**Supplementary materials**

**1. Average property value for protein sequences**

Average property value for any protein is calculated using following formula:

<Property>protein = (1)

Where, propertyi is the property value of each amino acid residue in a protein and N is the total number of amino acid residues in a protein.

**2. Change in property values upon amino acid substitutions**

The change in property value upon amino acid mutation is calculated using formula:

∆Pmutation = Psnp - Pwild-type  (2)

Where, Pwild-type and Psnp are the property values of wild-type and mutant residues, respectively and ∆Pmutation is the change in property due to mutation.

**3. Average property value for nucleic acid sequences**

Several experimental property values are reported for dinucleotides and hence, we split the DNA sequence into overlapping dinucleotides and the average property value is obtained using the formula

<Property>DNA = (3)

Where, N is the number of nucleotides in a sequence and prop(di) is the property value of dinucleotides.

**4. Example for computing property values and change due to mutation in DNA**

**Supplementary Figure S2** illustrates the steps for computing property values and change due to mutation in DNA

1. Input or upload Sequence in FASTA format (rs6647448).
2. Enter Mutation with position for example “C501T” here “C” is mutated to “T” at position 501.
3. Choose the property (Physicochemical properties, Conformational properties or Nucleotide content) and then click on submit.
4. The output shows the property along with unit and value. The results can be downloaded by clicking on “download now” button.

**5. Applications of the server**

We have illustrated the application of the server using an example to relate the change in amino acid properties with stability change upon mutation. The experimental change in stability values and the change in property values for a set of mutants are shown in **Supplementary Table S1**. The properties hydrophobicity, polarity, short and medium-range interaction energy, alpha helical propensity and beta-strand propensity showed a correlation of 0.81, -0.21, 0.14, 0.42 and 0.62, respectively. This analysis revealed that hydrophobicity (Hp) plays an important role to stability as reported in the literature [1]. Likewise, one can infer the importance of specific properties to protein structure and function such as change in folding rates [2], unfolding rates [3], aggregation propensity [4], protein disorder [5] and aggregation rate [6], disease causing mutations [7] and so on.

Further, the output of the tool can also be used as an input to the machine learning algorithms implemented in WEKA [8] with proper headers.

**6. Statistics of features**

Seq2feature computes various sequence based features such as 130 amino acid properties, 28 substitution matrices, 94 pairwise contact potentials from proteins and 16 physicochemical and 18 conformational properties as well as 8 nucleotide contents from DNA sequence. The statistics is presented in **Supplementary Table S2**

**7. Possible ways to reduce the number of features**

Several procedures have been followed to reduce the number of features in prediction algorithms: (i) reduce the features, which are related to each other. In this procedures if two properties show a correlation of more than any specific cutoff (E.g. 0.85) one of the properties is retained in the model [2], (ii) choose the properties depending on the nature of the problem, for example, consider physicochemical, energetic and conformational properties to relate the stability of a protein upon mutation at buried, partially buried and exposed regions [1], (iii) utilize feature selection algorithms such as “WrapperSubsetEval” and “Rankersearch” available in WEKA package [8], (iv) forward feature selection, starting with minimum number of properties and increase one by one till saturation [6] and (v) backward feature selection, starting with all properties and reduce one by one, which do not change the performance.

**References**

[1] Gromiha MM, Oobatake M, Kono H, Uedaira H, Sarai A. Role of structural and sequence information in the prediction of protein stability changes: comparison between buried and partially buried mutations. Protein Eng. 1999, 12(7):549-55.

[2] Chaudhary P, Naganathan AN, Gromiha MM. Folding RaCe: a robust method for predicting changes in protein folding rates upon point mutations. Bioinformatics. 2015, 31(13):2091-7

[3] Chaudhary P, Naganathan AN, Gromiha MM. Prediction of change in protein unfolding rates upon point mutations in two state proteins. Biochim Biophys Acta. 2016, 1864(9):1104-1109.

[4] Tartaglia, GGn Cavalli A, Pellarin R, Caflisch A. The role of aromaticity, exposed surface, and dipole moment in determining protein aggregation rates. Protein Sci, 2004, 13(7), 1939-1941.

[5] Anoosha P, Sakthivel R, Gromiha MM. Prediction of protein disorder on amino acid substitutions. Anal Biochem. 2015, 491:18-22

[6] Rawat P, Kumar S, Gromiha MM. An in-silico method for identifying aggregation rate enhancer and mitigator mutations in proteins. Int J Biol Macromol. 2018;118(Pt A):1157-1167.

[7] Kulandaisamy A, Priya SB, Sakthivel R, Frishman D, Gromiha MM. Statistical analysis of disease-causing and neutral mutations in human membrane proteins. Proteins. 2019, 87(6):452-466.

[8] Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH (2009) The WEKA Data mining software: an update. SIGKDD Explorations Newsleter 11, 10–18.

**Supplementary Tables**

**Supplementary Table S1**

**Relationship between experimental stability change and**

**Property change upon mutation obtained from Seq2Feature**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **mutation** | **Experimental G** | **Hp** | **Pb** | **Pa** | **Esm** | **P** |
| E49G | -1.7 | 0.79 | 0.38 | -0.94 | 0.2 | -49.9 |
| E49A | -0.3 | 1.64 | 0.46 | -0.09 | 0.24 | -49.9 |
| E49V | 3.2 | 3.32 | 1.33 | -0.45 | 0.09 | -49.77 |
| E49I | 8 | 3.93 | 1.23 | -0.43 | 0.03 | -49.8 |
| E49L | 6.2 | 2.78 | 0.93 | -0.3 | 0.16 | -49.77 |
| E49P | -0.6 | -0.35 | 0.18 | -0.94 | 0.08 | -48.32 |
| E49Y | 0 | 2.16 | 1.1 | -0.82 | -0.13 | -48.29 |
| E49F | 2.4 | 2.48 | 1.01 | -0.38 | -0.02 | -49.55 |
| E49W | 1.1 | 2.55 | 1 | -0.43 | -0.13 | -47.8 |
| E49H | 1.3 | 1.01 | 0.5 | -0.51 | 0.06 | 1.7 |
| E49K | -0.9 | -0.4 | 0.37 | -0.35 | -0.09 | -0.4 |
| E49N | -0.6 | 0.31 | 0.52 | -0.84 | 0.02 | -46.52 |
| E49Q | -2.5 | 0.37 | 0.73 | -0.4 | -0.04 | -46.37 |
| E49D | -0.3 | -0.31 | 0.17 | -0.5 | 0 | -0.2 |
| E49C | 2.2 | 2.89 | 0.82 | -0.81 | 0.21 | -48.42 |
| E49M | 4.5 | 2.21 | 0.68 | -0.06 | 0.14 | -48.47 |
| E49T | 0 | 0.71 | 0.82 | -0.68 | 0.09 | -48.24 |
| E49S | -1.4 | 0.27 | 0.38 | -0.74 | 0.14 | -48.23 |
| **Correlation** |  | **0.81** | **0.62** | **0.42** | **0.15** | **-0.21** |

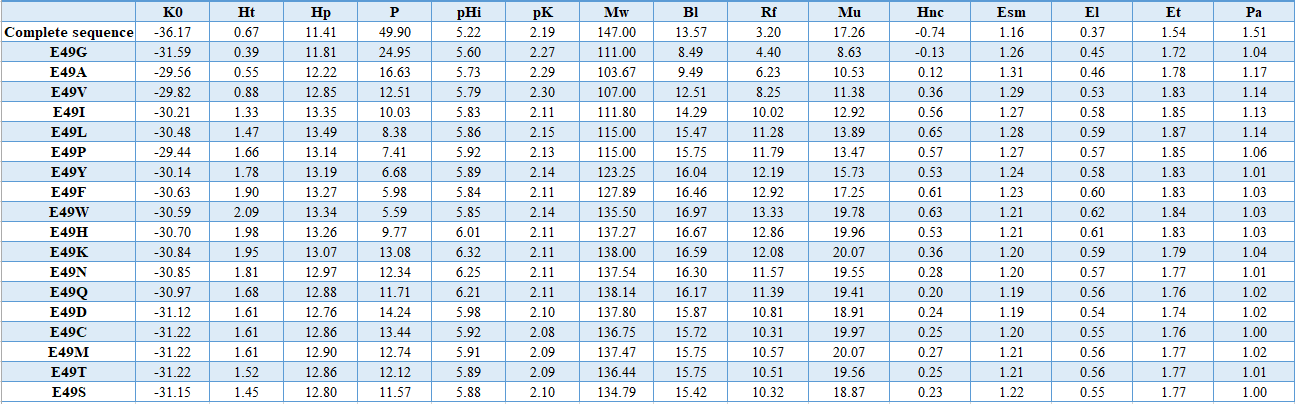
**Supplementary Table S2**

**Statistics on number of protein and DNA based features**

|  |  |  |  |
| --- | --- | --- | --- |
| Protein | | DNA | |
| Amino acid properties | 130 | Physicochemical properties | 16 |
| Substitution matrices | 28 | Conformational properties | 18 |
| Pairwise contact potential | 94 | Nucleotide content | 8 |

**Supplementary Table S3**

**Average property values for complete sequence and mutants obtained from Seq2Feature**



**Supplementary Table S4**

**Protein sequence based properties**

**Amino Acid properties**

|  |  |  |
| --- | --- | --- |
| 1 | K0 | Compressibility |
| 2 | Ht | Thermodynamic transfer hydrophobicity |
| 3 | Hp | Surrounding hydrophobicity |
| 4 | P | Polarity |
| 5 | pHi | Isoelectric point |
| 6 | pK' | Equilibrium constant with reference to the ionization property of COOH group |
| 7 | Mw | Molecular weight |
| 8 | Bl | Bulkiness |
| 9 | Rf | Chromatographic index |
| 10 | Mu | Refractive index |
| 11 | Hnc | Normalized consensus hydrophobicity |
| 12 | Esm | Short and medium range non-bonded energy |
| 13 | El | Long-rangenon-bondedenergy |
| 14 | Et | Total non-bonded energy |
| 15 | Pa | Alpha-helical tendency |
| 16 | Pb | Beta-helical tendency |
| 17 | Pt | Turn tendency |
| 18 | Pc | Coil tendency |
| 19 | Ca | Helical contact area |
| 20 | F | Mean rms fluctuation al displacement |
| 21 | Br | Buriedness |
| 22 | Ra | Solvent accessible reduction ratio |
| 23 | Ns | Average number of surrounding residues |
| 24 | aN | Power to beat the N-terminal of alphahelix |
| 25 | aC | Power to beat the C-terminal of alphahelix |
| 26 | aM | Power to beat the middle of alphahelix |
| 27 | V0 | Partial-specific volume |
| 28 | Nm | Average medium-range contacts |
| 29 | Nl | Average long-range contacts |
| 30 | Hgm | Combined surrounding hydrophobicity (globularandmembrane) |
| 31 | ASAD | Solvent accessible surface area for denatured protein |
| 32 | ASAN | Solvent accessible surface area for native protein |
| 33 | DASA | Solvent accessible surface area for protein unfolding |
| 34 | DGh | Gibbs free energy change of hydration for unfolding |
| 35 | GhD | Gibbs free energy change of hydration for denatured protein |
| 36 | GdN | Gibbs free energy change of hydration for native protein |
| 37 | DHh | Unfolding enthalpy change of hydration |
| 38 | -TDSh | Unfolding entropy change of hydration |
| 39 | DCph | Unfolding hydration heat capacity change |
| 40 | DGc | Unfolding Gibbs free energy change of chain |
| 41 | DHc | Unfolding enthalpy change of chain |
| 42 | -TDSc | Unfolding entropy change of chain |
| 43 | DGc | Unfolding Gibbs free energy change |
| 44 | DHc | Unfolding enthalpy change |
| 45 | -TDSc | Unfolding entropy change |
| 46 | v | Volume (number of non-hydrogen side chain atoms) |
| 47 | s | Shape (position of branch point in aside-chain) |
| 48 | f | Flexibility (number of side-chain dihedral angles) |
| 49 | Pf-s | Backbone dihedral probability |
| 50 | ARGP820101 | Hydrophobicity index (Argos et al., 1982) |
| 51 | BIGC670101 | Residue volume (Bigelow, 1967) |
| 52 | BIOV880102 | Information value for accessibility; average fraction 23% (Biou et al., 1988) |
| 53 | BURA740102 | Normalized frequency of extended structure (Burgess et al., 1974) |
| 54 | CHAM820101 | Polarizability parameter (Charton-Charton, 1982) |
| 55 | CHOC760101 | Residue accessible surface area in tripeptide (Chothia, 1976) |
| 56 | CHOP780202 | Normalized frequency of beta-sheet (Chou-Fasman, 1978b) |
| 57 | CHOP780204 | Normalized frequency of N-terminal helix (Chou-Fasman, 1978b) |
| 58 | DAYM780201 | Relative mutability (Dayhoff et al., 1978b) |
| 59 | EISD860101 | Solvation free energy (Eisenberg-McLachlan, 1986) |
| 60 | FASG760101 | Molecular weight (Fasman, 1976) |
| 61 | FASG760102 | Melting point (Fasman, 1976) |
| 62 | FAUJ830101 | Hydrophobic parameter pi (Fauchere-Pliska, 1983) |
| 63 | FAUJ880101 | Graph shape index (Fauchere et al., 1988) |
| 64 | FAUJ880106 | STERIMOL maximum width of the side chain (Fauchere et al., 1988) |
| 65 | GEIM800105 | Beta-strand indices (Geisow-Roberts, 1980) |
| 66 | GEIM800108 | Aperiodic indices (Geisow-Roberts, 1980) |
| 67 | GRAR740102 | Polarity (Grantham, 1974) |
| 68 | GRAR740103 | Volume (Grantham, 1974) |
| 69 | HOPA770101 | Hydration number (Hopfinger, 1971), Cited by Charton-Charton (1982) |
| 70 | ISOY800102 | Normalized relative frequency of extended structure (Isogai et al., 1980) |
| 71 | ISOY800103 | Normalized relative frequency of bend (Isogai et al., 1980) |
| 72 | ISOY800104 | Normalized relative frequency of bend R (Isogai et al., 1980) |
| 73 | JOND750101 | Hydrophobicity (Jones, 1975) |
| 74 | JUKT750101 | Amino acid distribution (Jukes et al., 1975) |
| 75 | KANM800102 | Average relative probability of beta-sheet (Kanehisa-Tsong, 1980) |
| 76 | KANM800103 | Average relative probability of inner helix (Kanehisa-Tsong, 1980) |
| 77 | KARP850101 | Flexibility parameter for no rigid neighbors (Karplus-Schulz, 1985) |
| 78 | KRIW710101 | Side chain interaction parameter (Krigbaum-Rubin, 1971) |
| 79 | KRIW790101 | Side chain interaction parameter (Krigbaum-Komoriya, 1979) |
| 80 | LAWE840101 | Transfer free energy, CHP/water (Lawson et al., 1984) |
| 81 | LEVM760101 | Hydrophobic parameter (Levitt, 1976) |
| 82 | LIFS790101 | Conformational preference for all beta-strands (Lifson-Sander, 1979) |
| 83 | LIFS790103 | Conformational preference for antiparallel beta-strands (Lifson-Sander, 1979) |
| 84 | MANP780101 | Average surrounding hydrophobicity (Manavalan-Ponnuswamy, 1978) |
| 85 | MAXF760102 | Normalized frequency of extended structure (Maxfield-Scheraga, 1976) |
| 86 | MAXF760106 | Normalized frequency of alpha region (Maxfield-Scheraga, 1976) |
| 87 | MIYS850101 | Effective partition energy (Miyazawa-Jernigan, 1985) |
| 88 | NAGK730102 | Normalized frequency of bata-structure (Nagano, 1973) |
| 89 | NAGK730103 | Normalized frequency of coil (Nagano, 1973) |
| 90 | NISK860101 | 14 A contact number (Nishikawa-Ooi, 1986) |
| 91 | OOBM770103 | Long range non-bonded energy per atom (Oobatake-Ooi, 1977) |
| 92 | OOBM850103 | Optimized transfer energy parameter (Oobatake et al., 1985) |
| 93 | OOBM850105 | Optimized side chain interaction parameter (Oobatake et al., 1985) |
| 94 | PONP800102 | Average gain in surrounding hydrophobicity (Ponnuswamy et al., 1980) |
| 95 | PONP800103 | Average gain ratio in surrounding hydrophobicity (Ponnuswamy et al., 1980) |
| 96 | PONP800107 | Accessibility reduction ratio (Ponnuswamy et al., 1980) |
| 97 | PRAM900104 | Relative frequency in reverse-turn (Prabhakaran, 1990) |
| 98 | RACS770101 | Average reduced distance for C-alpha (Rackovsky-Scheraga, 1977) |
| 99 | RACS770102 | Average reduced distance for side chain (Rackovsky-Scheraga, 1977) |
| 100 | RACS820113 | Value of theta(i) (Rackovsky-Scheraga, 1982) |
| 101 | ROBB760105 | Information measure for extended (Robson-Suzuki, 1976) |
| 102 | ROSM880102 | Side chain hydropathy, corrected for solvation (Roseman, 1988) |
| 103 | SIMZ760101 | Transfer free energy (Simon, 1976), Cited by Charton-Charton (1982) |
| 104 | WOEC730101 | Polar requirement (Woese, 1973) |
| 105 | YUTK870102 | Unfolding Gibbs energy in water, pH9.0 (Yutani et al., 1987) |
| 106 | ZIMJ680101 | Hydrophobicity (Zimmerman et al., 1968) |
| 107 | ZIMJ680102 | Bulkiness (Zimmerman et al., 1968) |
| 108 | ZIMJ680105 | RF rank (Zimmerman et al., 1968) |
| 109 | ONEK900102 | Helix formation parameters (delta delta G) (O'Neil-DeGrado, 1990) |
| 110 | VINM940101 | Normalized flexibility parameters (B-values), average (Vihinen et al., 1994) |
| 111 | NADH010101 | Hydropathy scale based on self-information values in the two-state model (5% accessibility) (Naderi-Manesh et al., 2001) |
| 112 | NADH010102 | Hydropathy scale based on self-information values in the two-state model (9% accessibility) (Naderi-Manesh et al., 2001) |
| 113 | NADH010104 | Hydropathy scale based on self-information values in the two-state model (20% accessibility) (Naderi-Manesh et al., 2001) |
| 114 | ADH010105 | Hydropathy scale based on self-information values in the two-state model (25% accessibility) (Naderi-Manesh et al., 2001) |
| 115 | FUKS010102 | Surface composition of amino acids in intracellular proteins of mesophiles (percent) (Fukuchi-Nishikawa, 2001) |
| 116 | FUKS010103 | Surface composition of amino acids in extracellular proteins of mesophiles (percent) (Fukuchi-Nishikawa, 2001) |
| 117 | KUHL950101 | Hydrophilicity scale (Kuhn et al., 1995) |
| 118 | ZHOH040101 | The stability scale from the knowledge-based atom-atom potential (Zhou-Zhou, 2004) |
| 119 | ZHOH040102 | The relative stability scale extracted from mutation experiments (Zhou-Zhou, 2004) |
| 120 | ZHOH040103 | Buriability (Zhou-Zhou, 2004) |
| 121 | PONJ960101 | Average volumes of residues (Pontius et al., 1996) |
| 122 | WOLR790101 | Hydrophobicity index (Wolfenden et al., 1979) |
| 123 | OLSK800101 | Average internal preferences (Olsen, 1980) |
| 124 | KIDA850101 | Hydrophobicity-related index (Kidera et al., 1985) |
| 125 | CORJ870102 | SWEIG index (Cornette et al., 1987) |
| 126 | CORJ870104 | PRILS index (Cornette et al., 1987) |
| 127 | CORJ870106 | ALTLS index (Cornette et al., 1987) |
| 128 | CORJ870108 | TOTLS index (Cornette et al., 1987) |
| 129 | MIYS990101 | Relative partition energies derived by the Bethe approximation (Miyazawa-Jernigan, 1999) |
| 130 | FASG890101 | Hydrophobicity index (Fasman, 1989) |

The numerical values are available at

https://www.iitm.ac.in/bioinfo/SBFE/Amino\_acid\_prop\_130.html

**Supplementary Table S5**

**Substitution matrices**

|  |  |  |
| --- | --- | --- |
| 1 | AZAE970101 | The single residue substitution matrix from interchanges of spatially neighbouring residues (Azarya-Sprinzak et al., 1997) |
| 2 | AZAE970102 | The substitution matrix derived from spatially conserved motifs (Azarya-Sprinzak et al., 1997) |
| 3 | GIAG010101 | Residue substitutions matrix from thermo/mesophilic to psychrophilic enzymes (Gianese et al., 2001) |
| 4 | HENS920101 | BLOSUM45 substitution matrix (Henikoff-Henikoff, 1992) |
| 5 | HENS920102 | BLOSUM62 substitution matrix (Henikoff-Henikoff, 1992) |
| 6 | HENS920103 | BLOSUM80 substitution matrix (Henikoff-Henikoff, 1992) |
| 7 | HENS920104 | BLOSUM50 substitution matrix (Henikoff-Henikoff, 1992) |
| 8 | KOSJ950101 | Context-dependent optimal substitution matrices for exposed helix (Koshi-Goldstein, 1995) |
| 9 | KOSJ950102 | Context-dependent optimal substitution matrices for exposed beta (Koshi-Goldstein, 1995) |
| 10 | KOSJ950103 | Context-dependent optimal substitution matrices for exposed turn (Koshi-Goldstein, 1995) |
| 11 | KOSJ950104 | Context-dependent optimal substitution matrices for exposed coil (Koshi-Goldstein, 1995) |
| 12 | KOSJ950105 | Context-dependent optimal substitution matrices for buried helix (Koshi-Goldstein, 1995) |
| 13 | KOSJ950106 | Context-dependent optimal substitution matrices for buried beta (Koshi-Goldstein, 1995) |
| 14 | KOSJ950107 | Context-dependent optimal substitution matrices for buried turn (Koshi-Goldstein, 1995) |
| 15 | KOSJ950108 | Context-dependent optimal substitution matrices for buried coil (Koshi-Goldstein, 1995) |
| 16 | KOSJ950109 | Context-dependent optimal substitution matrices for alpha helix (Koshi-Goldstein, 1995) |
| 17 | KOSJ950110 | Context-dependent optimal substitution matrices for beta sheet (Koshi-Goldstein, 1995) |
| 18 | KOSJ950111 | Context-dependent optimal substitution matrices for turn (Koshi-Goldstein, 1995) |
| 19 | KOSJ950112 | Context-dependent optimal substitution matrices for coil (Koshi-Goldstein, 1995) |
| 20 | KOSJ950113 | Context-dependent optimal substitution matrices for exposed residues (Koshi-Goldstein, 1995) |
| 21 | KOSJ950114 | Context-dependent optimal substitution matrices for buried residues (Koshi-Goldstein, 1995) |
| 22 | KOSJ950115 | Context-dependent optimal substitution matrices for all residues (Koshi-Goldstein, 1995) |
| 23 | MIYS930101 | Base-substitution-protein-stability matrix (Miyazawa-Jernigan, 1993) |
| 24 | MUET010101 | Non-symmetric substitution matrix (SLIM) for detection of homologous transmembrane proteins (Mueller et al., 2001) |
| 25 | OVEJ920102 | Environment-specific amino acid substitution matrix for alpha residues (Overington et al., 1992) |
| 26 | OVEJ920103 | Environment-specific amino acid substitution matrix for beta residues (Overington et al., 1992) |
| 27 | OVEJ920104 | Environment-specific amino acid substitution matrix for accessible residues (Overington et al., 1992) |
| 28 | OVEJ920105 | Environment-specific amino acid substitution matrix for inaccessible residues (Overington et al., 1992) |

Numerical values are available at <https://www.genome.jp/aaindex/>

**Supplementary Table S6**

**Pairwise properties and contact potentials**

|  |  |  |
| --- | --- | --- |
| 1 | TANS760101 | Statistical contact potential derived from 25 x-ray protein structures |
| 2 | TANS760102 | Number of contacts between side chains derived from 25 x-ray protein structures |
| 3 | ROBB790102 | Interaction energies derived from side chain contacts in the interiors of known protein structur |
| 4 | BRYS930101 | Distance-dependent statistical potential (only energies of contacts within 0-5 Angstrooms are in |
| 5 | THOP960101 | Mixed quasichemical and optimization-based protein contact potential |
| 6 | MIRL960101 | Statistical potential derived by the maximization of the harmonic mean of Z scores |
| 7 | VENM980101 | Statistical potential derived by the maximization of the perceptron criterion |
| 8 | BASU010101 | Optimization-based potential derived by the modified perceptron criterion |
| 9 | MIYS850102 | Quasichemical energy of transfer of amino acids from water to the protein environment |
| 10 | MIYS850103 | Quasichemical energy of interactions in an average buried environment |
| 11 | MIYS960101 | Quasichemical energy of transfer of amino acids from water to the protein environment |
| 12 | MIYS960102 | Quasichemical energy of interactions in an average buried environment |
| 13 | MIYS960103 | Number of contacts between side chains derived from 1168 x-ray protein structures |
| 14 | MIYS990106 | Quasichemical energy of transfer of amino acids from water to the protein environment |
| 15 | MIYS990107 | Quasichemical energy of interactions in an average buried environment |
| 16 | LIWA970101 | Modified version of the Miyazawa-Jernigan transfer energy |
| 17 | KESO980101 | Quasichemical transfer energy derived from interfacial regions of protein-protein complexes |
| 18 | KESO980102 | Quasichemical energy in an average protein environment derived from interfacial regions of prote |
| 19 | MOOG990101 | Quasichemical potential derived from interfacial regions of protein-protein complexes |
| 20 | BETM990101 | Modified version of the Miyazawa-Jernigan transfer energy |
| 21 | TOBD000101 | Optimization-derived potential obtained for small set of decoys |
| 22 | TOBD000102 | Optimization-derived potential obtained for large set of decoys |
| 23 | PARB960101 | Statistical contact potential derived by the quasichemical approximation |
| 24 | PARB960102 | Modified version of the Miyazawa-Jernigan transfer energy |
| 25 | KOLA930101 | Statistical potential derived by the quasichemical approximation |
| 26 | GODA950101 | Quasichemical statistical potential derived from buried contacts |
| 27 | SKOJ970101 | Statistical potential derived by the quasichemical approximation |
| 28 | SKOJ000101 | Statistical quasichemical potential with the partially composition-corrected pair scale |
| 29 | SKOJ000102 | Statistical quasichemical potential with the composition-corrected pair scale |
| 30 | BONM030101 | Quasichemical statistical potential for the antiparallel orientation of interacting side groups |
| 31 | BONM030102 | Quasichemical statistical potential for the intermediate orientation of interacting side groups |
| 32 | BONM030103 | Quasichemical statistical potential for the parallel orientation of interacting side groups |
| 33 | BONM030104 | Distances between centers of interacting side chains in the antiparallel orientation |
| 34 | BONM030105 | Distances between centers of interacting side chains in the intermediate orientation |
| 35 | BONM030106 | Distances between centers of interacting side chains in the parallel orientation |
| 36 | MICC010101 | Optimization-derived potential |
| 37 | SIMK990101 | Distance-dependent statistical potential (contacts within 0-5 Angstrooms) |
| 38 | SIMK990102 | Distance-dependent statistical potential (contacts within 5-7.5 Angstrooms) |
| 39 | SIMK990103 | Distance-dependent statistical potential (contacts within 7.5-10 Angstrooms) |
| 40 | SIMK990104 | Distance-dependent statistical potential (contacts within 10-12 Angstrooms) |
| 41 | SIMK990105 | Distance-dependent statistical potential (contacts longer than 12 Angstrooms) |
| 42 | ZHAC000101 | Environment-dependent residue contact energies (rows = helix, cols = helix) |
| 43 | ZHAC000102 | Environment-dependent residue contact energies (rows = helix, cols = strand) |
| 44 | ZHAC000103 | Environment-dependent residue contact energies (rows = helix, cols = coil) |
| 45 | ZHAC000104 | Environment-dependent residue contact energies (rows = strand, cols = strand) |
| 46 | ZHAC000105 | Environment-dependent residue contact energies (rows = strand, cols = coil) |
| 47 | ZHAC000106 | Environment-dependent residue contact energies (rows = coil, cols = coil) |

Numerical values are available at <https://www.genome.jp/aaindex/>

**Supplementary Table S7**

**DNA sequence based properties**

**Physicochemical properties**

|  |  |
| --- | --- |
| 1 | Stacking energy |
| 2 | Enthalpy |
| 3 | Entropy |
| 4 | Flexibility\_shift |
| 5 | Flexibility\_slide |
| 6 | Free energy |
| 7 | Melting Temperature |
| 8 | Mobility to bend towards major groove |
| 9 | Mobility to bend towards minor groove |
| 10 | Probability contacting nucleosome core |
| 11 | Rise stiffness |
| 12 | Roll stiffness |
| 13 | Shift stiffness |
| 14 | Slide stiffness |
| 15 | Tilt stiffness |
| 16 | Twist stiffness |

Numerical values are available at https://www.iitm.ac.in/bioinfo/SBFE/physico\_dna.html

**Conformational properties**

|  |  |
| --- | --- |
| 1 | Bend |
| 2 | Rise |
| 3 | Roll |
| 4 | Inclination |
| 5 | Major Groove Depth |
| 6 | Major Groove Distance |
| 7 | Major Groove Size |
| 8 | Major Groove Width |
| 9 | Minor Groove Depth |
| 10 | Minor Groove Distance |
| 11 | Minor Groove Size |
| 12 | Minor Groove Width |
| 13 | Shift |
| 14 | Propeller Twist |
| 15 | Slide |
| 16 | Tilt |
| 17 | Tip |
| 18 | Twist |

Numerical values are available at <https://www.iitm.ac.in/bioinfo/SBFE/conform_dna.html>

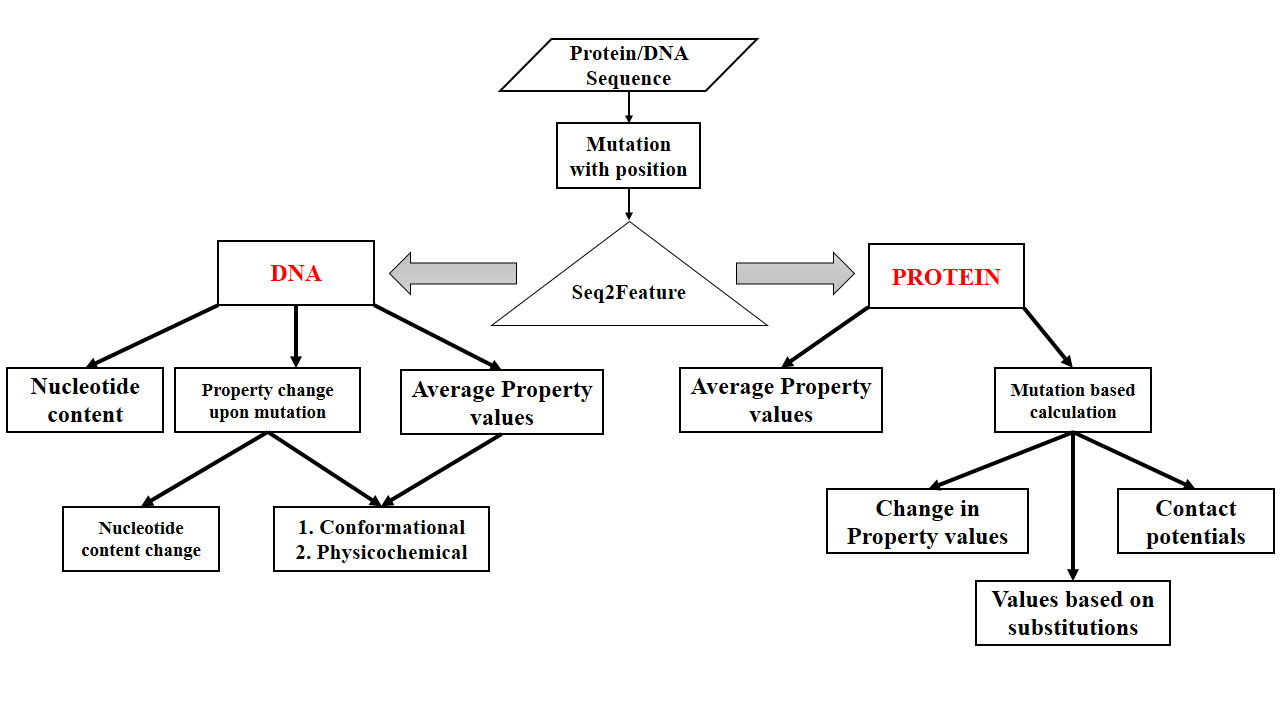
**Nucleotide content:**

|  |  |
| --- | --- |
| 1 | Adenine content |
| 2 | Cytosine content |
| 3 | GC content |
| 4 | Guanine content |
| 5 | Keto (GT) content |
| 6 | Purine (AG) content |
| 7 | Thymine content |
| 8 | Pyrimidine (CT) |

**Supplementary Figures**

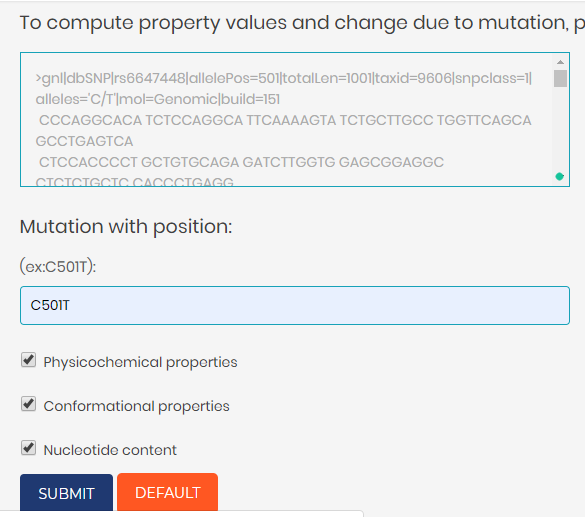
**Supplementary Figure S1**

**Flowchart of Seq2Feature**

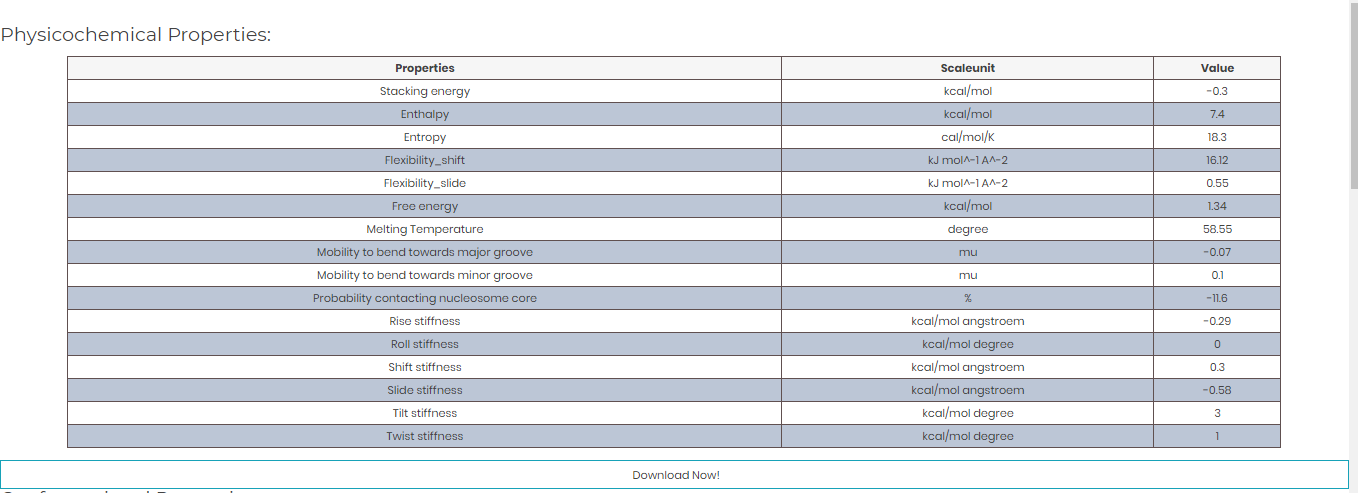


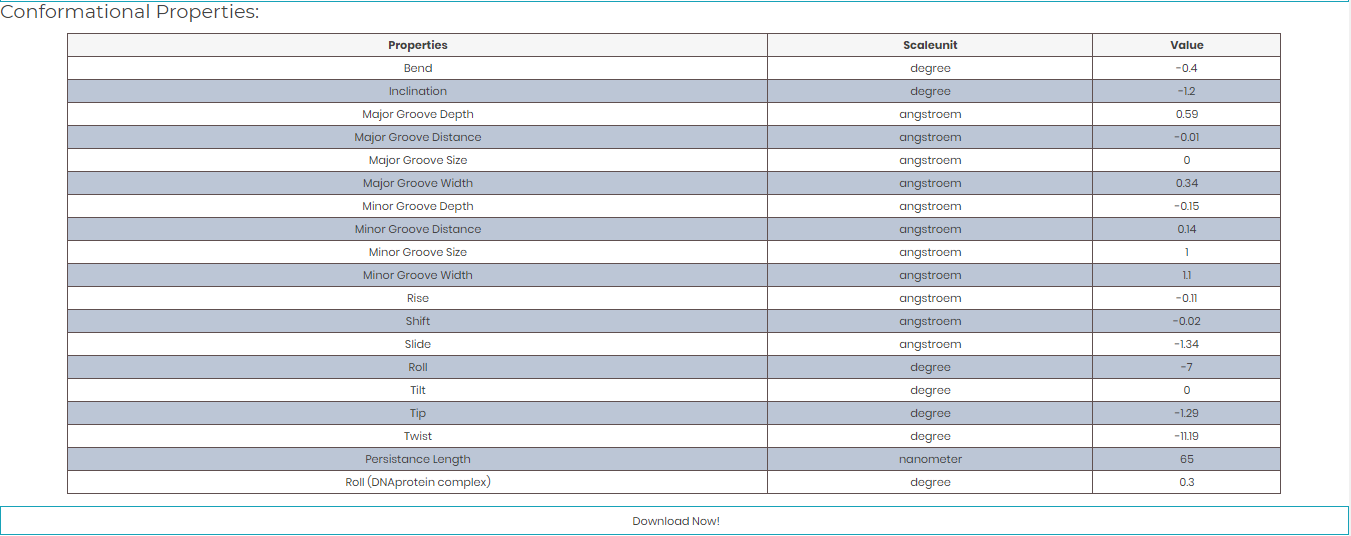
**Supplementary Figure S2**

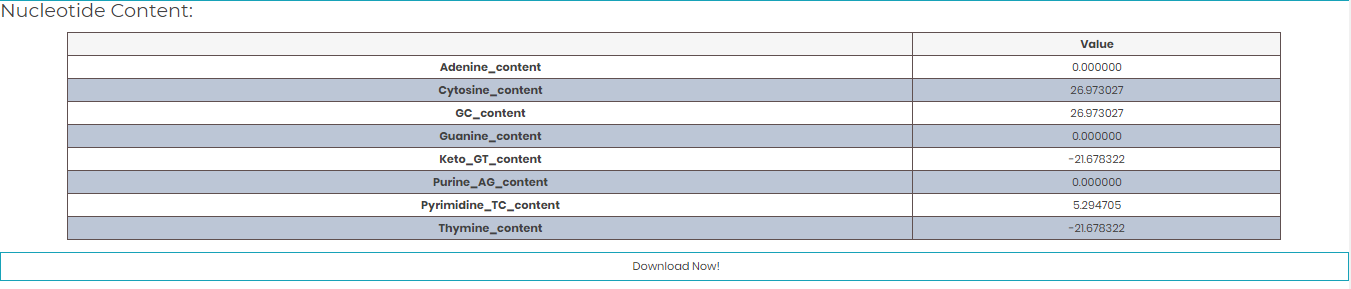
**Input DNA sequence with mutation**



Change in property values upon mutation





Change in nucleotide content upon mutation