ProtDCal-Suite: A web server for the numerical codification and functional analysis of proteins

Sandra Romero-Molina[1], Yasser B. Ruiz-Blanco[1*], James R. Green[2] and Elsa Sanchez-Garcia[1*]

[1] Computational Biochemistry, Center of Medical Biotechnology, University of Duisburg-Essen, Essen, North Rhine-Westphalia, 45117, Germany
[2] Systems and Computer Engineering, Carleton University, Ottawa, Ontario, K1S 5B6, Canada

Supplementary Information

Table of Content:

Table SM-1. Compendium of structural and chemical-physical amino acid properties.*

	Mw	**HP**	**IP**	**ECI**	**L1-9**	**Z1**	**Z2**	**Z3**	**ISA**	**Xi**	**Pa**	**Pb**	**Pt**	**ΔHf**	**Ap**
ALA	71	1.8	6.01	0.05	19.2	0.07	-1.73	0.09	62.9	-77.85	1.29	0.9	0.78	-433.66	202.42
ARG	156	-4.5	10.76	1.69	17.8	2.88	2.52	-3.44	52.98	108.86	0.96	0.99	0.88	-403.21	557.81
ASN	114	-3.5	5.41	1.31	21.72	3.22	1.45	0.84	17.87	-55.42	0.9	0.76	1.28	-466.91	377.84
ASP	115	-3.5	2.77	1.25	17.14	3.64	1.13	2.36	18.46	47.89	1.04	0.72	1.41	-518.1	360.26
CYS	103	2.5	5.07	0.15	18.83	0.71	-0.97	4.13	78.51	160.13	1.11	0.74	0.8	-425.69	236.80
GLN	128	-3.5	3.22	1.31	18.55	3.08	0.39	-0.07	19.53	134.68	1.44	0.75	1	-479.54	439.85
GLU	129	-3.5	5.65	1.36	17.31	2.18	0.53	-1.14	30.19	53.27	1.27	0.8	0.97	-531.69	417.46
GLY	57	-0.4	5.97	0.02	19.48	2.23	-5.36	0.3	19.93	-148.03	0.56	0.92	1.64	-420.86	172.08
HIS	137	-3.2	7.59	0.56	13.97	2.41	1.74	1.11	87.38	-4.57	1.22	1.08	0.69	-378.92	417.33
ILE	113	4.5	6.02	0.09	20.76	-4.44	-1.68	-1.03	149.77	-104.8	0.97	1.45	0.51	-449.27	309.12
LEU	113	3.8	5.98	0.01	17.65	-4.19	-1.03	-0.98	154.35	-148.5	1.3	1.02	0.59	-448.27	318.85
LYS	128	-3.9	9.74	0.53	17.05	2.84	1.41	-3.14	102.78	47.61	1.23	0.77	0.96	-446.97	409.91
MET	131	1.9	5.74	0.34	17.88	-2.49	-0.27	-0.41	132.22	46.37	1.47	0.97	0.39	-435.34	332.93
PHE	147	2.8	5.48	0.14	16.81	-4.92	1.3	0.45	189.42	47.67	1.07	1.32	0.58	-376.77	414.12
PRO	97	-1.6	6.48	0.16	18.55	-1.22	0.88	2.23	122.35	169.73	0.52	0.64	1.91	-422.17	261.24
SER	87	-0.8	5.68	0.56	18.91	1.96	-1.63	0.57	19.75	30.24	0.82	0.95	1.33	-479.75	265.01
THR	101	-0.7	5.87	0.65	17.15	0.92	-2.09	-1.4	59.44	46.04	0.82	1.21	1.03	-483.37	292.47
TRP	186	-0.9	5.89	1.08	20.94	-4.75	3.65	0.85	179.16	178.69	0.99	1.14	0.75	-365.49	530.87
TYR	163	-1.3	5.66	0.72	16.86	-1.39	2.32	0.01	132.16	49.11	0.72	1.25	1.05	-446.32	472.98
VAL	99	4.2	5.97	0.07	17.88	-2.69	-2.53	-1.29	120.91	-106.5	0.91	1.49	0.47	-434.3	276.26

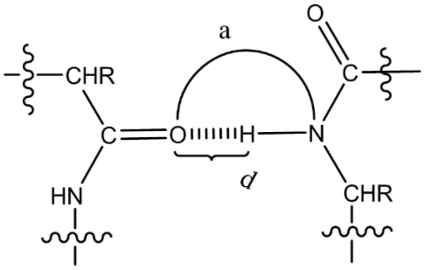
Mw Molar Weight [2]
HP Kyte-Doolitle's Hydrophobicity Scale [4]
IP Isoelectric Point [2]
ECI Electronic Charge Index[3]
L1-9 Compatibility parameter [5]
Z1 Composed parameter related with hydrophilicity [7]
Z2 Composed parameter related with steric features [7]
Z3 Composed parameter related with electronic features [7]

ISA Isotropic Surface Area [3]
Xi Compatibility parameter [5]
Pa Levitt's Probability of adopting alpha helix conformation [6]
Pb Levitt's Probability of adopting beta sheet conformation [6]
Pt Levitt's Probability of adopting beta turn conformation [6]
ΔHf(X) enthalpy of formation of the peptide: AAAA**X**AAAA. [5]
Ap Molecular area of non-carbon atoms in the sidechain

*This table was taken from the Supplementary Information of the ProtDCal manuscript.[1]

Table SM-2. Formulae and description of 3D-thermodynamics indices.*

Acronym	Formula	Description
Gc$_{(F)}$	$G_c(F)_i = RT(N-1)p_i \ln p_i$, $p_i = \left(\frac{3}{2\pi(i-1)3.8^2}\right)^{3/2} e^{-\frac{3r_i^2}{2(i-1)3.8^2}}$	Configurational free energy of a folded state. Index based on a “random-flight” model of the protein chain. [8] Where r_i represents the distance to the first residue in the chain.
W$_{(F)}$	$W_i^F = \sum_{j=1}^{N} \delta_{ij}^{ng} \delta_j^s N_j^w$	Number of water molecules close to a residue in a folded state.[9; 10] Where δ^{ng} takes value 1 if the pair of residues are neighbours, using a cutoff for the spatial distance (9.4 Å), or 0 otherwise. In the same way δ^s takes value 1 if the residue is superficial, using a cutoff for the solvent accessible surface area, or 0 otherwise. The parameters N^w represents the number of associated water molecules to the sidechain of a residue [11].
Gw$_{(F)}$	$G_w(F)_i = -TR\delta_{hyd} \ln \frac{W_i^F!}{(W_i^F - N_i^w)!}$	Free energy contribution of the entropy of the first shell of water molecules in a folded state [10]. δ_{hyd} takes value 1 if the residue has non-zero N_i^w, or zero otherwise.
Gs$_{(F)}$	$G_s(F)_i = H_i A_i^F$	Interfacial free energy contribution of a folded state. Where H_i is hydrophobicity in Kyte-Doolittle scale [4] and A^F is the solvent accessible surface area of a residue in a folded state.
ΔG_s	$\Delta G_{s\,i} = G_s(F)_i - G_s(U)_i$	Interfacial free energy variation.
HBd	$\Delta Hbd_i = 0.5\sum_{j=1}^{N}(\delta_{ij}^N + \delta_{ij}^O)$ Geometric definition of a H-bond:	Number of backbone’s hydrogen bonds. Where δ_{ij}^N takes value 1 if the Nitrogen atom of residue *i* is H-bonded with the Oxygen atom of residue *j* and 0 otherwise. In the same way δ_{ij}^O takes value one if the Oxygen atom of residue

	 $d < 2.5$ Å $a = 120º$ -- $180º$	*i* is H-bonded with the Nitrogen atom of residue *j* and zero otherwise.
ΔG_{el}	$$\Delta G_{el\ i} = -\frac{k_{el}}{2r^2}\sum_{j=1}^{N}\frac{q_i q_j r_i r_j}{r_{ij}}$$	Free energy contribution of the charge distribution within the protein. The parameters *q* are the Electronic Charge Indices of each residue [3]. Parameter k_{el}= 7.608.
ΔG_w	$$\Delta G_{w\ i} = k_w(G_w(F)_i - G_w(U)_i)$$	Folding free energy contribution of the entropy of the first shell of water molecules.
ΔG_{LJ}	$$\Delta G_{LJ\ i} = \frac{k_{LJ}}{2}\sum_{\substack{j=1;\\ \lvert j-i\rvert>1}}^{N}\left[\left(\frac{3.965}{r_{ij}}\right)^{12} - \left(\frac{3.965}{r_{ij}}\right)^{6}\right]$$	Residue-level Lennard-Jones interactions. Parameter k_{LJ} = 63.981.
ΔG_{tor}	$$\Delta G_{tor\ i} = k_{tor}[(\cos^2 2\phi_i - 1) + 0.256(\cos^2 2\psi_i - 1)]$$	Free energy contribution of backbone torsion angles. Parameter k_{tor}= 1.219.

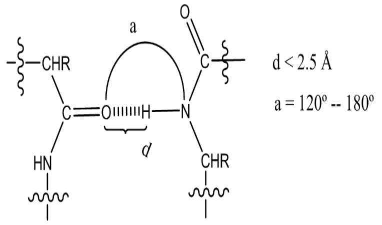
*This table was taken from the Supplementary Information of the ProtDCal manuscript.[1]

Table SM-3. Formulae and description of thermodynamics indices for protein sequences.*

Acronym	Formula	Description
W(U)	$W_i^U = \sum_{j=i-2}^{i+2} N_j^w$	Number of water molecules close to a residue in an unfolded state [10].
Gw(U)	$G_w(U)_i = -TR\delta_{hyd} \ln \frac{W_i^U!}{(W_i^U - N_i^w)!}$	Free energy contribution from the entropy of the first shell of water molecules in an unfolded state [10].
Gs(U)	$G_s(U)_i = H_i A_i^U$	Interfacial free energy contribution of an unfolded state

*This table was taken from the Supplementary Information of the ProtDCal manuscript.[1]

Table SM-4. Formulae and description of topographic indices.*

Acronym	Formula	Description
A_F	-	Solvent accessible surface area
ΔA	$\Delta A = A_F - A_U$	Buried area. Where A_U is the fully exposed surface area of each residue and A_F is the area in the folded state.
ΔA^{np}	$\Delta A^{np} = A^{np}{}_F - A^{np}{}_U$	Buried non-polar area. Here nitrogen atoms and oxygen atoms are excluded.
wSp	$wSp_i = \omega_i * \delta_i^s$	Weighted index of the solvent accessibility. Where ω represents any weighting property and the delta takes value 1 or 0 if the residue is considered superficial or internal respectively.
lnFD	$\ln FD_i = -\frac{\sum_{j;\|j-i\|>1}^{N} \|j-i\| / d_{ij}^3}{N-x}$	Logarithm of the Folding Degree. Where *d* is the spatial distance, *N* the length of the protein and *x* a parameter which takes value 2 for terminal residues and 3 for the others.
wR^2	$wRG_i^2 = \frac{w_i * d_i^2}{\sum_{i=1}^{N} w_i}$	Weighted Squared Radius. Where ω represents any weighting property and *d* is the spatial distance.
wΔHBd	$\Delta Hbd_i = \omega_i * (\delta_N + \delta_O)$ Geometric definition of a H-bond:  a O CHR C C O H N d HN CHR d < 2.5 Å a = 120º -- 180º	Weighted deficit or excess of the H-bond between the backbone atoms. Where δ_{ij}^N takes value 1 if the nitrogen atom of residue *i* is buried ($A_{(N)} < 0.01$Å) and is not H-bonded with any oxygen atom or 0 otherwise. In the same way δ_{ij}^O takes value 1 if the oxygen atom of residue *i* is buried ($A_{(O)} < 0.01$Å) and is not H-bonded with any nitrogen atom and 0 otherwise.
wNc	$wNc_i = 0.5\sum_{j\neq i}^{N} \omega_{ij}\delta_{ij}^c$	Weighted Number of Contact. Where δ_{ij}, takes value 1 when the contact conditions are fulfilled and 0 otherwise. A contacts is defined for pair of residues with spatial distances shorter than a cutoff *d* and topological distances longer than a cutoff *t*. The parameter ω_{ij} represents a weighting coefficient for each pair of residues. This parameter is computed as the product, $\omega_i\omega_j$, of the values, for each residue, of any property within a pool of 12 amino acid properties covering structural, physical-chemical features.
wFLC	$wFLC_i = \frac{\sum_{\|j-i\|\leq 4}^{N} \omega_{ij}\delta_{ij}^c}{\sum_{i=1}^{N}\sum_{j=1}^{N} \omega_{ij}\delta_{ij}^c}$	Weighted Fraction of Local Contacts. The parameters δ_{ij} and ω_{ij} means the same as previous but here the topological cutoff value is fixed in *t* = 1.
wNLC	$wNLC_i = 0.5\sum_{\|j-i\|\leq 4}^{N} \omega_{ij}\delta_{ij}^c$	Weighted Number of Local Contact The parameters δ_{ij} and ω_{ij} means the same as in *wNc* but here the topological cutoff value is fixed in *t* = 1.
wCO	$wCO_i = \frac{1}{2NN_c}\sum_{j\neq i}^{N} \omega_{ij}\delta^c{}_{ij}$	Weighted Relative Contact Order [12]. Where Nc represents the number of contacts in the protein.

wLCO	$wLCO_i = \frac{\sum_{j\neq i}^{N}\omega_{ij}\delta_{ij}^{c}}{N\sum_{j\neq i}^{N}\delta_{ij}^{c}}$	Weighted Local Contact Order. As difference with previous, the weighted contacts are divided by the same un-weighted local contact instead of all the contact in the protein.
wRWCO	$wRWCO_i = \frac{\sum_{j\neq i}^{N}\omega_{ij}\delta_{ij}^{c}}{N}$	Weighted Residue-Wise Contact Order [13].
wCTP	$wCTP_i = \frac{1}{2NN_c}\sum_{j\neq i}^{N}\omega_{ij}^{2}\delta_{ij}^{c}$	Weighted Chain Topology Parameter [14].
wCLQ	$wCLQ_i = \frac{\sum_{j<l}\delta_{ij}\delta_{il}\delta_{lj}\omega_{ij}\omega_{il}\omega_{lj}}{\sum_{j<l}\delta_{ij}\delta_{il}\omega_{ij}\omega_{il}}$	Weighted Cliquishness or Clustering Coefficient [15].
wPsi_H	$Psi\_H_i = \delta_i^{\psi H} * \omega_i$	Weighted Helix-like Psi angle. The delta takes value 1 if the angle is in the range [-77;-17] or 0 otherwise.
wPsi_S	$Psi\_S_i = \delta_i^{\psi S} * \omega_i$	Weighted Sheet-like Psi angle. The delta takes value 1 if the angle is in the range [94;154] or 0 otherwise.
wPsi_I	$Psi\_I_i = \delta_i^{\psi I} * \omega_i$	Weighted Irregular Psi angle. The delta takes value 1 if the angle is in one of the following ranges: [-180,-77), (-17;94), (154;180] or 0 otherwise.
wPhi_H	$Phi\_H_i = \delta_i^{\phi H} * \omega_i$	Weighted Helix like Phi angle. The delta takes value 1 if the angle is in the range [-87;-27] or 0 otherwise.
wPhi_S	$Phi\_S_i = \delta_i^{\phi S} * \omega_i$	Weighted Sheet like Phi angle. The delta takes value 1 if the angle is in the range [-159;-99] or 0 otherwise.
wPhi_I	$Phi\_I_i = \delta_i^{\phi I} * \omega_i$	Weighted Irregular Phi angle. The delta takes value 1 if the angle is in one of the following ranges: [-180,-159), (-99;-87), (-27;180] or 0 otherwise.
Phi	-	Phi dihedral angle
Psi	-	Psi dihedral angle
TCD	$wTCD_i = \frac{1}{2N^2}\sum_{j\neq i}^{N}\omega_{ij}\delta_{ij}^{c}$	Total Contact Distance [16].

*This table was taken from the Supplementary Information of the ProtDCal manuscript.[1]

Table SM-5. Weighting procedures (*Vicinity modifiers*) implemented in ProtDCal.*

Acronym	Formula	Description
AC_i^k	$AC_i^k = \sum_{j\geq 1}^{N} L_i L_j \delta(d_{ij}, k)$ $Condition:(d_{ij} = k)?\,\delta = 1:\delta = 0$	Autocorrelation. Where, L_x are the index values of residues *i* and *j* and *k* is a topological distance cutoff and *N* is the total number of residues.
GV_i^k	$GV_i^k = \frac{1}{N} \sum_{j=1;\, j\neq i}^{N} \frac{L_i L_j \delta(d_{ij}, k)}{d_{ij}}$	Gravitational
KH_i^m	$KH_i^m = \sum_{\alpha=1}^{A} \sqrt{\prod_{j=1}^{n_\alpha} L_{j\alpha}}$	Kier-Hall's connectivity-based operator. Where, *A* is the number of segments containing the residue *i*, with a maximum length of *m* residues, n_α is the number of residues in a sub-segment, $L_{j\alpha}$ is the index value of the residue *j* in the segment *α*.
ES_i	$ES_i = L_i + \Delta L_i = L_i + \sum_{j=1;\, j\neq i}^{N} \frac{L_i - L_j}{(d_{ij}+1)^2}$	Electro-topological state (E-state index). Where, L_i is the intrinsic state (index) of the i^{th} residue and ΔL_i is the field effect on the *ith* residue calculated as perturbation of the index value (L_i) of i^{th} residue by all other residues in the protein, d_{ij} is the topological distance between the i^{th} and the j^{th} residue, and N is the total number of residues.
IB_i^2	$IB_i^2 = (N-1)\sum_{j\neq i}^{N} a_{ij}\left(S_i S_j\right)^{-1/2}$ $S_i = L_i + \sum_{j\neq i}^{N} a_{ij} L_j$	Ivanciuc-Balaban. Where, a_{ij} represents th elements of the adjacency matrix, and *N* is the number of residues. The exponent 2 dues to the use of the exponent -1/2. Here the factor (N-1) represents the numbers of virtual bonds among residues.

*This table was taken from the Supplementary Information of the ProtDCal manuscript.[1]

Table SM-6. Summary of the definitions of residue-based groups.*

Acronym	Description
ALA	Represents all alanine residues contained in the protein
ARG	Represents all arginine residues contained in the protein.
ASN	Represents the all asparagine residues contained in the protein.
ASP	Represents the all aspartic residues contained in the protein.
CYS	Represents the all cysteine residues in the protein.
GLN	Represents the all glutamine residues in the protein.
GLU	Represents the all glutamic residues in the protein.
GLY	Represents the all glycine residues contained in the protein.
HIS	Represents the all histidine residues contained in the protein.
ILE	Represents the all isoleucine residues in the protein.
LEU	Represents the all leucine residues contained in the protein.
LYS	Represents the all lysine residues contained in the protein.
MET	Represents the all methionine residues contained in the protein.
PHE	Represents the all phenylalanine residues contained in the protein.
PRO	Represents the all proline residues contained in the protein.
SER	Represents the all Serine residues contained in the protein.
THR	Represents the all threonine residues contained in the protein.
TRP	Represents the all tryptophan residues contained in the protein.
TYR	Represents the all tyrosine residues contained in the protein.
VAL	Represents the all valine residues contained in the protein.

*This table was taken from the Supplementary Information of the ProtDCal manuscript.[1]

Table SM-7. Summary of the definitions of property-based groups.*

Acronym	Included Residues	Description
AHR	ALA, CYS, GLN, GLU, HIS, LEU, LYS, MET	Common residues in alpha helix motifs.
BSR	ILE, PHE, THR, TRP, TYR, VAL	Common residues in beta sheet motifs.
RTR	ASN, ASP, GLY, PRO, SER	Common residues in reverse turn motifs.
PCR	ARG, HIS, LYS	Positive-electric-charged residues.
NCR	ASP, GLU	Negative-electric-charged residues.
UCR	ASN, CYS, GLN, SER, THR, TYR	Uncharged residues.
ARM	HIS, PHE, TRP, TYR	Aromatic residues.
ALR	ALA, GLY, ILE, LEU, MET, PRO, VAL	Aliphatic residues.
UFR	GLY, PRO	Common residues promoting unfolding or distorted regions.
NPR	ALA, GLY, ILE, LEU, MET, PHE, PRO, TRP, VAL	Non-polar residues.
PLR	ARG, ASN, ASP, CYS, GLN, GLU, HIS, LYS, SER, THR, TYR	Polar residues.

*This table was taken from the Supplementary Information of the ProtDCal manuscript.[1]

Table SM-8. Summary of the definitions of topographic-based groups.*

Acronym	**Description**
HEX	All residues in alpha helix conformation
SHT	All residues in beta sheet conformation
TRN	All residues in reverse turn conformation
RCL	All residues in loops regions (Residues in TRN are excluded)
INT	Represents the all internal residues in the protein.
SUP	Represents all superficial residues contained in the protein.
PRT	The whole protein

*This table was taken from the Supplementary Information of the ProtDCal manuscript.[1]

Table SM-9. Aggregation operators: Distance invariants.*

Acronym	**Formula**	**Description**
N1	$N1 = \sum_{i=1}^{N} \lvert L_i \rvert$	Minkowski's norms (p = 1) Manhattan norm. Where L_i represents each index of the group of indices and N the number of indices in the group.
N2	$N2 = \sqrt{\sum_{i=1}^{N} \lvert L_i \rvert^2}$	Minkowski's norms (p = 2) Euclidean norm. Where L_i represents each index of the group of indices and N the number of indices in the group.
N3	$N3 = \sqrt[3]{\sum_{i=1}^{N} \lvert L_i \rvert^3}$	Minkowski's norms (p = 3). Where L_i represents each index of the group of indices and N the number of indices in the group.

*This table was taken from the Supplementary Information of the ProtDCal manuscript.[1]

Table SM-10. Aggregation operators: Means (first statistical moment) invariants.*

Acronym	Formula	Description
G	$G = \sqrt[N]{\prod_{i=1}^{N} L_i}$	Geometric Mean. Where N is the number of indices in the group.
Ar	$m_\alpha = \left(\frac{L_1^\alpha + L_2^\alpha + ... + L_N^\alpha}{N} \right)^{\frac{1}{\alpha}}$	Arithmetic Mean (potential with α = 1)
P2		Potential Mean (potential with α = 2)
P3		Potential Mean (potential with α = 3)
M		Harmonic Mean (potential with α = -1)

*This table was taken from the Supplementary Information of the ProtDCal manuscript.[1]

Table SM-11. Aggregation operators: Statistical (highest statistical moments) invariants.*

Acronym	Formula	Description
V	$V = \frac{\sum_{i=1}^{N}\left(L_i - \bar{L}\right)^2}{N - 1}$	Variance. Where N is the number of indices in the group.
S	$S = \frac{N(X_3)}{(N-1)(N-2)(DE)^3}$ $X_3 = \sum_{a=1}^{N}(L_a - \bar{L})^3$	Skewness. Where N is the number of indices in the group and $(DE)^3$ is the standard deviation raised to the 3rd power
K	$k = \frac{N(N+1)X_4 - 3(X_2)(X_2)(N-1)}{(N-1)(N-2)(N-3)(DE)^4}$ $X_j = \sum_{a=1}^{N}(L_a - \bar{L})^j$	Kurtosis. Where $(DE)^4$ is the standard deviation raised to the fourth power
DE	$DE = \sqrt{\frac{\left(\sum L_i - \bar{L}\right)^2}{N-1}}$	Standard Deviation
CV	$c_v = S/\bar{L}$	Variation Coefficient
RA	$RA = L_{\max} - L_{\min}$	Range
Q1	$P25 = \left[\frac{N}{4} + \frac{1}{2}\right]$	Percentile 25. Where N is the number of indices in the group.
Q2	$P50 = \left[\frac{N}{2} + \frac{1}{2}\right]$	Percentile 50. Where N is the number of indices in the group.
Q3	$P75 = \left[\frac{3N}{4} + \frac{1}{2}\right]$	Percentile 75. Where N is the number of indices in the group.
I50	I50 = P75 – P25	Inter-quartile Range
MX	L_i maximum	Maximum value of the group of indices.
MN	L_j minimum	Minimum value of the group of indices.

*This table was taken from the Supplementary Information of the ProtDCal manuscript.[1]

Table SM-12. Aggregation operators: Information-Theory-based invariants.*

Acronym	Formula (Equation)	Description
MI	$MI = -\sum_{i=1}^{K} \frac{N_k}{N} \log_2 \frac{N_k}{N}$	Mean Information Content. Where N_k is the number of indices in the same bin, K is the number of bins defined to compute the operator and N is the total number of indices in the group.
TI	$TI = N \log_2 N - \sum_{k=1}^{K} N_k \log_2 N_k$	Total Information Content.
SI	$SI = \frac{TI}{N \log_2 N}$	Standarized Infomation Content

*This table was taken from the Supplementary Information of the ProtDCal manuscript.[1]

References

1. Ruiz-Blanco YB, Paz W, Green J, Marrero-Ponce Y (2015) ProtDCal: A program to compute general-purpose-numerical descriptors for sequences and 3D-structures of proteins. BMC Bioinformatics 16:162.
2. Lehninger. Amino Acids, Peptides, and Proteins. (2005) Biochemistry. pp. 76-115.
3. Collantes ER, Dunn-III WJ (1995) Amino acid side chain descriptors for quantitative structure-activity relationship studies of peptide analogues. J Med Chem 38:2705-2713.
4. Kyte J, Doolitle RF (1982) A Simple Method for Displaying the Hydropathic Character of a Protein. . J Mol Biol 157:105-132.
5. S K, M K, J J (1999) Modeling of the amino acid side chain effects on peptide conformation. Bioorg Chem 27:434–442.
6. Levitt M (1978) Conformational Preferences of Amino Acids in Globular Proteins. Biochemistry 17.
7. Hellberg S. S, M., Skagerberg B., Wold, S. (1987) Peptide Quantitative Structure-Activity Relationship, a Multivariate Approach. J Med Chem 30:1126-1135. .
8. Kamide K, Dobashi T. Chapter 6: Statistical Mechanics and Excluded Volume of Polymer Chains. (2000) Physical Chemistry of Polymer Solutions Theoretical Background. Elsevier Science.
9. Ruiz-Blanco YB, García Y, Sotomayor-Torres CM, Marrero-Ponce Y (2010) New Set of 2D/3D Thermodynamic Indices for Proteins. A Formalism Based on “Molten Globule” Theory. Physics Procedia 8:63-72.
10. Ruiz-Blanco YB, Marrero-Ponce Y, Paz W, García Y, Salgado J (2013) Global Stability of Protein Folding from an Empirical Free Energy Function. Journal of Theoretical Biology 321:44-53.
11. Jiang L, Kuhlman B, Kortemme T, Baker D (2005) A “solvated rotamer” approach to modeling water-mediated hydrogen bonds at protein–protein interfaces. PROTEINS: Structure, Function, and Bioinformatics 58:893–904.
12. Plaxco KW, Simons KT, Baker D (1998) Contact Order, Transition State Placement and the Refolding Rates of Single Domain Proteins. J Mol Biol 277:985-994.
13. Burrage JSaK (2006) Predicting residue-wise contact orders in proteins by support vector regression. BMC Bioinformatics 425.
14. Nolting B, Schalike W, Hampel P, Grundig F, Gantert S, Sips N, Bandlow W, Qi PX (2003) Structural determinants of the rate of protein folding. J Theor Biol 223:299–307.
15. Micheletti C (2003) Prediction of Folding Rates and Transition-State Placement From Native-State Geometry. PROTEINS: Structure, Function, and Genetics 51:74–84.
16. Zhou H, Zhou Y (2002) Folding Rate Prediction Using Total Contact Distance. Biophysical Journal 82:458–463.