Balance of interactions contributing to stability of globular proteins

Fačkovec Boris[1,2], Vondrášek Jiří[1]

[1]Institute of Organic Chemistry and Biochemistry AS CR, v.v.i., Flemingovo nám. 2, 166 10 Prague 6, Czech Republic
[2]Faculty of Science, Charles University in Prague, Albertov 6, 128 43 Prague 2, Czech Republic
mailto: boris.fackovec@uochb.cas.cz

Introduction

- Structure prediction algorithms require fast evaluation of **free energy for a given structure**.
- Fast *in silico* prediction of stability change upon amino acid replacement is demanded by protein engineering community.
- Role of intramolecular interactions in stabilization of native state is still unsatisfactorily understood.

- Calculation of residue-residue interactions on even sub-chemical level of accuracy (achievable using benchmark quantum chemistry methods, e.g. CCSD(T)/CBS) is inaccurate due to error propagation.
- Entropy estimate from single structure is difficult.
- Effect of environment is not fully understood.
- Structure of denatured state is difficult to describe, random coil approximation is insufficient.
- Amount of consistent experimental data is insufficient for parametrization of models having more than about 20 parameters.

We hereby propose a model of protein stabilization as a sum of **physically meaningful** contributions.

Model

ASSUMPTIONS

- Native structure ensemble can be described by 1 structure
- Energetics of denatured state can be inferred from native structure
- Free energy contribution of a residue-residue interaction to stability is a monotonic function of interaction energy value for one interaction class

STABILITY DECOMPOSITION

- stabilization free energy is a sum of **1-body and 2-body amino acid contributions**

$$\Delta G_{fold} = \sum_{i=1}^{N} \Delta G_i^{(1)} + \sum_{i=1}^{N} \sum_{j=1}^{N} \Delta G_{ij}^{(2)} + \delta \Delta G^{(3)}$$

$\Delta G_i^{(1)}$= solv(denat) + conf.entropy - solv(native)
= a(type$_i$) + solv(native)
$\Delta G_{ij}^{(2)}$= resi-resi.int(native) - resi-resi.int(denat)
= b(type$_{ij}$) * int.energy$_{ij}$

- N is number of amino acids in protein
- a_i and b_{ij} are constants for particular type of amino acid of interaction respectively.
- In previous studies [1], we have found that **10 classes of interactions** are sufficient.
- **Solvation** energy in native state is calculated using a macroscopic model as **linear function of** non-polar, polar and charged **surface area**.

FINAL FORMULATION

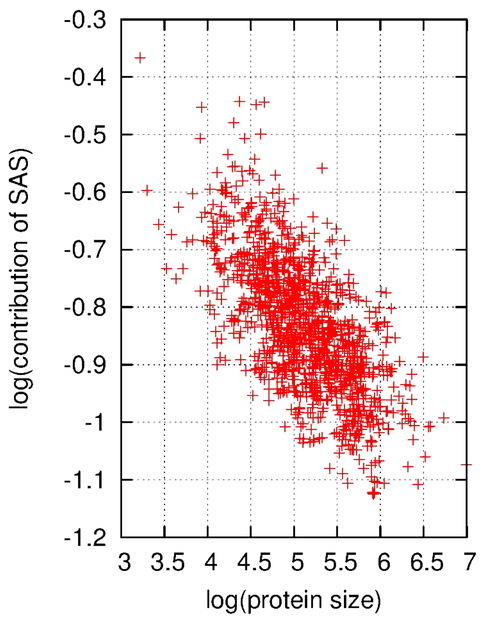
$$\Delta G_{fold} = \sum_{i=1}^{20} a_i \; n_i + \sum_{j=1}^{10} IE_j \; b_j + \sum_{k=1}^{3} SAS_k \; c_k$$

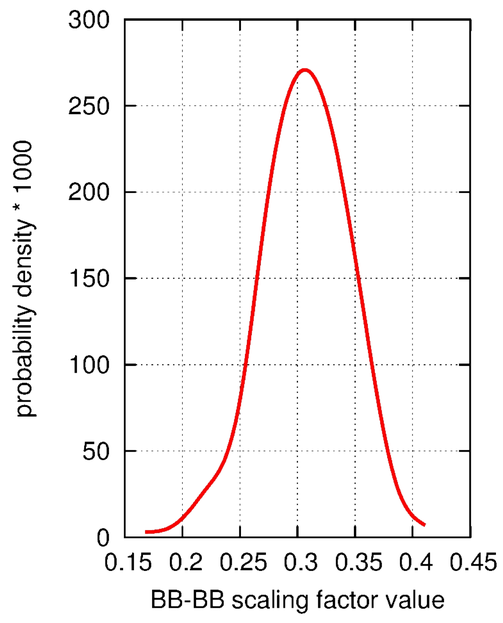
where n_i is number of amino acids of type i, IE_j sum of interaction energies of type j and SAS_k area of surface of type k.
- a_i, b_j and c_k constitute **33** parameters of the model, leter **scaling factors**
- IE_j and a_i are in kcal/mol. n_i and b_j dimensionless. SAS_k is in A^2 and c_k is in kcal/(mol.A^2)

Results & Discussion

scaling factor	value	stdev in %	contrib in %
GLY	-4.9	2.4	4.0
ALA	-6.6	1.8	5.8
VAL	-7.3	1.8	6.1
ILE	-7.4	1.9	4.8
CYS	-7.8	2.2	1.3
LEU	-8.2	1.6	8.6
MET	-9.2	1.9	1.7
PHE	-10.3	1.7	4.7
PRO	-16.7	1.8	8.8
THR	-7.9	1.6	4.9
SER	-8.2	1.5	5.5
TYR	-13.8	1.3	5.4
ASN	-14.8	1.0	7.3
TRP	-16.1	1.4	2.5
GLN	-17.0	0.90	7.4
LYS	-4.6	2.7	3.5
HIS	-5.3	2.9	1.3
ASP	-5.9	3.3	3.9
GLU	-6.6	3.2	5.3
ARG	-12.3	1.3	7.2
BB-BB	0.31	8.9	16.2
BB-CH	0.10	22	3.4
BB-PO	0.57	7.1	9.2
BB-NP	0.78	7.0	17.6
CH-CH	0.012	86	0.7
CH-PO	0.051	51	0.5
CH-NP	0.15	24	0.7
PO-PO	0.75	8.8	1.5
PO-NP	1.00	6.5	5.1
NP-NP	0.16	29	1.1
SASNP	0.43	13	6.3
SASPO	6.2	4.3	6.8
SASCH	19.9	1.0	30.9

TABLE: 1st column - types of scaling factors. 2nd column - their values. They are relative to each other, i.e. one of them must be set to arbitrary value. 3rd column - deviation across the structure set. 4th column - average **contribution to stabilization** (+ sign) or destabilization (- sign) of proteins.
- Setting largest 2-body scaling factor to 1 leads to 1-body scaling factors in the order of magnitude of solvation free energies.
- Expected values of 2-body scaling factors are between 0 and 1. They account for entropic compensation of interaction energies and persistence of interactions in denatured state.
- Surprisingly **low are scaling factors of charged** residues.





LEFT FIGURE: Notable is contribution of native state **surface solvation** (42%). Its average value 700 kcal/mol agrees with usual solvation free energies of proteins. Contribution of surface solvation to stability decreases with increasing protein size.
RIGHT FIGURE: Histogram of BB-BB scaling factor across proteins.

Methods

- X-ray structures, resolution better than 2A, no ligands, no DNA/RNA, 70% seq. identity removed
- Hydrogens were added (pH 5.5 - Histidine charged) and optimized in GROMACS using OPLS force field.
- Proteins were **fragmented** into backbones and sidechains. **Sidechains** were **classified** into 3 groups - non-polar (NP - Ala, Val, Ile, Cys, Leu, Met, Phe, Pro), polar (PO - Thr, Ser, Tyr, Asn, Trp, Gln) and charged (CH - Lys, His, Asp, Glu, Arg). 4 types of fragments = 10 types of interaction energies
- Interaction energy was calculated between each pair of (2 N) fragments - 4 N^2 interaction energies, which were divided into 10 **interaction energy matrices** [2].
- Surface area was calculated using algorithm implemented in g_sas routine of GROMACS package with default radius of probe (1.4A), non-polar surface defined by partial charge of atom between 0 and 0.35, polar between 0.35 an 0.65 and charged above 0.65.
- Parametrization of the model was done using **Courant-Fischer-Weyl** min-max **theorem**, from which it follows that expression $\frac{\mathbf{x}^T \mathbf{A}^T \mathbf{A} \mathbf{x}}{\mathbf{x}^T \mathbf{x}}$

has its minimum if $\mathbf{x}$ is eigenvector of $\mathbf{A}^T\mathbf{A}$ corresponding to its lowest eigenvalue.
- Exclusion of small (N<50) peptides led to matrix $\mathbf{A}$ of 1188 proteins times 33 calculated values per protein, $\mathbf{A}^T\mathbf{A}$ is symmetric matrix 33 times 33.
- Calculation of standard deviation of scaling factors was done by multiple calculations on different random subsets of 400 proteins from data set.

Conclusion

- Average absolute value of ΔG is **30 kcal/mol**, if average sum of stabilizing terms is **1700 kcal/mol**, root of mean square of **compensation** of stabilizing and destabilizing forces is **97.3%**.
- Model seems to be **transferable** and **well-balanced**.

- **Method** of parametrization can be used for different data sets.
- In future, we plan to develop **energy functions** for stability change prediction upon mutation and combine them with established methods (ERIS, FoldX etc.)

References:

[1] Decomposition of Intramolecular Interactions Between Amino-Acids in Globular Proteins - A Consequence for Structural Classes of Proteins and Methods of Their Classification, Fačkovec and Vondrášek, Chapter 4, ISBN: 978-953-307-280-7, 2011
[2] Identifying stabilizing key residues in proteins using interresidue interaction energy matrix, Bendová-Biedermannová et al., Proteins: Struct., Funct. & Bioinf, 2008

Acknowledgement:

This work was supported by Grant No. LC512 from the Ministry of Education, Youth and Sports (MSMT) of the Czech Republic. It was also a part of the research projects No. Z40550506 and MSM6198959216.