

June 6, 2016

Implicit solvent models in X-PLOR

Thomas Simonson, Savvas Polydorides, Francesco Villa, David Mignon

Laboratoire de Biochimie, Ecole Polytechnique
thomas.simonson@polytechnique.fr

Chapter 1

Generalized Born models

1.1 Introduction

The Generalized Born (GB) model [? ? ? ?] is an efficient and accurate implicit solvent model for biomolecular simulations and protein design. It describes the solvent around the biomolecule as a dielectric continuum. But the numerical complexities of an inhomogeneous solute/solvent dielectric system are swept away and replaced by approximate, efficient, analytical formulas. The model can be used either to determine the energy of a single structure or to generate multiple structures by molecular dynamics or simulated annealing. Several review articles describe the theoretical background, the performance, and the ongoing progress of the GB model; see eg [? ? ?]. Two GB variants have been implemented in X-PLOR [?] and CNS [?]. The first is termed GB/ACE (Schaefer & Karplus, *J. Phys. Chem.*, 1996, 100:1578), for “Analytical Continuum Electrostatics”; the second is termed GB/HCT, for “Hawkins, Cramer & Truhlar” (HCT, *Chem. Phys. Lett.*, 1995, 246:122). We emphasize that the GB solvation model describes the solvent response to the charges and Coulomb potential of the solute. Therefore, it is meaningless to use GB in a simulation or structure refinement where the ordinary electrostatics energy term is turned off.

The Theory section below reviews the GB/ACE and GB/HCT models. Expressions of the solvation energies and forces are given. This section can be skipped by those already familiar with the model. The following section, Syntax, gives the necessary syntax and the default options for using GB in X-PLOR.

1.2 Theory

1.2.1 GB energy

In the world of continuum electrostatics, a biomolecular solute is viewed as a set of (fractional) atomic charges in a cavity delimited by the solute surface, embedded in a high dielectric solvent medium [?]. The electrostatic energy E^{elec} is the sum of the Coulomb interaction energies between all solute charges and a solvation term ΔE^{solv} ; the latter includes the interaction energies of each solute charge with solvent (its “self-energy”), and a solvent-screening contribution to the interaction energies between solute charges:

$$E^{\text{elec}} = \sum_{i < j} \frac{q_i q_j}{r_{ij}} + \Delta E^{\text{solv}} \quad (1.1)$$

$$\Delta E^{\text{solv}} = \sum_i \Delta E_i^{\text{self}} + \sum_{i < j} \Delta E_{ij}^{\text{int}} \quad (1.2)$$

In the GB model, the solvent contribution $\Delta E_{ij}^{\text{int}}$ to the interaction energy between the charges q_i and q_j is approximated by [?]:

$$\Delta E_{ij}^{\text{int}} = -\frac{\tau q_i q_j}{(r_{ij}^2 + b_i b_j \exp[-r_{ij}^2/4b_i b_j])^{1/2}} \quad (1.3)$$

where r_{ij} is the distance between the charges, τ is given by

$$\tau = 1/\epsilon_p - 1/\epsilon_w \quad (1.4)$$

ϵ_p , ϵ_w are the protein and solvent dielectric constants, and b_i is the ‘solvation radius’ of charge i . By analogy to the case of a single charge in a spherical cavity, b_i is defined by

$$\Delta E_i^{\text{self}} = -\frac{\tau q_i^2}{2b_i} \quad (1.5)$$

where ΔE_i^{self} is the self-energy of charge i . By partitioning the solute into atomic volumes (following Lee & Richards, for example [?]), one can express the self-energy ΔE_i^{self} as a sum over all the solute atoms [? ?]:

$$\Delta E_i^{\text{self}} = -\frac{\tau q_i^2}{2R_i} + \tau q_i^2 \sum_{k \neq i} E_{ik}^{\text{self}} \quad (1.6)$$

where R_i is a constant atomic radius to be determined (close to the van der Waals radius) and E_{ik}^{self} is related to the integral of the electrostatic energy over the volume of atom k . Notice that the charges of the other atoms, q_k , do not appear here. The effect of these atoms is merely to exclude solvent from the vicinity of atom i [?].

The volume integral E_{ik} is approximated in two steps. The first step is to approximate the electric field by the ‘Coulombic field’ of charge i [?]. This is simply the unscreened field that would exist if q_i were in a vacuum; it radiates uniformly in all directions and falls off as $1/r^2$ with distance; the corresponding energy density is $1/r^4$. The next step is to calculate the integral of $1/r^4$ over the volume of atom k . The different GB variants do this in different ways. In GB/ACE, for example, Schaefer & Karplus assume the density of each solute atom is a gaussian centered at the atom’s position. The integral E_{ik} then has a tractable form, which can be approximated by interpolating between a Gaussian form at short ranges and a $1/r^4$ form at long range, leading to the Ansatz [?]:

$$E_{ik}^{\text{self}} = \frac{1}{\omega_{ik}} \exp(-r_{ik}^2/\sigma_{ik}^2) + \frac{V_k}{8\pi} \left(\frac{r_{ik}^3}{r_{ik}^4 + \mu_{ik}^4} \right)^4 \quad (1.7)$$

Here, ω_{ik} and μ_{ik} are simple functions of the atomic volume V_k , the atomic radii R_i , R_k ($= [3V_k/4\pi]^{1/3}$), and an adjustable “smoothing” parameter α that determines the width of the atomic gaussian distributions (see below). The atomic charges are taken directly from the existing force field. The adjustable parameters of the model are then the volumes V_k and the smoothing parameter α . Ionic strength is not included, although methods to do so have been proposed [? ?]. Volumes V_k can be either calculated using Voronoi polyhedra (using an external program [?] and reading them into X-PLOR), or assigned values from existing libraries [? ? ?]. Note that the V_k are considered to be constants, independent of the solute conformation. This is important to obtain tractable expressions for the GB forces (see below).

With the above self-energy approximations, ΔE_i^{self} can sometimes become positive, so that the (necessarily positive) solvation radius can no longer be defined by Eq. (1.5). Therefore, we use a definition proposed by Schaefer et al. [?]:

$$\begin{aligned} b_i &= -\frac{\tau q_i^2}{2\Delta E_i^{\text{self}}} & \text{if } \Delta E_i^{\text{self}} \leq E_{\min} = -\frac{\tau q_i^2}{2b_{\max}} \\ &= b_{\max} \left(2 - \frac{\Delta E_i^{\text{self}}}{E_{\min}} \right) & \text{if } \Delta E_i^{\text{self}} \geq E_{\min} \end{aligned} \quad (1.8)$$

Here, b_{max} is an upper limit for the solvation radius, which can be set to the largest linear dimension of the solute, for example. This definition leads to continuous energies and forces.

1.2.2 Calculation of forces

Interaction energy term

We first consider the GB ‘interaction’ term, on the far right of Eq. (1.2), and its gradient ∇_n with respect to the position of solute particle n . Noting that the solvation radii b_i, b_j depend on all the atomic positions and using the chain rule for differentiation, we have:

$$\nabla_n \sum_{i<j} \Delta E_{ij}^{\text{int}} = \sum_{i<j} \frac{\partial \Delta E_{ij}^{\text{int}}}{\partial r_{ij}} \nabla_n r_{ij} + \sum_{i<j} \frac{\partial \Delta E_{ij}^{\text{int}}}{\partial b_i} \nabla_n b_i + \sum_{i<j} \frac{\partial \Delta E_{ij}^{\text{int}}}{\partial b_j} \nabla_n b_j \quad (1.9)$$

Only terms with $i = n$ or $j = n$ contribute to the first sum on the right. The second sum can be written

$$\sum_{i<j} \frac{\partial \Delta E_{ij}^{\text{int}}}{\partial b_i} \nabla_n b_i = \frac{1}{2} \sum_i \left(\sum_{j \neq i} \frac{\partial \Delta E_{ij}^{\text{int}}}{\partial b_i} \right) \frac{\partial b_i}{\partial \Delta E_i^{\text{self}}} \nabla_n \Delta E_i^{\text{self}} \quad (1.10)$$

The quantity in parentheses will be denoted $dE_i^{\text{int,b}}$, since, for a given conformation, it depends only on i . The last quantity on the right can be written:

$$\begin{aligned} \nabla_n \Delta E_i^{\text{self}} &= \sum_{k \neq i} \nabla_n E_{ik}^{\text{self}} = \nabla_n E_{in}^{\text{self}} \quad \text{if } i \neq n \\ &= \sum_{k \neq n} \nabla_n E_{nk}^{\text{self}} \quad \text{if } i = n. \end{aligned} \quad (1.11)$$

Grouping the second and third terms on the right of (1.9) and rearranging the first, we obtain:

$$\nabla_n \sum_{i<j} \Delta E_{ij}^{\text{int}} = \sum_{i \neq n} \left(\frac{\partial \Delta E_{in}^{\text{int}}}{\partial r_{in}} + dE_n^{\text{int,b}} \frac{\partial b_n}{\partial \Delta E_n^{\text{self}}} \frac{\partial E_{ni}^{\text{self}}}{\partial r_{in}} + dE_i^{\text{int,b}} \frac{\partial b_i}{\partial \Delta E_i^{\text{self}}} \frac{\partial E_{in}^{\text{self}}}{\partial r_{in}} \right) \frac{r_n - r_i}{r_{in}} \quad (1.12)$$

with

$$dE_i^{\text{int,b}} = \sum_{j \neq i} \frac{\partial \Delta E_{ij}^{\text{int}}}{\partial b_i} \quad (1.13)$$

$$\begin{aligned} \frac{\partial b_n}{\partial \Delta E_n^{\text{self}}} &= -\frac{b_n}{\Delta E_n^{\text{self}}} && \text{if } \Delta E_n^{\text{self}} \leq E_{\min} = -\frac{\tau q_n^2}{2b_{\max}} \\ &= -\frac{b_{\max}}{E_{\min}} && \text{if } \Delta E_n^{\text{self}} \geq E_{\min} \end{aligned} \quad (1.14)$$

The quantities b_i and $dE_i^{\text{int,b}}$ can be ‘precalculated’, so that obtaining the force on atom n requires only a loop over all solute atoms. In (1.12), the derivatives of $\Delta E_{in}^{\text{int}}$ are the same for GB/ACE and GB/HCT:

$$\frac{1}{r_{in}} \frac{\partial \Delta E_{in}^{\text{int}}}{\partial r_{in}} = \frac{\tau q_i q_j}{\left[r_{ij}^2 + b_i b_j \exp\left(-\frac{r_{ij}^2}{4b_i b_j}\right) \right]^{3/2}} \left(1 - \frac{1}{4} \exp\left(-\frac{r_{ij}^2}{4b_i b_j}\right) \right) \quad (1.15)$$

$$dE_i^{\text{int,b}} = \sum_{j \neq i} \frac{\frac{1}{2} \tau q_i q_j b_j \exp\left(-\frac{r_{ij}^2}{4b_i b_j}\right)}{\left[r_{ij}^2 + b_i b_j \exp\left(-\frac{r_{ij}^2}{4b_i b_j}\right) \right]^{3/2}} \left(1 + \frac{r_{ij}^2}{4b_i b_j} \right) \quad (1.16)$$

GB/ACE self-energy term

The self-energy and the associated forces depend on the GB variant. With GB/ACE,

$$\frac{1}{r_{ij}} \frac{\partial E_{ij}^{\text{self}}}{\partial r_{ij}} = -\frac{2}{\omega_{ij} \sigma_{ij}^2} \exp\left(-\frac{r_{ij}^2}{\sigma_{ik}^2}\right) + \frac{V_j}{2\pi} \left(\frac{r_{ij}^{10}}{r_{ij}^4 + \mu_{ij}^4} \right)^5 (3(r_{ij}^4 + \mu_{ij}^4) - 4r_{ij}^4). \quad (1.17)$$

The parameters ω_{ij} , σ_{ij} , μ_{ij} are defined by:

$$\frac{1}{\omega_{ik}} = \frac{4}{3\pi\alpha_{ik}^3}(Q_{ik} - \arctan Q_{ik})\frac{1}{\alpha_{ik}R_k} \quad (1.18)$$

$$\sigma_{ik}^2 = \frac{3(Q_{ik} - \arctan Q_{ik})}{(3 + f_{ik})Q_{ik} - 4\arctan Q_{ik}}\alpha_{ik}^2 R_{ik}^2 \quad (1.19)$$

$$Q_{ik} = \frac{q_{ik}^2}{(2q_{ik}^2 + 1)^{1/2}} \quad (1.20)$$

$$f_{ik} = \frac{2}{q_{ik}^2 + 1} - \frac{1}{2q_{ik}^2 + 1} \quad (1.21)$$

$$q_{ik}^2 = \frac{\pi}{2} \left(\frac{\alpha_{ik} R_k}{R_i} \right)^2 \quad (1.22)$$

$$\alpha_{ik} = \text{Max}(\alpha, R_i/R_k) \quad (1.23)$$

$$\mu_{ik} = \frac{77\pi\sqrt{2}R_i}{512(1 - 2\pi^{3/2}\sigma_{ik}^3)\frac{R_i}{\omega_{ik}V_k}} \quad (1.24)$$

$$V_k = \frac{4}{3}\pi R_k^3 \quad (1.25)$$

GB/HCT self-energy

With GB/HCT, the self-energy contribution E_{ik}^{self} is given by [?]]

$$4E_{ik}^{\text{self}} = \frac{1}{L_{ik}} - \frac{1}{U_{ik}} + \frac{r_{ik}}{4} \left(\frac{1}{U_{ik}^2} - \frac{1}{L_{ik}^2} \right) + \frac{1}{2r_{ik}} \ln \frac{L_{ik}}{U_{ik}} + \frac{R_k^2}{4r_{ik}} \left(\frac{1}{L_{ik}^2} - \frac{1}{U_{ik}^2} \right), \quad (1.26)$$

where

$$\begin{aligned} L_{ik} &= 1 & \text{if } r_{ik} + R_k &\leq R_i, \\ L_{ik} &= R_i & \text{if } r_{ik} - R_k &\leq R_k < r_{ik} + R_k, \\ L_{ik} &= r_{ik} - R_k & \text{if } R_i &\leq R_k < r_{ik} - R_k, \end{aligned} \quad (1.27)$$

$$\begin{aligned} U_{ik} &= 1 & \text{if } r_{ik} + R_k &\leq R_i, \\ U_{ik} &= r_{ik} - R_k & \text{if } R_i &< r_{ik} + R_k. \end{aligned} \quad (1.28)$$

The corresponding gradient is given by:

$$\begin{aligned} \frac{4}{r_{ik}} \frac{\partial E_{ik}^{\text{self}}}{\partial r_{ik}} &= -\frac{1}{r_{ik}} \left(\frac{L'_{ik}}{L_{ik}^2} - \frac{U'_{ik}}{U_{ik}^2} \right) + \frac{1}{4r_{ik}} \left(\frac{1}{U_{ik}^2} - \frac{1}{L_{ik}^2} \right) - \frac{1}{2} \left(\frac{U'_{ik}}{U_{ik}^3} - \frac{L'_{ik}}{L_{ik}^3} \right) \\ &\quad - \frac{1}{2r_{ik}^3} \ln \frac{L_{ik}}{U_{ik}} + \frac{1}{2r_{ik}^2} \left(\frac{L'_{ik}}{L_{ik}} - \frac{U'_{ik}}{U_{ik}} \right) - \frac{R_k^2}{4r_{ik}^3} \left(\frac{1}{L_{ik}^2} - \frac{1}{U_{ik}^2} \right) - \frac{R_k^2}{2r_{ik}^2} \left(\frac{L'_{ik}}{L_{ik}^3} - \frac{U'_{ik}}{U_{ik}^3} \right) \end{aligned} \quad (1.29)$$

with $L'_{ik} = \partial L_{ik} / \partial r_{ik}$, $U'_{ik} = \partial U_{ik} / \partial r_{ik}$. The radii R_k are calculated from the atomic volumes as in Eq. (1.25), then reduced by a scaling factor $S_k \leq 1$ which depends only on the chemical type of atom k . Reasonable values are given in Table 1 of [?].

This basic model was modified by Onufriev et al [?] to improve performance for proteins. The self-energy in Eq. (1.6) is replaced by:

$$\Delta E_i^{\text{self}} = -\frac{\tau q_i^2}{2b_i} \quad (1.30)$$

$$b_i = \left[(R_i - \rho_0)^{-1} - \lambda \sum_{k \neq i} E_{ik}^{\text{self}} \right]^{-1} - \delta \quad (1.31)$$

In other words, the atomic radius R_i is reduced by a constant offset ρ_0 , the self-energy contribution E_{ik}^{self} is scaled by a constant factor λ , and the solvation radius b_i is reduced by a constant offset δ . The values $\lambda = 1.4$, $\rho_0 = 0.09$ Å and $\delta = 0.15$ Å were used in [?].

1.2.3 Pairs of interacting groups

In structure refinement, it is often necessary to use a model in which different parts of the macromolecule are artificially duplicated, for example a protein side chain that is disordered and occupies multiple positions in a crystal structure. To allow for these situations, both X-PLOR [?] and CNS [?] view the system formally as a set of “pairs of interacting groups”. Usually, there is only one such pair: the macromolecule interacting with itself:

$$M \leftrightarrow M,$$

where M is the macromolecule and \leftrightarrow indicates an interaction. In the case of a single disordered protein side chain thought to have two main conformations, one would normally consider a protein P with two copies of the side chain: S_1 and S_2 , leading to the following pairs of interacting groups:

$$\begin{aligned} P \setminus \{S_1, S_2\} &\leftrightarrow P \setminus \{S_1, S_2\} \\ P \setminus \{S_1, S_2\} &\leftrightarrow S_1; \text{ weight of } 1/2 \\ P \setminus \{S_1, S_2\} &\leftrightarrow S_2; \text{ weight of } 1/2, \end{aligned}$$

where $P \setminus \{S_1, S_2\}$ represents the protein without the disordered side chain and the protein- S interactions are weighted by $1/2$ because there are two copies of S . The two copies of S do not interact with each other. This formalism is implemented in X-PLOR through the **constraints interaction** statement (for an example, see the gbttests/testfirst.inp test case).

The same formalism applies to the GB energy terms. If the interacting groups are denoted A_p, B_p with $p = 1, N$, their pairs take the form $P_p = A_p \times B_p = \{(i, j); i \in A_p, j \in B_p\}$. There are N pairs of groups P_p and each has a weight w_p . The GB interaction and self energies take the form:

$$\Delta E^{\text{int}} = \frac{1}{2} \sum_{p=1}^N w_p \left(\sum_{i \in A_p, j \in B_p} \Delta E_{ij}^{\text{int}} \right) \quad (1.32)$$

$$\Delta E^{\text{self}} = \sum_{p=1}^N w_p \sum_{i \in A_p, j \in B_p} \left(-\frac{\tau q_i^2}{2R_i} \delta_{ij} + \tau q_i^2 E_{ij}^{\text{self}} \right) \quad (1.33)$$

These equations generalize Eqs. (1.3), (1.6), which correspond to a single pair $P_1 = M \times M$, M being the whole macromolecule. The factor $\frac{1}{2}$ in Eq. (1.32) corrects for double counting of i, j and j, i terms; δ_{ij} is the Kronecker symbol.

1.2.4 Crystal symmetry

The implementation of crystal symmetry is described below and in a published article [?]. The system is now assumed to have n_G symmetry elements, which are isometries of the form

$$S : \underline{r} \rightarrow \underline{R}\underline{r} + \underline{\rho} \quad (1.34)$$

\underline{R} is a rotation or an inversion with respect to a plane or a point, and $\underline{\rho}$ is a translation vector. The total solvation energy now involves a sum over symmetry images; the solvation energy E per asymmetric unit is

$$E = \frac{1}{2n_G} \sum_{iS} \sum_{jS'} g(S\underline{r}_i, S'\underline{r}_j) = \frac{1}{2} \sum_{ijS} g(\underline{r}_i, S\underline{r}_j) \quad (1.35)$$

where n_G is the order of the symmetry group (which is infinite for an infinite crystal). In practice, the infinite summation over all crystal translations can be truncated with

a minimum image convention [?], since the total electrostatic interaction energy (Coulomb plus solvation) is rather short-ranged, in contrast to the Coulomb energy alone.

To obtain the solvation forces, we use the relations

$$\begin{aligned}\nabla_n g(\underline{r}_n, S\underline{r}_j) &= g'(\underline{r}_n, S\underline{r}_j) \frac{\underline{r}_n - S\underline{r}_j}{|\underline{r}_n - S\underline{r}_j|} \\ \nabla_n g(\underline{r}_i, S\underline{r}_n) &= R^{-1} g'(\underline{r}_i, S\underline{r}_n) \frac{\underline{r}_i - S\underline{r}_n}{|\underline{r}_i - S\underline{r}_n|} \\ \nabla_n g(\underline{r}_n, S\underline{r}_n) &= 2g'(\underline{r}_n, S\underline{r}_n) \frac{\underline{r}_n - S\underline{r}_n}{|\underline{r}_n - S\underline{r}_n|}\end{aligned}$$

Here, $g'(\underline{r}_i, \underline{r}_j)$ represents differentiation of $g_{ij} = g(\underline{r}_i, \underline{r}_j)$ considered as a function of the scalar variable $r_{ij} = |\underline{r}_i - \underline{r}_j|$. The gradient of the solvation energy takes the form

$$\begin{aligned}\nabla_i E &= \sum_{i \leq j, S} \lambda_{ij} g'(r_{ij}) \frac{\underline{r}_i - \underline{r}_j}{r_{ij}} \\ &+ \sum_{j \leq i, S} R^{-1} \lambda_{ij} g'(r_{Ij}) \frac{\underline{r}_I - \underline{r}_j}{r_{Ij}}\end{aligned}\tag{1.36}$$

The indices I, J correspond to the images of the particles i, j under S .

The energy and forces can be accumulated by summing over the interacting pairs (i, j) where $i \leq j$ [?]. While processing the (i, j) term, we do two things: 1) we accumulate the contribution of j to the force on i ('direct' contribution); 2) we calculate and set aside $G_{ij} = \lambda_{ij} R^{-1} g'(r_{ij}) \frac{\underline{r}_i - \underline{r}_I}{r_{ij}}$, which represents the contribution of i to the force on j ('scatter' contribution). In the vectorized code of CNS or X-PLOR, once the loop over all j is finished, the G_{ij} are 'scattered' [?], or added to the appropriate atomic forces, F_j .

1.3 Syntax

1.3.1 GB energy terms

The GB solvation energy is divided into four terms: direct self-energy and interaction energy terms, and direct and self-energy terms corresponding to interactions with

symmetry images:

$$E_{GBSOLV} = E_{GBSELF} + E_{GBINT} + E_{PGBS} + E_{PGBI}$$

They are available to the user through the variables \$GBSE, \$GBIN, \$PGBS, and \$PGBI. They are activated by the **flags** statement in the usual way:

```
flags include gbse gbin pgbs pgbi end
```

They are inactive by default.

1.3.2 Setting the GB options

All the parameters of the GB solvent model are under user control, with sensible defaults. The setup of the atomic volumes is described further on. The other GB parameters are set up with the **nbonds** subcommand:

NBONDS <nbonds-statement> | <gborn-nbonds-statement> **END**

applies to electrostatic, van der Waals, and GB energy terms.

<gborn-nbonds-statement> ::=

GBACE | **GBHCT** Exclusive flags activating the GB/ACE or the GB/HCT model. Default: inactive.

WEPS=<real> Solvent dielectric constant. Default: 1 if GB is inactive, 80 if GB is active.

SMOOTH=<real> Determines the atomic widths in GB/ACE; denoted α in Eq. (1.23). Default: 1.

LAMBDA=<real> Scaling factor for solvation radii in GB/HCT; denoted λ in Eq. (1.31). Default: 1.

OFFSET=<real> Offset for atomic radii in GB/HCT; denoted ρ_0 in Eq. (1.31). Default: 0.

1.3.3 Setting up atomic volumes for GB

Two approaches can be used:

Volume libraries

Two sets of ‘standard’ atomic volumes are available for proteins, in two force field parameter files: **param19.gb.pro** and **paramber.gb.inp**, located in \$GBXPLO/g-btoppar (see File Organization, below). These volumes are automatically read along with the other force field parameters. The first set was developed by Schaefer and coworkers [?] and modified and tested for protein simulations by Calimet et al [?], and is meant to be used with the Charmm19 topology (toph19.pro) and parameter set. The second was developed and tested by our group [?] and is meant to be used with the Amber all-atom force field [?]. Other volume libraries are available in the literature and can be formatted for X-PLOR, for example nucleic acid libraries [?].

The syntax of the **NONBonded** subcommand is modified accordingly:

NONB <type> <real> <real> <real> <real> [<real> <real>]

reads the Lennard-Jones parameters for a specified chemical type, as before; the first pair of reals is ϵ , σ ; the second pair is ϵ , σ for 1–4 non-bonded interactions. The last two reals are V , the atomic volume (Eqs. 1.6, 1.25), and S , the scaling parameter used for the HCT solvation radius (see text following Eq. 1.29). If the last two reals are omitted, V and S will both be set to 9999. Thus, for applications not using GB, there is backward compatibility with X-PLOR parameter files not set up for GB. But for applications using GB, V must be included in the parameter file for both GB/ACE and GB/HCT, and S must be included for GB/HCT.

Volumes calculated with an external program

In some cases, it may be desirable to calculate the atomic volumes corresponding to a particular family of conformations and/or proteins, instead of relying on ‘standard’ values [?]. The standard GB/ACE volumes were obtained from atomic Voronoi volumes calculated for a large set of protein structures, then averaged over each chemical type [?], then reduced by a factor of 0.9 to account for systematic errors in the GB/ACE self-energy approximation [?]. Several programs have the capability to calculate Voronoi volumes for each individual atom of a particular protein (eg the VORONOI package of Fred Richards). If these are then stored in a particular field of a PDB coordinate file (for example the field normally used for the temperature factors, WMAIN), this information can be read into X-PLOR using the **coordinate** statement, then made available to the GB routines internally. To

do this, the volumes must be copied into the RMSD array, then averaged over each chemical type using the **parameter reduce** statement:

```

coor @volumes.pdb                                {read coordinate file with      }
                                                {atomic volumes in wmain field  }

vector do (rmsd = wmain) (all)                   {copy into rmsd field          }

flags exclude * include gbse gbin end           {activate GB energy terms, so   }
                                                {GB parameters will be reduced }

parameter reduce selection=(all)                 {average volumes over          }
  overwrite=true mode=average end               {each chemical type            }
end
flags include bonds angl dihe impr vdw elec     {reactivate the other terms}

```

The atomic volumes, suitably averaged, are then available for GB calculations.

1.3.4 Examples

Minimization with GB/ACE

```

coordinates @protein.pdb
parameter
nbonds
tolerance=0.25 atom cdie trunc
nbxmod=5 vswitch e14fac=1. cutnb=15. ctonnb=13. ctofnb=14.
eps=1. weps=80. smooth=1.3 gbace {GB options}
end
end
flags include gbse gbin end
minimize powell nstep=50 end

```

Molecular dynamics with GB/HCT

```

remarks Asparagine MD with GB/HCT

```

```

remarks    this file: dyna.inp

topology
  @GBXPLO:gbtoppar/amber/topamber.inp    {Amber topology file      }
  @GBXPLO:gbtoppar/amber/patches.pro      {N- and C-terminal patches }
end                                           {for Amber force field    }
parameter
  @GBXPLO:gbtoppar/amber/paramber.gb.inp {Amber parameter file     }
end                                           {including GB parameters  }
segment
name="ASN1"
molecule name=ASN number=1 end
end
patch NASN refe=nil=(resid 1) end
patch CASN refe=nil=(resid 1) end
parameter
nbonds
  atom cdie trunc
  e14fac=0.8333333                      ! use this to reproduce amber elec
  cutnb 500. ctonnb 480. ctofnb 490.    ! essentially no cutoff
  tolerance=100.                        ! only build the nonbonded list once
  nbxmod 5 vswitch
  wmin=1.0
end
end
parameters nbonds
  EPS=1. WEPS=80. GBHCT                  ! GB parameters
  offset=0.09 lambda=1.33                ! GB parameters
end end
coord @volumes.pdb                      ! PDB with volumes in wmain
vector do (RMSD = wmain) (all)           ! copy into rmsd
vector do (rmsd = rmsd * 0.9) (all)       ! reduce volumes by 10%
flags include gbse gbin end
parameter reduce selection=(all) overwrite=true mode=average end end

```

```
coord @asn.pdb

! Now run constant energy dynamics; random initial velocities
vector do (vx = maxwell(250)) (all)
vector do (vy = maxwell(250)) (all)
vector do (vz = maxwell(250)) (all)
dynamics verlet
    nstep=500000 timest=0.001 {ps}      ! 500 ps dynamics
    iasvel=current                      ! current velocities
    nprint=250 iprfrq=250               ! statistics output
end
stop
```

1.4 Installation and testing

1.4.1 File organization

Currently, the GB source files, test files, and documentation are stored separately from the rest of the X-PLOR distribution, although this is likely to change in the future. The top of the GB directory tree is named “gbxplor”. We assume it is in the topmost directory of an existing X-PLOR distribution, and assigned to the environment variable \$GBXPLO. It contains a Readme file and the four following subdirectories:

- **gbsource** contains the additional or modified source code for the GB model. The new or modified source files are:

File	Description
gborn.s	The main source module for GB.
gborn.fcm	Common blocks for GB.
energy.s	Main energy routines; modified to set up and call GB routines.
ener.fcm	Common blocks for energy routines; modified to include GB terms.
nbonds.s	Set up of nonbonded options; modified to parse and set up GB options.
nbonds.fcm	Common blocks containing nonbonded and GB options.
parmio.s	Force field parameter reader; modified to read and set up GB parameters.
param.fcm	Common blocks for force field parameters; modified to include GB parameters.

- **gbtest** contains test files with the necessary data and shell scripts.
- **gbtoppar** contains protein parameter files for use with GB/ACE and GB/HCT.
- **gbdoc** contains this documentation.

1.4.2 Installing

If the files described here were obtained in the form of an archive **gbxplor.tar.gz**, simply unpack it by “`gunzip -c gbxplor.tar.gz | tar xvf -`”. **Make a backup copy of the original X-PLOR source directory**; copy the new or modified GB source code modules from `$GBXPLO/gbsource` into the main source directory `$SOURCE`, and follow the usual X-PLOR installation procedure. Typically, this means precompiling, then compiling the new modules, as well as any other modules that make use of `ener.fcm`, `nbonds.fcm`, or `param.fcm`.

1.4.3 Testing

Test files are in the directory `$GBXPLO/gbtest`. The tests consist of energy minimization and dynamics of an asparagine molecule and of the protein thioredoxin. The shell script `$GBXPLO/gbtest/runtests.com` can be used to execute all the tests (requiring about five minutes of CPU time on an 800MHz Intel PIII). `$GBXPLO/gbtest/difftests.com` compares the output to the output files provided with the code.

Chapter 2

Generalized Born in a protein design context

2.1 Residue GB and the protein design context

In protein design, we compute interaction energies between residue pairs, to be stored in an energy matrix [?]. If the energy function is pairwise additive, each I, J element of the energy matrix will not depend on residues other than I and J . However, in continuum electrostatics, the energy is a many-body function. Thus, the GB interaction between two atoms i, j has the form:

$$E_{ij}^{\text{int}} = E_{ij}^{\text{int}}(\mathbf{r}_i, \mathbf{r}_j) = \frac{\tau q_i q_j}{(r_{ij}^2 + b_i b_j \exp[-r_{ij}^2/4b_i b_j])^{1/2}} \quad (2.1)$$

It depends on the whole set of solute atomic coordinates through the atomic solvation radii,

$$b_i = b_i(\mathbf{r}_1, \dots, \mathbf{r}_N) \quad (2.2)$$

Thus, an atomic-pairwise decomposition of screening energies is not possible. Nevertheless, Archontis & Simonson showed how to modify the GB formulation to employ *residue-pairwise* solvation energies [?].

We define the self-energy contribution of a pair of residues I and J :

$$E_{IJ}^{\text{self}} = \sum_{i \in I, j \in J} E_{ij}^{\text{self}} \quad (2.3)$$

The total self-energy can be written

$$E^{\text{self}} = \sum_I E_I^{\text{self}} = \sum_I \sum_J E_{IJ}^{\text{self}} \quad (2.4)$$

where E_I^{self} is the self-energy of residue I and each sum runs over all residues.

We now define *residue* solvation radii B_I , by analogy to the atomic solvation radii above [?]:

$$E_I^{\text{self}} = \tau \sum_{i \in I} \frac{q_i^2}{2b_i} \stackrel{\text{def}}{=} \tau \sum_{i \in I} \frac{q_i^2}{2B_I} \quad (2.5)$$

Thus, B_I is a harmonic average of the atomic solvation radii b_i weighted by the squared charges:

$$\sum_{i \in I} \frac{1}{B_I} = \sum_{i \in I} \frac{q_i^2}{b_i} \quad (2.6)$$

Finally, we consider the GB interaction energy between residues I, J . With the usual GB formulation, we would have

$$\Delta E_{IJ}^{\text{int}} = -\tau \sum_{i \in I, j \in J} \frac{q_i q_j}{(r_{ij}^2 + b_i b_j \exp[-r_{ij}^2/4b_i b_j])^{1/2}} \quad (2.7)$$

Here, however, we introduce a new approximation for $\Delta E_{IJ}^{\text{int}}$, where the residue solvation radii replace the atomic ones:

$$\Delta E_{IJ}^{\text{int}} = -\tau \sum_{i \in I, j \in J, i \neq j} \frac{q_i q_j}{(r_{ij}^2 + B_I B_J \exp[-r_{ij}^2/4B_I B_J])^{1/2}} \quad (2.8)$$

This new approximation for the GB interaction is referred to as “**residue GB**” [?]. For now, we assume we are using this method.

For a given structure and set of interatomic distances, the interaction energy $\Delta E_{IJ}^{\text{int}}$ is a slowly varying function of the quantity $B_I B_J$, which we denote B . The B -dependency can be approximated by fitting $\Delta E_{IJ}^{\text{int}}(B)$ to a generalized polynomial as follows [?]:

$$\Delta E_{IJ}^{\text{int}}(B) \approx c_1^{IJ} + c_2^{IJ} B + c_3^{IJ} B^2 + c_4^{IJ} B^{-\frac{1}{2}} + c_5^{IJ} B^{-\frac{3}{2}} \quad (2.9)$$

The c_n are constant fitting coefficients. This approximation holds for a large range of B values. The *fitting coefficients* c_n depend only the sets of coordinates $\mathbf{r}_i \in I$ and $\mathbf{r}_j \in J$: they can be computed without information on the rest of the solute structure. In this sense, residue GB is residue-pairwise additive.

2.2 X-PLOR syntax: the pick gbfit statement

To characterize the GB interaction between a pair of residues in a protein design context, we can compute the fitting coefficients c_n introduced above, using the following statement:

```
pick gbfit (<selection 1>) (<selection 2>)
```

The GB interaction between the two selections is computed as a function of $B = B_I B_J$, where B_I , B_J are the residue B values of the two selections. The resulting fitting coefficients are stored in the user variables \$GBFIT1, \$GBFIT2, ..., \$GBFIT5. Currently, XPLOR performs the fit using a collection of B values uniformly distributed between 1 and 150 Å. In the future, more options will be added to increase user control over the fit.

2.3 Examples

The following example shows how to compute the solvation radius B_I for a given residue.

```
Remarks Compute the residue B value for a given residue
topology
  @xplor3.9/toppar/amber2xplor/lib/masses_parm99.rtf
  @xplor3.9/toppar/amber2xplor/lib/amino_parm99SB.bbunif.rtf
end
parameters @xplor3.9/toppar/amber2xplor/lib/parm99SB.gbAA.prm end
parameter nbonds
  atom trunc cdie eps=1.0 weps=80.0 e14fac=0.833333333333
  cutnb 500. ctonnb 480. ctofnb 490. nbxmod 5
end end
segment chain sequence SER ALA CYS end end end
coord @data/3pept.pdb
parameter nbonds gbhct bato end end      { activate GB      }
flags exclude * include gbse gbint end
constraints interaction (known) (known) end
```

```

vector show (bsolv) (known)           {default values are 1.}
energy end                             {energy command leads }
vector show (bsolv) (known)           {to updated values   }
vector do (rmsd = charge*charge/bsolv)(resid 1)
vector show sum (rmsd) (resid 1)
eval ($ccob=$result)                  {residue B is the     }
vector do (rmsd=charge*charge) (resid 1) {harmonic average of }
vector show sum (rmsd) (resid 1)       {atomic b's, weighted }
eval ($cc=$result)                    {by the atomic charges}
eval ($bsolv=$cc/$ccob)
stop

```

The next example shows how to compute fitting coefficients and write out information for an energy matrix.

Remarks Fit GB interaction energy between residues I, J as a function of B topology

```

@explor3.9/toppar/amber2xplor/lib/masses_parm99.rtf
@explor3.9/toppar/amber2xplor/lib/amino_parm99SB.bbunif.rtf
end
parameters @explor3.9/toppar/amber2xplor/lib/parm99SB.gbAA.prm end
parameter nbonds
  atom trunc cdie eps=1.0 weps=80.0 e14fac=0.8333333333
  cutnb 500. ctonnb 480. ctofnb 490. nbxmod 5
end end
segment chain sequence SER ALA CYS end end end
coor @data/3pept.pdb                  {compute and fit     }
pick gbfit (resid 1) (resid 2)        {GB interaction;     }
display $GBFIT1 $GBFIT2 $GBFIT3 $GBFIT4 $GBFIT5 {display coefficients}
stop                                  {in matrix format    }

```

Notice that the atomic solvation radii for a specific selection are updated when the GB self energy (\$GBSE) is computed with the bato parameter:

```

parameter nbonds gbhct bato end end

```

Chapter 3

Modelling dispersion interactions

3.1 Theory

The solute-solvent interaction consists of the electrostatic part where the atomic charges in the low dielectric cavity (solute) interact with the high dielectric surrounding medium (solvent) as described by the Generalized Born (GB) model; and the non electrostatic (non polar) part which describes the cavity formation and the van der Waals solute-solvent interaction.

3.1.1 Solute-Solvent van der Waals Dispersion Model

In spirit of the Weeks-Chandler-Andersen (WCA) repulsive/attractive decomposition of the nonpolar contribution to the solvation free energy [?] we model the solute-solvent van der Waals dispersion interactions using the attractive part of the Lennard-Jones potential.

Following the continuum solute-solvent van der Waals (vdW) energy model of Gallicchio et al. [?] the average vdW dispersion interaction of atom i with water is given by the integral of the attractive LJ potential between atom i and the oxygen atom of the water molecule, over the solvent volume, where the water number density ρ_w is assumed constant.

$$\Delta G^{\text{DI}} = \sum_i \bar{U}_i^{\text{vdw}} \quad (3.1)$$

$$\bar{U}_i^{\text{vdW}} = -\rho_w \int_{\text{solv}} \frac{4\epsilon_{iw}\sigma_{iw}^6}{|r - r_i|^6} d^3r \quad (3.2)$$

ϵ_{iw} and σ_{iw} are the LJ potential parameters for the solute atom-water oxygen pair. The total solute-solvent vdW dispersion interaction is given by the sum of the individual vdW interactions of all atoms of the solute. The integral in the above equation can be re-written as the difference of the integral over the whole space and the solute region outside the vdW radius R_i . In other words, the solvation free energy of an isolated atom i fully solvated, is reduced by the presence of all surrounding solute atoms j .

$$\bar{U}_i^{\text{vdW}} = -4\epsilon_{ij}\sigma_{ij}^6\rho_w \left(\frac{4\pi}{3R_i^3} - \int_{r>R_i}^{\text{solu}} \frac{1}{|r - r_i|^6} d^3r \right) \quad (3.3)$$

$$\bar{U}_i^{\text{vdW}} = -\frac{f_i}{R_i^3} + f_i \left(\frac{3}{4\pi} \int_{r>R_i}^{\text{solu}} \frac{1}{|r - r_i|^6} d^3r \right), \quad f_i = \frac{16\pi}{3}\epsilon_{ij}\sigma_{ij}^6 \quad (3.4)$$

The integral of $1/r^6$ over the solute region is approximated by two contributions as proposed by Onufriev [?] and computed analytically. The main integral I_i^{vdW} over the atomic vdW spheres and the correction integral I_i^{neck} over the “neck” shaped free space regions formed between pairs of vdW spheres.

$$\frac{3}{4\pi} \int_{r>R_i}^{\text{solu}} \frac{1}{|r - r_i|^6} d^3r \approx I_i^{\text{vdW}} + I_i^{\text{neck}} \quad (3.5)$$

$$\approx I_i^{\text{vdW}} + I_i^{\text{neck}}$$

$$\begin{aligned} I_i^{\text{vdW}} &= \frac{3}{4\pi} \int_{r>R_i}^{\text{solu/vdw}} \frac{1}{|r - r_i|^6} d^3r = \sum_{j \neq i} I_{ij}^{\text{vdw}}(r_{ij}, R_i, S^{\text{vdw}} R_j) \\ I_i^{\text{neck}} &= \frac{3}{4\pi} \int_{r>R_i}^{\text{solu/neck}} \frac{1}{|r - r_i|^6} d^3r = \frac{3}{4\pi} S^{\text{neck}} \sum_{j \neq i} I_{ij}^{\text{neck}}(r_{ij}, R_i, R_j) \end{aligned} \quad (3.6)$$

Both terms of eq. 3.5 are computed by the sum $\sum_{j \neq i} I_{ij}$ of all atom pairwise interactions expressed by the following analytical conditional functions:

$$I_{ij}^{\text{vdW}}(R_i, R_j, r_{ij}) = \begin{cases} \frac{R_j^3}{(r_{ij}^2 - R_j^2)^3}, & \text{if } r_{ij} \geq R_i + R_j \\ \frac{1}{16r_{ij}} \left(\frac{r_{ij} + 3R_j}{(r_{ij} + R_j)^3} + \frac{3(R_j^2 - R_i^2 - (r_{ij} - R_i)^2) + 2r_{ij}R_i}{R_i^4} \right), & \text{otherwise} \end{cases} \quad (3.7)$$

$$I_{ij}^{neck}(R_i, R_j, r_{ij}) = \begin{cases} A_{ij}(r_{ij} - B_{ij})^4(R_i + R_j + 2R_w - r_{ij})^4, & \text{if } B_{ij} < r_{ij} < R_i + R_j + 2R_w \\ 0, & \text{otherwise} \end{cases} \quad (3.8)$$

The neck term parameters A and B themselves depend on the atomic vdW radii of each pair and the water probe radius R_w . Finally the total solute-solvent vdW dispersion interaction is given by

$$\Delta G^{DI} = \sum_i -\frac{f_i}{R_i^3} + \sum_i \sum_{j \neq i} f_i(I_{ij}^{vdw} + \frac{3}{4\pi} S^{neck} I_{ij}^{neck}) \quad (3.9)$$

3.1.2 Gaussian Nonpolar Solvent Model

The Lazaridis and Karplus (LK) model [?] expresses the total solvation free energy of a particular molecular conformation as a sum over contributions from individual groups of atoms, as follows:

$$\begin{aligned} \Delta G^{LK} &= \sum_i \Delta G_i^{solv} \\ \Delta G_i^{solv} &= G_i^{ref} - \sum_{j \neq i} \int_{V_j} f_i(r_{ij}) dV \\ &= G_i^{ref} - \sum_{j \neq i} f_i(r_{ij}) V_j \end{aligned} \quad (3.10)$$

Each contribution reflects the change in the solvation free energy due to the transfer of the corresponding group from the unfolded (fully solvated) to the folded (partially solvated or buried) conformation. This transfer is accompanied by a partial or total replacement of the surrounding high dielectric solvent by the less polar solute medium, a change in the solvent orientation around the solute and a modification in the solute-solvent interactions. The solvation energy of a fully solvent exposed group i is given by an empirically determined reference value G_i^{ref} . The same group inside the solute is screened from solvent by the surrounding groups each contributing to a reduction in the solvation energy of group i . This reduction is expressed by the integral over the volume of the surrounding solute groups of a

gaussian energy density function :

$$f_i(r_{ij}) = \frac{G_i^{\text{free}}}{2\pi^{3/2}\lambda_i r_{ij}^2} e^{-\left(\frac{r_{ij}-R_i}{\lambda_i}\right)^2} \quad (3.11)$$

which depends on the distance r_{ij} , the vdW radius R_i , the gaussian correlation length λ_i and G_i^{free} . This last parameter is such that, when group i is fully buried the total solvation energy becomes zero. In the LK model, the integrals are approximated by the product of the density function of group i and the atomic volume of the surrounding solute group j :

$$\Delta G^{\text{LK}} = \sum_i G_i^{\text{ref}} - \sum_i \sum_{j \neq i} f_i(r_{ij}) V_j \quad (3.12)$$

3.2 Implementation

The solute-solvent vdW dispersion interaction was inserted at the level of non bonded Generalized Born (GB) evaluation with an individual subroutine. After an energy call the parameters of the model described above are read and stored. That is, type and atom-based van der Waals radii (R^{vdw}) the solvent type (oxygen atom for water) and parameters A , B if the neck correction term is turned on. In a first stage the van der Waals interaction of each isolated atom of the solute with the continuum solvent is accumulated to give the “self” part of the dispersion energy. The factor at the numerator of Eq. (3.4) depends on the water density number and the Lennard-Jones B coefficient for each solute-solvent atom pair. At a second stage, the interaction of all solute atoms j surrounding each atom i , is evaluated and summed, to give the “interaction” part which accounts for the replacement of the surrounding solvent in the isolated atom state by solute. As shown in Eq. (3.6) van der Waals radii of the surrounding atoms are scaled down by a factor S^{vdw} . The neck term correction of the simplistic representation of the solute as van der Waals spheres, accounts for omitted space between atomic vdW spheres within the solute boundary. The total neck contribution is scaled by a factor S^{neck} (Eq. 3.6). The parameters A and B present in Eq. (3.8) depend on the pair of $R_i^{\text{vdw}}, R_j^{\text{vdw}}$ and the water probe radius R_w . A set of A and B values has been evaluated on a 2-dimensional equally spaced grid ($R_i^{\text{vdw}} \times R_j^{\text{vdw}}$) using the numerical method NSR6 developed by Onufriev et. al [?]. If the van der Waals radii of a pair do not

coincide with a grid node then A and B parameters are obtained by cubic spline interpolation. The second derivative values of neck parameters are also read from the parameter file.

The solute-solvent van der Waals dispersion interaction term is pairwise decomposable, the derivative has analytical form and the forces are readily obtained by

$$\begin{aligned}\frac{\partial \Delta G_i^{\text{DI}}}{\partial x_i} &= \sum_{j \neq i} f_i \frac{\partial I_{ij}^{\text{vdW}}}{\partial x_i} + \frac{3}{4\pi} S^{\text{neck}} f_i \frac{\partial I_{ij}^{\text{neck}}}{\partial x_i} \\ &= \sum_{j \neq i} \left\{ f_i \frac{\partial I_{ij}^{\text{vdW}}}{\partial r_{ij}} + \frac{3}{4\pi} S^{\text{neck}} f_i \frac{\partial I_{ij}^{\text{neck}}}{\partial r_{ij}} \right\} \frac{(x_i - x_j)}{r_{ij}}\end{aligned}\quad (3.13)$$

$$\frac{\partial I_{ij}^{\text{vdW}}}{\partial r_{ij}} = \begin{cases} \sum_{j \neq i} \frac{-6r_{ij}R_j^3}{(r_{ij}^2 - R_j^2)^4}, & \text{if } r_{ij} \geq R_i + R_j \\ \sum_{j \neq i} \frac{-2}{16r_{ij}} \left(\frac{r_{ij} + 4R_j}{(r_{ij} + R_j)^4} + \frac{3r_{ij} - 4R_i}{R_i^4} \right) - \frac{1}{16r_{ij}^2} I_{ij}^{\text{vdW}}(r_{ij} < R_i + R_j), & \text{otherwise} \end{cases}\quad (3.14)$$

$$\frac{\partial I_{ij}^{\text{neck}}}{\partial r_{ij}} = \begin{cases} 4A_{ij}(r_{ij} - B_{ij})^3(R_i + R_j + 2R_w - r_{ij})^3(R_i + R_j + 2(R_w - r_{ij}) + B_{ij}), & \text{if } B_{ij} < r_{ij} < R_i + R_j + 2R_w \\ 0, & \text{otherwise} \end{cases}\quad (3.15)$$

The Lazaridis-Karplus gaussian nonpolar solvent model is put within the non-bonding intramolecular interactions calculation path with a separate subroutine. At first the type-based parameters of the model G^{ref} , G^{free} and λ are being read and stored in arrays. A separate type-based array is needed for use of the model with AMBER, because it was initially developed for CHARMM and there is a poor correspondence between carbon atoms types CT and CH1E/CH2E/CH3E. The nonpolar part of the solvation free energy is computed in two steps. First, we compute the sum of atomic solvation free energies (G^{ref}) in their isolated-state completely surrounded by solvent (first term of Eq. (3.16)); and then we subtract the sum of the desolvation (replacement of solvent by solute) of each atom from all remaining solute atoms j (second term of Eq. (3.16)). All atom pairs contribute to the LK model, so 1-2 and 1-3 pairs excluded from van der Waals and electrostatic interactions, are taken into account. All hydrogen atoms are considered part of the

heavy atom they are attached to, and are excluded from the calculation, as in the initial model.

$$\Delta G^{\text{LK}} = \sum_i G_i^{\text{ref}} - \sum_i \sum_{j>i} (f_i(r_{ij})V_j + f_j(r_{ij})V_i) \quad (3.16)$$

When using the constraint interaction command to compute the interaction energy between two selected groups of atoms, the first term of equation (3.16) is evaluated only for those atoms which belong in both groups. If the two groups do not share any atoms, only the second term is computed describing the desolvation of the first group from the second and the desolvation of the second from the first. The command *cons inte (resid R₁ or resid R₂)(resid R₁ or resid R₂) end* computes the solvation energy of both residues R₁ and R₂ as follows

$$\Delta G^{\text{LK}}(R_1, R_2) = \sum_{i \in R_1, R_2} G_i^{\text{ref}} - \sum_{i \in R_1, R_2} \sum_{i < j \in R_1, R_2} (f_i(r_{ij})V_j + f_j(r_{ij})V_i) \quad (3.17)$$

But, the command *cons inte (resid R₁)(resid R₁ or resid R₂) end* computes the solvation energy of residue R₁ in the presence of residue R₂ correctly taking into account the desolvation of R₁ from R₂ but it also includes the desolvation of residue R₂ from residue R₁. To eliminate the latter undesirable contribution we remove $f_j(r_{ij})$, $j \in R_2$ by setting the atom-based parameters of residue R₂ to zero, with the command *parameters GNSP (resid R₂) 0.0 0.0 1.0 end*. Now the solvation energy of residue R₁ in the presence of R₂ is given by

$$\Delta G^{\text{LK}}(R_1, R_1 - R_2) = \sum_{i \in R_1} G_i^{\text{ref}} - \sum_{i \in R_1} \sum_{i < j \in R_1, R_2} f_i(r_{ij})V_j \quad (3.18)$$

The solvation free energy depends on the distance between pairs of atoms (r_{ij}), and its derivative has the following analytical form

$$\frac{\partial \Delta G^{\text{LK}}}{\partial x_i} = \sum_{j \neq i} \frac{2}{r_{ij}} \left\{ \left(\frac{1}{r_{ij}} + \frac{r_{ij} - R_i}{\lambda_i^2} \right) f_i V_j + \left(\frac{1}{r_{ij}} + \frac{r_{ij} - R_j}{\lambda_j^2} \right) f_j V_i \right\} (x_i - x_j) \quad (3.19)$$

3.3 Syntax

3.3.1 Solute-solvent van der Waals dispersion energy

The dispersion term is assigned the variable name **\$GBDI** and is activated by the flag statement: **flags include GBDI end**

Setting the GBDI options

NBONDS <nonds-statement>|<gborn-nonds-statement>|<gbdi-nonds-statement> **END**

<gbdi-nonds-statement>:==

GBDI GBDN Flags activating the main GBDI term and the neck contribution. Default: inactive.

WTYPE=<string> Solvent chemical type. Default: OW (TIP3P oxygen atom).

WRHO=<real> Solvent density number. Default: 1.

SGBDI=<real> R_j^{vdW} scaling factor. Default: 1.

SNECK=<real> Neck term scaling factor. Default: 1.

RWAT=<real> Water probe radius. Default: 1.4.

Setting up the parameters

The type-based parameters of the vdW dispersion model are being set with a parameter statement.

PARAMeter {<parameter-statement>} **END**

<parameter-statement>:==

DSPN <RvdW-statement><neckAB-statement> **END**

<RvdW-statement>:==

GNOD

<integer> <real> ... <real> defines the size of the grid and assigns a vdW radius foreach node of the grid.

<neckAB-statement>:==

NCKA

<real> ... <real> assigns a value of the neck-A parameter to all nodes of a row in the $(R_i^{vdW} \times R_j^{vdW})$ grid.

NCKB

<real> ... <real> assigns a value of the neck-B parameter to all nodes of a row in the $(R_i^{vdW} \times R_j^{vdW})$ grid.

NC2A

<real> ... <real> assigns a value of the second derivative of neck-A parameter

to all nodes of a row in the $(R_i^{\text{vdW}} \times R_j^{\text{vdW}})$ grid, used for cubic spline interpolation.
NC2B

<real> ... <real> assigns a value of the second derivative of neck-B parameter to all nodes of a row in the $(R_i^{\text{vdW}} \times R_j^{\text{vdW}})$ grid, used for cubic spline interpolation.

Neck parameters calculated with an external program

The analytical expression of the neck correction (Eq. 3.6) to the $1/|r|^6$ integral over the solute space (sum of atomic van der Waals spheres) uses parameters A_{ij} B_{ij} , which depend on the van der Waals radii of the atom pair and the water probe radius.

3.3.2 Lazaridis-Karplus interaction energy

The LK term is assigned the variable name **\$GNSM** and is activated by the flag statement: **flags include GNSM end.**

Setting GNSM

NBONDS <nbonds-statement>|<gborn-nbonds-statement>|<gbdi-nbonds-statement>|<gnsn-nbonds-statement>END

<gnsn-nbonds-statement>:==

GNSM Flag activating the GNSM term. Default: inactive.

Setting up the parameters

The type- and atom-based parameters of the LK model are being set with a parameter statement.

PARAMeter {<parameter-statement>} END

<parameter-statement>:==

GNSP

<type> <real> <real> <real> adds G^{ref} , G^{free} and λ parameters for the atom type to the parameter database.

`<selection> <real> <real> <real>` adds G^{ref} , G^{free} and λ parameters for the selected atoms to the parameter database.

3.3.3 Examples

Minimization and MD with GB/DI/LK

```

1 topology
2 @xplor3.9/toppar/amber2xplor/lib/masses_parm99.rtf
3 @xplor3.9/toppar/amber2xplor/lib/amino_parm99SB.bbunif.rtf
4 @xplor3.9/toppar/amber2xplor/lib/solvents.rtf
5 @xplor3.9/toppar/amber2xplor/lib/ions.rtf
6 end
7
8 parameters
9 @parm99SB.plus.prm                                !{ plus DI and LK parameters }
10 end
11
12 structure @allh_model.psf end
13 coordinates @allh_model.pdb
14
15 @LK_charmm2amber.str    !{ type conversion from charmm19 to amber99 }
16
17 parameter
18     nbonds
19     atom trunc cdie eps=1 e14fac=0.8333333333
20     ctonnb=97. ctofnb=98. cutnb=99. nbxmod=5 toler=100.
21     gbhct eps=1. weps=80.
22     gbdi wtype = OW sgbdi=0.6211 wrho=0.033428    !{ GBDI parameters }
23     gbdn sneck=0.4058 rwat=1.4                    !{ GBDN parameters }
24     gnsml                                           !{ GNSM option }
25 end
26 end
27
28 flags include gbse gbin gbdi gnsml end

```

```
29
30 energy end
31
32 minimize powell nstep=50 end
33
34 energy end
35
36 display $gbse $gbin $gbdi $gnsn
37
38 vector do (vx = maxwell(250)) (all)
39 vector do (vy = maxwell(250)) (all)
40 vector do (vz = maxwell(250)) (all)
41
42 dynamics verlet
43 nstep=1000 timest=0.001 iasvel=current
44 nprint=250 iprfrq=250
45 end
46
47 stop
```

Chapter 4

Some data structures for protein design