ADSBET2: Automated Determination of Salt-Bridge Energy-Terms version 2

Arnab Nayek#, Parth Sarthi Sen Gupta#, Shyamashree Banerjee#, Vishma Pratap Sur, Pratyay Seth, Sunit Das, Rifat Nawaz Ul Islam, & Amal Kumar Bandyopadhyay*

[1]Department of Biotechnology, The University of Burdwan, Golapbag, Burdwan, 713104, West Bengal, India; Amal Kumar Bandyopadhyay – Email: akbanerjee@biotech.buruniv.ac.in; Phone - +91-342-2657231(O), 9474723882(M), Fax - +91-3422657231; *Corresponding author
#Authors equally contributed

Abstract:
Component (bridge: $\Delta\Delta G_{brd}$, background: $\Delta\Delta G_{prot}$, desolvation: $\Delta\Delta G_{dsolv}$) and net ($\Delta\Delta G_{net}$) energy-terms of salt-bridge-structure (*SBS*) are auto-generated by the program ADSBET that makes use of general purpose Adaptive Poison Boltzmann Solver (*APBS*) method. While the procedure reports gross energy terms (*Kcal Mol^{-1}*), report on bond-multiplicity corrected normalized energy-terms (*Kcal Mol^{-1} Bond^{-1}*) along with their accessibility (*ASA*) in monomer, isolated-*SBS* (*ISBS*) and networked-*SBS* (*NSBS*) format would be very useful for statistical comparison among *SBS*s and understanding their location in protein structure. In this end, ADSBET2 potentially incorporates these features along with additional model for side-chain. Gross and normalized energy-terms are redirected in monomer, *ISBS* and *NSBS* format along with their *ASA* informations. It works on any number of *SBS*s for any number of structure files present in a database. Taken together, ADSBET2 has been suitable for statistical analyses of *SBSs* energetics and finds applications in protein engineering and structural bioinformatics.

Availability: ADSBET2 is freely available at http://sourceforge.net/projects/ADSBET2/ for all users.

Keywords: Energy terms, Networked, Isolated, Salt Bridge, ASA, Kcal Mol^{-1}Bond^{-1}

Background:

Salt-bridge, electrostatic interactions between partial charges of side-chain of acidic and basic residues, plays crucial role in stability and packing of protein. It is more so for proteins adapted in extreme environments [1]. Use of general purpose integrators of Poison Boltzmann Equation such as open-source Adaptive Poison Boltzmann Solver (*APBS*) [2] is the only means over experimental pKa or double-mutational cycle methods [3] to compute overall components (i.e. bridge-energy: $\Delta\Delta G_{brd}$, background-energy: $\Delta\Delta G_{prot}$, desolvation-energy: $\Delta\Delta G_{dsolv}$) and net-energy ($\Delta\Delta G_{net}$) terms of *SBS* [4, 5]. However, relative to other applications of the solver, computation of energy-terms of *SBS* is highly labor intensive that includes determination of protein-specific a] and residue specific salt bridges or ion-pairs [6, 7], b] grid-points, c] grid center and d] hydrophobic-isosteres-mediated mutation-files (five per *SBS*) from original protein-charge-radius (*PQR*) file [8] prior computation. Further in post-computation scenario, great deals of manual involvement are also necessary to obtain i] energy-terms from reaction field energies and relevant choice of partial atomic charges [9] and ii] side-chains-specific accessibility of *SBS*. While ADSBET [10] performs all these pre and post-run computations, it has few short-comings. Firstly it does not report bond multiplicity corrected normalized energy-terms ($EnergyBond^{-1}SBS^{-1}$) in monomer, *ISBS* and *NSBS* format which would be necessary for comparison among *SBS*s

present in a database. While, for example, comparable net energy-terms for *ISBS* (in *S1* and *S2* of **Figure 1**: FII) are *E1* and *E2/2* and that for *NSBS* (in *N1* and *N2* of **Figure 1**: FII) are (*E3+E4*) and (*E5/2+E6/2*) respectively, ADSBET only reports *E1*, *E2*, *E3*, *E4*, *E5* and *E6* in monomer format. Secondly it does not inform side-chain model-specific accessibility **[10]** which would be crucial for understanding surface and core location of *SBS* **[5, 6]** (**Figure 1**: *p*). Finally as far as models for side-chains of

interacting partner of *SBS* are concerned, only a less popular model (i.e. *MDL-1*) is available in ADSBET **[10]** (**Figure 1**). In this context, ADSBET2 potentially includes all these features along with popular additional model of side chain (i.e. *MDL-2*). When opted, it reports side-chain *ASA* by the use of either of three alternative methods.
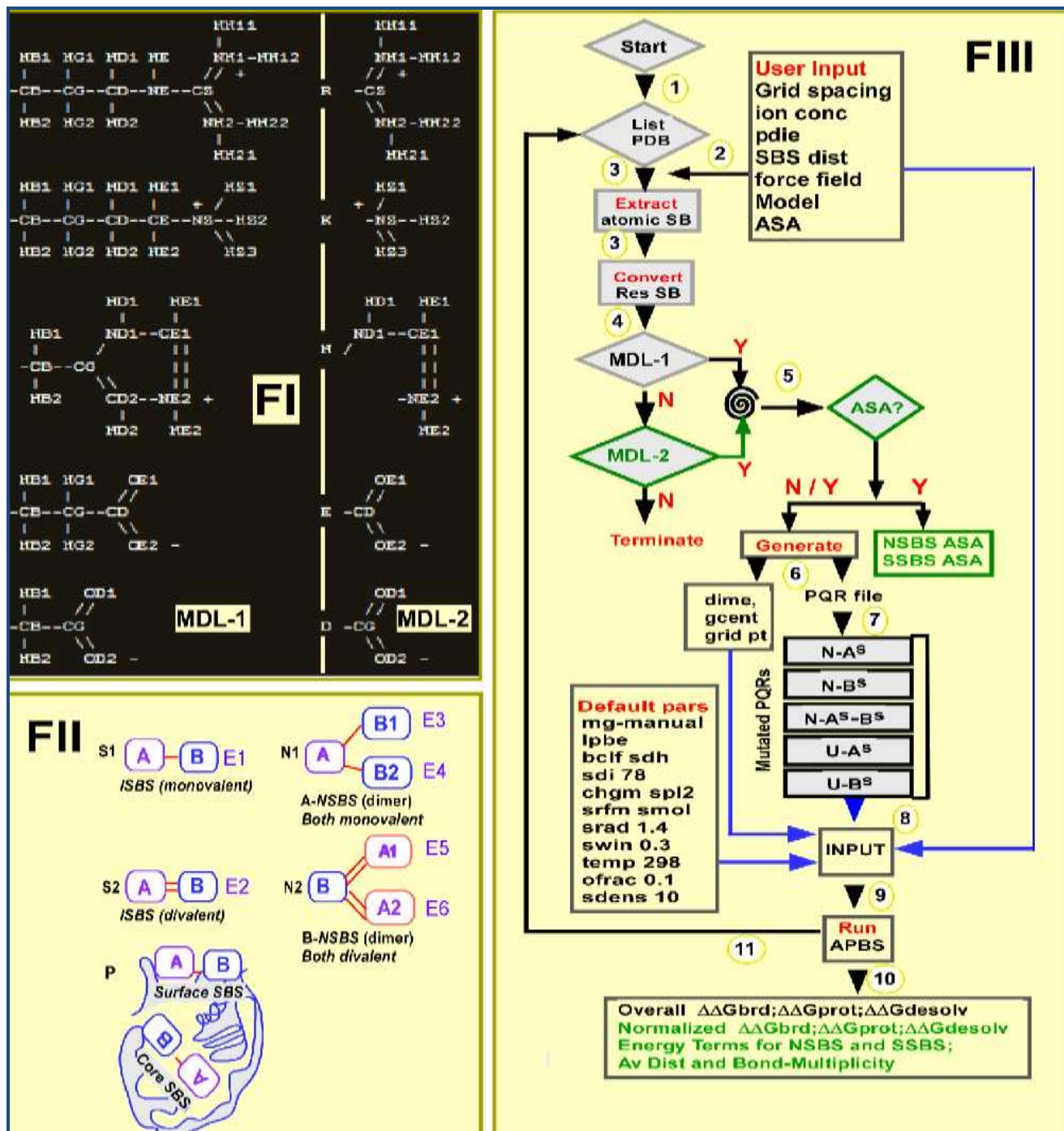

Figure 1: Model-based side-chains (FI), different SBS (FII) and flow-chart (FIII) for ADSBET2. **MDL-1** and **MDL-2** are used in model-1 and model-2 respectively. *S1* (energy *E1*) and *N1* (energy *E3* and *E4*) are single-bonded isolated and networked *SBS*. *S2* (energy *E2*) and *N2* (energy *E5* and *E6*) are multiple-bonded isolated and networked *SBS*. **P** shows core and surface location of *SBS*. In FIII, *N-A^s* and *U-A^s* are acidic side-chains in folded and unfolded state of protein respectively. Additional features in ADSBET2 over its earlier version are shown in green color.

BIOINFORMATION

Method of computation of $\Delta\Delta G_{brd}$, $\Delta\Delta G_{prot}$, $\Delta\Delta G_{dsolv}$ and $\Delta\Delta G_{net}$ remains similar as earlier **[10]**. Detailed methods also precede analytical result of each item of outputs. Normalized energy-terms of each monomeric *SBS* are obtained by dividing these terms by their bond-multiplicity (**Figure 1**: FII, *S2*, and *N2*). When opted, accessibility (hence surface and core location) of *SBS* is computed. Overall and normalized energy-terms for *NSBS* are obtained by summing energy-terms of all monomeric cases having either a common acidic (*A-NSBS*: FII *N1*) or basic (*B-NSBS*: FII *N2*) partner in a *SBS* (thus forming dimer, trimer etc). Apart from *MDL-1* of ADSBET **[10]**, ADSBET2 utilizes additional most-popular model for side chains i.e. *MDL-2* (**Figure 1**: FI). Operational details of the program includes following sequence of events: a] make list of x-ray files in PDB format (**Figure 1**: FIII), b] uptake user's parameters from command line, c] extract atomic *SBS* and then convert them into residue-specific *SBS*, d] select model of side-chain, e] select accessibility mode and method, f] Generate force-field dependent *PQR*-file and protein-specific *APBS*-parameters (such as dime, gcent, grid-points, ionic-strength and pH), g] generate 5 relevant mutated-*POR*-files per *SBS*, h] prepare input-file for *APBS* using user, default, generated parameters and mutated *PQR*s, i] run *APBS*, j] redirect output and k] repeat the process for other *SBSs* and *PDB*s.

Figure 2: Total and normalized energy terms along with named outputs. Bond-multiplicity, Av DIST and ASA are also reported. ddGprot-I **[4]** and ddGprot-II **[5]** (hence ddGnet-I and ddGnet-II respectively) are background energies of model based side chains of non-*SBS* residues.

Program input

Poison Boltzmann Equation solvers (such as *APBS*) require multi-parameters input for generation of reaction field energy **[2]**. Thus, apart from *PDB*-files in the working directory users need to input parameters (**Figure 1**: FIII*: User Input) such as model for side-chain (F1), *ASA* method, grid-spacing, pH, mobile ion concentration (in Molar), dielectric constant of protein, salt-bridge or ion-pair distance and force-field. Protein specific parameters (such as dime, gcent and grid-points) are auto-generated by the program and other parameters are used as default (**Figure 1**: FIII: Default pars) if not mentioned otherwise.

Program output

Details of *SBS*-energetics and accessibility for any number of *SBS* in any number of *PDB*s are redirected into two different named-outputs. Unlike ADSBET **[10]**, ADSBET2 redirects most relevant model-based additional output (**Figure 1**: FIII: Green color parts) on overall and normalized energetics in monomer, *ISBS* and *NSBS* format along with their core and surface location informations (**Figure 2**).

Caveats and future development:

The program which is interpreted by AWK programming language is tested to run in the OS: CYGWIN (32 bit). We further are working to develop GUI-based version of the program.

Conclusion:

ADSBET2, interpreted by AWK programming language, produces model-based overall and normalized *SBS*-energetics in monomer, ISBS and NSBS formats along with their accessibility information for any member of X-ray structure files with any number of *SBS*s in them. These results are useful for statistical comparison of SBS-energetics of candidate salt-bridges or ion-pairs present in a database.

References:

[1] Bandyopadhyay AK *et al. Extremophiles* 2007 **11**: 615 [PMID: 17406782]

[2] Baker NA *et al. Proc Natl Acad Sci USA.* 2001 **98**: 10037 [PMID: 11517324]

[3] Bosshard HR *et al. J Mol Recog.* 2004 **17**: 1 [PMID: 14872533]

[4] Nayek A *et al. Plos one* 2014 **9**: e93862 [PMID: 24743799]

[5] Kumar S & Nussinov R, *J Mol Biol.* 1999 **293**: 1241 [PMID: 10547298]

[6] Gupta PS *et al. Bioinformation* 2014 **10**: 105 [PMID: 24616564]

[7] Gupta PS *et al. Bioinformation* 2015 **11**: 39 [PMID: 25780279]

[8] Dolinsky TJ *et al. Nucleic Acids Res.* 2004 **32**: W665 [PMID: 15215472]

[9] Hendsch ZS & Tidor B, *Protein Sci.* 1994 **3**: 211 [PMID: 8003958]

[10] Nayek A *et al. IJIPLS* 2015 **5**: 28