

# A Short Tutorial of Calculating $\Delta\Delta G$

Ying Liu

November 18, 2008

## Rationale of $\Delta\Delta G$

$\Delta\Delta G$  is based on  $\Delta G$ :

$$\Delta G = G_f - G_u$$

where  $G_f$  is the free energy of the folded protein, and  $G_u$  is that of the unfolded protein. Then  $\Delta G$  is the free energy change involved in protein folding.

$\Delta\Delta G$  arises when mutation occurs:

$$\Delta\Delta G = \Delta G^{\text{MT}} - \Delta G^{\text{WT}}$$

where  $\Delta G^{\text{MT}}$  is the folding energy of the mutant, and  $\Delta G^{\text{WT}}$  is the folding energy of wildtype protein. Thus  $\Delta\Delta G$  is a measure of impact of a certain mutation on protein folding.

Our calculation is based on Bueno et al. (2007). It is highly recommended that you read the paper before you proceed.

## Calculation

Suppose a certain residue  $i$  is mutated to  $j$  in a protein, and let the subscript denote the mutation status, whereas the superscript denote the folding status. Then

$$\begin{aligned}\Delta\Delta G &= \Delta G_j - \Delta G_i \\ &= (G_j^f - G_j^u) - (G_i^f - G_i^u) \\ &= (G_j^f - G_i^f) - (G_j^u - G_i^u)\end{aligned}\tag{*}$$

and

$$G = E_{\text{vdw}} + G_{\text{elec}} + \sigma_{\text{hyd}}\text{ASA} - TS_{\text{conf}}\tag{**}$$

So we will calculate the each term in equation (\*\*) for each  $G$  in equation (\*).

Next we perform calculation on Barnase as an example.

---

## Folded

### Materials

We obtained the structures of Barnase from the Protein Data Bank Berman et al. (2000), both the wildtype (PDB id: 1BNI) and the I51V mutant (PDB id: 1BSA); i.e., the 51st residue, Ile, in the wildtype, is mutated to Val in the mutant.

### Methods

$E_{\text{vdw}}$  and  $G_{\text{elec}}$  can be calculated with NAMD. The instructions can be found in the Appendix section. The side chain accessibility surface area (ASA) can be calculated using the GetArea webserver (Fraczkiewicz and Braun (1998)<http://curie.utmb.edu/getarea.html>). You can use  $34 \text{ cal mol}^{-1} \text{ \AA}^{-2}$  for  $\sigma_{\text{hyd}}$ , as used in Bueno et al. (2007). The entropy  $TS_{\text{conf}}$  can be calculated based on the solvent-exposed ratio and the side-chain entropy listed in Fraczkiewicz and Braun (1998), which is described as the pseudocode below:

```
entropy is initially set to 0
for each residue i in the protein
  if exposed_ratio[i] is greater than 20%
    then entropy <- entropy + (entropy of AA type) * (1 - exposed_ratio[i])
return entropy
```

Note that the entropy of AA type can be looked up in Table 1 of Lee et al. (1994). Also note that this table only has the side-chain entropy for 17 amino acids, whereas Ala, Gly and Pro has 0 side-chain entropy.

In case that the mutant does not have a corresponding file in PDB, you can use the SCWRL homology modeling server Canutescu et al. (2003)<http://www1.jcsg.org/scripts/prod/scwrl/serve.cgi> to generate a mutant model, and add an extra column to the end of each row to be consistent with the standard PDB format. Use this file to feed the GetArea webserver.

## Unfolded

### Materials

We will assume that the energy difference between unfolded wildtype and mutant can be estimated by the energy difference between tri-peptide chain Ala-X-Ala, where X is the residue where mutation occurs. In our example, we will study the energy difference between Ala-Ile-Ala and Ala-Val-Ala. The tripeptides can be constructed with VMD or PyMol.

---

## Methods

Like in the folded case, the  $E_{\text{vdw}}$  and  $G_{\text{elec}}$  terms can be calculated with NAMD. ASA of amino acids in unfolded state can be found in Table 1 of Zielenkiewicz and Saenger (1992). The entropy term can be found in Table 1 of Lee et al. (1994).

## Calculating $E_{\text{vdw}}$ and $G_{\text{elec}}$ from NAMD

Here is a very concise description of using NAMD; for more information, please refer to the NAMD tutorial and user's guide.

### What We Need to Run NAMD

We need five files:

File Type	File Suffix	File Generation
pdb file	*.pdb	from PDB
topology file	top_*.inp	VMD installation directory (VMD/plugins/noarch/tcl/readcharmmtop1.0)
protein structure file	.psf	generated with pdb file and topology file
parameter file	par_*.inp	VMD installation directory (VMD/plugins/noarch/tcl/readcharmmpar1.0)
configuration file	*.conf	modified from an example configuration file

Table 1: Five types of files involved in NAMD.

You have to set up the configuration file to run NAMD. Some important settings include:

- the location of .pdb, .psf and .par files;
- the number of steps you wish to run.

The command line is

```
namd2 filename.conf > filename.log
```

Then you can use the energies from the log file.

## Acknowledgement

This tutorial is made possible by Dr. Marta Bueno, Prof. Carlos Camacho, and Dr. Alpay Temiz.

---

ETITLE:	TS ELECT TOTAL	BOND VDW TEMP	ANGLE BOUNDARY TOTAL2	DIHED MISC TOTAL3	IMPRP KINETIC TEMPAV
G					
ENERGY:	0	1913.4792	3008.0536	358.3486	10.7302
	-18281.1014	5217.3380	0.0000	0.0000	0.0000
0	-7773.1518	0.0000	-7773.1518	-7773.1518	0.0000

Figure 1: A snapshot of the NAMD log file.

## References

- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000). The protein data bank. *Nucleic Acids Res*, 28(1):235–242.
- Bueno, M., Camacho, C. J., and Sancho, J. (2007). Simple estimate of the free energy change due to aliphatic mutations: superior predictions based on first principles. *Proteins*, 68(4):850–862.
- Canutescu, A. A., Shelenkov, A. A., and Dunbrack, R. L. (2003). A graph-theory algorithm for rapid protein side-chain prediction. *Protein Sci*, 12(9):2001–2014.
- Fraczkiewicz, R. and Braun, W. (1998). Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules. *Journal of Computational Chemistry*, 19:319–333.
- Lee, K. H., Xie, D., Freire, E., and Amzel, L. M. (1994). Estimation of changes in side chain configurational entropy in binding and folding: general methods and application to helix formation. *Proteins*, 20(1):68–84.
- Zielenkiewicz, P. and Saenger, W. (1992). Residue solvent accessibilities in the unfolded polypeptide chain. *Biophys J*, 63(6):1483–1486.