

MEICPS: substitution mutations to engineer intracellular protein stability

Boojala V.B. Reddy, P. Ramesh and Shrish Tiwari

Centre for Cellular and Molecular Biology, Hyderabad 500007, India

Received on June 2, 1997; revised on August 21, 1997; accepted on August 28, 1997

Abstract

Summary: In MEICPS, results from earlier analyses are utilized to suggest possible substitution point mutations to engineer intracellular stability using a given sequence or structure of the protein.

Availability: From bvbreddy@ccmb.ap.nic.in. This program needs data from other software, PSA and SSTRUC, available from sali@tamika.rockefeller.edu and tom@cryst.bioc.cam.ac.uk, respectively.

Contact: bvbreddy@ccmb.ap.nic.in

It has been suggested that sequence-dependent properties, global structural features and location in the intracellular environment determine the *in vivo* stability of proteins (Rogers *et al.*, 1986; Rechsteiner and Rogers, 1996). From our earlier analysis of sequence data of a set of stable proteins (*in vivo* half-life ≥ 16 h) versus less stable proteins (*in vivo* half-life ≤ 5 h), it was proposed that the overall composition of the dipeptide sequence in a protein determines its intracellular stability (Guruprasad *et al.*, 1990). We have further shown that, for a given sequence, the dipeptide occurrence could be used to distinguish short-lived from stable proteins, and suggest this as a method to predict *in vivo* stability.

We classified 400 combinations of dipeptides into three classes—stabilizing (Stb), destabilizing (Dst) and normal (Nor)—and studied their structural distribution in a set of 303 non-homologous best resolved (≤ 2.0 Å) protein structures (Reddy, 1993; 1996). A significant difference in the overall frequency of occurrence of Stb and Dst dipeptides in different secondary structural regions was observed. The sensitive dipeptides (Stb + Dst) are less frequent in β -strands and more frequent in coils. A high frequency of occurrence of Stb is observed in the regions closer to the molecular surface compared to the Dst and Nor dipeptides. Significantly high dipole interactions are observed in the Dst dipeptides. The studies indicate that although the Dst dipeptides are more hydrophilic in nature, they are localized significantly more in the buried regions of protein structures; on the other hand, Stb are more hydrophobic in nature, but relatively more accessible to the solvent.

In MEICPS, all the sequence- and structure-dependent observations have been coded to suggest possible substitution

point mutations to increase the intracellular stability of a given structure or sequence.

The given PDB format structure file is used to calculate secondary structure type and Ooi values using the SSTRUC program (Smith, 1989). The solvent-accessible contact area of each dipeptide was calculated using the PSA (Sali and Blundell, 1990). The fraction of each class of dipeptides in the three different ranges of percentage accessibility contact area was computed. These calculations were carried out for dipeptide side-chains, main chains, polar side-chains, non-polar side-chains and total atoms separately for each class of dipeptides (Reddy, 1996).

We have used the secondary structure types, solvent accessibility and Ooi values, each in three subclasses, to define the structural environment of a dipeptide. Based on the Stb, Dst and Nor dipeptide fractional probabilities (Reddy, 1996), the dipeptide structural propensity (DSP) values are calculated. The sum of these is normalized to 100 and the proportional percentage of each dipeptide class is taken as the corresponding DSP value. This is repeated for all the dipeptide subclasses in each structural parameter (see Table 1).

In the Ooi values, there are two groups of data, one for 8 Å and another for 14 Å radius; 50% weight is given to each of these groups to calculate the DSP. Similarly, there are five groups of data in the case of the solvent accessibility parameter; correspondingly, 20% weight is given to each of these groups to calculate DSP (Table 1).

The program uses the sequence data file in SWISS-PROT format and the structure data file in PDB format to suggest substituted point mutations (Figure 1).

Using the sequence data file: if only the sequence of amino acids is to be used to suggest point mutations, we first identify the Dst dipeptides by using the dipeptide instability weight values (DIWV) given in Reddy (1996). We note the change in protein instability index (PII) with all the possible changes in the N-terminal or C-terminal residues to convert the Dst to an Stb dipeptide (Guruprasad *et al.*, 1990). For every Dst dipeptide, the program suggests the best possible substitution mutation(s) with the corresponding PII values (Example 1). **Using the structure data file:** the structural propensity index (SPI) is calculated by summing all the DSP values corresponding to their respective structural environment. Thus, for each

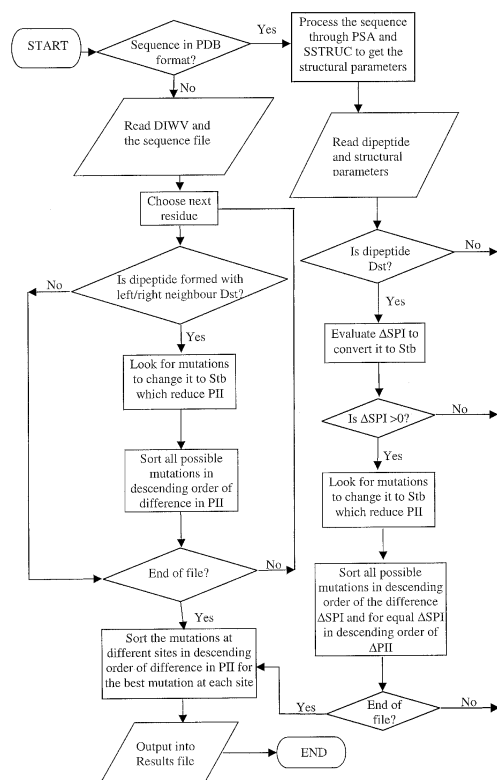


Fig. 1. A simplified flow diagram of the computer program MEICPS.

EXAMPLE 1

Mutations suggested to improve the stability of the protein.
PII of DCOR_MOUSE with 461 residues is 52.16

Code Pos Res Suggested mutations, decrease in PII

DCOR_MOUSE 41 FYV (I 3.30)(W 3.30)(A 3.30)(G 3.30)
(T 3.30)(L 3.30)(K 3.30)
DCOR_MOUSE 13 HIL (F 3.21)(F 3.21)(G 2.99)(P 2.99)
DCOR_MOUSE 12 CHI (G 2.47)
DCOR_MOUSE 289 VNI (G 2.21)
DCOR_MOUSE 290 NII (G 2.21)(F 1.99)(W 1.99)(T 1.99)(P 1.91)
...Continued

EXAMPLE 2

Mutations to improve the protein structural environment and stability.
PII of 8ACN with 753 residues is 26.68

Code Pos Res Mutations, change in SPI, decrease in PII

8ACN 316 SEL (G 13.75 0.71)(F 7.16 0.71)
8ACN 523 DSS (W 13.75 0.71)
8ACN 611 QEF (Q 13.75 0.50)(F 13.75 0.42)
8ACN 301 PDS (W 13.75 0.42)
8ACN 522 KDS (W 13.75 0.42)
...Continue

SPI: Structural Propensity Index
PII: Protein Stability Index

Examples 1 and 2.

dipeptide we calculate SPI-Stb, SPI-Dst and SPI-Nor, irre-

Table 1. Dipeptide structural propensity values

3. Solvent accessibility				
a) All atom accessibility				
Type	<7%	7-20%	>20%	
Stb	5.8	6.1	7.1	
Dst	7.3	6.9	6.4	
Nor	6.9	7.0	6.5	
b) Non-polar side-chain accessibility				
Stb	6.2	6.5	7.3	
Dst	7.0	6.9	6.2	
Nor	6.8	6.6	6.4	
c) Polar side-chain accessibility				
Stb	5.7	6.1	7.1	
Dst	7.4	7.0	6.4	
Nor	6.9	6.9	6.5	
d) Side-chain accessibility				
Stb	6.0	6.3	7.0	
Dst	7.1	6.9	6.5	
Nor	6.9	6.9	6.5	
e) Main-chain accessibility				
Stb	5.1	6.8	7.9	
Dst	8.2	6.2	5.6	
Nor	6.8	7.0	6.5	

spective of their class. If the SPI-Dst is significantly less than SPI-Stb, a point mutation is suggested to convert the Dst to Stb and with a maximum possible decrease in PII (Example 2).

The program suggests all possible substitution mutations and puts them in decreasing order of preference. The user chooses which do not have an effect on the basic function or activity of the protein and which are experimentally feasible.

Acknowledgement

The authors are grateful to DIC-Bioinformatics (CCMB) for computational facilities.

References

- Guruprasad, K., Reddy, B.V.B. and Pandit, M.W. (1990) Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting *in vivo* stability of a protein from its primary sequence. *Protein Eng.*, **4**, 155–161.
- Rechsteiner, M. and Rogers, S.W. (1996) PEST sequences and regulation by proteolysis. *Trends Biochem. Sci.*, **21**, 267–271.
- Reddy, B.V.B. (1993) Correlation between protein stability and its dipeptide composition: studies on their structural and biological implications. *Protein Eng.*, **6**(Suppl.), 23–23.
- Reddy, B.V.B. (1996) Structural distributions of dipeptides that are identified to be determinants of intracellular protein stability. *J. Biomol. Struct. Dyn.*, **14**, 201–210.
- Rogers, S., Wells, R. and Rechsteiner, M. (1986) Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis. *Science*, **234**, 364–368.
- Sali, A. and Blundell, T.L. (1990) Definitions of general topological equivalence in protein structures. *J. Mol. Biol.*, **212**, 403–442.
- Smith, D. (1989) *SSTRUC A Computer Program*. Department of Crystallography, Birkbeck College, University of London.