Chapter 1

An introduction to continuum electrostatics

For both major subjects of this work, folding and titration, continuum electrostatics is of crucial importance. Therefore, I will give in this chapter a basic introduction to the concepts and methods of continuum electrostatics that were applied here.

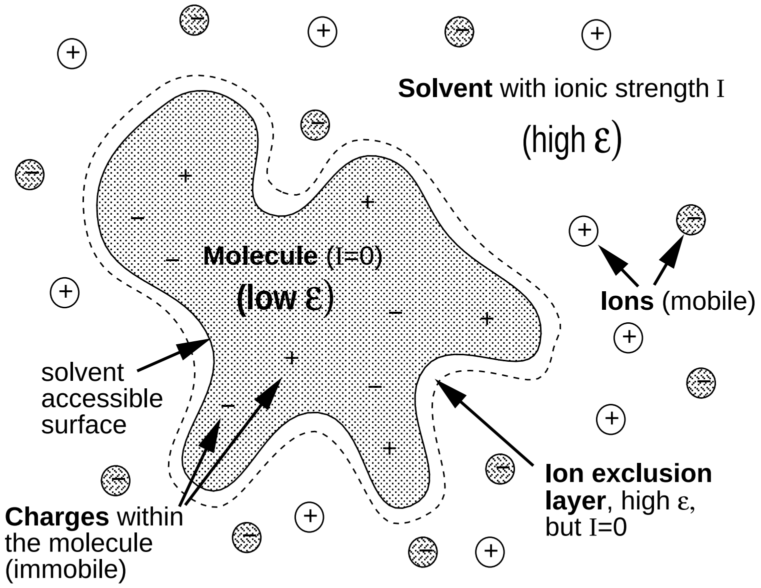

Figure 1.1: A molecule in a heterogenous dielectric medium.

Figure 1.1 shows a large molecule in an ionic solution. In the electrostatic continuum approach, the solvent is not represented by explicit solvent molecules, but implicitly by a medium with a high dielectric constant ε. The interior of the molecule, separated from the solvent by the solvent accessible surface, is assigned a low dielectric constant. At a first glance, one should expect this constant to be unity (as in vacuum) since the atoms of the molecule are all represented explicitly in this model. However, electronic and nuclear polarization effects not considered explicitly by the model may cause a higher dielectric constant to be more appropriate. The value used for the dielectric constant within the molecule is subject of intense discussion (Warshel & Russel, 1984; Warshel & Åqvist, 1991; Honig & Nicholls, 1995; Warshel *et al.*, 1997; Rabenstein *et al.*, 1998a; Ullmann & Knapp, 1999). I will provide more details of this discussion in connection with the applications described in the following chapters (see especially section 3.2.3.5).

1.1 The Poisson-Boltzmann equation

1.1.1 Derivation

To describe the electrostatic interaction of the molecular system depicted in Figure 1.1 mathematically, the Poisson-Boltzmann equation is used. It is derived from the Coulomb potential:

$$\phi(\vec{r}) = \sum_i \frac{q_i}{\varepsilon|\vec{r}-\vec{r}_i|} \tag{1.1}$$

where the sum runs over all point charges q_i at the position $\vec{r}_i$. ε is the dielectric constant. The Coulomb potential can also be formulated for the charge density ρ instead of point charges:

$$\phi(\vec{r}) = \int_V \frac{\rho(\vec{r'})d\vec{r'}}{\varepsilon|\vec{r}-\vec{r'}|} \tag{1.2}$$

Applying the Laplace operator to both sides of this equation results after a few transformations in the Poisson equation:

$$\Delta\phi(\vec{r}) = -4\pi\frac{\rho(\vec{r})}{\varepsilon} \tag{1.3}$$

If ε is not constant but depends on the position $\vec{r}$, the Poisson equation adopts the following form:

$$\nabla\cdot[\varepsilon(\vec{r})\nabla\phi(\vec{r})] = -4\pi\rho(\vec{r}) \tag{1.4}$$

To describe the effect of the ions in the solvent, the Debye-Hückel theory is needed, which I will not describe in detail here. The mobile ions in the solution arrange themselves Boltzmann distributed in the electric field similar to gas atoms in a gravitation potential. The resulting charge density $\rho_{ion}(\vec{r})$ of the ions is added to the charge density $\rho(\vec{r})$ without ions:

$$\rho_{ion}(\vec{r}) = \sum_s c_s(\vec{r})q_s e^{-\beta q_s\phi(\vec{r})} \tag{1.5}$$

The sum runs over all kind of ions s. c_s is the original concentration of ion s, q_s is the charge of ion s. β equals $(k_{\mathrm{B}}T)^{-1}$. Adding of the ionic charge density to the Poisson equation leads to the Poisson-Boltzmann equation (PBE):

$$\nabla\cdot[\varepsilon(\vec{r})\nabla\phi(\vec{r})] + 4\pi\sum_s c_s(\vec{r})q_s e^{-\beta q_s\phi(\vec{r})} = -4\pi\rho(\vec{r}) \tag{1.6}$$

The Debye-Hückel theory requires overall electroneutrality:

$$\sum_s c_s(\vec{r})q_s = 0 \tag{1.7}$$

In the special case of only monovalent ions, the Poisson-Boltzmann equation can therefore be written elegantly with sinh (sinh(x) is defined as $\frac{1}{2}(e^x - e^{-x})$):

$$\nabla\cdot[\varepsilon(\vec{r})\nabla\phi(\vec{r})] - 8\pi c_s(\vec{r})\sinh[\beta \mathrm{e}\phi(\vec{r})] = -4\pi\rho(\vec{r}) \tag{1.8}$$

where e is the unit elementary charge.

To simplify the numerical solution, the Poisson-Boltzmann equation can be linearized by expanding the exponentials:

$$\sum_s c_s(\vec{r})q_s e^{-\beta q_s\phi(\vec{r})} \approx \sum_s c_s(\vec{r})q_s - \beta\sum_s c_s(\vec{r})q_s^2\phi(\vec{r}) \tag{1.9}$$

$\sum_s c_s(\vec{r})q_s$ is zero due to electroneutrality. $\sum_s c_s(\vec{r})q_s^2$ is twice the ionic strength $I(\vec{r}) = \frac{1}{2}\sum_s c_s(\vec{r})q_s^2$. Together with eq 1.6, this results in the linearized PBE:

$$\nabla \cdot [\varepsilon(\vec{r})\nabla\phi(\vec{r})] - 8\pi\beta I(\vec{r})\phi(\vec{r}) = -4\pi\rho(\vec{r}) \tag{1.10}$$

For a more detailed description of the deviation of the PBE and the Debye-Hückel theory, see chapter 15-1 (pp 328–340) of McQuarrie (1976) or chapter 18 (pp 321–339) of Hill (1986).

1.1.2 Numerical solution

Analytical solutions of the PBE exist only for simple geometries (Kirkwood, 1934; Tanford & Kirkwood, 1957; Daune, 1997). For complex geometries, solutions can be obtained by numerical methods. Most often, the PBE is solved by finite difference methods (Warwicker & Watson, 1982; Nicholls & Honig, 1991; Honig & Nicholls, 1995). A finite difference method is also applied in this work. However, there are also more elegant numerical methods such as boundary element methods (Sklenar *et al.*, 1990; Zauhar & Varnek, 1996) or multigrid-based methods (Holst *et al.*, 1994; Holst & Saied, 1995) to solve the PBE. The reason for nevertheless preferring a finite-difference method here is merely the practical availability of a suitable suite of programs (Bashford & Gerwert, 1992; Bashford, 1997) that has been thoroughly tested and proven to yield reliable results in numerous applications. So far, there have been only a few applications of boundary element methods to calculate protonation behavior of proteins (Ripoll *et al.*, 1996; Juffer *et al.*, 1997; Vila *et al.*, 1998).

1.1.2.1 Finite difference method

To apply the finite difference method, all relevant physical quantities (charge q, dielectric constant ε, ionic strength I, electrostatic potential ϕ) are mapped on a cubic grid with grid constant l (Fig. 1.2).

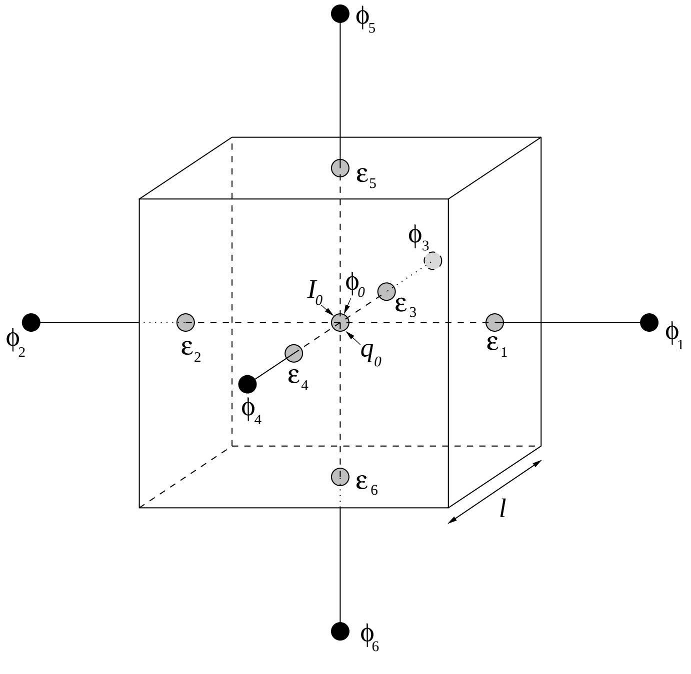

Figure 1.2: Part of the grid used to solve the PBE.

The linearized PBE (eq 1.10) is integrated over the volume V of one cubic grid element:

$$\int_V \nabla \cdot [\varepsilon(\vec{r}) \nabla \phi(\vec{r})] d\vec{r} - \int_V 8\pi\beta I(\vec{r}) \phi(\vec{r}) d\vec{r} = \int_V -4\pi\rho(\vec{r}) d\vec{r} \tag{1.11}$$

While the second and third integrals are easily resolvable, the first one is more difficult. In a first step, it is transformed into a surface integral using Gauß' theorem:

$$\int_A \varepsilon(\vec{r}) \nabla \phi(\vec{r}) dA - 8\pi\beta I_0 \phi_0 l^3 = -4\pi q_0 \tag{1.12}$$

The surface integral is now calculated separately for all six sides of the cubic grid element. In doing so, the gradient of the electrostatic potential $\nabla\phi$ is substituted by its finite difference form in the respective direction:

$$\sum_{i=1}^{6} \frac{\varepsilon_i (\phi_i - \phi_0) l^2}{l} - 8\pi\beta I_0 \phi_0 l^3 = -4\pi q_0 \tag{1.13}$$

Eq. 1.13 is simplified and rearranged to yield a finite difference expression for ϕ_0:

$$\phi_0 = \frac{(\sum_{i=1}^{6} \varepsilon_i \phi_i) + \frac{4\pi q_0}{l}}{(\sum_{i=1}^{6} \varepsilon_i) + 8\pi\beta I_0 l^2} \tag{1.14}$$

Starting from arbitrary values, the electrostatic potential is iteratively calculated for each grid point according to eq 1.14 until a convergence criterion is met. For the details of this procedure see Nicholls and Honig (1991) and chapter 17.3 (pp 762–768) of Press *et al.* (1992).

1.1.2.2 Focusing

There is a problem at the borders of the grid, since the grid points at the border have less than six neighbor points. If the grid is much larger than the molecule, so that the border grid points are far away from the charges in the molecule, ϕ outside of the grid can be set to zero (or better to a value according to the Debye-Hückel sum (Klapper *et al.*, 1986)). Due to computer limitations, the resolution of a large grid has to be poor. To enhance resolution, additional calculations with a smaller high-resolution grid can be performed (Klapper *et al.*, 1986). This grid is centered on the region of interest. On its boundaries, the electrostatic potential is set to a value interpolated from the calculation with the larger grid. Such a focusing step can be repeated if necessary.

1.1.2.3 Grid artifact

An electrostatic energy G_E can be calculated from the electrostatic potential ϕ:

$$G_E = \frac{1}{2} \sum_i q_i \phi_i \tag{1.15}$$

The sum runs over all atoms i. q_i is the atomic partial charge of atom i and ϕ_i the electrostatic potential at the position of atom i.

The electrostatic energy of a classical point charge in its own electrostatic potential, the so-called self energy, is infinite large. These singularities are avoided if the PBE is solved on a grid, since a point charge is smeared over the cubic grid element where it is localized. However, arbitrary values, depending only on the grid resolution and position, remain from the infinite self energy of the singularities. This grid artifact is called *grid energy*. There are several possibilities to get rid of the grid energy. The easiest one is not to try to calculate absolute electrostatic energies but energy differences between systems where the grid energy is equal, so that the grid energy vanishes in the difference. In chapter 3, this method has been applied, and it will be discussed there in more detail.

Another approach is to use a singularity free charge distribution or to use Green's theorem (Beroza & Fredkin, 1996). The boundary element methods reported earlier do not need a grid and therefore are not afflicted with the grid artifact at all.

1.2 The Analytical Continuum Solvent (ACS)

If Poisson-Boltzmann electrostatics is used for molecular dynamics (MD) or Monte Carlo (MC) simulation of large molecules, even the fastest methods to solve the PBE (Hoffmann *et al.*, 1998) are too slow for solving problems like long-term protein dynamics or even protein folding. To overcome this problem, Schaefer and Karplus (1996) developed the so-called analytical continuum electrostatics (ACE) potential, which approximates the potential given by solving the Poisson equation, but needs much less computational effort. Similar to other analytical continuum electrostatics methods (Sklenar *et al.*, 1990; Still *et al.*, 1990; Schaefer & Froemmel, 1990; Gilson & Honig, 1991), the self-energy part of ACE is based on the integrated field concept, whose basic idea dates back to the work of Born (1920). ACE combines the self-energy potential with the generalized Born equation for charge–charge interaction (Klopman, 1967; Constanciel & Contreras, 1984; Still *et al.*, 1990). By adding a non-polar free energy of solvation term, ACE is extended to the analytical continuum solvent (ACS) (Schaefer *et al.*, 1998).

The following is an introduction into the basics of ACE/ACS according to Schaefer and Karplus (1996). Applications of ACS will be described in the last part of chapter 2.

1.2.1 Integral formulation of the electrostatic energy

The starting point of the ACE approach is the description of the electrostatic energy G_E in terms of the energy density $u(\vec{r})$ rather than of the electrostatic potential as the solution of the Poisson equation. The energy density $u(\vec{r})$ of an electrostatic field generated by a charge distribution $\rho(\vec{r})$ can be expressed in terms of the electric displacement vector

$$\vec{D}(\vec{r}) = \varepsilon(\vec{r})\vec{E}(\vec{r}) \tag{1.16}$$

as follows:

$$u(\vec{r}) = \frac{1}{8\pi\varepsilon(\vec{r})}\vec{D}^2(\vec{r}) \tag{1.17}$$

By integrating the energy density over the full space, the electrostatic energy G_E is obtained (Born, 1920):

$$G_E = \int_V u(\vec{r})d\vec{r} = \frac{1}{8\pi}\int_V \frac{1}{\varepsilon(\vec{r})}\vec{D}^2(\vec{r})d\vec{r} \tag{1.18}$$

This expression can be derived from the well known expression for the electrostatic energy of a charge distribution $\rho(\vec{r})$ in the electrostatic potential $\phi(\vec{r})$:

$$G_E = \frac{1}{2}\int_V \rho(\vec{r})\phi(\vec{r})d\vec{r} \tag{1.19}$$

The discrete form of eq 1.19 was already introduced as eq 1.15. According to the Poisson equation 1.4, the charge density $\rho(\vec{r})$ in eq 1.19 is substituted by $-\frac{1}{4\pi}\nabla\cdot[\varepsilon(\vec{r})\nabla\phi(\vec{r})]$:

$$G_E = -\frac{1}{8\pi}\int\limits_V \phi(\vec{r})\nabla\cdot[\varepsilon(\vec{r})\nabla\phi(\vec{r})]d\vec{r} \tag{1.20}$$

Integration by parts leads to

$$G_E = -\frac{1}{8\pi}[\phi(\vec{r})\varepsilon(\vec{r})\nabla\phi(\vec{r})]_V + \frac{1}{8\pi}\int\limits_V \varepsilon(\vec{r})\nabla\phi(\vec{r})\nabla\phi(\vec{r})d\vec{r} \tag{1.21}$$

The first term vanishes due to the vanishing electrostatic potential $\phi(\vec{r})$ at the (infinite far away) boundaries of the volume V. With $\vec{E}(\vec{r}) = -\nabla\phi(\vec{r})$ and with eq 1.16, the following expression is obtained:

$$G_E = \frac{1}{8\pi}\int\limits_V \varepsilon(\vec{r})\vec{E}(\vec{r})\vec{E}(\vec{r})d\vec{r} = \frac{1}{8\pi}\int\limits_V \frac{1}{\varepsilon(\vec{r})}\vec{D}(\vec{r})\vec{D}(\vec{r})d\vec{r} \tag{1.22}$$

This matches the expression in eq 1.18.

Corresponding to the situation described before (Figure 1.1), I split now the integral expression of the electrostatic energy (eq 1.18) in two parts, one is the integral over the volume of the solute (*i. e.* the protein) V_p with the low dielectric constant ε_p, the other is the integral over the remaining volume of the solvent V_s with the high dielectric constant ε_s.

$$G_E = \frac{1}{8\pi\varepsilon_s}\int\limits_{V_s} \vec{D}^2(\vec{r})d\vec{r} + \frac{1}{8\pi\varepsilon_p}\int\limits_{V_p} \vec{D}^2(\vec{r})d\vec{r} \tag{1.23}$$

The so-called reduced dielectric constant $\Delta\varepsilon$ is defined by the following expression:

$$\frac{1}{\Delta\varepsilon} = \frac{1}{\varepsilon_p} - \frac{1}{\varepsilon_s} \tag{1.24}$$

Since $\varepsilon_p < \varepsilon_s$, the reduced dielectric constant $\Delta\varepsilon$ is always positive. Using the reduced dielectric constant $\Delta\varepsilon$, eq 1.23 can be rewritten extending the first integral over the full space V:

$$G_E = \frac{1}{8\pi\varepsilon_s}\int\limits_{V} \vec{D}^2(\vec{r})d\vec{r} + \frac{1}{8\pi\Delta\varepsilon}\int\limits_{V_p} \vec{D}^2(\vec{r})d\vec{r} \tag{1.25}$$

This expression is still exact. An approximation is introduced by assuming that the first term corresponds to the situation in a homogenous dielectric medium, where the dielectric displacement is described by the simple Coulomb field. This assumption is not exact since the dielectric displacement $\vec{D}$ in both integrals must satisfy the boundary conditions on the electric field at the interface between solute and solvent, *i. e.* the tangential component of the electric field $\vec{E} = \vec{D}/\varepsilon$ and the normal component of the dielectric displacement $\vec{D}$ do not change when passing through the solute–solvent boundary. This means that the field lines are somewhat distorted at the dielectric boundary. This distortion is not represented in the Coulomb field in a homogenous dielectric medium. The relative error introduced by the approximation is, however, not larger than a few percent (Schaefer & Froemmel, 1990). Such a simplifying approximation is not possible for the second integral, which is much more difficult to evaluate.

Given the situation of discrete atoms that bear point charges as atomic partial charges, the interaction terms of atoms with themselves G_i^{self} can be distinguished from interaction terms between different atoms G_{ij}^{int}:

$$G_E = \sum_i \left(G_i^{self} + \sum_{j>i} G_{ij}^{int} \right) \tag{1.26}$$

where both sums run up to the total number of atoms. G_{ij}^{int} and G_i^{self} can be calculated according to eq 1.25 (applying the approximation of the dielectric displacement in the first integral by the Coulomb field):

$$G_{ij}^{int} = \frac{q_i q_j}{\varepsilon_s |\vec{r}_{ij}|} + \frac{1}{8\pi\Delta\varepsilon} \int\limits_{V_p} \vec{D}_i(\vec{r}) \vec{D}_j(\vec{r}) d\vec{r} \tag{1.27}$$

$$G_i^{self} = \frac{q_i^2}{2\varepsilon_s R_i} + \frac{1}{8\pi\Delta\varepsilon} \int\limits_{V_p} \vec{D}_i^2(\vec{r}) d\vec{r} \tag{1.28}$$

The self energy G_i^{self} of a point charge yields a diverging energy contribution. To avoid this singularity, a trick is applied in eq 1.28: The atom is not longer considered as a point, but as a sphere with radius R_i, *e. g.* the van-der-Waals radius. The point charge q_i is distributed over the surface of this sphere. The first term in eq 1.28 is obtained for this special distribution of the charge q_i and is called the Born energy term (Born, 1920).

1.2.2 Solvation energy

The energy of placing a molecule from a homogenous dielectric medium with the same dielectric constant ε_p for solvent and solute into a heterogenous dielectric medium with ε_p for the solute and a higher dielectric constant ε_s for the solvent, is called here *solvation energy* ΔG_E. This energy is identical to the electrostatic part of the conventional solvation energy if ε_p is unity, which is the dielectric constant of vacuum. (This energy, however, does not include the non-polar parts of the solvation energy.) ΔG_E is calculated by taking the difference between the electrostatic energy in the heterogenous dielectric medium G_E (according to eq 1.26) and the electrostatic energy in the homogenous dielectric medium G_{hom}:

$$\Delta G_E = G_E - G_{hom} \tag{1.29}$$

If the Born formula is also applied for the calculation of self energies in the homogenous dielectric medium, G_{hom} is calculated as follows:

$$G_{hom} = \sum_i \left(\frac{q_i^2}{2\varepsilon_p R_i} + \sum_{j>i} \frac{q_i q_j}{\varepsilon_p |\vec{r}_{ij}|} \right) \tag{1.30}$$

Again, the energy is splitted into a self energy and an interaction term:

$$\Delta G_E = \Delta G^{self} + \Delta G^{int} = \sum_i \left(\Delta G_i^{self} + \sum_{j>i} \Delta G_{ij}^{int} \right) \tag{1.31}$$

Taking into account eq 1.27 and eq 1.28, ΔG_{ij}^{int} and ΔG_i^{self} resolve to

$$\Delta G_{ij}^{int} = \frac{q_i q_j}{\varepsilon_s |\vec{r}_{ij}|} + \frac{1}{8\pi\Delta\varepsilon} \int\limits_{V_p} \vec{D}_i(\vec{r}) \vec{D}_j(\vec{r}) d\vec{r} - \frac{q_i q_j}{\varepsilon_p |\vec{r}_{ij}|} \tag{1.32}$$

$$\Delta G_i^{self} = \frac{q_i^2}{2\varepsilon_s R_i} + \frac{1}{8\pi\Delta\varepsilon} \int\limits_{V_p} \vec{D}_i^2(\vec{r}) d\vec{r} - \frac{q_i^2}{2\varepsilon_p R_i} \tag{1.33}$$

With the expression for the reduced dielectric constant eq 1.24, this can be written shorter as

$$\Delta G_{ij}^{int} = -\frac{q_i q_j}{\Delta\varepsilon|\vec{r}_{ij}|} + \frac{1}{8\pi\Delta\varepsilon}\int\limits_{V_p} \vec{D}_i(\vec{r})\vec{D}_j(\vec{r})d\vec{r} \tag{1.34}$$

$$\Delta G_i^{self} = -\frac{q_i^2}{2\Delta\varepsilon R_i} + \frac{1}{8\pi\Delta\varepsilon}\int\limits_{V_p} \vec{D}_i^2(\vec{r})d\vec{r} \tag{1.35}$$

1.2.3 Generalized Born approximation

In the spirit of the Born energy term, ΔG_i^{self} can be formally written as an analogous Born energy term with a yet unknown effective Born radius b_i, which accounts for the volume covered by the solute surrounding atom i:

$$\Delta G_i^{self} = -\frac{q_i^2}{2\Delta\varepsilon b_i} \tag{1.36}$$

If ΔG_i^{self} is already calculated (e. g. by solving the PBE or by applying the analytical approximation described in the next sections), the effective Born radius b_i is given by the following equation:

$$b_i = -\frac{q_i^2}{2\Delta\varepsilon\Delta G_i^{self}} \tag{1.37}$$

In the same way, ΔG_{ij}^{int} can be written with an effective interaction distance R_{ij}^{solv} accounting for the influence of the solute volume:

$$\Delta G_{ij}^{int} = -\frac{q_i q_j}{\Delta\varepsilon R_{ij}^{solv}} \tag{1.38}$$

Putting together eq 1.37 and eq 1.38 with eq 1.31 yields the so-called generalized Born equation (Still *et al.*, 1990):

$$\Delta G_E = -\sum_i \left(\frac{q_i^2}{2\Delta\varepsilon b_i} + \sum_{j>i} \frac{q_i q_j}{\Delta\varepsilon R_{ij}^{solv}} \right) \tag{1.39}$$

What I call generalized Born approximation here, is the approximation of R_{ij}^{solv} by a function of b_i and b_j and the real distance between q_i and q_j, $|\vec{r}_{ij}|$. To derive such a function, we will carry out a *Gedankenexperiment.* Imagine a charge q_k with the effective Born radius b_k. This charge has a self energy of

$$\Delta G_k^{self} = -\frac{q_k^2}{2\Delta\varepsilon b_k} \tag{1.40}$$

Now we divide the charge q_k into two smaller ones, q_i and q_j, with effective Born radii b_i and b_j, and separate the two new charges by an infinitesimal small distance. The self energy of the two new charges plus their interaction energy should be the same as the self energy of the one old charge before:

$$-\frac{q_k^2}{2\Delta\varepsilon b_k} = -\frac{q_i^2}{2\Delta\varepsilon b_i} - \frac{q_j^2}{2\Delta\varepsilon b_j} - \frac{q_i q_j}{\Delta\varepsilon R_{ij}^{solv}} \tag{1.41}$$

The effective Born radii b_i, b_j, and b_k can be assumed to be equal since the effective Born radius mainly accounts for the geometry of the solute and all three charges are only infinitesimal separated.

If we set $R_{ij}^{solv} = b_i = b_j = b_k$, eq 1.41 is fulfilled since $q_k = q_i + q_j$. Thus, R_{ij}^{solv} approaches the effective Born radius for small distances between q_i and q_j.

In the next step of the *Gedankenexperiment*, we separate the two charges by a large distance compared to the size of the solute. ΔG_{ij}^{int} will then be mainly determined by the simple Coulomb energy. This means that R_{ij}^{solv} approaches the real distance $|\vec{r}_{ij}|$.

A suitable function to connect these two extreme cases was proposed by Still *et al.* (1990):

$$R_{ij}^{solv} = \sqrt{|\vec{r}_{ij}|^2 + b_i b_j \exp(-|\vec{r}_{ij}|^2/4b_i b_j)} \tag{1.42}$$

This function approaches $|\vec{r}_{ij}|$ if the distance $|\vec{r}_{ij}|$ is large compared to the effective Born radii, and approaches the geometric mean of the two Born radii b_i and b_j for a small distance $|\vec{r}_{ij}|$. The function is not only in agreement with the results of our *Gedankenexperiment*, but is also based on some other considerations (Still *et al.*, 1990), which I have not mentioned here.

The expression for ΔG_E can now be written with only one sum over all possible pairs ij (including pairs where $i = j$). To account for the double counting of interaction energies, a factor of $1/2$ has to be included for the interaction energies.

$$\Delta G_E = -\frac{1}{2\Delta\varepsilon} \sum_{ij} \frac{q_i q_j}{\sqrt{|\vec{r}_{ij}|^2 + b_i b_j \exp(-|\vec{r}_{ij}|^2/4b_i b_j)}} \tag{1.43}$$

However, to really calculate ΔG_E, a way to determine the self energy ΔG_i^{self} has still to be found.

1.2.4 Pairwise self energy potential

To evaluate the integral in eq 1.35, a molecular density function $P_S(\vec{r})$ is introduced:

$$P_S(\vec{r}) = \begin{cases} 1 & : \quad \vec{r} \text{ is inside solute volume} \\ 0 & : \quad \text{otherwise} \end{cases} \tag{1.44}$$

Eq 1.35 can now be written with the integral over all space:

$$\Delta G_i^{self} = -\frac{q_i^2}{2\Delta\varepsilon R_i} + \frac{1}{8\pi\Delta\varepsilon} \int_V \vec{D}_i^2(\vec{r}) P_S(\vec{r}) d\vec{r} \tag{1.45}$$

The molecular density function $P_S(\vec{r})$ is now splitted into a sum of atomic density functions. For each atom k, there is one atomic density function $P_k(\vec{r})$ describing the volume distribution of the atom.

$$P_S(\vec{r}) = \sum_k P_k(\vec{r}) \tag{1.46}$$

The contribution of atom k to the self energy of atom i, G_{ik}^{self}, is given by

$$G_{ik}^{self} = \frac{1}{8\pi\Delta\varepsilon} \int_V \vec{D}_i^2(\vec{r}) P_k(\vec{r}) d\vec{r} \tag{1.47}$$

The sum of G_{ik}^{self} for all atoms k leads to the second term in eq 1.45, so that this equation can be rewritten as

$$\Delta G_i^{self} = -\frac{q_i^2}{2\Delta\varepsilon R_i} + \sum_{k \neq i} G_{ik}^{self} \tag{1.48}$$

The restriction $k \neq i$ is in principle not necessary since it does not change the exact result. However, due to approximations applied later, the explicit exclusion improves the approximated result. For details see Schaefer and Karplus (1996).

The contribution G_{ik}^{self} to the self energy is always positive. Since the dielectric displacement $\vec{D}_i$ generally decreases with increasing distance from the charge i generating the displacement, G_{ik}^{self} represents an effective repulsion between atoms i and k. This repulsion is caused by the contribution of atom k to the solute volume, preventing atom i from interacting with the solvent directly.

1.2.5 Atomic charge and volume representation

In practice, it is not easy to find a suitable set of atomic density functions $P_k(\vec{r})$. Simply representing atomic volumes by spheres, leads to overlaps of bonded atom pairs, but at the same time there are still cavities between closely neighbored non-bonded atoms. Doing it this way, the solute dielectric medium is very discontinuous. To flatten these discontinuities, both the charge distribution of atom i, $\rho_i(\vec{r})$ and the atomic density function of atom k, $P_k(\vec{r})$, are represented by three-dimensional Gaussians:

$$\rho_i(\vec{r}) = q_i \pi^{-3/2} \hat{R}_i^{-3} \exp - \frac{(\vec{r}-\vec{r}_i)^2}{\hat{R}_i^2} \quad ; \qquad \hat{R}_i = \frac{R_i}{\sqrt{\pi/2}} \tag{1.49}$$

$$P_k(\vec{r}) = \frac{4}{3\sqrt{\pi}\alpha^3} \exp - \frac{(\vec{r}-\vec{r}_i)^2}{(\alpha \tilde{R}_k)^2} \tag{1.50}$$

R_i is, as before, the van-der-Waals radius of atom i. The width parameter $\hat{R}_i$ of the charge distribution is chosen such that the Born self energy of the Gaussian-distributed charge is the same as if the charge was equally distributed on a sphere of radius R_i. The width parameter $\tilde{R}_k$ of the volume distribution is the effective atom radius derived from the average solvent-inaccessible volume contribution $\tilde{V}_k$ of the different atom types in protein structures from the PDB. The Gaussians are normalized such that

$$q_i = \int_V \rho_i(\vec{r}) d\vec{r} \tag{1.51}$$

and

$$\tilde{V}_k = \frac{4\pi \tilde{R}_k^3}{3} = \int_V P_k(\vec{r}) d\vec{r} \tag{1.52}$$

The volume $\tilde{V}_k$ does not depend on the smoothing parameter α, which is introduced to control the width of the individual atomic volume functions $P_k(\vec{r})$. With a value of $\alpha = 1.0$, the fluctuations of the volume density are still too strong to yield reasonable results. A value for α in the range from 1.2 to 1.8 leads to a sufficiently smooth volume density but does not yet greatly alter the contour of the protein as a whole.

1.2.6 Evaluation of the self energy

To solve the integral in eq 1.47, again the dielectric displacement $\vec{D}_i$ is approximated by the Coulomb field, which means neglecting the reaction field contribution to the energy density of the electric field. For details and discussion of the introduced error, see Schaefer and Karplus (1996). In a homogeneous dielectric with dielectric constant ε, the electrostatic potential ϕ_i of a Gaussian charge distribution $\rho_i(\vec{r})$, see eq 1.49, is given by

$$\phi_i(\vec{r}) = q_i \frac{\operatorname{erf}\left(|\vec{r}-\vec{r}_i|/\hat{R}_i\right)}{\varepsilon|\vec{r}-\vec{r}_i|} \tag{1.53}$$

where erf denotes the error function:

$$\mathrm{erf}(z) = \frac{2}{\sqrt{\pi}} \int\limits_{0}^{z} e^{-t^2}\, dt \tag{1.54}$$

Based on the Coulomb field approximation, the dielectric displacement of the Gaussian charge distribution is

$$\vec{D}_i(\vec{r}) = -q_i \vec{\nabla} \frac{\mathrm{erf}\left(|\vec{r}-\vec{r}_i|/\hat{R}_i\right)}{|\vec{r}-\vec{r}_i|} \tag{1.55}$$

Insertion of eq 1.55 and eq 1.50 into eq 1.47 leads to the following expression for the energy G_{ik}^{self} that atom k contributes to the self energy of atom i:

$$G_{ik}^{self} = \frac{q_i^2}{6\pi^{3/2}\Delta\varepsilon\alpha^3} \int\limits_{V} \left(\vec{\nabla} \frac{\mathrm{erf}\left(|\vec{r}-\vec{r}_i|/\hat{R}_i\right)}{|\vec{r}-\vec{r}_i|} \right)^2 \exp\left(-\frac{(\vec{r}-\vec{r}_k)^2}{\left(\alpha\tilde{R}_k\right)^2} \right) d\vec{r} \tag{1.56}$$

This integral is not analytically solvable. However, a suitable approximation can be used if G_{ik}^{self} is finite and monotonically decreasing with the distance $|\vec{r}_{ik}|$. These assumptions fail in the limit $R_i \to 0$ and if the ratio $R_i/(\alpha\tilde{R}_k)$ is larger than a critical value that is close to unity. By numerically integrating eq 1.56, it turns out that at short range ($|\vec{r}_{ik}| < \alpha\tilde{R}_k$), G_{ik}^{self} decreases like a Gaussian, whereas it approaches $q_i^2\tilde{V}_k/(8\pi\Delta\varepsilon|\vec{r}_{ik}|^4)$ at long range. This behavior leads to the following *Ansatz*:

$$G_{ik}^{self} = \frac{q_i}{\Delta\varepsilon\omega_{ik}} \exp\left(-\frac{|\vec{r}_{ik}|^2}{\sigma_{ik}^2} \right) + \frac{q_i^2\tilde{V}_k}{8\pi\Delta\varepsilon} \left(\frac{|\vec{r}_{ik}|^3}{|\vec{r}_{ik}|^4 + \mu_{ik}^4} \right)^4 \tag{1.57}$$

The parameters ω_{ik} and σ_{ik} determine the height and width of the Gaussian that approximates G_{ik}^{self} in the short-range domain. The first term in eq 1.57 becomes negligible for large $|\vec{r}_{ik}|$, and the second term vanishes at $|\vec{r}_{ik}| = 0$ due to the introduction of the parameter μ_{ik}. The parameters in eq 1.57 are determined by calculating analytically the exact value for G_{ik}^{self} and the second derivative $\partial^2 G_{ik}^{self}/\partial|\vec{r}_{ik}|^2$ at $|\vec{r}_{ik}| = 0$ using eq 1.56. This is possible since $|\vec{r}_{ik}| = 0$ means that $\vec{r}_i = \vec{r}_k$. At $|\vec{r}_{ik}| = 0$, the second term of eq 1.57 as well as its second derivative vanishes, so that the parameters ω_{ik} and σ_{ik} can be determined as follows:

$$\omega_{ik} = \frac{3\pi\alpha_{ik}^4\tilde{R}_k}{4(Q_{ik} - \arctan Q_{ik})} \tag{1.58}$$

$$\sigma_{ik} = \sqrt{\frac{3\alpha_{ik}^2\tilde{R}_k^2(Q_{ik} - \arctan Q_{ik})}{(3+f_{ik})Q_{ik} - 4\arctan Q_{ik}}} \tag{1.59}$$

with

$$Q_{ik} = \frac{q_{ik}^2}{\sqrt{2q_{ik}^2+1}}; \qquad f_{ik} = \frac{2}{q_{ik}^2+1} - \frac{1}{2q_{ik}^2+1}; \qquad q_{ik}^2 = \frac{\pi}{2}\left(\frac{\alpha_{ik}\tilde{R}_k}{R_i}\right)^2; \qquad \alpha_{ik} = \max\left(\alpha, \frac{R_i}{\tilde{R}_k}\right) \tag{1.60}$$

The long-range parameter μ_{ik} is determined by making sure that the value of the Born energy term of a charge q_i in a homogenous dielectric medium with dielectric constant ε_p is maintained. This yields

$$\mu_{ik} = \frac{77\pi\sqrt{2}R_i}{512\left(1 - \frac{2\pi^{3/2}\sigma_{ik}^3 R_i}{\omega_{ik}\tilde{V}_k}\right)} \tag{1.61}$$

1.2.7 Non-polar solvation term

To extend ACE to ACS, a term for the non-polar parts of the solvation energy G^{np} is added to the electrostatic solvation energy:

$$G^{np} = \sum_i G_i^{np} = 4\pi\sigma \sum_i (R_i + R_s)^2 \frac{R_i}{b_i} \tag{1.62}$$

R_s is the radius of a water probe sphere. The empirical solvation parameter σ has the dimension $\mathrm{J}/(\mathrm{mol}\,\mathring{\mathrm{A}}^2)$, which appears in conventional solvent-accessible surface approximations of the hydrophobic effect. It is typically set to values of about 10 to $12\,\mathrm{J}/(\mathrm{mol}\,\mathring{\mathrm{A}}^2)$. All other parameters in eq 1.62 are known from the electrostatics part, so that the calculation is straight forward.

ACS was introduced by Schaefer *et al.* (1998). In their paper, they applied also some other minor improvements of the basic ACE potential, which I will not describe here. Unfortunately, the paper itself (Schaefer *et al.*, 1998) does not give a lot of details, either. The publication of a more extensive description is in preparation.