H235		0	0	11	3	2	1	3	0
5.	K101		0	0	1	7	5	1	4	2
6.	K102		0	0	0	3	8	1	4	4
7.	K103		0	0	2	4	6	1	5	2
8.	L100		1	9	6	2	0	1	1	0
9.	L228		0	0	1	8	8	1	2	0
10.	P095		0	6	9	0	2	1	2	0
11.	P225		0	0	1	11	7	1	0	0
12.	P226		0	0	7	7	4	1	1	0
13.	P236		0	0	9	6	4	1	0	0
14.	V106		0	2	13	2	2	1	0	0
15.	V179		0	0	1	9	7	1	2	0
16.	W229		0	7	7	5	0	1	0	0
17.	Y181		0	6	8	3	1	1	1	0
18.	Y183		0	5	11	2	1	1	0	0
19.	Y188		1	7	5	4	1	1	1	0
20.	Y318		5	9	4	1	0	1	0	0
			12	58	113	87	69	20	30	11

Table 5 Shows the number of stable proteins obtained on mutation of a specific AAR

S. No.	Mutated AARs	Number of stable designer proteins
1.	E138_B	15
2.	F227	19
3.	G190	16
4.	H235	16
5.	K101	13
6.	K102	11
7.	K103	12
8.	L100	18
9.	L228	17
10.	P095	17
11.	P225	19
12.	P226	18
13.	P236	19
14.	V106	19
15.	V179	17
16.	W229	19
17.	Y181	18
18.	Y183	19
19.	Y188	18
20.	Y318	19

The following observations are derived from Table 6, and as previously said, 339 stable and 41 unstable designer proteins are found. Of the 339 stable designer protein, 12 highly stable designed proteins are obtained on the mutation by A, D, G, E, S and T. Their ΔΔG values thus obtained are between -4.0 and -3.0. Of these 12 designed proteins it is observed that the maximum number (03 each) of most stable proteins are obtained when mutated by G and S. No regular pattern of impact of mutation by a specific property of is obtained. 58 proteins having ΔΔG values between -3.0 and -2.0 are obtained. Of these highest number (07 each) of designed proteins, within this stability range, is obtained when mutated by N and T. 113 proteins having ΔΔG values between -2.0 and -1.0 are obtained. These can be classified as moderately stable. 87 proteins having ΔΔG values between -2.0 and -1.0 are obtained and these can be classified relatively less stable. 99 designer protein having ΔΔG values between -0.5 and 0.5 are obtained, and the stability of these cannot be justified as the ΔΔG values are within the noise range so should be considered indifferent or neutral. A total of 11 highly unstable designed proteins are obtained on the mutation by E, I, L M, and P. Their ΔΔG values thus obtained are greater than 0.5. In this case the mutation caused by hydrophobic has given the most unstable designed protein.

As can be seen from the Table 7, the designer proteins that resulted from the mutations generated by A, C, D, G, H, K, Q, N, S, and T are all more stable than the original 4G1Q protein. On mutation by L and Y, highest number (08) of unstable

Table 6 Classification of designed proteins

S. No.	Mutation by AA residue	ΔΔG	Stable than 4G1Q					4G1Q	Unstable than 4G1Q	
			− 4.0–≤ 3.0	− 3.0–≤ 2.0	− 2.0–≤ − 1.0	− 1.0–≤ − 0.5	− 0.5 < 0.0	0	− 0.0–≤ 0.5	> 0.5
1.	A		2	4	5	6	3	0	0	0
2.	C		0	0	10	5	5	0	0	0
3.	D		2	5	6	2	3	0	2	0
4.	E		1	6	7	0	1	1	3	1
5.	F		0	0	6	6	6	1	1	0
6.	G		3	5	5	5	1	1	0	0
7.	H		0	3	8	6	2	1	0	0
8.	I		0	1	2	4	6	0	3	4
9.	K		0	3	9	2	3	3	0	0
10.	L		0	1	2	3	5	2	4	4
11.	M		0	0	3	7	6	0	3	1
12.	N		0	7	4	5	4	0	0	0
13.	P		0	2	7	2	4	4	0	1
14.	Q		0	6	5	7	2	0	0	0
15.	R		0	1	8	2	7	0	2	0
16.	S		3	6	8	3	0	0	0	0
17.	T		1	7	4	7	1	0	0	0
18.	V		0	1	5	4	3	2	4	0
19.	W		0	0	5	5	6	1	3	0
20.	Y		0	0	4	6	1	4	5	0
			12	58	113	87	69	20	30	11

Table 7 Shows the number of stable and unstable proteins obtained on mutation by a specific AAR

S. No.	Mutation by AAR	Number of stable designer proteins	Number of unstable designer proteins
1.	A	20	0
2.	C	20	0
3.	D	18	2
4.	E	15	4
5.	F	18	1
6.	G	19	0
7.	H	19	0
8.	I	13	7
9.	K	17	0
10.	L	11	7
11.	M	16	4
12.	N	20	0
13.	P	15	1
14.	Q	20	0
15.	R	18	2
16.	S	20	0
17.	T	20	0
18.	V	13	5
19.	W	16	3
20.	Y	11	5

Table 8 Mutated and mutation by AARs yielding most unstable designer proteins

S. No.	Mutation in AAR	Mutation by AAR
1.	K101	I
2.	E138_B	M
3.	K102	I
4.	K102	L
5.	K101	L
6.	K103	I
7.	K102	P
8.	E138_B	I
9.	E138_B	L
10.	K102	E
11.	K103	L

designer proteins, suggesting I and L follow, relatively better than other AARs, the trend of natural phenomena wherein mutation causes instability in the protein.

The comparative stability analyses reveals that the following combinations give the top 11 most unstable *de novo* designed proteins and are presented in Table 8.

From the Table 8 it is observed that mutation of combinations K102-I/L/P/E give most unstable proteins.

4 Conclusions

The study has given surprising results and a higher number of stable designer proteins were obtained on mutation. As the work take cares of single point mutation and nothing else, the results are non-traditional. However, the environment at each position should be considered. If interacting molecules are not present in the model, such as at a known zinc-binding site, then a seemingly favourable mutation will not be favourable in reality.

A position that has a lot of negative ΔΔGs could mean that this position evolved a destabilizing residue because it is necessary for its catalytic activity, for binding another molecule, or because of another functionally relevant reason.

Moreover, it must be kept in mind that this quantifies a single-point mutation. Sometimes sufficient stability can only be attained by various interrelated changes. Only one mutation can be predicted by ΔΔG at a time. It is a must to induce the mutations and run some relax reiterations in order to determine if multiple mutations would have a cumulative effect on stability. It takes longer much time calculate even almost exact ΔΔG.

Acknowledgements The Authors would like to thank Dr. Subhash Basak and Dr. Tanmoy Chakraborty, Conveners, 8th IUWMC, 2022 for accepting the submission for presentation.

Author contributions The work was performed by NS and LS, while SG and NJ helped in refining the interpretations and contributed in manuscript preparation.

Funding Not applicable.

Data availability All provided with the manuscript.

Declarations

Competing interests The authors declare no competing interests.

Ethical approval Not applicable.

6. References

1. P.D. Tanford, Adv. Protein Chem. **121**, 121 (1968)
2. P. Gainza, S. Wehrle, A. Van Hall-Beauvais et al., De novo design of protein interactions with learned surface fingerprints. Nature **617**, 176 (2023). https://doi.org/10.1038/s41586-023-05993-x
3. W.-W. Liao, M. Asri, J. Ebler, D. Doerr, M. Haukness, G. Hickey, S. Lu, J.K. Lucas, J. Monlong, H.J. Abel, S. Buonaiuto, X.H. Chang, H. Cheng, J. Chu, V. Colonna, J.M. Eizenga, X. Feng, C. Fischer, R.S. Fulton, S. Garg, C. Groza, A. Guarracino, W.T. Harvey, S. Heumos, K. Howe, M. Jain, T.-Y. Lu, C. Markello, F.J. Martin, M.W. Mitchell, K.M. Munson, M.N. Mwaniki, A.M. Novak, H.E. Olsen, T. Pesout, D. Porubsky, P. Prins, J.A. Sibbesen, C. Tomlinson, F. Villani, M.R. Vollger,

H.P.R. Consortium, G. Bourque, M.J. Chaisson, P. Flicek, A.M. Phillippy, J.M. Zook, E.E. Eichler, T. Miga, E. Wang, B. Garrison, D. Paten, E.D Haussler, K.H Jarvis, D. Paten, E.D Haussler, K.H Jarvis, Nature. 617, 312 (2023)

4. R.A. Langan, S.E. Boyken, A.H. Ng, J.A. Samson, G. Dods, A.M. Westbrook, T.H. Nguyen, M.J. Lajoie, Z. Chen, S. Berger, V.K. Mulligan, J.E. Dueber, W.R.P. Novak, H. El-Samad, D. Baker, Nature **572**, 205 (2019)
5. K.E. Dunn, F. Dannenberg, T.E. Ouldridge, M. Kwiatkowska, A.J. Turberfield, J. Bath, Nature **525**, 82 (2015)
6. J.P. Renaud, C.W. Chung, U.H. Danielson, U. Egner, M. Hennig, R.E. Hubbard, H. Nar, Nat. Rev. Drug Discov. **15**, 679 (2016)
7. E.O. Freed, Nat. Rev. Microbiol. **13**, 484 (2015)
8. J.M. Llibre, B. Clotet, Expert Rev. Anti-Infect. Ther. **10**(5), 557 (2012)
9. H.R. Rangel, D.J. Garzaro, AIDS Rev. **21**(4), 200–211 (2019)
10. WHO, HIV Drug Resistance Report 2019. World Health Organization
11. S.C. Basak, M. Vračko, *Big data analytics in chemoinformatics and bioinformatics: with applications to computer-aided Drug design, cancer biology emerging pathogens and computational toxicology* (Elsevier, Amsterdam, 2022)
12. P. Dean, Drug design in the 1990s. Nat. Biotechnol. **15**, 1018 (1997)
13. J. Lyu, J.J. Irwin, B.K. Shoichet, Nat. Chem. Biol. **19**, 712 (2023)
14. T.A. Collier, T.J. Piggot, J.R. Allison, Methods Mol. Biol. **2072**, 311 (2020)
15. L. Verlet, Phys. Rev. **159**, 98 (1967)
16. T. Blundell, D. Carney, S. Gardner, F. Hayes, B. Howlin, T. Hubbard, J. Overington, D.A. Singh, B.L. Sibanda, M. Sutcliffe, Eur. J. Biochem. **172**, 513 (1988)
17. T. Schwede, J. Kopp, N. Guex, M.C. Peitsch, Nucleic Acids Res. **31**, 3381 (2003)
18. Jothi, Protein Pept. Lett. **19**, 1191 (2012)
19. L. Loewe, W.G. Hill, Philos. Trans. Royal Soc. B Biol. Sci. **365**, 1153 (2010)
20. D.E.V. Pires, D.B. Ascher, T.L. Blundell, Nucleic Acids Res. **42**, W314 (2014)
21. K.P. Tan, T.R. Kanitkar, C.K. Kwoh, M.S. Madhusudhan, Front. Mol. Biosci. **8**, 646288 (2021)
22. S. Iqbal, F. Ge, F. Li, T. Akutsu, Y. Zheng, R.B. Gasser, D.J. Yu, G.I. Webb, J. Song, J. Chem. Inf. Model **62**, 4270 (2022)
23. S. Iqbal, D. Hoksza, E. Pérez-Palma, P. May, J.B. Jespersen, S.S. Ahmed, Z.T. Rifat, H.O. Heyne, M.S. Rahman, J.R. Cottrell, F.F. Wagner, M.J. Daly, A.J. Campbell, D. Lal, Nucleic Acids Res. **48**, W132 (2021)
24. C.H.M. Rodrigues, D.E.V. Pires, D.B. Ascher, Protein Sci. **30**, 60 (2021)
25. https://www.rcsb.org/structure/4g1q
26. D. Kuroda, J. Bauman, J. Challa et al., Snapshot of the equilibrium dynamics of a drug bound to HIV-1 reverse transcriptase. Nat. Chem. **5**, 174 (2013)

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

Authors and Affiliations

Laxmi Sule[1] · Swagata Gupta[2] · Nilanjana Jain[3] · Nitin S. Sapre[4]

✉ Nitin S. Sapre
nsapre@sgsits.ac.in; sukusap@yahoo.com

Laxmi Sule
laxmisule3@gmail.com

Swagata Gupta
swagataagupta@yahoo.co.uk; swagatagupta28@gmail.com

Nilanjana Jain
nilanjanamjain@gmail.com

[1] Department of Applied Chemistry, SGSITS, Indore, Madhya Pradesh, India

[2] Department of Chemistry, Govt. Holkar (Model Autonomous) Science College, Indore, Madhya Pradesh, India

[3] Department of Chemistry, Guest Faculty, Govt. College, Udaynagar, Dewas, Madhya Pradesh, India

[4] Department of Chemistry, SGSITS, Indore, Madhya Pradesh, India