OriginalResearchArticle Bioinformatics

International Journal of Pharma and Bio Sciences

ISSN
0975-6299

INSILICO ANALYSIS ON C2H2 ZINC-FINGER INTERACTIONS AND EFFECT OF CONSERVATIVE SUBSTITUTION

M VIDHYA[*1]

[1] *DCDS, Indian Institute of Science, Bangalore -560054.*

ABSTRACT

Zinc Finger protein motifs vary in structure and sequence among different superfamilies. They possess stable scaffold and perform highly specialized functions. These transcription factors are stable and rarely undergo conformational changes upon target binding.The multiple finger-like protrusions make tandem contacts with target molecules. The binding properties of these proteins depend on amino acid sequence of finger domains, the linker between fingers, higher order structure formed and the number of fingers making tandem contacts. They are highly versatile in modes of binding, even within the same class. They are found to bind to varying targets as DNA and other proteins.The study analyses variations exhibited in the binding of the C2H2 domain of the protein variant. Binding protein carrying conservatively substituted residues in place of target binding positions are studied. The differences in residues after conservative substitution were statistically differentiable, during clustering of atoms in the protein.

KEYWORDS : *Zif268, Dynamics, Simulation, Residue, Clustering, Regression*

M VIDHYA*

DCDS, Indian Institute of Science, Bangalore -560054

***Corresponding Author**

Received on: 05-01-2017
Revised and Accepted on : 15-02-2017
DOI: http://dx.doi.org/10.22376/ijpbs.2017.8.2.b151-159

INTRODUCTION

Biological macromolecules display great variation in functions performed in an organism. The underlying structure and effective function of participating molecules provide basis for physical and biological normalcy, during development and existence of an organism. Minor changes in residues are found to induce major variations in cellular mechanisms. Mutational variations in molecular systems remain conservative or non-conservative, where changes are hardly or significantly reflected through the biological function.[1] Proteins are highly versatile and bind to numerous molecules including DNA, other proteins and small molecules in numerous ways. Conformation of a protein is guided by the chemical basis of the sequence of the protein. Proteins display a variety of binding modes, which define various orientations of binding partners relative to each other which can be cooperative with multiple binding; or non-cooperative. Interface interactions guide the binding specificity and affinity of interacting partners, with the protein. Binding specificity differentiates every interaction and interacting molecules; with the strength of binding analyzed using the equilibrium constant. Presence of more than one ligand, which bind to protein are found to affect each other's binding reciprocally.[2] Changes occurring in proteins, involve amino acid variations and mutations which form major determinant in altering protein functions. They affect protein folding and stability, the interactions that the protein is involved in, and protein function. Eventually, the expression of the protein and sub cellular localization is also effected.[3,4] Effects of mutation also vary and can be insignificant or detrimental, based on the effect on molecular function. Mutational effects are subject to evolutionary selection pressure. Mutation analysis based on the effects, guide induction or restoration of mutational changes for favourable biological functioning in organisms or avert the deleterious effects.[5,6] Mutations which may be random or sporadic and follow somatic selection, are found to contribute to cancer by activating or inactivating protein function as in case of oncogenes.[7,1] Macromolecular interactions including nucleic acid or protein molecules involve, novel methods adopted by proteins to bind to specific partners. The molecular structure of Zinc Finger (ZnF) motifs, are characterized based on variations regarding structure and binding modes and affinities. Studying the integration of Zinc atoms in single or in multiples, the highly versatile binding modes are exhibited even between members of same class. These motifs rarely undergo conformational changes, as molecules are stable even when bound to target molecules.[8] Zinc-fingers exhibit variation in coordinating Zinc atom by varying the number and type of amino acids involved in Zinc coordination. The proteins are highly specific in recognition of target based on the base pairs. Mutagenesis based on site directed and phage display selection have enhanced the knowledge of binding patterns obtained from X-ray crystallography and NMR studies.[9] Zinc fingers of ZIF268 family were analyzed based on their binding to DNA. Complexes were found to differ in their experimental stability properties. They shared the same overall three-dimensional (3D) structure and exhibited similarity during simulation. DNA binding of proteins reflected various energy states corresponding to dynamically different states. Coupling of energy between residues, change in response to protein or DNA sequence variation; modulates selectivity in recognition and relative importance of various regions during binding. The energy network of the protein, is found to stabilize protein conformations that recognize and bind DNA. Various conformations are found to be effective in the internal dynamics. Zinc fingers are found to exhibit specificity and selectivity for DNA sequences based on variations in energy and dynamic properties while binding to different partners.[10] Zinc finger protein classes are predominated by Cys2His2 zinc fingers, which bind DNA in eukaryote transcription process. Several fingers are found to make tandem contacts along the DNA. With a conserved beta-beta-alpha structure, amino acids on the surface of the helix form the recognition structure for interaction with DNA. Beyond the modular structure, they possess the ability to recognise and bind varying targets with high specificity. The protein exhibits specific patterns in interactions, which are found to be recurring in side-chain based interactions. Modularity however has not limited the target recognition capacity based on a single “recognition code” which enable design of optimal protein for any desired target binding.[11] The 3D structures of Zif268-DNA complexes reveal residues at critical positions -1, 2, 3 and 6 in the start of recognition helix, making maximum contacts to the DNA stretch from helix 1.[12,13] Variation in sequence of protein carried out for analysis of binding specificity were verified through phage display approaches for preparing a library of variants randomizing the critical amino acids of first zinc finger.[12] DNA sequences were selected based on affinity which was driven by the variation in base in the region recognized by the protein. This specifically bound the protein to a specific site in the DNA. Comparison was made with crystal structures of the complexes between wild type and mutated proteins, with target sequences retaining overall similarity. Dissociation constants were used and the complexes are differentiated. Specific positions are identified in the protein as residues at -1, 2, 3 and 6 position of the recognition helix which interdependently participate in DNA recognition.[14,15,16] Presence of particular set of residues in various positions, enable specific interaction and strong predictive power for proteins in recognition of target. Arginine at position -1 during target recognition show high specificity for interaction with Guanine of target DNA. Base recognition is enabled with each amino acid present at binding position, identifying the single base specifically based on context dependent manner. It involves side chain-side chain interactions, sequence dependent conformation flexibility of DNA, variaion of interaction due to presence of water, mutual interaction effects of residues or transmitted effects due to other local changes or interactions.[16]The dissociation constant of proteins with similar binding residues at the recognition helix as DSNR in 1A1F and wild type 1A1G, forming base contacts with GACC and GCGT sequences of target DNA respectively; are determined as 0.019(Kd, nM) for 1A1F complex and 1.8 (Kd, nM) for 1A1G complex. Recognition of DNA by Cys2His2 zinc fingers, enable identification of specific interactions among zinc finger proteins based on phage display. Molecular dynamic simulation reveals, variations in RMSF (Residue-based Root Mean Square Fluctuations). For the proteins, calculation is carried out on the C-alpha atoms. DNA fluctuation, were studied based on the

fluctuation of the C1 atoms in the scale of nm. While total Protein Flexibility (RMSF sum (nm)) was found to be 7.18 and Total Complex Flexibility (RMSF sum (nm)) was 9.49 for 1A1F complex, 1A1G exhibited total protein flexibility as 6.05 and 8.70 for total complex flexibility.[17,18] The network of interacting components in proteins have been analyzed to differentiate various aspects of macromolecules, including structure flexibility, domain folding, recurring structural patterns, the important residues involved in folding, fluctuation in residues, side chain clusters etc. [19-23] Analysis of macromolecular systems, is done based on interactions among the network components, for characterization of the whole system along with its components.

MATERIALS AND METHODS

Macromolecular Structure

Input coordinates for the structure of Zinc finger protein variant(Zif268), bound to DNA at GACC site was taken from PDB. The DSNR active site variant was obtained as X-ray diffraction structure with resolution 2.1 Å. The quality of atomic model was determined based on R values where, the values were found to be comparable to a typical value of 0.20. The free R value was found to be 0.270 while the R-value work was found to be reduced to 0.225. R values obtained for the complexes are much lesser than random fit value of 0.63 and higher than 0 indicating, the molecular conformation is not randomized and neither provides a perfect fit.[24] Compound 1A1G, which is also DSNR binding site variant of ZIF268 bound to GCGT site of DNA. The resolution of the wild type molecule was obtained at 1.9 Å. The free R value of 0.275 and R-value work of 0.215 indicated normalcy of values in case of typical structures and structures are marginally organized based on atomic conformations of the molecule. Molecules carried mutations from Arginine to Aspartic acid at 118th position (R118D), Aspartic acid to Serine at residue position 120 (D120S), and Glutamic acid to Asparagine in 121st position (E121N). The structures were energy minimized to obtain most stable conformation of the complex with DNA. Structural characterization of zinc finger complexes is done based on physical properties, active site presence and molecular properties. Analyses of structural complexes done with Ramachandran plot, express extent of agreement of backbone conformations of all residues. The analysis verifies correspondence of known allowed areas in the Ramachandran plot, and the values if they are within the expected ranges for well refined structures. The evaluation of structures using Ramachandran plot, verified deviations that the backbone conformations of all residues in protein was in correspondence to known allowed areas within expected range for well-defined structures.[25,26] The variants (DSNR) of Zif268 and other variants were differentiable based on energetic coupling of residues during MD simulations.

Molecular Dynamics

Complexes exhibiting specificity in interaction with changes in interacting components of the target, as bases of DNA were taken for analysis. 3D structures of the protein were obtained from structure repository for protein structures, PDB at RCSB protein databank. The coordinates for the molecules were taken from the data entries1A1F and 1A1G from the repository. Both the molecules were energy minimized. RMSD was calculated using Swiss PDB viewer.[27] Differences were found to be below 1 Å, between separated DNAs and amongst protein complexes. The complexes were subjected to geometric optimization through simple gradient methods and conjugate gradient methods, with the frequency 1 for 1000 respectively; by simulating molecules based on the molecular mechanical force field AMBER03,[28] using Gromacs,[29] MD package. The radius cut off of 1 Å was used during energy minimization. Optimization is performed through energy minimization, by subjecting the molecule to 100 picoseconds of simulation. Deformation of molecule is reduced and major shape based variations were avoided; by restraining the atoms to a fixed reference position, before effectuating MD simulations. Position restraining is carried out by subjecting the molecule to coupling as Berendsen thermostat. The reference temperature is kept at 300K. Protein, DNA, ion and solvent groups are included in the analysis. Lincs algorithm,[30] is adopted for constraining all bonds to provide stability during further simulations. The starting configuration after minimization is kept constrained, and the bonds restricted from rotation over more than 90 degrees in a single step. Complexes are subjected to MD simulation under the force field AMBER03. Water molecules are included explicitly, in the simulation box before simulations are carried out for a time period of 10ns. Complexes are analyzed based on the thermodynamic energy contribution, both after complexation and simulation of the complexes. The average energy variation during MD simulation among non-bonded interactions and total energy are determined. Variations that occur are examined for van der Waals and electrostatic interaction, relative to the variation in total energy of complexes during simulation.

Residue Interaction Network

Proteins are capable of performing vast array of functions with high specificity, which is identified to be directly related to their structural conformations. The conformations are totally dependent on complex interactions among its constituent amino acid residues. Versatility of proteins, their dynamics and structural specifications, are calculated through analysis of interaction among residues, with graph theoretical approaches. This uses amino acid residue interaction networks (RIN) or residue interaction graphs (RIG), the protein contact networks (PCN) or residue contact networks (RCN). In an interaction network (RIN), the PCN and RCN are constructed based on C-alpha atom of each amino acid residue as a node. The Euclidean distance dE between C-alpha atoms are then analyzed to determine presence of an edge between two residues. If dE between C-alpha atoms of the pair of amino acid falls within a given cut-off distance, an edge is perceived to exist.[31]Complexes analyzed for number of nodes and edges, indicated presence of 107 Nodes and 417 Edges. The interactions were analyzed with closest network policy, with multiple interaction types in the molecular complex. Heteroatoms are retained but water molecules eliminated during analysis.Residue Interaction Network is further analyzed to determine the characteristics of the cluster. Average clustering coefficient of clusters containing number of node from 2 to 12 are prepared with complex structural conformation including initial coordinates prior to dynamics, structural

conformations of complexes assumed at different stages of dynamics simulations till 10ns.[32]Evaluation of the clustering coefficient for different groups containing different number of nodes, are compared based on the mean obtained for cluster from 0 to 10 ns. The standard deviation of values, mean obtained in clusters with particular number of nodes, the variance standard deviation, population standard deviaion are determined to differentiate variaions among clusters, containing varying number of nodes and their comparison during simulation. Regression analysis was done to determine the correlation among the mean values obtained for clustering coefficient, for each group containing 2 to 12 nodes in cluster, for each of the complex conformations of zinc finger variants. Molecules are analyzed using visualization tool UCSF chimera, and interactions at the binding site were determined. R-squared values are calculated to determine how close the data are to the regression line obtained. The comparison of error variance relative to variance of dependent variable, the mean values of average clustering coefficient calculated on the basis of number of nodes in cluster, are determined.[3]

RESULTS AND DISCUSSION

Macromolecular structure

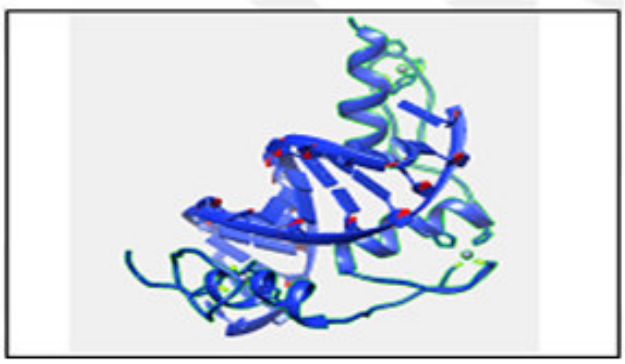

Figure 1
1A1F Znf268 - DSNR site

Structural conformations of ZIF268 variants taken from PDB, analyzed for their atomic model

Based on R values, indicated comparable R values to the typical value of 0.20. Free R values of 0.270 and 0.215 was found for 1A1F and 1A1G respectively. The molecular conformations are taken for analysis in case of both Zif268 variants, and are identified to be reasonably fitting protein ligand complex.Both the complexes contained Cys2His2 (C2H2) scaffold. The binding residues included at the binding site DSNR (Asp-Ser-Asp-Arg) identified the target DNA specifically. [Fig 1]The Zif268 variant 1A1F possessed a molecular weight of 9948.66 Da for the protein molecule and 6655.41 Da, for the nucleic acid chains bound. Protein molecule consisted of 84 amino acids and 22 nucleic acid molecules identified at the experimental pH of 7.5. Comparison of the molecular conformation information indicate, absence of starting and end of chain residues for protein. Starting position Met and Glu were found missing and terminal Arg, Gln, Lys and Asp were found to be complete in comparison to PDB seqres records. Residue at position 123, Thr123 was found to have alternate conformations. Three active sites were identified in protein chain as active site 1 with residues (A: CYS107- A:CYS112- A:HIS125- A:HIS129), active site 2 with residues (A:CYS137- A:CYS140- A:HIS153- A:HIS157) and active site 3 formed with residues. The variant of Zif268 1A1G, carried molecular weights 9948.66 Da and 6655.41 Da for protein and nucleic acid subunits respectively, which are identified at an experimental pH of 8.0. The molecule contained 155 water molecules and few residues like Met and Glu are missing in the start of the protein chain. Arg, Gln, Lys and Asp are found to be missing at the end of the chain. The other parts of the molecule including the DNA bases are found to be complete. Three active sites were identified in protein chain as active site 1 (A:CYS107- A:CYS112- A:HIS125- A:HIS129), active site 2 (A:CYS137- A:CYS140- A:HIS153- A:157), active site 3 (A:CYS165- A:CYS168- A:HIS181- A:HIS185). Ramachandran score for the structures are found to be within expected range for well refined structures.They were obtined as :- -0.589 for 1A1F complex and -0.641 1A1G complex.The average B-factor occupancy for 1A1F coordinates was found to be 36.352 which exhibited a standard deviaion of 9.175. RMS is found to be 37.492. The macromolecule exhibited an occupancy of 0.998. The radius of gyration is found to be 15.5 Å. In case of 1A1G complex, the B-factor was 34.495 with a standard deviaion of 19.043. The RMS is predicted as 39.402. An average occupancy of 1.0 is identified. The complex exhibited a radius of gyration of 15.6 Å.Both the complexes exhibited small variaions in their binding parameter

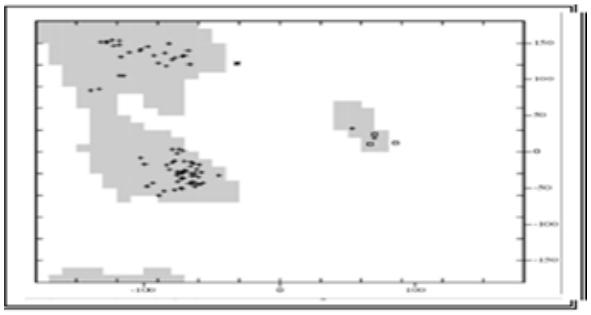

Figure 2
Zif268 - 1A1F variant Ramachandran Plot analysis

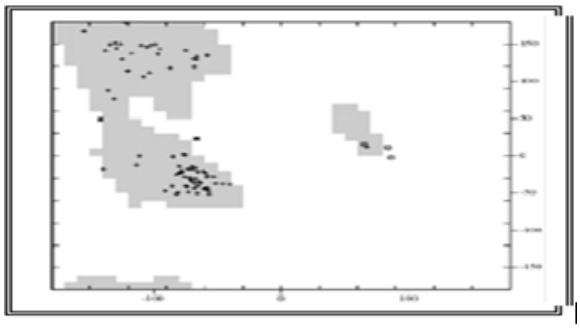

Figure 3
Zif268- 1A1G variant Ramachandran plot analysis

From the set of residues verified based on Ramachandran plot obtained for complex 1A1F, 79 of the residues out of 84 were verified. 78 residues are found to be in core region and one residue found to be an outlier which amount to 1.3 percent. No residues were found to be missing any atom. Analysis of structure of 1A1G protein in complex revealed two outliers, and 77 residues in core regions in the complex of protein-DNA. 79 out of 84 residues are checked in the complex and percentage outliers found to be 2.5. No atoms were found to be missing in any of the residues. [Fig 2, 3]

Molecular Dynamics

based on change in interacting components of the .

target, the bases of DNA. Coordinates of Zif268 variants 1A1F and 1A1G are energy minimized for further analysis. RMSD indicated differences below 1Å between separated DNAs and amongst protein complexes.The variants of protein DNA complexestaken, shared the same overall three-dimensional structure. These variaions were identified as thermodynamic differences, exhibited by complexes during simulation. The results obtained after 10ns of MD simulation for the geometrically optimized structures, are analyzed for total energy, van der Waals and electrostatic contributions. [Table1] The complexes exhibited specificity in interaction

PROTEIN	LJ(SR)	Coulomb(SR)	Total Energy
1A1F	-3327.52	-40614.1	-28651.6
1A1G	-2886.08	-41329.2	-28480.7

Table 1
Energy contribution of non-bonded interaction in complexes of Zif268

Analysis reveals, change in total energy is greatly influenced by contribution from electrostatic energy in each complex. The average energy variation in non-bonded energy, exhibited by the short range (SR) interactions as the vdW (van der Waals-LJ) and electrostatic (coulomb), in comparison to total energy obtained for various complexes, reveal higher distinction between vdW and electrostatic energy among the complexes. The molecule is found to optimize itself during binding with less of van der Waals variaion, in comparison to relatively higher negative energy variation obtained for electrostatic energy.Docked complex structures of Zif268 variants mutated at their binding site residues, the 'DSNR' site and energy minimized are taken for analysis. A dataset of complexes, is prepared with mutations in the Asp residue at 118^{th} position, Serine in 120^{th} position, Asparagine in the 21^{st} position and Arginine in 124^{th} position. Mutations are carried out individually, maintaining conservation in residues. Complexes with conservative substituents are prepared. Mutations are carried out in all the four positions in the recognition helix as 118^{th}, 120^{th} , 121^{st}, and 124^{th} position. Mutated complexes contained Gln, Glu, Asn in 118^{th} position in place of Asp; Asn, Cys, Gln, Thr and Tyr in place of Ser at 120^{th} position, Gln in place of Asn at 121^{st} position and His and Lys replacing Arg at 124^{th} position in complexes prepared from both Zif268 variants 1A1F and 1A1G variants.

Residue Interaction Network

Network of interacting residues, is studied for each structure in the complex dataset, and statistical analysis based on average clustering coefficient is done to determine the clustering variation exhibited by clusters comprising varying number of nodes. The number of nodes range from 2 to 12, among various structural conformations obtained from simulations. Network of interacting residues RIN are prepared for the complexes 1A1F and 1A1G using closest neighbour network policy forming multiple interactions. Complex network of both the molecules are analyzed and the network is identified, to contain 207 nodes and 417 edges.Detailed analysis of the complex network focusing on RIN is carried out for both Zif268 variants. RIN created using Cytoscape analysis software through integration with visualization software UCSF Chimera 1.10.2 is analyzed.Network of interaction residues is prepared for complexes of Zif268 variants and their mutant complexes in Cytoscape 3.0. Network of complexes are taken as undirected network for analysis of conformation based on interactions of components in network. Average clustering coefficient of nodes analyzed for each of the complex networks. Clustering coefficient of groups are differentiated based on number of nodes in clusters ranging from 2 to 12, show variations for each conservative substitution.Complexes with similar cluster groups contain substituents Gln, Glu at 118, His, Lys at 124th position, Asn substituted complex at 120 position and substitution of Asn by Gln in 121 position. The substitutions at different positions showed similar variation with respect to network clustering coefficient. Substitution of Serine at position 120 with Tyr provided RIN coordinates whose average clustering coefficient was differentiable from clusters of all other complexes, contained different substitutions at different positions. The average clustering coefficient is higher for all clusters with varying number of nodes, in case of Tyrosine substitution of Serine, at 120th position. For complexes with similar substitutions, the position of mutation brought about changes in average clustering coefficient which is reflected in their mean values as exhibited in case of replacement of Asp at 118th position and Ser at 120th position. Asn and Gln at 120th position are highly similar and provide similar clustering in RIN. Substitution of Asn with Gln provides great variability in RIN network in comparison to Gln substituted at 118th or 120th position. Cys substitution at 120th position exhibit close networking interaction as shown by Gln substituent at the same position.The plugin StructureViz in Cytoscape linked visualization of biological networks in Cytoscape to molecular structures loaded in Chimera. Molecular Structures are associated in the complex using node attributes.Incase of substitutions in Zif268 variant 1A1G, clustering in mutant complexes exhibit similar clustering with number of nodes 2,3 in all the cases

.

Mean		118Asp				120Ser			121Asn	124Arg	
	Gln	Glu	Asn	Asn	Cys	Gln	Thr	Tyr	Gln	His	Lys
2	0.56016	0.56016	0.56016	0.56016	0.57405	0.57405	0.57405	0.58794	0.56016	0.59257	0.59257
3	0.51917	0.51917	0.52661	0.53190	0.53190	0.53190	0.52727	0.53653	0.53190	0.51570	0.51570
4	0.50799	0.50799	0.50425	0.50852	0.50601	0.50959	0.50746	0.52156	0.50501	0.51105	0.51105
5	0.54847	0.54847	0.54589	0.54268	0.54404	0.54268	0.54443	0.55700	0.54404	0.53954	0.54092
6	0.54281	0.54248	0.54382	0.54536	0.54487	0.54536	0.54145	0.55744	0.54487	0.54140	0.54065
7	0.43581	0.43561	0.43528	0.43958	0.43925	0.43958	0.43465	0.44989	0.43925	0.43665	0.43652
8	0.41413	0.41280	0.41293	0.41144	0.41188	0.41144	0.41074	0.42254	0.41165	0.41190	0.41264
9	0.34578	0.34672	0.34672	0.34828	0.34822	0.34810	0.34733	0.35879	0.34913	0.34782	0.34753
10	0.33137	0.33137	0.33137	0.33147	0.33151	0.33120	0.33141	0.33675	0.33120	0.33112	0.33120
11	0.31419	0.31419	0.31419	0.31285	0.31290	0.31285	0.31358	0.32237	0.31285	0.31202	0.31202
12	0.29764	0.29764	0.29764	0.30270	0.30270	0.30270	0.29954	0.31122	0.30312	0.30124	0.30124

Table 2

1A1F: clustering coefficient for cluster with mutated protein binding residues

Mean		118Asp				120Ser			121Asn	124Arg	
	Gln	Glu	Asn	Asn	Cys	Gln	Thr	Tyr	Gln	His	Lys
2	0.50916	0.50916	0.50916	0.50916	0.50916	0.50916	0.50916	0.50916	0.50916	0.50916	0.50916
3	0.42594	0.42594	0.42594	0.42594	0.42594	0.42594	0.42594	0.43057	0.42594	0.42594	0.42594
4	0.42058	0.41595	0.41595	0.42058	0.42058	0.42058	0.42058	0.43447	0.42058	0.42058	0.42058
5	0.43565	0.43219	0.43200	0.43679	0.43565	0.43810	0.43565	0.44544	0.43565	0.43679	0.43679
6	0.45788	0.45664	0.45695	0.45723	0.45788	0.45720	0.45788	0.46896	0.45762	0.45585	0.45723
7	0.40811	0.40723	0.40674	0.40833	0.40811	0.40856	0.40811	0.42063	0.40922	0.40961	0.40887
8	0.38689	0.38292	0.38400	0.38616	0.38689	0.38616	0.38689	0.39667	0.38681	0.38727	0.38713
9	0.35602	0.35662	0.35662	0.35602	0.35602	0.35602	0.35602	0.36661	0.35602	0.35863	0.35937
10	0.31856	0.31781	0.31781	0.31856	0.31856	0.31860	0.31856	0.29262	0.31856	0.31771	0.31743
11	0.29085	0.29068	0.29089	0.29085	0.29085	0.29085	0.29085	0.30236	0.29100	0.29252	0.29260
12	0.28720	0.31781	0.28468	0.28720	0.28720	0.28720	0.28720	0.29573	0.28720	0.28720	0.28720

Table 3

1A1G: Clustering coefficient for cluster with mutated protein binding residues

The clustering of residues is not similar in case of substitution of Ser with Tyr at 120th position, while the other substitutions showed lesser variation, in terms of differences in values exhibited by average clustering coefficient for clusters containing varying number of nodes.The change of mean of average clustering coefficient, is studied based on regression model for data. Correlation of number of nodes in cluster against deviation in mean of average clustering coefficient is taken for each of the mutants. Results indicate, good fit for models based on linear regression. Change in clustering coefficient for different clusters, involving conservative mutations at particular positions is evident in their linear regression analysis and also varied the R2 values.Analysis of results indicate, similar variations in regression analysis for both Zif268 variants with respect to regression analysis and R-squared values. Structures with substitutions at 118th position as Gln and Glu, along with 121 Asn to Gln mutations are found to be relatively more differentiable based on their average clustering coefficient involving clusters formed by 2 to 12 nodes. Asparagine substitutions are differentiable at 118 Asp and 120 Serine positions. Position 120 mutants, Cys Gln Thr, and mutations of Arg 124 to His and Lys indicate similar changes, during best fit analysis. Variation analysis of similarity among substituent clustering coefficient in complex of Zif268 variants, reveal better fit for clustering coefficient values, reflecting higher similarity among complexes. [Table 2, Table 3]

		1A1F		1A1G	
Position	Substitution	Regression equation	R2	Regression equation	R2
118 Asp	Gln	-0.03x + 0.61	0.88	-0.02x + 0.51	0.87
118 Asp	Glu	-0.03x + 0.61	0.88	-0.02x + 0.50	0.84
118 Asp	Asn	-0.03x + 0.61	0.89	-0.02x + 0.51	0.87
120 Ser	Asn	-0.03x + 0.62	0.89	-0.02x + 0.51	0.87
120 Ser	Cys	-0.03x + 0.62	0.9	-0.02x + 0.51	0.87
120 Ser	Gln	-0.03x + 0.62	0.9	-0.02x + 0.51	0.87
120 Ser	Thr	-0.03x + 0.62	0.9	-0.02x + 0.51	0.87
120 Ser	Tyr	-0.03x + 0.62	0.89	-0.02x + 0.52	0.84
121 Asn	Gln	-0.03x + 0.61	0.89	-0.02x + 0.51	0.87
124 Arg	His	-0.03x + 0.62	0.9	-0.02x + 0.51	0.87

Table 4

Regression analysis and R-squared analysis of Zif268 substituents

The trend line equation is used to compare dependent and independent variables. Change in residues at specific positions along with the replacements, reflected the linear regression. [Table 4] R2 computed, determined minor variations in regression with each mutation at specific positions. The coefficient of determination R-squared indicated variations, showed higher similarity or closeness in case of similar and more conservative substitutions at given position. The substitutions at different position indicated variation in linear regression analysis.

CONCLUSION

Macromolecules are primarily focused, for interference in the biological system to implement changes, that will be beneficial to the proper functioning of the system. Proteins are focused on, for their versatility, specificity and molecular properties which enable easy and effective manipulation or induction of desirable change. High specificity identified among Zinc-finger protein variants for their target, and differences in intra-protein energy networks identified for the protein, share overall structural and sequence similarity. Static and dynamic interactions, driven by potential and kinetic energy whose variations are controlled by bonded and non-bonded interactions differentiate proteins based on their properties.Zif268 protein which are identified to be highly specific in their interaction and bind with high affinity to their target, show stability in dynamic interactions in complex over a period of time. The differences exhibited in complexes are differentiable based on their average clustering coefficient, the position and residue, that form substituents at specific positions.Variations among the R-squared value, indicate differences in spite of high similarity in structure and sequence among the residues. Conservative mutations in specific positions are found to affect residue interactions based on position and substituent residues, which contribute to overall protein conformation and stability of the molecular complex.

FUNDING/ACKNOWLEDGEMENT

The resources and financial support for the study was provided by the Department of Science and Technology, INDIA at Biomolecular Computation Laboratory, DCDS (Bioinformatics Centre), IISc, Bangalore, under project SR/WOS-A/LS-85/2010; is acknowledged. The generous support for carrying out the study at Biomolecular Computation Laboratory, DCDS, IISc, Bangalore, is acknowledged. The support and guidance provided by Dr. Debnath Pal, Associate Professor, Biomolecular Computation Laboratory, DCDS, IISc, Bangalore; is thankfully acknowledged.

CONFLICT OF INTEREST

Conflict of interest declared none.

REFERENCES

1. Reva B, Antipin Y, Sander C. Predicting the functional impact of protein mutations: application to cancer genomics. Nucleic acids research. 2011 Jul 3:gkr407.
2. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. Protein function. 2002.
3. Tiede S, Cantz M, Spranger J, Braulke T. Missense mutation in the N‐acetylglucosamine‐1‐phosphotransferase gene (GNPTA) in a patient with mucolipidosis II induces changes in the size and cellular distribution of GNPTG. Human mutation. 2006 Aug 1;27(8):830-1.
4. Saha S, Chatterjee P. Protein function prediction from protein interaction network using physico-chemical properties of amino acids. International Journal of pharmacy and Biological Sciences, India. 2014;4(2):55-65.
5. Cairns J. Mutation selection and the natural history of cancer. Nature. 1975 May 15;255(5505):197-200.
6. Hanahan D, Weinberg RA. The hallmarks of cancer. cell. 2000 Jan 7;100(1):57-70.
7. Croce CM. Oncogenes and cancer. New England Journal of Medicine. 2008 Jan 31;358(5):502-11.
8. Klug A. The discovery of zinc fingers and their applications in gene regulation and genome manipulation. Annual review of biochemistry. 2010 Jul 7;79:213-31.
9. Leon O, Roth M. Zinc fingers: DNA binding and protein-protein interactions. Biological research. 2000;33(1):21-30.
10. Torella R, Moroni E, Caselle M, Morra G, Colombo G. Investigating dynamic and energetic determinants of protein nucleic acid recognition: analysis of the zinc finger zif268-DNA complexes. BMC structural biology. 2010 Nov 24;10(1):42.
11. Wolfe SA, Nekludova L, Pabo CO. DNA recognition by Cys2His2 zinc finger proteins. Annual review of biophysics and biomolecular structure. 2000 Jun;29(1):183-212.
12. Rebar EJ, Pabo CO. Zinc finger phage: affinity selection of fingers with new DNA-binding specificities. Science-AAAS-Weekly Paper Edition-including Guide to Scientific Information. 1994 Feb 4;263(5147):671-3.
13. Elrod-Erickson M, Benson TE, Pabo CO. High-resolution structures of variant Zif268–DNA complexes: implications for understanding zinc finger–DNA recognition. Structure. 1998 Apr 15;6(4):451-64.
14. Choo Y, Klug A. Physical basis of a protein-DNA recognition code. Current opinion in structural biology. 1997 Feb 1;7(1):117-25.
15. Desjarlais JR, Berg JM. Redesigning the DNA‐binding specificity of a zinc finger protein: A data base‐guided approach. Proteins: Structure, Function, and Bioinformatics. 1992 Feb 1;12(2):101-4.
16. Desjarlais JR, Berg JM. Toward rules relating zinc finger protein sequences and DNA binding site preferences. Proceedings of the National Academy of Sciences. 1992 Aug 15;89(16):7345-9.
17. Zhao D, Huang Z. Recognition code of ZNF191 (243-368) and its interaction with DNA. Bioinorganic chemistry and applications. 2015 Sep 20;2015.
18. Amitai G, Shemesh A, Sitbon E, Shklar M, Netanely D, Venger I, Pietrokovski S. Network analysis of protein structures identifies functional residues. Journal of molecular biology. 2004 Dec 3;344(4):1135-46.
19. Jacobs DJ, Rader AJ, Kuhn LA, Thorpe MF. Protein flexibility predictions using graph theory. Proteins: Structure, Function, and Bioinformatics. 2001 Aug 1;44(2):150-65.
20. Dokholyan NV, Li L, Ding F, Shakhnovich EI. Topological determinants of protein folding. Proceedings of the National Academy of Sciences. 2002 Jun 25;99(13):8637-41.

21. Wangikar PP, Tendulkar AV, Ramya S, Mali DN, Sarawagi S. Functional sites in protein families uncovered via an objective and automated graph theoretic approach. Journal of molecular biology. 2003 Feb 21;326(3):955-78.
22. Vendruscolo M, Dokholyan NV, Paci E, Karplus M. Small-world view of the amino acids that play a key role in protein folding. Physical Review E. 2002 Jun 25;65(6):061910.
23. Atilgan AR, Akan P, Baysal C. Small-world communication of residues and significance for protein dynamics. Biophysical journal. 2004 Jan 31;86(1):85-91.
24. Studio D. version 2.5, Accelrys. Inc., San Diego, CA. 2009;92121.
25. Ravcheev DA, Li X, Latif H, Zengler K, Leyn SA, Korostelev YD, Kazakov AE, Novichkov PS, Osterman AL, Rodionov DA. Transcriptional regulation of central carbon and energy metabolism in bacteria by redox-responsive repressor Rex. Journal of bacteriology. 2012 Mar 1;194(5):1145-57.
26. Vriend G. WHAT IF: a molecular modeling and drug design program. Journal of molecular graphics. 1990 Mar 1;8(1):52-6.
27. Guex N, Peitsch MC. SWISS‐MODEL and the Swiss‐Pdb Viewer: an environment for comparative protein modeling. electrophoresis. 1997 Jan 1;18(15):2714-23.
28. Duan Y, Wu C, Chowdhury S, Lee MC, Xiong G, Zhang W, Yang R, Cieplak P, Luo R, Lee T, Caldwell J. A point‐charge force field for molecular mechanics simulations of proteins based on condensed‐phase quantum mechanical calculations. Journal of computational chemistry. 2003 Dec 1;24(16):1999-2012.
29. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ. GROMACS: fast, flexible, and free. Journal of computational chemistry. 2005 Dec 1;26(16):1701-18.
30. Hess B, Bekker H, Berendsen HJ, Fraaije JG. LINCS: a linear constraint solver for molecular simulations. Journal of computational chemistry. 1997 Sep 1;18(12):1463-72.
31. Grewal RK, Mitra D, Roy S. Mapping networks of light–dark transition in LOV photoreceptors. Bioinformatics. 2015 Nov 15;31(22):3608-16.
32. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome research. 2003 Nov 1;13(11):2498-504.

Reviewers of this article

Dr. Amab Roy

Senior Research Scientist, Indian Institute of Science, Bangalore - 560012, India

Mr. Anubrata Paul M.Sc. Biotech (Research)

Department of Biotechnology, Natural Products Research Laboratory, Centre for Drug Design Discovery & Development (C-4D) , SRM University, Delhi-NCR,Sonepat.

Prof.Dr.K.Suriaprabha

Asst. Editor , International Journal of Pharma and Bio sciences.

Prof.P.Muthuprasanna

Managing Editor , International Journal of Pharma and Bio sciences.

We sincerely thank the above reviewers for peer reviewing the manuscript