August 2008 Volume 2 Issue 2

Trade Science Inc.

BioTechnology

An Indian Journal

FULL PAPER

BTAIJ, 2(2), 2008 [106-110]

STABPRO-A computational approach to protein stabilization

Padma V.Iyer [1]*, G.Raman[2]

[1]Food Engineering and Technology Department, University Institute of Chemical Technology, Nathalal Parikh Marg, Matunga, Mumbai-400019, (INDIA)

Tel.: 91-022-24145616; Fax: 91-022-24145614

[2]Directorate of Systems (WZ), New Custom House, Ballard Estate, Mumbai-400001, (INDIA)

E-mail: iyer_padma@rediffmail.com

Received: 16th June, 2008 ; Accepted: 21st June, 2008

ABSTRACT

The present work envisages an approach to predict the effect of planned mutations in the amino acid sequences on stability of proteins using an application software named STABPRO, which was developed by the authors. The software predicts the effect of planned mutations on stability of a protein based on parameters such as hydrophobicity, steric volume, aliphatic-aromatic character, polar and non-polar residues that were present before and after the planned mutation. Further, four enzymes reported in literature for mutations were taken up for the demonstration of the software. It was found that the software correctly gave a quantitative prediction of the results given in literature.

KEYWORDS

Protein stability;
Protein engineering;
Mutation;
Conformation;
Bioinformatics;
STABPRO.

1. INTRODUCTION

Biological polymers like protein are characterized by complex and hierarchically organized structures. Structural diversity of proteins coupled with their functional versatility has made the study of protein structure and folding a central area of modern biology. A particular polypeptide chain folds to its native conformation in a few milliseconds via directed pathways ensuring an operative protein. On the other hand, this conformation exhibits a marginal stability equivalent to the energy of a few hydrogen bonds. The conformational stability defined as a free energy change ΔG configuration for the transitions folded state - unfolded state is in the order of 30-60 KJ/mol at room temperature[1]. The biological significance of the low ΔG configuration is the requirement for balance between rigidity and flexibility. Rigidity determines the protein stability while flexibility is important for protein function.

Discovering the relationship between protein sequence and conformation is a fascinating theoretical problem of fundamental importance[2]. Protein structures are more highly conserved than sequences during evolution. Evolutionary studies have found subtle changes in amino acid sequence between thermophilic enzymes and their mesophilic counterparts. These subtle differences do not affect the function of the protein as much as their stability. The only disadvantage that enzymes face, compared to chemical catalysts, is their inherent instability. Presently stable proteins are obtained by isolation from extremophiles (thermophilic enzymes), mutation of enzymes obtained from mesophiles or by the physical stabilization of mesophilic enzymes by methods like immobilization, medium engineering etc.

A thermophilic protein differs from its mesophilic counterpart by small differences in its primary structure, the three-dimensional structure of the two being largely similar. Operating temperatures for thermophilic enzymes is usually about 20-30°C higher corresponding to an increase in stability by 5-7 kcal/mol, which is a small change in ΔG. A few salt bridges, several additional hydrogen bonds or seven to ten additional CH_3-groups in the hydrophobic nucleus of the protein can contribute to this small change in ΔG[3]. Thus we see that subtle differences existing between the primary structures of mesophilic and thermophilic enzymes can lead to a huge difference in operational stability of the enzymes.

Protein engineering refers to the use of mutations to introduce changes into protein molecules. The primary goal of protein engineering is to create a novel protein that possesses some improved property by changing the amino acid sequence of the existing protein resulting in a protein which may fold differently, have an alteration in activity or may be bestowed with increased or decreased stability. The design of proteins with enhanced stability with retention of similar activity is one of the major goals of protein engineering. Mutation of protein to obtain another protein of interest remains a difficult task, even with the great deal of information on structure activity relationships. The predominant approach to a stable biocatalyst design by protein engineering has focused on the modification of existing mesophilic enzyme sequences based on thermophilic enzyme sequences. This approach has yielded great results because there are certain characteristics of thermophilic enzyme sequences, which have been found to contribute towards their stability. Thus amino acid residues prone to deactivation by hydrolysis, deamination, oxidation etc but not in the essential active state region can be replaced by amino acids having similar spatial arrangement, charge and other characteristics resulting in a protein having same activity/function and increased stability. Some basic factors contributing to protein stability are given in TABLE 1. A compilation of amino acid properties and steric parameters is given in TABLE 2, all of which are important for amino acid replacement towards enhanced activity, stability or both.

The possibility of correctly identifying the amino acid to be changed and the selection of a residue which is most suitable replacement in terms of stabilization as well as activity, is quite difficult. This can be made easier using a software which, on a primary level, correctly predicts the stabilization/destabilization effect of a residue change based on parameters like hydrophobicity, aromatic-aliphatic character, steric volume occupied etc.

TABLE 1: Factors contributing to protein stability

Features	Reasons for stability
Low number of cysteines and methionines	Lesser tendency of SH-group oxidation
Higher content of Arginine and lower content of lysine	Arginine helps in screening internal hydrophobic part of molecule from surrounding
Lower number of polar amino acids like Serine or threonine	Polar groups in protein interior makes protein thermodynamically unstable
Increase in hydrophobicity	Hydrophobic interactions strengthen with temperature (up to 70°C)
High aliphatic character	Aliphatic hydrophobic interactions increase with temperature while aromatic ones do not increase.

TABLE 2: Properties of amino acids

Amino acid	Codes	Polarity	Charge	Aliphatic/ Aromatic	pKa	Volume (A3)	Kyte-doolittle index
Alanine	Ala A	nonpolar	-	Aliphatic	-	89	1.8
Cysteine	Cys C	polar	uncharged	Aliphatic	8.4	109	2.5
Aspartic acid	Asp D	polar	charged	Aliphatic	3.9	111	-3.5
Glutamic acid	Glu E	polar	charged	Aliphatic	4.1	138	-3.5
Phenylalanine	Phe F	nonpolar	-	Aromatic	-	190	2.8
Glycine	Gly G	polar	uncharged	Aliphatic	-	60	-0.4
Histidine	His H	polar	charged	-	6.1	153	-3.2
Isoleucine	Ile I	nonpolar	-	Aliphatic	-	167	4.5
Lysine	Lys K	polar	charged	Aliphatic	10.8	169	-3.9
Leucine	Leu L	nonpolar	-	Aliphatic	-	167	3.8
Methionine	Met M	nonpolar	-	Aliphatic	-	163	1.9
Asparagine	Asn N	polar	uncharged	Aliphatic	-	114	-3.5
Proline	Pro P	nonpolar	-	Aliphatic	-	113	-1.6
Glutamine	Gln Q	polar	uncharged	Aliphatic	-	144	-3.5
Arginine	Arg R	polar	charged	Aliphatic	12.5	174	-4.5
Serine	Ser S	polar	uncharged	Aliphatic	-	89	-0.8
Threonine	Thr T	polar	uncharged	Aliphatic	-	116	-0.7
Valine	Val V	nonpolar	-	Aliphatic	-	140	4.2
Tryptophan	Trp W	nonpolar	-	Aromatic	-	228	-0.9
Tyrosine	Tyr Y	polar	uncharged	-	10.1	194	-1.3

The present work involves development of software STABRO to predict stability as a result of mutation based on the abovementioned general parameters of protein stability. STABPRO results, in combination with thermodynamic considerations, can assist in making a planned mutation giving a protein with higher stability and efficiency.

2. EXPERIMENTAL

The software has been developed in free version of Visual Basic Express Edition (running on the .NET

FULL PAPER

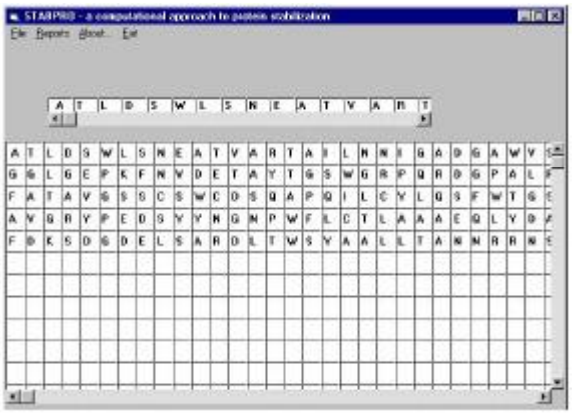

Figure 1: Screenshot showing the protein data bank file read by STABPRO and its display in the format given

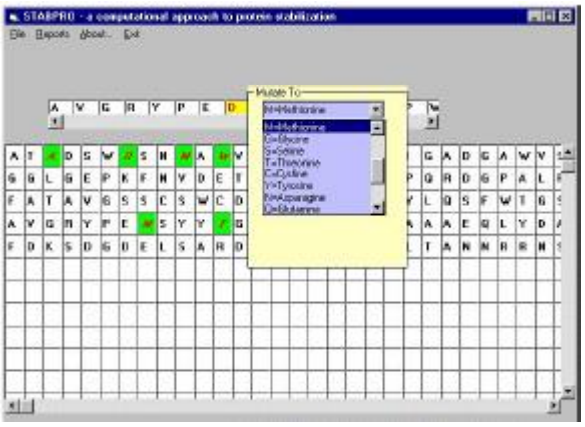

Figure 2: Screenshot indicating the selecting of a particular residue and selection of the residue to which the mutation is planned. The residues mutated have been indicated in italics

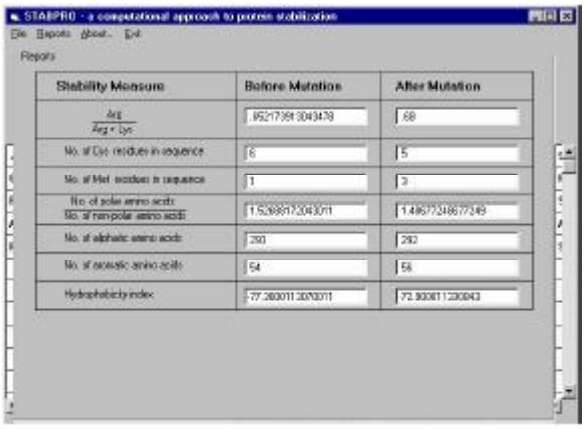

Figure 3: Screenshot indicating the results of the computation before and after planned mutation giving the values of the stability parameters

Framework). The software basically accepts sequence data from a PDB (Protein Database) file. Sequences can be imported from the PDB database using the 'Load Sequence' option. Single alphabet codes for the amino acid sequence is taken and arranged. The amino acid sequence is displayed in the top grid as well as in the bottom grid. Figure 2 indicates how the residue to be mutated is selected in the top grid and a drop-down box displaying the amino acid list is displayed. From the drop-down box the amino acid to which it has to be changed is selected e.g.: Val to Thr at position 138. So at V (Val) 138 the drop-down box is displayed containing all the amino acids from which T (Thr) is selected. In the lower grid the letter T is displayed distinctly to indicate the change. Multiple amino acid changes can be tried and the software will give the cumulative effect of all the changes. 'Reports' option will give the calculated values of the stability parameters BEFORE and AFTER the mutation. The output can be saved into HTML format, plain text data or .jpg file.

3. RESULTS AND DISCUSSION

As can be seen in, using data obtained from literature the utility of STABPRO was checked.

The parameters for stability of protein have been identified as

- **Arg/(Arg+Lys):** There is often a higher number of

TABLE 3: Effect of planned mutation results from STABPRO for T4 lysozyme, Lactate dehydrogenase, Staphylococcal nuclease and Ferredoxin before the mutations and after the mutation as given in text

Stability indicator	T4 lysozyme		Lactate dehydrogenase		Staphylococcal nuclease		Ferredoxin	
	Before	After	Before	After	Before	After	Before	After
Arg/(Arg+Lys)	0.5	0.5	0.4615	0.4358	0.1785	0.1785	0.1428	0.25
No.of Cys residues	0	0	5	5	0	0	10	10
No.of Met residues	5	5	7	7	4	3	1	1
Ratio of polar to non-polar esidues	1.38	1.35	1.262	1.294	1.84	1.94	1.304	1.304
No.of aliphatic amino acids	86	87	181	181	69	71	44	43
No.of aromatic amino acids	14	14	21	21	11	9	9	9
Hydrophobicity	-66.5	-61.5	-0.3	-4.6	-132.5	-143.5	-27.1	-33.1
Effect of mutation	Stabilization		Destabilization		Stabilization		Stabilization	

arginine residues and lower number of lysine residues in thermophilic proteins. Both residues are generally localized on surface of the protein and have contact with water. Also they have large hydrocarbon moieties due to which contact with water results in instability. Arginine however has one -CH_2 group less and the guanidine group of arginine owing to its large steric volume helps in screening the internal hydrophobic cavities. It has similar characteristics as lysine and hence higher number of arginine residues and lower number of lysine residues can give an increase in stability[3].

- **Number of cysteine (Cys) residues and methionine (Met) residues:** Cysteine and methionine are prone to oxidation of their -SH groups. Thus a decrease in the number of these amino acids can decrease destabilization due to -SH oxidation. Therefore, lesser number of cysteine residues in a protein or decreased accessibility of the residues to the solvent will result in decreased propensity to oxidation type inactivation. Alcohol dehydrogenase from *Bacillus stearothermophilus* has same number of cysteine residues as its mesophilic counterpart but all its SH-groups are localized inside the protein globule and are hence unreactive in oxidation reactions[4].
- **Ratio of the number of polar to non-polar residues:** Hydrophobic index and ratio of polar to non-polar residues are important in hydrophobic interactions. Hydrophobic interactions are the only interactions that increase in strength with temperature at least up to 60-70°C[5].
- **Number of aliphatic amino acids:** Generally the number of aliphatic amino acids is directly related to protein stability. Ikai (1980) developed a parameter called aliphatic index for quantitative estimation of aliphatic amino acid content and correlated it to stability of thermophilic proteins. Aliphatic index of thermophilic proteins is much higher than that of mesophilic proteins. It was found that while aromatic hydrophobicity changed very little with temperature, aliphatic hydrophobicity increased with temperature[6].
- **Number of aromatic amino acids:** In some thermophilic proteins, a decrease in aromatic amino acid residues is observed corresponding to an increase in aliphatic amino acid residues such that the overall hydrophobicity of the protein remains the same. This results in better stability, as aliphatic amino acids are important in terms of both energy and geometric criteria of stabilization.
- **Hydrophobicity:** Thermophilic proteins show enhanced stability in correlation to their hydrophobicity. When point mutations of amino acid residues with those having higher hydrophobicity was done, it resulted in better stability. For example, glutamic acid residue located in the interior of tryptophan synthase, when replaced with more hydrophobic residues like tyrosine, valine or methionine resulted in stabilization[7]
- **Compactness of packing:** Thermophilic proteins have more compact packing compared to their mesophilic counterparts. A compact structure keeps internal water out of the hydrophobic nucleus of the protein thus improving its stability. Replacement of an amino acid by a bulkier amino acid, which is chemically similar (in terms of charge, aliphatic, polar character and steric volume etc), and which does not change the polypeptide backbone conformation greatly and results in increased stability of the protein. Thus, we see that alanine and threonine, owing to their more compact structure, occur more frequently in thermophilic proteins[3].
- **Electrostatic interactions:** Formation of one or two additional salt bridges in the interior of the protein can account for 5-7kcal/mol decrease in free energy of a thermophilic protein as compared to its mesophilic counterpart lacking these salt bridges. When *Neurospora crassa* tyrosinase isozymes were studied, it was found that stable isozymes had an extra intramolecular salt bridge compared to more labile ones[8].

The role of the above mentioned parameters have been discussed in the following four cases. The results obtained have been tabulated in TABLE 3 wherein the software indicates the beneficial/non-beneficial nature of mutation.

Consider the case of T4 lysozyme wherein, Ile3[9] and Thr157[10] on mutation to Ala results in an increase in hydrophobicity, decrease in interior cavities and formation of hydrogen bonds and salt bridge interactions-all leading to an increase in stability. From the results, it can be seen that the polar amino acid content decreases, aliphatic amino acid content increases and hydrophobicity also increases. Thus the software correctly pre-

FULL PAPER

dicts this to be a beneficial mutation towards stability.

Similarly, in case of *Staphylococcus sp*. nuclease Leu, Val, Tyr, Phe, Met and Ile residues are mutated to Ala/Gly. These lead to increased aliphatic characteristics and elimination of internal cavities[11]. STABPRO gives comparative values, which are highly indicative of the stabilizing effect.

In case of Ferredoxin, conversion of Gly, Ser, Lys, Asp to Ala, Thr, Arg, Glu respectively leads to overall compactness of the protein due to increase in the volume occupied. Also, there is an increase in aliphatic character contributing to stability[12].

Consider again the case of lactate dehydrogenase, wherein a mutation of Glu, Arg, Thr, Ala (as found in thermophilic enzymes) to Asp, Lys, Ser, Gly respectively (as found in mesophilic enzymes) leads to destabilization. Values from STABPRO show a decrease in the Arg/(Arg+Lys) value, a decrease in the hydrophobicity and an increase in polar amino acid content, which are all indicative of lower stability.

4. CONCLUSION

Thus, STABPRO shows prediction of stability based on the abovementioned stability indicators. This approach is entirely based on incorporating the structural features of thermophilic enzymes. Thus, it is seen that STABPRO is helpful in predicting the stabilization or destabilization effect of planned mutations on protein stability. Further work for a complete prediction of protein stability incorporating the effect of thermodynamics is in progress.

5. REFERENCES

[1] P.L.Prevalov; Adv.Prot.Chem., **33**, 167-241 **(1979)**.
[2] M.Levitt, A.Warshel; Nature, **253**, 694-698 **(1995)**.
[3] Mozhaev, Martinek; Enzyme and Microbial Technology, **6**, 50-59 **(1984)**.
[4] J.Bridgen, E.Kolb, E.Harris; FEBS Letters, **33**, 1-3 **(1973)**.
[5] J.F.Brandts; J.Amer.Chem.Soc., **86**, 4302-4314 **(1964)**.
[6] A.Ikai; J.Biochem.(Tokyo), **88**, 1895-1898 **(1980)**.
[7] K.Yutani, K.Ogasahara, M.Suzuki, Y.Sugino; J. Biochem.(Tokyo), **87**, 117-121**(1980)**.
[8] C.Ruegg, D.Ammer, K.Lerch; J.Biol Chem., **257**, 6420-6426 **(1982)**.
[9] M.Matsumara, W.Becktel, B.Matthews; Nature, **334**, 406-410 **(1988)**.
[10] Alber, S.Dao-pin, K.Wilson, J.Wozniak, S.Cook, B.Matthews; Nature, **330**, 41-46 **(1987)**.
[11] D.Shortle, E.Stites, K.Meeker; Biochemistry, **29**, 8033-8041 **(1990)**.
[12] P.Argos, G.Rossmann, M.Grau, H.Zuber, G.Frank, J.Tratschin; Biochemistry, **18**, 5698-5703 **(1979)**.