International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

Identification of Protein Engineering Strategies of Insulin Receptor Tyrosine Kinase - A Protein by Using Hotspot Wizard Webserver

Kaushick Mishra[1], Soumendra Nath Talapatra[2]

[1]Ph. D Scholar, Department of Biotechnology, School of Life Sciences, Seacom Skills University, Kendradangal, Shantiniketan, Birbhum – 731236, West Bengal, India
Corresponding Author Email: *kaushick30[at]gmail.com*
Phone: +91 – 9903277646

[2]School of Life Sciences, Seacom Skills University, Kendradangal, Shantiniketan, Birbhum – 731236, West Bengal, India

Abstract: ***Insulin receptor (IR) tyrosine kinase (Tyr - K) A is well - known larger, multi - sphere and integral membrane protein. The well - established metabolic abnormality is the insulin resistance in which patients are found with non - insulin - dependent diabetes mellitus. The objective of the present study was to identify the protein engineering strategy as per hot spots for protein stability and site - specific mutation of IR - Tyr - K A protein by using an online tool, HotSpot Wizard (version, 3.1). The prediction results through an output interface where functional hot spots, stability hot spots (structural flexibility), correlated hot spots and stability hot spots (sequence consensus) recorded. Moreover, average B - factor along with mutability rate and score were also obtained. In conclusion, pocket identification and mutability prediction of IR - Tyr - K A could be led to detect structural alternation mainly in insulin resistance and space for ligand binding pocket in new drug discovery for disease prevention. This prediction is suggested to compare with experimental hotspots for IR - Tyr - K A protein as per disease therapy after new drug design.***

Keywords: Drug design, Hotspots design, Insulin receptor (IR) tyrosine kinase, Protein engineering strategies, Site directed mutagenesis

1. Introduction

Many studies indicated that in individuals with insulin resistance, the expression of the insulin receptor (IR) observes reduction or absent. [1 - 3] According to Whitehead et al., [4] insulin resistance - associated with insulin receptor substrate - 1 (IRS - 1) mutations, and an increasing phosphorylation of this substrate on serine residues may lead to the reduction of its tyrosine phosphorylation, ultimately develop lower insulin signaling. [3, 5] It was noticed in the insulin signalling pathway that the initial critical node is the receptor itself (IR), which is created two isoforms such as IRA and IRB through the alternative splicing. [3]

The insulin receptor (IR) is a heterotetrametric structure consisting of two α and β subunits, which are connected by disulphide bonds. Insulin connects towards the α subunit of the IR and triggers the tyrosine kinase in the β subunit. [6] Once the tyrosine kinase of IR is stimulated, it promoted autophosphorylation of the β subunit, where phosphorylation observed on three tyrosine residues (Tyr - 1158, Tyr - 1162, and Tyr - 1163), which is essential for intensification of the activity of kinase. [7]

In general, protein - protein interaction indicated residues of $\Delta\Delta G \geq 2$kcal/mol, is known as hot spot. [8] Moreover, certain residues in protein - protein interactions, also termed as hot spots. These residues have unique and variety of energetic properties, can be designed as an important target of protein - protein complex. [9] Some experiments resulted that only a small subset of contact residues observed significant binding free energy. [10] These residues have been termed 'hot spots' and if mutated then they can disrupt the interaction. [11, 12]

In recent study, Yunn et al. [13] developed an aptamer, named IR - A43, which binds to the IR, and validated that IR - A43 and insulin bind to the insulin receptor with mutual positive cooperativity. While IR - A43 individually is inactive, in the presence of insulin, it potentiates autophosphorylation and downstream signaling of the IR. As per recent research interest, several computational tools for protein engineering have been developed especially for the detection for tunnel and cavity, smart libraries, mutation positions, functions, etc. [14 - 20]

A computational prediction was performed for IR - Tyr - K A protein to detect of hot spots for engineered protein stability, cavity and tunnels, catalytic activity, substrate specificity, enantioselectivity and site directed mutagenesis by using HotSpot Wizards, version 3.0.

2. Materials and Methods

The IR - Tyr - K A protein, the crystal structure of protein (PDB ID: 3ekk) was selected and incorporated the ID in the search box of the input interface of HotSpot Wizard (version 3.0) online tool developed by Sumbalova et al. [19] In this automated prediction study, the process structure was determined. HotSpot Wizard 3.0 is an online tool for automatic detection of protein hot spots for engineered proteins' stability, cavity and tunnels, catalytic activity, substrate specificity, enantioselectivity and site directed mutagenesis [15 - 17] In other words, this tool can be utilized for the annotation of protein structures. This present online server includes sequence, structural and evolutionary information obtained from 3 databases and 20 computational tools. [19] According to Bendl et al. [17] and Sebestova et al., [21] this online tool integrates annotated residues, which can be known

easily for mutagenesis and designed for suitable codons for each implemented strategy. Ultimately, this software helps in comprehensive annotations of protein structures and engineering with the stable design of site - specific mutations and targeted libraries.

In the present study, this software was calculated automatically hotspots for function, stability, correlated and consensus sequences for IR - Tyr - K A (Fig.1). Bendl et al. [17] have developed the workflow steps in HotSpot Wizard, the calculation is based on the particular protein annotations, mutagenesis hot spots and smart library design as first, second and final phases respectively.

For statistical analysis, Z scoring values were obtained for each computational tools such as DCA (Direct Coupling analysis), ELSC (Explicit Likelihood of Subset Variation), McBASC (McLachlan Based Substitution correlation), MI (Mutual Information), aMIc (All Microarray Clustering), OMES (Observed Minus Expected Squared) and SCA (Statistical Coupling Analysis).

3. Results

Fig 1 depicts three - dimensional (3D) ribbon structure of studied protein as IR - Tyr - K A (PDB ID: 3ekk) in which chain A obtained 980 - 1586 residues and 1 - 2668 atoms.

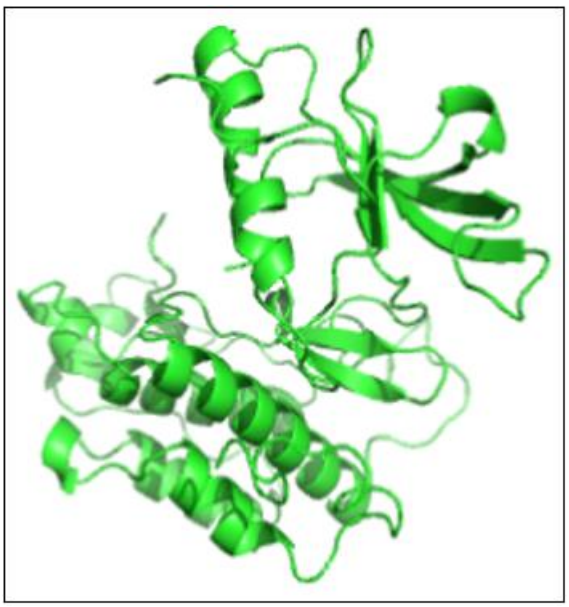


Figure 1: 3D ribbon structure of studied protein as IR - Tyr - K A (PDB ID: 3ekk)

Fig 2 (A - D) evaluates the results from output interface through Hotspots wizard for four separate prediction data such as functional hot spots, stability hot spots (structural flexibility), correlated hot spots and stability hot spots (sequence consensus).
In functional hot spots, the data were obtained for activity, substrate specificity and selectivity, and this step identified residues, which were forming catalytic pocket or accessible tunnel that were not directly participated in the catalysis or located at the evolutionary - conserved position.

For stability hot spots (structural flexibility), the prediction was done to identify the residues in flexible structure, which is observed mainly residues with highest B - factors (Table 1).

In case of the study of correlated hot spots, the data were obtained same as functional hot spots along with the identification of correlated position through consensus approach resulted data from other computational tools viz. DCA (Direct Coupling analysis), ELSC (Explicit Likelihood of Subset Variation), McBASC (McLachlan Based Substitution correlation), MI (Mutual Information), aMIc (All Microarray Clustering), OMES (Observed Minus Expected Squared) and SCA (Statistical Coupling Analysis).

For stability hot spots (sequence consensus), consensus design is an important strategy for the stabilization of proteins. It helps amino acid conservation in sets of homologous protein to identify likely beneficial as well as deleterious mutations of the target protein.

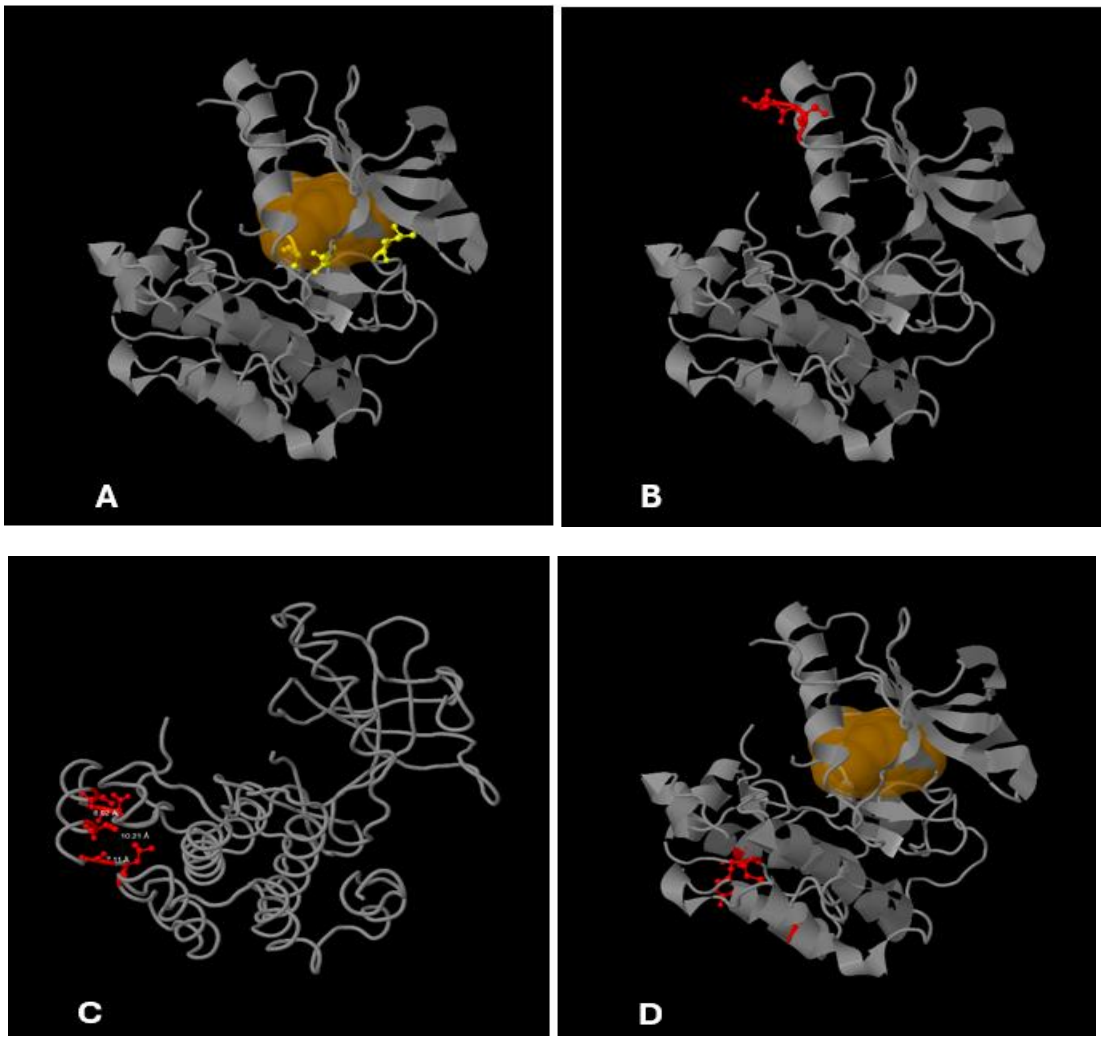


Figure 2: A. Functional hotspot; B. Stability hotspot for structural flexibility; C. Correlated hotspot and D. Sequence consensus of protein as IR - Tyr - K A

Table 1: Study of functional hotspots of protein as IR - Tyr - K A

Studied Protein	Chains	Residues & position	Secondary structures	Pockets & tunnels	Average B - factor (in Å^2)	Mutability rate & score
3ekk	A	Gly & 1149	Loop (L)	1 (catalytic), 2	25.30	High unreliable & 6
	A	Val & 1060	Loop (L)	1 (catalytic), 4	16.73	Moderate unreliable & 5
	A	Ala & 1078	Loop (L)	1 (catalytic), 10	16.53	Moderate unreliable & 5

In Table 2, consensus z - scoring value was obtained for different parameters such as aMIc (2.61 2.41, 2.29, 3.05, and 2.92), DCA (1.26, 1.88, 2.20, 2.55, and 2.08), ELSC (14.12, 16.57, 10.74, 6.48, and 10.61), McBASC (0.50, 0.42, 0.49, 0.54, and 0.51), MI (0.23, 0.59, 0.55, 0.43 and 0.61), OMES (7.56, 23.35, 16.54, 16.56 and 15.47) and SCA (- 0.58, 1.45, 1.53, 0.80 and 0.21) were obtained through this tool for IR - Tyr - K A protein as per correlated residues.

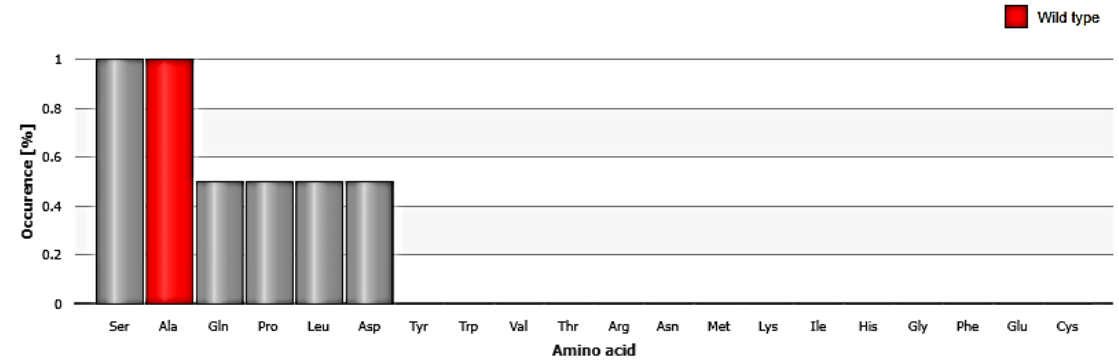
Table 2: Values obtained from different tools for functional hot spots of IR - Tyr - K A

Chains	Consensus z - scoring values						
	aMIc	DCA	ELSC	McBASC	MI	OMES	SCA
A	2.61	1.26	14.12	0.50	0.23	7.56	- 0.58
A	2.41	1.88	16.57	0.42	0.59	23.35	1.45
A	2.29	2.20	10.74	0.49	0.55	16.54	1.53
	3.05	2.55	6.48	0.54	0.43	16.56	0.80
	2.92	2.08	10.61	0.51	0.61	15.47	0.21

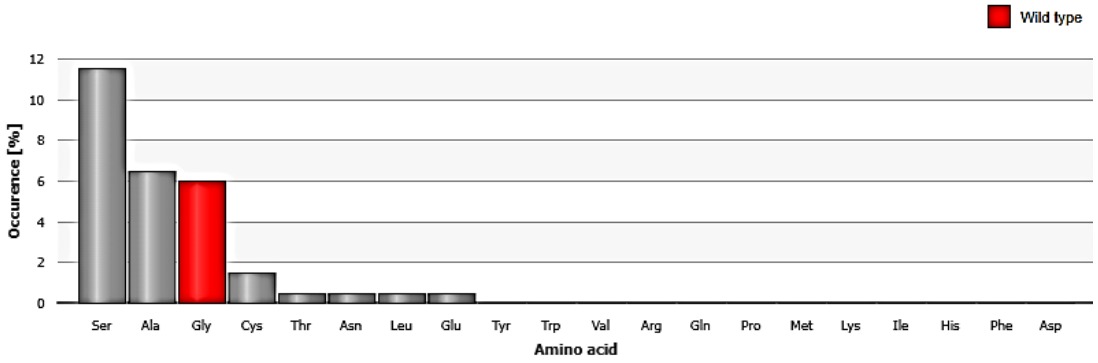
A = Direct Coupling analysis, ELSC = Explicit Likelihood of Subset Variation, McBASC = McLachlan Based Substitution correlation, MI = Mutual Information, aMIc = All Microarray Clustering, OMES = Observed Minus Expected Squared, SCA = Statistical Coupling Analysis

In Fig.3, it was obtained that the amino acid residues fulfilling the criterion of minimal frequency in the multiple sequence alignment. The wild type variety was observed Ser (100%), Gly (60%) and Val (150%) as per positions of different amino acids frequencies of IR - Tyr - K A protein (Fig.3A, B and C).

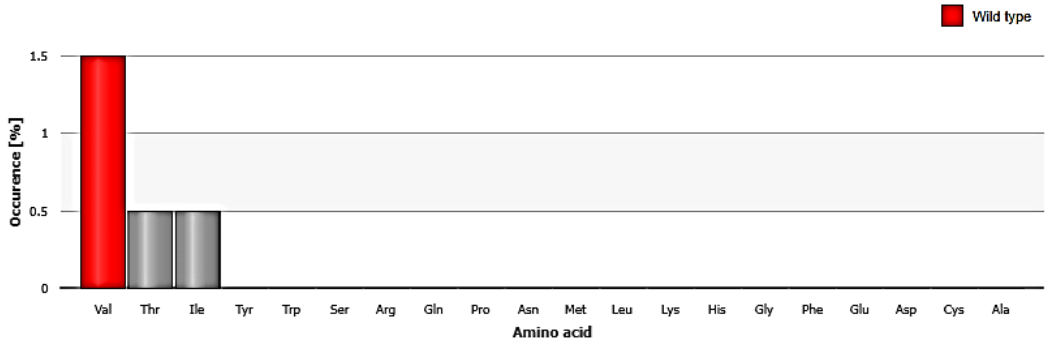
Fig 4 predicts that mutational landscape, which mainly showed the estimation of the probability in relation to preservation of protein function for individual substitution at a particular site of IR - Tyr - K A protein. It was obtained that higher deleterious mutation in Fig 4 C, followed by Fig 4 B and Fig 4A.



Number of gaps: 192 (96 %)
Total number of sequences: 200

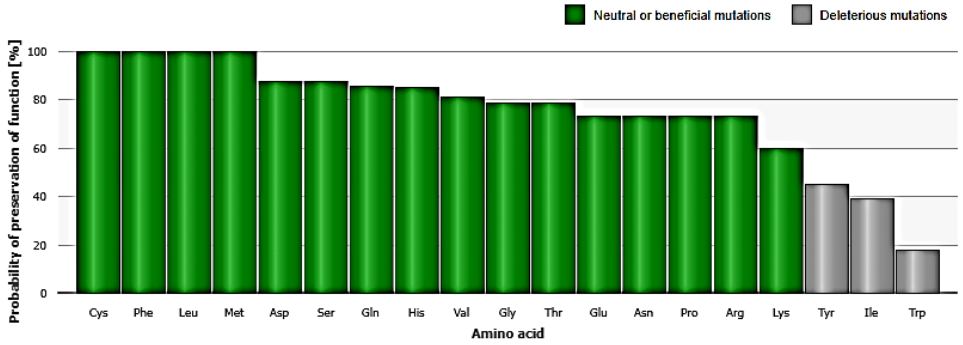


Number of gaps: 145 (72.5 %)
Total number of sequences: 200



Number of gaps: 195 (97.5 %)
Total number of sequences: 200

Figure 3: Amino acids frequencies as per positions of IR - Tyr - K A



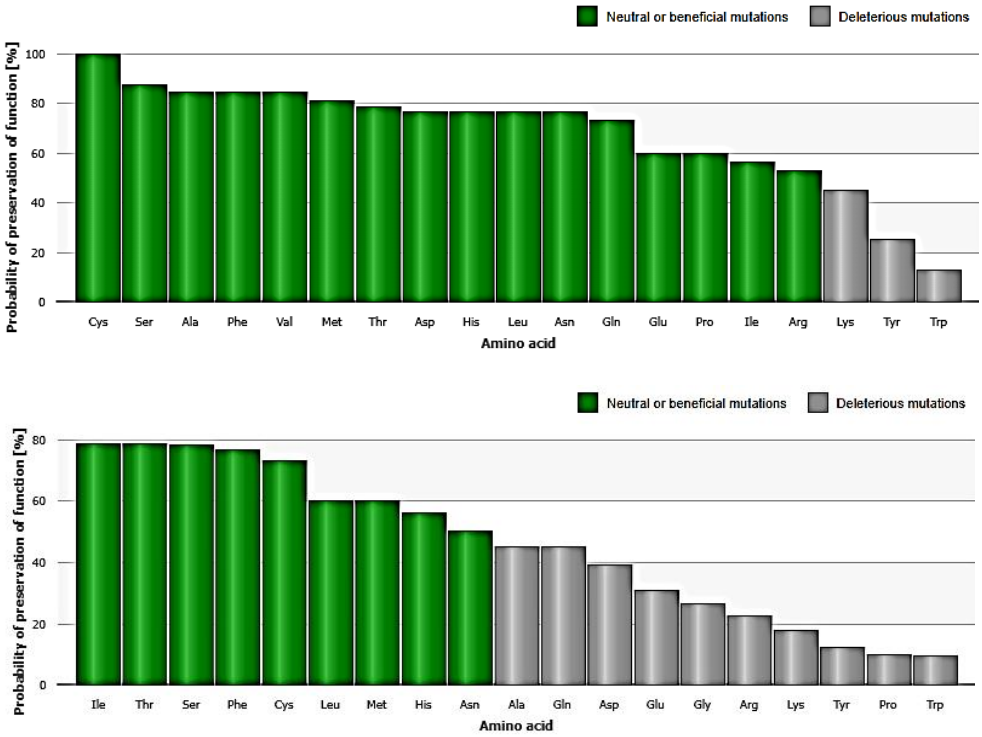


Figure 4: Amino acids mutability landscape of IR - Tyr - K A

4. Discussion

In the present prediction, wild - type and mutated consensus sequences were obtained based on hot spots. [22] Bendl et al. [17] supported the concept of molecular mechanisms of studied protein, which still unclear that how the sequences of protein encode the exact function? Still have not yet answered of this question. [23, 24] It was documented that experimental evolution work suffered major problems when occurred by several irregular study of mutagenesis and detecting of large sequence libraries to evaluate the mutational landscape and proteins showed important structural and functional properties. [17, 24, 25]

An experimental study by Cama et al. [26] indicated that patient as heterozygous for a mutation substituting Isoleucine for Methionine at position 1153. Met1153 was found in the intracellular domain of the receptor near the cluster of tyrosine phosphorylation sites at positions 1158, 1162, and 1163. Finally, the mutant receptor expressed in NIH - 3T3 cells demonstrated that the Ile1153 - mutation impaired the ability of insulin to stimulate autophosphorylation of solubilized IR. Additionally, the mutation impaired the ability of insulin to stimulate the receptor tyrosine kinase activity to phosphorylate an artificial substrate [poly (Glu - Tyr)].

Although, the prediction of different hotspots can be facilitated drug designing process and development. It was suggested that the starting point of a binding site of a receptor in the hotspots might be granted to analyze the docking of ligands. [27] Thus, the prediction of hotspots is a suitable tool to identify exact functional mechanisms of specific protein of interest to identify mutant residue (s) in relation to cause of disease and new drug discovery. [18, 20, 28]

5. Conclusions

It is concluded that HotSpot Wizard (version 3.0) is an online computational tool, which helped easily to obtain results for IR - Tyr - K A protein through protein engineering protocol by the integration of several inbuilt databases derived from other bioinformatics tools and all the data generated within short duration to prevent tedious jobs of experiment. This tool also helped to know mutagenic residues of the studied protein without prior knowledge of computational biology to set up input interface. The parameters like pocket identification and mutability prediction of this protein can lead to know structural alternation of particular in disease diagnosis as well as space for ligand binding pocket in new drug discoveries. The present prediction work is suggested to compare with experimental hotspots for this protein related to drug - ability assessment to prevent diabetes.

Acknowledgement
Authors convey thanks to the developer for the online tool used in this study.

Conflict of interest
Authors declare no conflict of interest.

References

[1] Taylor SI, Accili D, Imai Y. Insulin resistance or insulin deficiency. Which is the primary cause of NIDDM?. Diabetes.1994; 43 (6): 735 - 740.
[2] Foti D, Chiefari E, Fedele M, Iuliano R, Brunetti L, Paonessa F, et al. Lack of the architectural factor HMGA1 causes insulin resistance and diabetes in humans and mice. Nature Medicine.2005; 11 (7): 765 - 773.
[3] Escribano O, Beneit N, Rubio - Longás C, López - Pastor AR, Gómez - Hernández A. The role of insulin

receptor isoforms in diabetes and its metabolic and vascular complications. Journal of Diabetes Research.2017; 2017: 1403206.

[4] Whitehead JP, Humphreys P, Krook A, Jackson R, Hayward A, Lewis H, et al. Molecular scanning of the insulin receptor substrate 1 gene in subjects with severe insulin resistance: detection and functional analysis of a naturally occurring mutation in a YMXM motif. Diabetes.1998; 47 (5): 837 - 839.

[5] Morino K, Petersen KF, Dufour S, Befroy D, Frattini J, Shatzkes N, et al. Reduced mitochondrial density and increased IRS - 1 serine phosphorylation in muscle of insulin - resistant offspring of type 2 diabetic parents. The Journal of Clinical Investigation.2005; 115 (12): 3587 - 3593.

[6] Saini V. Molecular mechanisms of insulin resistance in type 2 diabetes mellitus. *World* Journal of Diabetes.2010; 1 (3): 68 - 75.

[7] White MF, Shoelson SE, Keutmann H, Kahn CR. A cascade of tyrosine autophosphorylation in the beta - subunit activates the phosphotransferase of the insulin receptor. The Journal of Biological Chemistry.1988; 263 (6): 2969 - 2980.

[8] DeLano WL. Unraveling hot spots in binding interfaces: progress and challenges. Current Opinion in Structural Biology.2002; 12 (1): 14 - 20.

[9] Morrow JK, Zhang S. Computational prediction of protein hot spot residues. Current pharmaceutical design.2012; 18 (9): 1255 - 1265.

[10] Verkhivker GM, Bouzida D, Gehlhaar DK, Rejto PA, Freer ST, Rose PW. Computational detection of the binding - site hot spot at the remodeled human growth hormone - receptor interface. Proteins.2003; 53 (2): 201 - 219.

[11] Moreira IS, Fernandes PA, Ramos MJ. Hot spots - - a review of the protein - protein interface determinant amino - acid residues. Proteins.2007; 68 (4): 803 - 812.

[12] Lise S, Buchan D, Pontil M, Jones DT. Predictions of hot spot residues at protein - protein interfaces using support vector machines. PloS one.2011; 6 (2): e16774.

[13] Yunn NO, Park M, Park S, Lee J, Noh J, Shin E, et al. A hotspot for enhancing insulin receptor activation revealed by a conformation - specific allosteric aptamer. Nucleic Acids Research.2021; 49 (2): 700 - 712.

[14] Bednar D, Beerens K, Sebestova E, Bendl J, Khare S, Chaloupkova R, et al. FireProt: Energy - and evolution - based computational design of thermostable multiple - point mutants. PLoS Computational Biology.2015; 11 (11): e1004556.

[15] Pavelka A, Chovancova E, Damborsky J. HotSpot Wizard: a web server for identification of hot spots in protein engineering. Nucleic Acids Research.2009; 37 (Web Server issue): W376 - W383.

[16] Prokop Z, Sato Y, Brezovsky J, Mozga T, Chaloupkova R, Koudelakova T, et al. Enantioselectivity of haloalkane dehalogenases and its modulation by surface loop engineering. Angewandte Chemie (International ed. in English).2010; 49 (35): 6111 - 6115.

[17] Bendl J, Stourac J, Sebestova E, Vavra O, Musil M, Brezovsky J, Damborsky J. HotSpot Wizard 2.0: automated design of site - specific mutations and smart libraries in protein engineering. Nucleic Acids Research.2016; 44 (W1): W479 - W487.

[18] Talukdar P, Talapatra SN. Oxy - haemoglobin protein engineering: An automated design for hotspots stability, site - specific mutations and smart libraries by using HotSpot Wizard 2.0 software. International Journal of Advanced Research in Computer Science.2017; 8: 220 - 228.

[19] Sumbalova L, Stourac J, Martinek T, Bednar D, Damborsky J. HotSpot Wizard 3.0: web server for automated design of mutations and smart libraries based on sequence input information. Nucleic Acids Research.2018; 46 (W1): W356 - W362.

[20] Chakravarty S, Talapatra SN. Protein engineering strategies of DNA Gyrase B: An approach through Hotspot Wizard online tool. Internation Journal of Science and Research.2024; 13 (3): 1884 - 1890.

[21] Sebestova E, Bendl J, Brezovsky J, Damborsky J. Computational tools for designing smart libraries. Methods in Molecular Biology (Clifton, N. J.).2014; 1179: 291 - 314.

[22] Richter SN, Giaretta G, Comuzzi V, Leo E, Mitchenall LA, Fisher LM, et al. Hot - spot consensus of fluoroquinolone - mediated DNA cleavage by Gram - negative and Gram - positive type II DNA topoisomerases. Nucleic Acids Research.2007; 35 (18): 6075 - 6085.

[23] Romero PA, Arnold FH. Exploring protein fitness landscapes by directed evolution. Nature reviews. Molecular Cell Biology.2009; 10 (12): 866 - 876.

[24] Currin A, Swainston N, Day PJ, Kell DB. Synthetic biology for the directed evolution of protein biocatalysts: navigating sequence space intelligently. Chemical Society Reviews.2015; 44 (5): 1172 - 1239.

[25] Acevedo - Rocha C, Reetz M, Nov Y. Economical analysis of saturation mutagenesis experiments. Scientific Reports.2015; 5: 10654.

[26] Cama A, De La Luz Sierra M, Ottini L, Kadowaki T, Gorden P, Imperato - Mcginley J, et al. A mutation in the tyrosine kinase domain of the insulin receptor associated with insulin resistance in an obese woman, The Journal of Clinical Endocrinology & Metabolism.1991; 73 (4): 894 - 901.

[27] González - Ruiz D, Gohlke H. Targeting protein - protein interactions with small molecules: challenges and perspectives for computational binding epitope detection and ligand finding. Current Medicinal Chemistry.2006; 13 (22): 2607 - 2625.

[28] Sil K, Talapatra SN. Prediction of site directed mutagenesis of acetylcholinesterase by using Hotspot Wizard tool. Internation Journal of Science and Research.2024; 13 (9): 1597 - 1600.