MMGBSA: Thermodynamics of Biomolecular Systems

The MMGBSA approach employs molecular mechanics, the generalized Born model and solvent accessibility method to elicit free energies from structural information circumventing the computational complexity of free energy simulations. The MMGBSA approach is parameterized within the additivity approximation [1] wherein the net free energy change is treated as a sum of a comprehensive set of individual energy components, each with a physical basis. The approach, when used with modified solvation parameters viz. m2GB model [2], is expeditious and fairly reliable for studying the energetics of bimolecular systems. Especially important is the application of this approach to determine the binding free energies in biomolecular complexes e.g protein-DNA, protein-drug and DNA-drug complexes.

The statistical mechanical theory [3,4] of binding affinities in aqueous media is presented below.

$$[P]_{aq} + [L]_{aq} = [P^*L^*]_{aq} \quad (1)$$

P and L are the reactants and P*L* is the product of binding in aqueous medium. The superscript '*' denotes structural changes accompanying binding. At equilibrium

$$\mu_{P.aq} + \mu_{L.aq} = \mu_{P^*L^*.aq} \quad (2)$$

$\mu_{P.aq}$ is the chemical potential of species P in the solvent medium (partial molar Gibbs free energy) and $\mu^o_{P.aq}$ is its standard chemical potential i.e. under conditions of 1 bar in gaseous state and 1 molar (designated as C^o) in liquid state.

$$\mu^o_{P.aq} + RT \ln (\gamma_P C_P/C^o) + \mu^o_{L.aq} + RT \ln (\gamma_L C_L/C^o)$$
$$= \mu^o_{P^*L^*.aq} + RT \ln (\gamma_{P^*L^*} C_{P^*L^*}/C^o) \quad (3)$$

where γ_P is the activity coefficient of species P and C_P its concentration. The standard molar Gibbs free energy of the reaction (standard absolute molar Gibbs free energy of binding) is

$$\Delta G^o_{aq} = \mu^o_{P^*L^*.aq} - (\mu^o_{P.aq} + \mu^o_{L.aq})$$
$$= - RT \ln [\gamma_{P^*L^*} C_{P^*L^*} C^o /(\gamma_P C_P)(\gamma_L C_L)] = - RT \ln K_{eq.aq} \quad (4)$$

In terms of canonical partition functions (Q)

$$\Delta G^o_{aq} = \Delta A^o_{aq} + P\Delta V^o_{aq} = - RT \ln K_{eq.aq} =$$
$$- RT \ln [\{Q_{P^*L^*.aq}/(N_A Q_w)\}/\{(Q_{P.aq}/(N_A Q_w))(Q_{L.aq}/(N_A Q_w))\}] + P\Delta V^o_{aq} \quad (5)$$

$$\Delta G^o_{aq} = - RT \ln [\{Q^{tr}_{P^*L^*} Q^{rot}_{P^*L^*} Z^{int}_{P^*L^*.aq} Q^{el}_{P^*L^*} N_A Q_w\}/$$
$$\{(Q^{tr}_P Q^{rot}_P Z^{int}_{P.aq} Q^{el}_P)(Q^{tr}_L Q^{rot}_L Z^{int}_{L.aq} Q^{el}_L)\}] + P\Delta V^o_{aq} \quad (6)$$

ΔA^o is the standard Helmholtz free energy of the reaction. N_A, the Avogadro number, in the above equation originates in expressing partition functions Q as molar partition functions and $P\Delta V^o_{aq}$ is the pressure-volume correction to Helmholtz free energy in the solvent medium. Q_w denotes the partition function for pure solvent (water). Z^{int} is the configurational partition function.

It includes contributions from vibrations and internal motions as well as solvation (hydration) effects. The translational and rotational terms have been separated out.

$$Z^{int}_{P.aq} = \int\ldots\int \exp\{(-E(\mathbf{X}^{N}_{P},\mathbf{X}^{M}_{W})/k_BT\}\, d\mathbf{X}^{N}_{P}\, d\mathbf{X}^{M}_{W} = \langle \exp(E(\mathbf{X}^{N}_{P},\mathbf{X}^{M}_{W})/k_BT\rangle \quad (7)$$

$\mathbf{X}^{N}_{P}$ and $\mathbf{X}^{M}_{W}$ represent the configurational space accessible to the solute P and solvent W respectively, in the presence of each other. $E(\mathbf{X}^{N}_{P},\mathbf{X}^{M}_{W})$ denotes the total potential energy of the system describing non-idealities. It includes intramolecular interactions within the solute P and solvent W as well as intermolecular interactions between the solute and the solvent. k_BT is the product of Boltzmann constant and temperature (in Kelvin).

$$Q^{el}_{P} \sim 1 \text{ (for non-covalent associations)} \quad (8)$$

The standard free energy change accompanying binding may be written as a sum of external (translational and rotational) and internal (intramolecular, intermolecular and solvation) contributions.

$$\Delta G^{o} = -RT \ln [Q^{tr}_{P*L*} N_A/(Q^{tr}_{P}Q^{tr}_{L})] - RT \ln [Q^{rot}_{P*L*}/(Q^{rot}_{P}Q^{rot}_{L})]$$
$$- RT \ln [(Z^{int}_{P*L*.aq}Q_w)/(Z^{int}_{P.aq}Z^{int}_{L.aq})] + P\Delta V^{o}_{aq} \quad (9)$$

Eq. (9) is an exact expression for evaluating binding free energies for non-covalent associations in aqueous medium. The first two terms on the right hand side of eq. (9) can be computed analytically. The third term is accessible to free energy molecular simulations configured in the canonical ensemble such as the perturbation method, thermodynamic integration, potential of mean force method etc. [5], albeit they are computationally expensive even for a single ligand and not practical in a high through-put sense even on supercomputers.

In the following, we consider some simplifications to bring the binding free energy computations into feasibility domain. The molecular translational partition function of P is

$$q^{tr}_{P} = V/\Lambda^{3}_{P} = V/(h^2/2\pi m_P k_B T)^{3/2} \quad (10)$$

The molar partition function of P is

$$Q^{tr}_{P} = (q^{tr}_{P})^{NA} \quad (11)$$

Note that the volume V has been included in the translational part consistent with ideal gas statistical mechanics. This would require that the Z^{int} be divided by V to quantify non-idealities (excess free energies). The translational part of the free energy in eq (9) is now given by the Sackur-Tetrode [6] equivalent as

$$\Delta G^{o}_{tr} = -RT \ln [(N_A/V)(\Lambda^{3}_{P}\Lambda^{3}_{L}/\Lambda^{3}_{P*L*})]$$
$$= -RT \ln [(N_A/V)(h^2/2\pi k_B T)^{3/2}\{m_{P*L*}/(m_P m_L)\}^{3/2}] \quad (12)$$

The expression in the square brackets in eq (12) is dimensionless. (N_A/V) may be replaced by a concentration term ensuring that upon transfer to aqueous medium standard free energies are recovered with the reference state anchored to a molar concentration of unity. Note that this expression is the same whether in gas phase or liquid phase provided the translational and rotational motions of the solute are unaffected by the solvent. This will be true only in a

continuum, friction-less solvent influencing the position dependent potential energy but not the velocity dependent kinetic energy of the solute. Hence in a transfer process (an experiment involving transfer of species P from one phase to another phase such as from gas phase to liquid phase or octanol to water etc.), this term cancels out. In binding processes however, no such cancellation occurs. Also if P, L and P*L* could be seen as a collection of non-bonded mono-atomic particles, then again the translational partition function for each species could be written as a product of the individual partition functions of the constituent atoms and since the number of atoms is conserved during binding, these terms would cancel out. Again, this is not so for polyatomic species where the mass in translational partition function m_P (= Σ_i m_i) is evaluated as a sum of the masses of the constituent atoms. It is recommended that Sackur-Tetrode equation be applied not in aqueous medium directly where it is invalid but upon transfer to vacuum via a suitable thermodynamic cycle.

Similar arguments apply to the rotational partition functions. Separating the rotational part from internal motions implies working under rigid rotor approximation.

$$\Delta G^{o}_{rot} = - RT \ln [(\sigma_P\sigma_L/\sigma_{P*L*})(1/(8\pi^2))(h^2/2\pi k_B T)^{3/2} x$$
$$\{(I^a_{P*L*} I^b_{P*L*} I^c_{P*L*})/(I^a_P I^b_P I^c_P I^a_L I^b_L I^c_L)\}^{1/2}] \quad (13)$$

I^a_P , I^b_P and I^c_P are the components of moments of inertia of species P along the principal axes and σ_P its symmetry number. Murray and Verdonk [6] brought out the importance of rotational and translational entropies lost by small molecules on binding to proteins.

$$\Delta G^{o} = \Delta G^{o}_{tr} + \Delta G^{o}_{rot} - RT \ln [(Z^{int}_{P*L*.aq} Q_w V)/(Z^{int}_{P.aq} Z^{int}_{L.aq})] + P\Delta V^{o}_{aq} \quad (14)$$

Free energy contributions from internal motions that are coupled to solvent are best handled via molecular simulations. Separating the two will amount to an approximation.

$$Z^{int}_{P.aq} = Z_P^{vib.conf} . Z_P^{solvn.}$$
$$Z^{int}_{P.aq} = \int\ldots\int \exp \{(-E(\mathbf{X}^N_P, \mathbf{X}^M_W)/k_B T\}\, d\mathbf{X}^N_P\, d\mathbf{X}^M_W =$$
$$[\int..\int \exp \{(-E(\mathbf{X}^N_P)/k_B T\}\, d\mathbf{X}^N_P\,] \times [\int..\int \exp \{(-E(\mathbf{X}_P^{Nfixed}; \mathbf{X}^M_W)/k_B T\}\, d\mathbf{X}^M_W] \quad (15)$$

Equations similar to (15) can be written for L and P*L* and converted to excess free energies. Such a separation allows

$$\Delta G^{o} = \Delta G^{o}_{tr} + \Delta G^{o}_{rot} + \Delta G^{o}_{int} + \Delta G^{o}_{solvn.} \quad (16)$$

Eq. (16) forms the theoretical basis for the additivity assumed in free energy computations as employed in master equation methods [7,8]. The $P\Delta V^{o}_{aq}$ term in equation (9) is often neglected in liquid-state work. If eqs (15) and (16) are employed for each structure generated according to Boltzmann distribution either via molecular dynamics or Metropolis Monte Carlo and averages computed with a suitably calibrated model for solvation energy for each structure, the results are expected to correspond to eq. (9) which is exact.

Recent advances in free energy methodology offer two attractive methods viz. the MMPBSA [9-11] and the MMGBSA [12-14], which utilize the structural information emanating

from molecular dynamics simulations to develop estimates of binding free energies using equations (15) and (16) above, in a *post facto* analysis of the trajectories on each structure followed by energy component averaging. The essence is to generate structures with explicit solvent and transfer these to continuum solvent for energy evaluations thus rendering the free energy problem computationally tractable. A practical implementation of the above free energy methodology involves computation of average intramolecular energy (internal energy / enthalpy), corresponding entropies, solvation free energies of the solute along the MD trajectories of the free and bound protein and ligand.

$$\Delta G^{o}_{int} = \Delta H^{o}_{int} - T\Delta S^{o}_{int} \quad (17)$$

$$\Delta H^{o}_{int} = \Delta H^{o}_{intermolecular} + \Delta H^{o}_{intramolecular}$$

$$\Delta H^{o}_{intermolecular} = \Delta H^{o}_{el} + \Delta H^{o}_{vdW} =$$

$$< \Delta E^{o}_{intermolecular} > = < \Delta E^{o}_{el} > + < \Delta E^{o}_{vdW} > \quad (18)$$

$$\Delta H^{o}_{intramolecular} = < \Delta E^{o}_{intramolecular} > \quad (19)$$

where, ΔE^{o}_{el} , ΔE^{o}_{vdW} represent the electrostatic and van der Waals components of the intermolecular interaction energy between the protein and the inhibitor and $\Delta E^{o}_{intramolecular}$ represents changes in the intramolecular energy which includes both bonded and non-bonded terms as described by a force field for the protein and the inhibitor upon binding. All these quantities can be computed from a force field either for a fixed structure (from minimization studies) or for an ensemble of structures from MD simulations.

$$\Delta S^{o}_{int} = \Delta S^{o}_{vib,config} \quad (20)$$

$\Delta S^{o}_{vib,\ config}$ can be calculated by normal mode analysis for energy minimized structures (ΔS^{o}_{vib}) or by quasi harmonic approximation introduced by Karplus and Kushick [15] and subsequently extended and adapted to MD simulations by Schlitter [16] and van Gunsteren [17].

To account for structural deformation upon binding, we include adaptation expense which accounts for changes in the intramolecular energetics explicitly in ΔG^{o}_{int}, and it is calculated as the difference in the free energies of the bound and unbound states of the protein and the inhibitor in the presence of the solvent.

In the MMGBSA or MMPBSA models, the solvation free energies are computed as

$$\Delta G^{o}_{solvn} = \Delta G^{o}_{GBSA} = \Delta G^{o}_{GB} + \Delta G^{o}_{SA} \quad (21)$$

where ΔG^{o}_{GB} refers to the electrostatic component of solvation while ΔG^{o}_{SA} is the non-electrostatic contribution, called cavitation energy in literature [18]. The defining equation

employed for evaluating the electrostatic contribution to the solvation free energy [12] with the MMGBSA model is

$$G^{0}_{el.solvn} = -166\,(1 - 1/\in) \sum_{i=1}^{n} \sum_{j=1}^{n} q_i q_j / f_{m2GB} \qquad (22)$$

$$\Delta G^{o}_{GB} = G^{0}_{el.solvn}\,(\text{final state}) - G^{0}_{el.solvn}\,(\text{initial state}) \qquad (23)$$

Similar equations are formulated to deal with added salt effects at the Debye-Huckel level [19]. Small ions associated with the biomolecular target and the ligand to maintain electroneutrality are dealt with explicitly in simulations and processed as part of the solute. The non-electrostatic (nel) contributions to the solvation free energy [20] are computed as a function of the solvent accessible (SA) surface area [21]

$$G^{0}_{nel.solvn} = \gamma_{nel}\,\Delta A \qquad (24)$$

$$\Delta G^{o}_{SA} = G^{0}_{nel.solvn}\,(\text{final state}) - G^{0}_{nel.solvn}\,(\text{initial state}) \qquad (25)$$

The quantity γ_{nel} has been assigned a value of 7.2 cal/mol/Å^2 [22]. This may be considered [19] as a resultant of +47cal/mol/Å^2 from the cavity term [23] and –39.8 cal/mol/Å^2 from van der Waals interactions of the solute and the solvent [24]. This separation is only for the purpose of interpretation and does not alter the free energy estimates. Thus, a combination of equations (9), (10), (14) and (17) yields the absolute binding free energies. The governing equation (16) for estimation of free energy change upon binding is

$$\Delta G^{o} = \Delta G^{o}_{tr} + \Delta G^{o}_{rot} + \Delta G^{o}_{int.} + \Delta G^{o}_{solvn.}$$

The internal and solvation energy components in eq. 16 can be described as,

$$\Delta G^{o}_{int} + \Delta G^{o}_{solvn.} =$$

$$\Delta G^{o}_{adapt} + \Delta G^{o}_{vc} + \Delta G^{o}_{el} + \Delta G^{o}_{vdW} + \Delta G^{o}_{solvn,el} + \Delta G^{o}_{solvn,nel} \qquad (26)$$

with ΔG^{o}_{adapt} and vibrational configurational entropy changes identified as $\Delta G^{o}_{intramolecular}$ and $\Delta G^{o}_{solvn,nel}$ with $\Delta G^{o}_{cav.}$ and $\Delta G^{o}_{net\ el}$ as a sum of ΔG^{o}_{el} and $\Delta G^{o}_{solvn,el}$. These components can be computed via the MMGBSA methodology. The thermodynamic cycle employed to construct the standard free energies of protein-inhibitor binding in solution is illustrated in Figure 5. Building on eq. (16) and (26), the net binding process is decomposed into six steps and the corresponding binding free energy is calculated as a sum of five components:

$$\Delta G^{o}_{net} = \Delta G_{vdw} + \Delta G_{netel} + \Delta G_{cav} + \Delta G_{adpt} + \Delta G_{trvc} \qquad (27)$$

In a phenomenological view, equation (16) may be rearranged (eq. 27) and the net binding free energy may be considered to be a sum of the free energy changes due to the following terms: (i) van der Waals interactions between the protein and the inhibitor indicating the

influence of shape complementarities and packing effects; (ii) net electrostatics which includes interactions between partial or full charges, hydrogen bonds and electrostatics of desolvation upon binding and added salt effects, (iii) cavitation effects, which account for change in size and shape of solvent cavity on complexation giving rise to water reorganization, a component of which, originating from nonpolar sources, is the hydrophobic effect. Here the nonelectrostatics of desolvation of both polar and nonpolar atoms is accounted for in the cavitation term; (iv) the deformation expense (i.e. the intramolecular contributions due to structural variations upon complexation), (v) translational, rotational and vibrational, configurational entropy losses.

References:

1. K.A. Dill, J. Biol. Chem. **1997,** *272,* 701.
2. Jayaram, B.; Liu, Y.; Beveridge, D.L. *J. Chem. Phys*. **1998**, *109*, 1465.
3. McQuarrie, D.A. *Statistical Mechanics*, Harper and Row Publishers, New York. **1976**.
4. Atkins, P.W.; Paula, J. *Physical Chemistry* Oxford University Press, New York .**2002**.
5. Gilson, M.K.; Given, V.; Bush, B.L.; McCammon, J.A. *Biophys. J*. **1997**, *72*, 1047.
6. Murray, C. W.; Verdonk, M. L.*J. Comput. Aided. Mol. Des.* **2002***, 16,* 741.
7. Ajay; Murcko, M.A. *J. Med. Chem*. **1995**, *38*, 4953.
8. Aqvist, J. *J. Comp. Chem*. **1996**, *17*, 1587.
9. Honig, B.; Nicholls, A. *Science*. **1995**, *268*, 1144.
10. Swanson, J.M.J.; Henchman, R.H.; McCammon, J.A. *Biophy. J.* **2004**, *86*, 67.
11. Srinivasan, J.; Cheatham III, T.E.; Cieplak, P.; Kollman, P.A.; Case, D.A. *J. Am. Chem. Soc.* **1998**, *120*, 9401.
12. Jayaram, B.; Sprous, D.; Beveridge, D.L. *J. Phys. Chem. B*. **1998**, *102*, 9571.
13. Bashford, D; Case, D.A. *Annu. Rev. Phys. Chem*. **2000**, *51*, 129.
14. Zou, X.; Sun, Y.; Kuntz, I.D. *J. Am. Chem. Soc.* **1999**, *121*, 8033.
15. Karplus, M.; Kushick, J.N. *Macromolecules*. **1981**, *14,* 325.
16. Schlitter, J. *Chem . Phy. Lett*. **1993**, *215*, 617.
17. Scahfer, H.; Mark, A.E., vanGunsteren. *J. Chem. Phys*. **2000**, *113,* 7809.
18. Still, W.C.; Tempczyk, A.; Hawley, R.C.; Hendrickson, T.J. *J. Am. Chem. Soc*. **1990**, *112*, 6127.
19. Jayaram, B.; McConnell, K.J.; Dixit, S.B.; Beveridge, D.L. *J. Comput. Phy.* **1999**, *151,* 333.
20. Jayaram, B.; Liu, Y.; Beveridge, D.L. *J. Chem. Phys*. **1998**, *109*, 1465.
21. Lee, K.; Richards, F.M. *J. Mol. Biol*. **1971**, *55,* 379.
22. McCammon, J.A.; Harvey, S.C. *Dynamics of proteins and nucleic acids.* Cambridge University Press, Cambridge. **1987**.
23. Sharp, K.A.; Nicholls, A.; Fine, R.F.; Honig, B. *Science*. **1991**, *252*, 106.
24. Ohtaki, H.; Fukushima, N. *J. Sol. Chem*. **1992**, *21*, 23.