

## Chapter 2 Method

Protonation reactions comprise the energetics of loss or addition of a proton, whereas redox reactions describe the loss or addition of an electron. The equilibria of *protonation and redox reactions* are governed by the  $pK_a$  and the  $E_{redox}^0$  values of participating molecules, respectively. The derivation of  $pK_a$  and  $E_{redox}^0$  values from corresponding equilibria is documented in section 2.1.1 and 2.1.2.

In the present study  $pK_a$  and  $E_{redox}^0$  values were computed from an indirect rather than from a direct evaluation of the reaction free energy ( $\Delta G_R$ ). The underlying basis to obtain  $\Delta G_R$  in that way is provided by two *thermodynamic cycles* (see Scheme 2.1 and 2.1). Along these cycles educts were transferred from the condensed phase into the gas phase ( $\Delta G_{solv}$  (educts)). There, either of the two considered chemical reactions were triggered ( $\Delta G_{gas}$ ). Finally, the products were released into the solution ( $\Delta G_{solv}$  (products)). Thus, thermodynamic cycles split the overall reaction into three steps. Chapter 2.1.1 and 2.1.2 explain thermodynamic cycles for protonation and redox reactions, respectively and provide equations that define the quantitative relationship between  $\Delta G_R$  and  $\Delta G_{solv}$  (products/educts) and  $\Delta G_{gas}$ .

Neither the solvation free energy of the proton  $\Delta G_{solv}(H^+)$  nor the potential of the normal hydrogen electrode, which contribute to computed  $pK_a$  and  $E^0$  values, respectively, were estimated in the scope of the present study. Due to inconsistent experimental data I<sup>[40, 42, 43, 45, 46, 49, 54, 84, 85, 87-90]</sup> used a proton solvation free energy, within the experimental error range, which then minimized the root-mean-square value between experimental and computed  $pK_a$  values. Section 2.5 conveys the framework of two experimental<sup>[84, 85]</sup> and one theoretical studies<sup>[89]</sup> to determine  $\Delta G_{solv}(H^+)$  in detail. One-electron reduction potentials were estimated using the recommended literature value of the standard (normal) hydrogen electrode. The derivation of this potential by Reiss and Heller is explained in chapter 2.6.

Throughout the present work gas phase free energies were determined by means of *ab initio quantum chemical (QC)* methods. The employment of QC methods required the breakage of  $\Delta G_{gas}$  into the electronic energy, the vibrational energy and the thermal correction. These energy terms were estimated in so called *single point* computations on optimized geometries. Proton affinity (PA) and electron affinity (EA) describe the enthalpic contributions to  $\Delta G_{gas}$  of a protonation and a redox reaction, respectively. PA and EA values were computed because they can be directly compared to experimental results. The definitions of PA and EA are given in chapter 2.1.1 and 2.1.2, respectively. Equations to convert the enthalpic values into reaction free energies are also provided therein.

PA values for closed shell systems were estimated by means of QC computations with DFT functionals. Molecules considered for one- electron reduction potentials commute between an open shell radical and a closed shell non-radical state. I tested the performance of the DFT functional B3LYP and of the post Hartree-Fock method G3MP2 to yield accurate EA values for organic compounds. Chemical reactions, with the concomitant occurrence of proton and electron transfer were also modeled by means of the G3MP2 method. Chapter 2.2 conveys the basics of DFT and G3MP2. Therein, a showcase for the importance of exchange-correlation expressions within a quantum chemical method is given for the DFT approach.

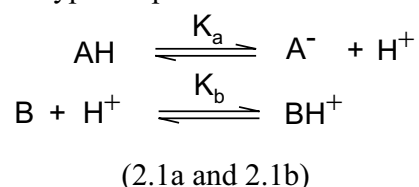
To obtain  $\text{pK}_a$  and  $E_{redox}^0$  values gas phase free energies were to be combined with solvation free energies. In the present study solvation free energies were estimated in a *two-step* computational approach. First, atomic partial charges were determined in the vacuum by the QC method B3LYP/6-31G\*\*. Second, solvation free energies were determined solving the Poisson-Boltzmann equation with these atomic partial charges. This procedure covers only the electrostatic part instead of the total solvation free energy. It will be shown in section 2.3 that in cases where the continuum is simulated with ionic strength equal to 0, this equation is reduced to the much simpler Poisson equation (see chapter 2.4). Using the above procedure to estimate electrostatic solvation free energies requires three sets of molecule dependent parameters to be determined: optimized atomic coordinates, partial atomic charges and vdW radii of solvation. Chapter 2.3 introduces methods to derive partial atomic charges from quantum mechanical computations. Chapter 2.4 documents the derivation of the Poisson and the Poisson-Boltzmann equation. The chosen set of vdW radii is documented in the result part in section 2.7.

## 2.1 Reaction free energies

### 2.1.1 Computation of acid-base equilibria

A titratable molecule or group is defined to have a certain potential to donate a proton into or uptake a proton from the surrounding environment. The protonation state of a titratable group in solution or within an ensemble of other titratable groups is quantitatively described by the  $pK_a$  value. The relationship between the  $pK_a$  value and the reaction free energy is derived in the following section. Subsequently I will explain how  $\Delta G_R^\circ$  can be expressed through computable terms.

The equations below illustrate two types of protonation reactions:



From the above equations the equilibrium constant  $K_a$  and  $K_b$  are derived by the law of mass action:

$$K_a = \frac{[\text{A}^-][\text{H}^+]}{[\text{AH}]} \quad K_b = \frac{[\text{B}][\text{H}^+]}{[\text{BH}^+]}$$

(2.2.a and 2.2b)

The pH value is defined as the negative decadic logarithms of the proton concentration in solution :

$$\text{pH} = -\log [\text{H}^+] \quad (2.3)$$

The  $pK_a$  of an acid is defined as the negative decadic logarithm of the equilibrium constant  $K_a$

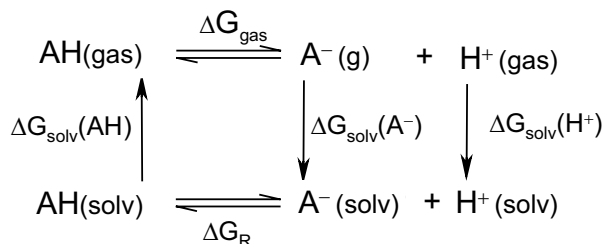
$$pK_a = -\log K_a \quad (2.4)$$

Taking the equilibrium of eq 2.2a and using the definitions of eqs. 2.3 and 2.4 we obtain the Henderson-Hasselbalch equation:

$$pH = pK_a + \log \frac{[\text{A}^-]}{[\text{AH}]} \quad (2.5)$$

Eq. 2.6 (see below) links the standard reaction free energy  $\Delta G_R^\circ$  of an acid-base equilibrium with the  $pK_a$  value, where R is the gas constant, T the absolute temperature in Kelvin and 2.303 converts the decadic logarithm into the natural logarithm.

$$pK_a = -\log(K_a) = \frac{\Delta G_R^\circ}{2.303RT} \quad (2.6)$$



**Scheme 2.1.** Thermodynamic cycle connecting gas (g) and solvent (solv) phase for the computation of absolute  $\text{pK}_{\text{a}}$  values of the titratable group AH.

The above thermodynamic cycle (scheme 2.1) illustrates the derivation of  $\Delta G_{\text{R}}$  in a protonation/deprotonation reaction from three components: the solvation free energy of educts ( $\Delta G_{\text{solv}}(\text{AH})$ ), the reaction free energy in the vacuum ( $\Delta G_{\text{gas}}$ ) and the solvation free energy of products ( $\Delta G_{\text{solv}}(\text{AH})$ ). If we follow the thermodynamic cycle in the indicated manner (see scheme 2.1) we can derive the following expression for  $\Delta G_{\text{R}}$ .

$$\Delta G_{\text{R}} = \Delta G_{\text{gas}} + \Delta G_{\text{solv}}(\text{A}^{-}) + \Delta G_{\text{solv}}(\text{H}^{+}) - \Delta G_{\text{solv}}(\text{AH}) \quad (2.7)$$

$\Delta G_{\text{gas}}$  can be expressed from three components as displayed in the next equation:

$$\Delta G_{\text{gas}} = G_{\text{gas}}(\text{A}^{-}) + \Delta G_{\text{gas}}(\text{H}^{+}) - G_{\text{gas}}(\text{AH}) \quad (2.8)$$

$G_{\text{gas}}(\text{AH})$  and  $G_{\text{gas}}(\text{A}^{-})$  in equation (2.2.8) are now expressed as a sum of three computable contributions.

$$\Delta G_{\text{gas}}^{298\text{K}} = E_0 + \text{ZPVE} + \text{EVE}^{298\text{K}} = E_0 + G_{\text{vib}}^{298\text{K}} \quad (2.9)$$

To obtain accurate  $\text{pK}_{\text{a}}$  values I computed for both protonation states in the gas phase the zero point electronic energy  $E_0$ , the zero point vibrational energy ZPVE and the excess vibrational free energy  $\text{EVE}^{298\text{K}}$  at 298 K. Modeling  $\text{pK}_{\text{a}}$  values along the thermodynamic cycle given in the above scheme requires the accurate evaluation of  $\Delta G_{\text{gas}}$  and  $\Delta \Delta G_{\text{solv}}$ . To establish a reliable computational protocol to predict  $\text{pK}_{\text{a}}$  values for a broad range of acid-base equilibria we analyzed the performance of different QC methods to model accurately gas phase energetics. Therefore we compared for a number of compounds (see 3.1 and 3.3) experimental and computed data. Experimental results in vacuum are expressed in proton affinities (PA), which are the enthalpic contributions to the reaction free energies. The PA is defined as:

$$\text{PA} \equiv G_{\text{gas}}(\text{A}^{-}) + E_{\text{gas}}(\text{H}^{+}) - G_{\text{gas}}(\text{AH}) \quad (2.10)$$

I used for  $E_{\text{gas}}(\text{H}^{+}) = 2.5 \text{ RT}$  (with  $\text{R} \cdot 298 \text{ K} = 1.48 \text{ kcal/mol}$ ).  $1.5 \text{ RT}$  is the enthalpy contribution of the three translational degrees of freedom of  $\text{H}^{+}$  and  $1 \text{ RT} = \text{pV}$  is the excess mechanical energy required to generate two particles ( $\text{A}^{-}$ ,  $\text{H}^{+}$ ) out of one (AH) as estimated from the ideal gas model. To obtain the gas phase free energy  $\Delta G_{\text{gas}}$  entering equation (2.8), which is employed to calculate absolute  $\text{pK}_{\text{a}}$  values, the entropy contribution  $-\text{T} \Delta \text{S}(\text{H}^{+})$  of the free proton estimated from the ideal gas model must be added to the enthalpy term  $E_{\text{gas}}(\text{H}^{+})$  in equation (3). Hence, we have

$$\Delta G_{gas}(H^+) = E_{gas}(H^+) - 298K\Delta S(H^+) = 2.5RT298K - 7.76kcal * mol^{-1} = -6.28kcal * mol^{-1} \quad (2.11)$$

Note that in some publications the measured PA may refer to free energies. According to Eq. (2.11) and (2.10) the corresponding PA value is than lower by 7.76 kcal/mol. In the present study PA values were exclusively used in terms of enthalpies as defined by Eq. (2.10).

$\Delta G_{solv}(A^-)$  and  $\Delta G_{solv}(AH)$  can be summarized into  $\Delta\Delta G_{solv}$ , as shown by the following two equations:

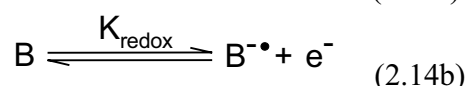
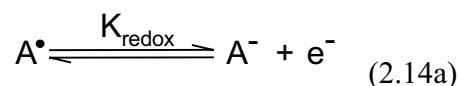
$$\Delta G_R = \Delta G_{gas} + \Delta\Delta G_{solv}(AH^*) + \Delta G_{solv}(H^+) \quad (2.12)$$

$$\text{with } \Delta\Delta G_{solv}(AH^*) = \Delta G_{solv}(A^\#) - \Delta G_{solv}(AH^*) \quad (2.13)$$

Note that for titratable compounds that are neutral (carboxylic acids, benzoic acids, imides, phenols) we have  $AH^* \equiv AH$  and  $A^\# \equiv A^-$ . For compounds that are cationic (imidazoles and pyridines) in the protonated state we have  $AH^* \equiv AH^+$  and  $A^\# \equiv A$ . Using the above definitions and equations from 2.6 to 2.13 the  $pK_a$  value of an organic compound can now be calculated.

### 2.1.2 Computation of redox equilibria

In analogy to a titratable group one defines a redox-active molecule or group by the ability to uptake an electron from or deliver an electron into the surrounding environment. The fundamental quantity to describe the redox behavior of a molecule is the standard redox potential  $E_{redox}^0$ . In the current thesis one-electron reduction of molecules between a neutral and an anionic state were estimated. In redox-reactions molecules commute between a closed shell (non-radical) and an open shell (radical) state. The current study provides examples with neutral radical states (phenoxides) and anionic radical states (quinones) (see section 3.2). The two cases are illustrated by the subsequent equations:



From the above equations the equilibrium constant  $K_a$  and  $K_b$  for a redox reaction can be derived by the law of mass action:

$$K_a = \frac{[A^-]}{[A^\bullet] [e^-]} \quad K_b = \frac{[B^{\bullet-}]}{[B] [e^-]} \quad (2.15a \text{ and } 2.15b)$$

In analogy to the definition of the pH value and the  $pK_a$  value one defines the solution redox potential  $E_{solv}$  and the standard redox potential  $E_{redox}^0$  of the redox couple  $A_{oxd}/A_{red}$ , as:

$$E_{solv} = -\frac{RT}{F} \ln [e^-] \quad (2.16)$$

$$E_{redox}^0 = \frac{RT}{F} \ln K_a \quad (2.17)$$

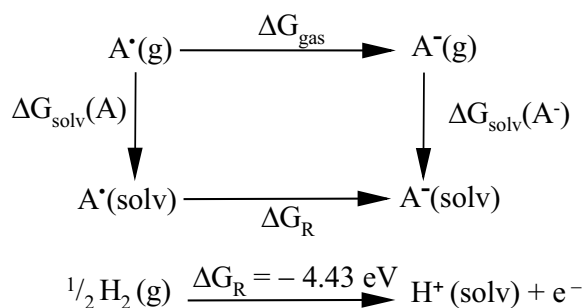
Using the above two equations expression we now derive the Nernst law, which is the analog to the Henderson-Hasselbalch equation (see eq. 2.5) for titratable groups:

$$E_{sol} = E^{\circ} + \frac{RT}{F} \ln \left[ \frac{A^{\bullet}}{A^{-}} \right] \quad (2.18)$$

$E_{redox}^{\circ}$  is related to the standard reaction free energy of a one-electron reduction potential via the following equation:

$$\Delta G_R^{\circ} = -FE_{redox}^{\circ} \quad (2.19)$$

$\Delta G_{redox}(\text{solv})$  represents the reaction free energy of the electron transfer reaction  $A + e^{-} \leftrightarrow A^{-}$  in solution.  $F$  is the Faraday constant, which is equal to  $23.06 \text{ kcal mol}^{-1} \text{ V}^{-1}$  and  $E_{redox}^{\circ}$  the redox potential under standard conditions.



**Scheme 2.2:** Thermodynamic cycle connecting gas (g) and solvent (solv) phase for the computation of redox potentials in solution of the redox-active group A.

To calculate  $\Delta G_{redox}(\text{solv})$  and  $E^{\circ}$  the above thermodynamic cycle was employed. It shows the transfer of a molecule of interest from the oxidized state into the reduced state via the vacuum, so that  $\Delta G_R$  can be formulated as a sum of three components:

$$\Delta G_{redox} = G(A^{-}) - G(A) - \Delta G_{\text{NHE}} \quad (2.20)$$

Substitution of eq. 2.20 into eq. 2.19 leads then to a modified expression for the standard redox potential:

$$FE_{redox}^{\circ} = \Delta G_{\text{gas}} - \Delta \Delta G_{\text{solv}} + \Delta G_{\text{NHE}} \quad (2.21)$$

$\Delta \Delta G_{\text{solv}}$  is the free energy difference in the solvation free energy between the reduced and oxidized state and  $\Delta G_{\text{NHE}}$  the potential of the normal hydrogen electrode (NHE). As mentioned in the introduction  $E_{redox}^{\circ}$  values are relative values that are referred to as the potential of the standard hydrogen electrode. In the present work all computed  $E_{redox}^{\circ}$  values are given relative to the potential of the standard hydrogen electrode. A potential of  $-4.43 \text{ eV}$  has been determined experimentally for the NHE. This value is commonly used in connection with  $E_{redox}^{\circ}$  values. To compare  $E^{\circ}$  values vs. NHE to data measured against the standard calomel electrode (SCE)  $0.241 \text{ eV}$  have to be subtracted. The NHE donates the electrons necessary to transfer a substance from the oxidized state to the reduced state. The solvation energy of the electron equals zero and is therefore not included in eq. 2.21.  $\Delta G_{\text{gas}}$  and  $\Delta \Delta G_{\text{solv}}$  in the above equation are expressed as indicated by the two subsequent equations:

$$\Delta G_{\text{gas}}(A^{-}/A) = \Delta G_{\text{gas}}(A^{-}) - \Delta G_{\text{gas}}(A) \quad (2.22)$$

$$\Delta\Delta G_{\text{solv}}(\text{A}^-/\text{A}) = \Delta G_{\text{solv}}(\text{A}^-) - \Delta G_{\text{solv}}(\text{A}) \quad (2.23)$$

Modeling  $E_{\text{redox}}^0$  values via the thermodynamic cycle given above requires an accurate evaluation of  $\Delta G_{\text{gas}}$  and  $\Delta\Delta G_{\text{solv}}$ . To establish a reliable computational protocol to predict  $E_{\text{redox}}^0$  values for yet unknown one-electron reduction potentials we used in the present applications QC methods that fit to a broad spectrum of experimental gas phase results. Experimental results in vacuum are expressed in EA values. A conventional definition of EA (adiabatic EA) is in terms of zero temperature enthalpic gas phase energies  $H_{\text{gas}}^{\text{T}=0\text{K}}$  between the two charge states  $\text{A}^-/\text{A}$  yielding.

$$\text{EA}(\text{A}^-/\text{A}) = -\Delta H_{\text{gas}}^{\text{T}=0\text{K}}(\text{A}^-/\text{A}) = H_{\text{gas}}^{\text{T}=0\text{K}}(\text{A}) - H_{\text{gas}}^{\text{T}=0\text{K}}(\text{A}^-). \quad (2.24)$$

The enthalpy is composed of electronic energy  $E_0$ , zero-point vibration energy ZPVE and for non-vanishing temperature also of translational rotational vibrational (TRV) contributions  $H_{\text{TRV}}^{\text{T}=298\text{K}}$  according to

$$H_{\text{gas}}^{\text{T}=298\text{K}} = E_0 + \text{ZPVE} + H_{\text{TRV}}^{\text{T}=298\text{K}} \quad (2.25)$$

Similarly the gas phase free energy needed to compute the redox potential is given as sum of electronic energy  $E_0$ , zero-point vibrational energy ZPVE and the temperature dependent TRV contributions to free energy according to

$$G_{\text{gas}}^{\text{T}=298\text{K}} = E_0 + G_{\text{non-electronic}}^{\text{T}=298\text{K}} = E_0 + \text{ZPVE} + G_{\text{TRV}}^{\text{T}=298\text{K}} \quad (2.26)$$

In agreement with other theoretical groups that evaluated one-electron reduction potentials and  $\text{pK}_a$  values I applied a thermodynamic cycle (scheme 2.2). High-level QC methods were used to estimate gas phase EA and  $\Delta G_{\text{gas}}$ . Using equations 2.19 to 2.26 the  $E_{\text{redox}}^0$  value of an organic compound can now be calculated.

## 2.2 Quantum chemical methods

As stated in the general introduction gas phase free energies were computed by means of suitable QC methods. QC methods were applied because the breaking and forming of a covalent bond (computation of  $pK_a$  values) and the release and uptake of an access electron (computation of one-electron reduction potentials) cannot be described by means of molecular mechanics.

The definitions of PA and EA values (see equation 2.10 and 2.24) comprise electronic and vibrational energy differences between different protonation or redox states of a considered molecule, which are the enthalpic contribution to  $\Delta G_{\text{gas}}(pK_a)$  and  $\Delta G_{\text{gas}}(\text{redox})$ , respectively. For the estimation of reaction free energies thermal correction are added as an additional term (see eqs. 2.9 and 2.26). The computation of the electronic energies is the demanding part and requires the usage of more elaborated QC methods than the estimation of vibrational energies or thermal corrections.

The current thesis documents accurately computed PA values for a spectrum of organic molecules based on the DFT functionals B3LYP<sup>[75, 77]</sup> functional, Becke-half&half (Becke<sup>1/2</sup>),<sup>[76]</sup> which is a mixture between pure Hartree-Fock and DFT and the post Hartree-Fock method Gaussian-3-Møller-Plesset -2 (G3MP2).<sup>[78]</sup> Even for modern computational quantum chemistry the quantitative description of EA values is especially demanding.<sup>[79, 80]</sup> The correlation energy of the unpaired electron in the radical state exceeds the correlation energy of the paired electrons. As a result, sophisticated QC methods are required to obtain a balanced description of both electronic states. Hence, major research attempts were carried out in the past to find improved expressions for the correlation energy. The complete failure of the DFT functional Becke<sup>(1/2)</sup> (see chapter 2.2.4) and the partial failure of B3LYP (see chapter 2.2.4) to reproduce quantitatively experimental EA values (see section 3.2) support the above statement. For the chosen set of compounds only G3MP2 (see section 2.2.5) was able to estimate EA values in the experimental error range (see section 3.2).

This section intends to convey the basics of quantum mechanics and quantum chemistry, which are necessary to understand general aspects and ideas (not mathematical details) in scientific studies that have the potential to lead to improved results on PA and EA values. It will be shown that difficulties in modern computational quantum chemistry have their origin in the structure of the Schrödinger equation, which is the base of quantum mechanics and quantum chemistry. The basic problem, which confronts quantum chemists is summarized in a statement by Szabo and Oslund: "Finding and describing approximate solutions to the electronic Schrödinger equation has been a major preoccupation of quantum chemists since the birth of quantum mechanics".<sup>[91]</sup> According to the previous statement only approximate solutions of the Schrödinger equation can be found for many electron systems by means of QC methods. Consequently a compromise between the demands of accurateness and the technical and computational feasibility has to be found.

The infeasibility of theoretical approaches to solve the Schrödinger equation directly for many-particle systems launched the development of theoretical approaches, which describe the characteristics of many-particle systems by the antisymmetrized product of single interacting particles (electrons). A major attempt of these theories is the search for expressions that describe the interaction energy between those particles. In a mean field approximation like Hartree Fock



(see chapter 2.2.2) the electron-electron interaction is described by an averaged potential i.e. one electron moves in an averaged potential created by the other electrons. The deviation in energy of a N-particle system from the energy obtained using a mean field approximation is termed correlation energy.<sup>[91]</sup>

The structure of the Schrödinger equation renders impossible straightforward QC solutions for many-electron systems. Most QC methods were tested for anorganic molecules only, thus a major attempt of this thesis was to do similar tests for organic molecules, which were considered here. The considered systems of the current thesis are small compared to macro molecular biological systems as for example heme. Hence, even CPU time and memory demanding QC methods were tested.

### 2.2.1 The basics to understand quantum chemical methods:

The time independent, non-relativistic Schrödinger equation for a many electron system can be written as:<sup>[91]</sup>

$$H|\Phi\rangle = E|\Phi\rangle \quad (2.27)$$

H is the Hamiltonian operator for a system of nuclei and electrons described by position vectors  $\vec{R}_A$  and  $\vec{r}_i$ , respectively.  $\Phi$  is the wave function of the many particle system and E is the energy of the eigenfunction. In SI units the Schrödinger equation of the hydrogen atom can be written as:

$$-\frac{\hbar^2}{2m} \nabla_r^2 \Phi(\vec{r}) - \frac{e^2}{4\pi\epsilon_0 r} \Phi(\vec{r}) = E \Phi(\vec{r}) \quad (2.28)$$

In eq. 2.28  $-\frac{\hbar^2}{2m} \nabla_r^2$  is the kinetic energy of the electron system,  $-\frac{e^2}{4\pi\epsilon_0 r}$  is the potential energy and E the energy of the system. In QC computations the Schrödinger equation is expressed in atomic units. In chapter 2.1.1 of “Modern Quantum Chemistry” Szabo and Oslund convert explicitly the Schrödinger of the hydrogen atom from SI units to atomic units. Expressed in atomic units equation 2.28 looks like:<sup>[91]</sup>

$$\left( -\frac{1}{2} \nabla_r^2 - \frac{1}{r} \right) \Phi(\vec{r}) = E \Phi(\vec{r}) \quad (2.29)$$

Equation 2.29 can be converted to equation 2.27. The Hamiltonian for N electrons and M nuclei of equation 2.27 can be expressed as:

$$H = -\sum_{i=1}^N \frac{1}{2} \nabla_i^2 - \sum_{A=1}^M \frac{1}{2M_A} \nabla_A^2 - \sum_{i=1}^N \sum_{A=1}^M \frac{Z_A}{r_{iA}} + \sum_{i=1}^N \sum_{j>i}^N \frac{1}{r_{ij}} + \sum_{A=1}^M \sum_{B>A}^M \frac{Z_A Z_B}{R_{AB}} \quad (2.30)$$

In eq. 2.30 the distance between the *i*-th electron and A-th nucleus is  $\vec{r}_{iA} = |\vec{r}_i - \vec{R}_{iA}|$ , the distance between the *i*-th and the *j*-th electron is  $\vec{r}_{ij} = |\vec{r}_i - \vec{r}_j|$  and the distance between the *A*-th and the *B*-th nucleus is  $\vec{R}_{ij} = |\vec{R}_i - \vec{R}_j|$ .  $M_A$  is the ratio of the mass of nucleus A to the mass of an electron and  $Z_a$  is the atomic number of nucleus A. The Laplacian operators  $\nabla_i^2$  and  $\nabla_A^2$  involve differentiation with respect to the coordinates of the *i*-th electron and *A*-th nucleus, respectively. The first term in the above equation is the operator for the kinetic energy of the electrons; the second is the operator for the kinetic energy of the nucleus; the third term represents the Coloumb

attraction between electrons and nuclei; the fourth and fifth term represents the repulsion between electrons and between nuclei, respectively.

The **Born-Oppenheimer approximation** considers electrons as moving particles in the field of fixed nuclei. Due to significant differences in mass between electron ( $m_{el} \sim 10^{-31}$  kg) and nucleus ( $m_{nuc} \sim 10^{-27}$  kg), the electronic degrees of freedom can be considered to respond instantaneously to any change in the nuclear configuration i.e their wave functions correspond always to a stationary state.<sup>[92, 93]</sup> Hence, the interaction between nuclei and electrons is modified due to the motion of the nuclei only adiabatically and does not cause transitions between different stationary states.<sup>[93]</sup> Application of the Born-Oppenheimer approximation enables one to formulate the electronic Hamiltonian. The time independent electronic Schrödinger equation describing the motion of electrons in the electrostatic field of the stationary nuclei comprises only term one, three and four of eq. 2.30, where the electronic Hamiltonian and the electronic wave functions depend only parametrically on the coordinates of the nuclei:<sup>[91]</sup>

$$H_{elec}(\bar{R}_{nuc}) \Phi_a(\bar{r}, \{\bar{R}_{nuc}\}) = E_{a,elec}(\bar{R}_{nuc}) \Phi_a(\bar{r}, \{\bar{R}_{nuc}\}) \quad (2.31)$$

with the electronic Hamiltonian:

$$H_{elec}(\bar{R}_{nuc}) = -\sum_{i=1}^N \frac{1}{2} \nabla_i^2 - \sum_{i=1}^N \sum_{A=1}^M \frac{Z_A}{r_{iA}} + \sum_{i=1}^N \sum_{j>i}^N \frac{1}{r_{ij}} \quad (2.32)$$

The three terms on the right side of eq. 2.32 were already introduced in eq. 2.30. The total energy for fixed nuclei includes also the mutual nuclear repulsion. Therefore,

$$E_{tot} = E_{elec} + \sum \sum \frac{Z_A Z_B}{R_{AB}} \quad (2.33)$$

Solutions to a nuclear Schrödinger equation describe the vibration, rotation and translation of a molecule. The nuclear Schrödinger equation describes the motion of the nuclei in the mean field of the electrons, which is obtained by averaging over the electronic wave functions:

$$H_{nuc}(\bar{R}_{elec}) \Phi_{nuc}(\bar{r}_{nuc}, \{\bar{R}_{elec}\}) = E_{nuc} \Phi_{nuc}(\bar{r}_{nuc}, \{\bar{R}_{elec}\}) \quad (2.34)$$

The nuclear Hamiltonian (see equation below) describes the kinetic potential of the nuclei ( $\sum_{A=1}^M \frac{1}{2M_A} \nabla_A^2$ ), the mutual Coulomb interaction of the nuclei ( $\sum_{A=1}^N \sum_{B>A}^M \frac{Z_A Z_B}{\bar{R}_{AB}}$ ) and the interaction of the nuclei with the electronic charge distribution corresponding to the actual configuration of the nuclei ( $E_{elec}(\{\bar{R}_A\})$ ):

$$H_{nuc}(\bar{R}_{elec}) = -\sum_{A=1}^M \frac{1}{2M_A} \nabla_A^2 + \sum_{A=1}^N \sum_{B>A}^M \frac{Z_A Z_B}{\bar{R}_{AB}} + E_{elec}(\{\bar{R}_A\}) \quad (2.35)$$

To cover the full energetics of a system the electronic and nuclear **eigenfunctions** have to be calculated. Hence in the current study on pK<sub>a</sub> and one-electron reduction potentials electronic energies were computed independently from vibrational energies and their results were summed to obtain reaction free energies in the vacuum.

**Antisymmetry Principle and Spin Orbitals.** The square of the wave function but not the function itself is a physical observable. For a two-electron system the quantity ( $|\Psi(\bar{x}_1, \bar{x}_2)|^2 d\bar{x}_1, d\bar{x}_2$ ) expresses the probability of finding electron 1 in volume  $\bar{x}_1, \bar{x}_1 + d\bar{x}_1$  and simultaneously electron 2 in volume  $\bar{x}_2, \bar{x}_2 + d\bar{x}_2$ . Of course the meaning of the square of the

wave function can be generalized to a many electron system, so that the probability of finding  $N$  electrons in the respective space elements between  $a$  and  $b$  is given as:

$$\int_{a_1}^{b_1} \int_{a_2}^{b_2} \cdots \int_{a_n}^{b_n} |\Psi(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n)|^2 d\bar{x}_1, d\bar{x}_2, \dots, d\bar{x}_n \quad (2.36)$$

To describe a many-electron system the wave function must not distinguish between different electrons, because they are indistinguishable particles. That means that if we consider a two-electron system than the probability of finding electron 1 or 2 at some point in space must be identical. Both of the following two equations fulfill this condition:

$$\Psi(x, y, z, x', y', z') = \Psi(x', y', z', x, y, z) \quad (2.37)$$

$$\Psi(x, y, z, x', y', z') = -\Psi(x', y', z', x, y, z) \quad (2.38)$$

In the first case we say that the wave function is symmetric with respect to an exchange of electrons. In the second case, we say the wave function is antisymmetric. Both symmetric and antisymmetric wave functions lead to probabilities ( $|\Psi|^2 d\bar{x}$ ) that are symmetrical. This means that the probabilities are unchanged when the electrons exchange their positions. **For electrons wave functions are antisymmetric.** This statement cannot be derived from any other law of nature and therefore it is itself a law of nature. This principle is also known as the **Pauli exclusion principle** or the **antisymmetry principle**. It holds for two-electron systems as well as for many-electron systems. Eq. 2.39 is the generalization of eq. 2.38: <sup>[91, 92, 94]</sup>

$$\Psi(\dots, x_k, y_k, z_k, \dots, x_m, y_m, z_m, \dots) = -\Psi(\dots, x_m, y_m, z_m, \dots, x_k, y_k, z_k, \dots) \quad (2.39)$$

The electronic Hamiltonian (see eq. 2.32) depends only on the spatial coordinates of the electron. To completely describe **the motion and the occupied phase space of the electron** the spin of the electron needs to be introduced. In the context of the non-relativistic Schrödinger equation this can be accomplished with the introduction of the two spin functions  $\alpha(\omega)$  and  $\beta(\omega)$ , which correspond to the spin up and spin down state, respectively, which are functions of the unspecified spin variable  $\omega$ . It only has to be ensured that two spin functions are complete and orthonormal with respect to each other. (For further information see also: Szabo and Oslund "Modern Quantum Chemistry" chapter 2.1.3.).<sup>[91]</sup>

After introducing the spin function the electron will depend on the three spatial coordinates and one spin variable  $\omega$ . The coordinate  $\bar{x}$  is used in the following as the abbreviation for three spatial coordinates and one spin degree of orbital one:

$$\bar{x} = \{\bar{r}, \omega\} \quad (2.40)$$

A **spin orbital** is defined as a wave function for a single electron. A spatial orbital depends only on three spatial coordinates and its square  $|\Psi_i(\bar{r})|^2$  describes the probability to find the electron in a small volume element  $\bar{r}, \bar{r} + d\bar{r}$  surrounding  $\bar{r}$ . Spin orbitals of single-particles include the description of the spin of the electron. Hence, spin orbitals depend on three coordinates and have either  $\alpha$  or  $\beta$  spin.

$$\chi(\bar{x}) = \begin{cases} \Psi(\bar{r}) \alpha(\omega) \\ \Psi(\bar{r}) \beta(\omega) \end{cases} \quad (2.41)$$

As the wave function of spatial orbitals the wave function of spin orbitals fulfills the Pauli antisymmetry principle:

$$\Phi(\bar{r}_1, \omega_1, \bar{r}_2, \omega_2, \bar{r}_3, \omega_3, \dots) = -\Phi(\bar{r}_1, \omega_1, \bar{r}_3, \omega_3, \bar{r}_2, \omega_2, \dots) \quad (2.42)$$

**Quantum Mechanics and Quantum Chemistry.** Even the reduced form of the Schrödinger equation, the electronic Schrödinger equation, cannot be solved for many-electron systems without further assumptions. Additional approximations make the Schrödinger equation applicable for many-electron systems, which is a prerequisite to perform quantum chemical calculations: Wave function based theories (based on the Hartree-Fock approach) and density functional theory (DFT) evaluate the properties of molecular systems from antisymmetrized products of  $N$  single electrons.<sup>[91, 94]</sup>

**In 1929 Slater** introduced a mathematical tool to obtain a correctly antisymmetrized wave function of a many-electron system from a antisymmetrized linear combination of the wave functions of  $N$  one-electron spin orbitals. This expression is referred to in the scientific literature as the Slater determinant. The Slater determinant for a two-electron system can be written as follows:

$$\Psi(\bar{x}_1, \bar{x}_2) = 2^{-1/2} \begin{vmatrix} \chi_1(\bar{x}_1) & \chi_2(\bar{x}_1) \\ \chi_1(\bar{x}_2) & \chi_2(\bar{x}_2) \end{vmatrix} \quad (2.43)$$

$2^{-1/2}$  is the normalization factor. The minus sign in  $\frac{1}{\sqrt{2}} [\chi_i(\bar{x}_1)\chi_j(\bar{x}_2) - \chi_i(\bar{x}_2)\chi_j(\bar{x}_1)]$  insures that the linear combination yields a  $\Psi(\bar{x}_1, \bar{x}_2)$  that is antisymmetric with respect to the interchange of the coordinates of electron one and two. For a system of  $N$ -electrons the normalization factor has the form  $(N!)^{-1/2}$ . A many-electron wave function can be obtained from the following Slater determinant.

$$\Psi(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_N) = (N!)^{-1/2} \begin{vmatrix} \chi_1(\bar{x}_1) & \chi_2(\bar{x}_1) & \dots & \chi_N(\bar{x}_1) \\ \chi_1(\bar{x}_2) & \chi_2(\bar{x}_2) & \dots & \chi_N(\bar{x}_2) \\ \vdots & \vdots & \dots & \vdots \\ \chi_1(\bar{x}_N) & \chi_2(\bar{x}_N) & \dots & \chi_N(\bar{x}_N) \end{vmatrix} \quad (2.44)$$

Interchanging the coordinates of any two electrons corresponds to an interchange of two rows of the Slater determinant, which changes the sign of the determinant. Hence the Slater determinant fulfills the requirements of the Pauli exclusion or antisymmetry principle. Many electron- wave functions can approximately be described from a many-electron Slater determinant.

### 2.2.2 Hartree-Fock theory:

Within *Hartree-Fock theory* a systematical mathematical approach is given to compute the lowest energy wave function. According to the **variation principle** the ground state wave function is the one with the lowest energy. The wave function obtained within the Hartree-Fock scheme is antisymmetric, because of the use of a Slater determinant. The following two equations are the basic equations in the Hartree-Fock theory:<sup>[91]</sup>

$$f(i) \chi(\bar{r}_i, \omega) = \varepsilon \chi(\bar{r}_i, \omega) \quad (2.45)$$

where  $f(i)$  is an effective one-electron operator the so-called Fock operator, which can be written as:

$$f(i) = -\frac{1}{2} \nabla_i^2 - \sum_{A=1}^M \frac{Z_A}{r_{iA}} + v^{HF}(i) \quad (2.46)$$

Comparison of the Fock operator (eq. 2.46) with the electronic Hamiltonian (eq. 2.32) reveals that the expressions for the kinetic energy and the electron-nuclear attraction are identical. The

electron-electron repulsion in the electronic Hamiltonian is replaced by an average potential  $v^{\text{HF}}(\mathbf{i})$ . The Hartree-Fock theory assumes that single electrons of an atomic or molecular system are moving in an averaged potential of other electrons.

A systematic variation of the spin orbitals reveals the wave function with the lowest energy. From the variation of the spin orbitals constraining them to be orthonormal the following integro-differential equation, with an explicit expression of the Fock operator is obtained:

$$\left( -\frac{1}{2}\nabla_1^2 - \sum_A \frac{Z_A}{r_{1A}} \right) \chi_a(1) + \sum_{b \neq a} \left[ \int |\chi_b(2)|^2 d\bar{x}_2 \bar{r}_{12}^{-1} \right] \chi_a(1) - \sum_{b \neq a} \left[ \int \chi_b^*(2) d\bar{x}_2 \chi_a(2) r^{-12} \right] \chi_b(1) = \varepsilon_a \chi_a(1) \quad (2.47)$$

The second term in the above equation is the Coulomb term, the third one is the exchange term. The Coulomb term represents the electrostatic repulsion between electron 1 in orbital ( $\chi_a$ ) and electron 2 in orbital ( $\chi_b$ ) and replaces the instantaneous potential between electron 1 and 2 by an average potential because of the integration over the coordinate of electron 2. This integration reduces the two-electron Coulomb potential to a one-electron potential. From eq. 2.47 the Coulomb operator is  $J_b(1)$  defined as:

$$J_b(1) = \int |\chi_b(\bar{r}_2, \omega_2)|^2 (d\bar{r}_2, d\omega_2) \bar{r}_{12}^{-1} \quad (2.48)$$

Since the operator  $J_b(1)$  depends solely on the coordinates of electron 1; it is a local operator, that describes the Coulomb coupling of the electron in orbital  $\chi_a$  with the remaining electrons. From the second term in eq. 2.47 the exchange operator  $K_b(1)$  is defined as:

$$K_b(1) \chi_a(\bar{r}_1, \omega) = \left[ \int \chi_b^*(\bar{r}_2, \omega_2) \bar{r}_{12}^{-1} \chi_a(\bar{r}_2, \omega_2) dx_2 \right] \chi_b(\bar{r}_1, \omega) \quad (2.49)$$

The exchange term in eq. 2.47 is a direct mathematical consequence of the antisymmetric nature of the Slater determinant. Unlike for the Coulomb term there is no obvious physical interpretation for the exchange term.  $K_b(2)$  exchanges the coordinates of the electrons in the two spin orbitals  $\chi_a$  and  $\chi_b$ . Unlike the local Coulomb operator the exchange operator is a nonlocal operator, because the operating of  $K_b(1)$  on  $\chi_a(\bar{r}_1, \omega)$  depends on the value of  $\chi_a$  not only at  $\bar{r}_1$  but at all other points in space  $r \bar{r}_2$ . The Hartree-Fock potential can now be written explicitly as:

$$v^{\text{HF}}(1) = \left( \sum_b (J_b(1) - K_b(1)) \right) \quad (2.50)$$

and the corresponding Hartree-Fock operator is then defined as:

$$f(1) = h(1) + v^{\text{HF}}(1) \quad (2.51)$$

where  $h(1)$  is a core-Hamiltonian operator:

$$h(1) = -\frac{1}{2}\nabla_1^2 + \sum_A \frac{Z_A}{r_{1A}} \quad (2.51a)$$

, which describes the kinetic potential (first term) and the electron-nuclei attraction (second term). The Hartree Fock potential  $v^{\text{HF}}(\mathbf{i})$  depends on the spin orbitals of the other electrons. Hence, the HF equation is nonlinear and must be solved iteratively. The procedure for solving the Hartree-Fock equation is called the self-consistent-field (SCF) method. The solution of the Hartree-Fock eigenvalue problem (see eq. 2.45) yields a set of orthonormal Hartree-Fock spin orbitals  $\{\chi_k\}$  with orbital energies  $\{\varepsilon_k\}$ . The  $N$  spin orbitals with lowest energies are called the occupied or hole spin orbitals, whereas the remaining orbitals obtained from the SCF calculation are called virtual or unoccupied orbitals. The Slater determinant formed from the occupied orbitals is the

HF ground state wave function and is the best variational approximation to the ground state of the system, of the single determinant form.

A single Slater determinant formed from the set  $\{\chi_\alpha\}$  of occupied orbitals is the variational HF ground state. The larger and more complete the set of basis functions  $\{\chi_\alpha\}$  the greater is the degree of flexibility in the expansion for the spin orbitals and the lower will be the energy of the **expectation value**  $E_0 = \langle \psi_0 | H | \psi_0 \rangle$ . Larger basis sets lower the HF energy  $E_0$  until a limit is reached, called the **Hartree-Fock limit**. In practice any finite value of the number of basis set functions  $K$  will lead to an energy above the Hartree-Fock limit. <sup>[91]</sup>

In a computational QC approach the HF equation for spin orbitals is transformed into an equation of spatial orbitals by performing suitable traces over the spin variables:

$$f(\vec{r}_i)\Psi(\vec{r}_i) = E_i\Psi(\vec{r}_i) \quad (2.52)$$

The Fock operator with its local and non-local operator, is also converted from spin to spatial orbitals. The complete mathematical procedure is shown in Szabo and Oslund : “Modern Quantum Chemistry”.<sup>[91]</sup> No mathematical procedure exists at the moment to solve the above equation exactly. **Roothaan** introduced a mathematical tool to convert the above integro-differential equation into a set of algebraic equations, which can be solved by standard matrix techniques. Thereby the unknown spatial orbitals  $\Psi_i(\vec{r}_i)$  are expanded into a set of  $K$  known basis function  $\{\phi_\mu\}$ :

$$\Psi_i(\vec{r}_i) = \sum_{\mu=1}^K C_{\mu i} \phi_\mu \quad i=1,2,\dots,K \quad (2.53)$$

To make the task computational feasible  $\{\Phi\}$  has to be incomplete, which implies that a finite expansion only approximates the “true” spatial orbitals. The quality of the solution depends on the completeness of the basis. With the above set of basis functions the HF equation for spatial orbitals can now be written as:

$$f(1) \sum_v C_{vi} \phi_v(1) = \varepsilon_i \sum_v C_{vi} \phi_v(1) \quad (2.54)$$

This integro-differential equation is converted into a matrix equation by multiplying the left side by  $\phi_\mu^*(1)$  and integrating over the electronic coordinates. In this way the overlap matrix **S** and the Fock matrix **F** are obtained:

$$S_{\mu\nu} = \int \phi_\mu^*(1) \phi_\nu(1) dr_1 \quad (2.55)$$

$$F_{\mu\nu} = \int \phi_\mu^*(1) f(1) \phi_\nu(1) dr_1 \quad (2.56)$$

Using the above definitions of the overlap matrix **S** and the Fock operator **F** the integrated form of the Hartree-Fock equation can be expressed as:

$$\sum_v F_{\mu\nu} C_{vi} = \varepsilon_i \sum_v S_{\mu\nu} C_{vi} \quad i = 1,2,\dots,K \quad (2.57)$$

Eq. 2.57 can be turned into a compact single matrix equation and solved using the laws of matrix calculations:

$$\mathbf{FC} = \mathbf{SC}\boldsymbol{\varepsilon} \quad (2.58)$$

In the Hartree-Fock scheme each electron is influenced by the attractive electron-nuclei and repulsive electron-electron potential. The one-electron spatial part of the spin orbital is an atomic orbital (AO). These AOs are described as hydrogen like orbitals for whom the assumption holds that the electron distribution is spherically symmetric around the nucleus. Consequently the potential that the electron  $i$  experiences in the field of other electrons is also spherically symmetric. The angular parts of the potential and of the AOs will be described by spherical harmonics. Molecular orbitals (MOs) are then obtained as a linear combination of the atomic orbitals (LCAO). Because of the antisymmetry principle each molecular orbital can hold no more than two electrons. Two different types of basis sets are currently in use in QC calculations. Slater type orbitals (STO) are exponential function with clear physical meaning, because they mimic the exact eigenfunction of the hydrogen atom. Slater type orbitals describe correctly the behavior in the proximity of the nuclei, which is known as the *cusp* behavior and also the exponential decay in the tail regions, when  $\bar{r}$  approaches infinity. The general form of Slater type orbitals is given in the following equation:

$$\phi^{STO} = N_{STO} r^{n-1} e^{(-\zeta r)} Y_{lm}(\Theta, \phi) \quad (2.59)$$

In eq. 2.59  $N_{STO}$  is the normalization factor for Slater type orbitals,  $n$  is the principal quantum number,  $\zeta$  is the orbital exponent and  $Y_{lm}$  is the spherical harmonic that describes the angular part of the orbital function. For the 1s hydrogen AO the spherical harmonic  $Y_{lm}(\Theta, \phi) = 1$ . Therefore, the exact 1s STO for the hydrogen atom centered in the origin of the coordinate system  $\mathbf{R}$  has the form:

$$\phi(\bar{r} - \bar{R}) = \left( \frac{\zeta^3}{\pi} \right)^{1/2} e^{-\zeta |\bar{r} - \bar{R}|} \quad (2.60)$$

Beside STOs Gaussian type orbitals (GTO) are frequently used in quantum chemical calculations. Using Roothaan's method to formulate the spatial Hartree-Fock equation we have to make a choice at the beginning of the computation about the applied basis set, which is used to express the molecular orbitals of the system. GTOs are used because they lead to simpler integral expressions than the STOs. The general form of a GTO has the following form:

$$\phi^{GTO} = N_{GTO} x^l y^m z^n e^{-\alpha r^2} \quad (2.61)$$

$N$  is the Normalization factor, which ensures that  $\langle \phi_\mu | \phi_\mu \rangle = 1$ .  $\alpha$  represents the orbital exponent, which determines how compact ( $\alpha$  large) or diffuse ( $\alpha$  small) the resulting function is.  $l = 1 + m + n$  is used to specify if the AO is a s-function ( $l = 0$ ), p-function ( $l = 1$ ) d-function ( $l = 3$ ) etc. For  $l > 1$  the number of cartesian functions of a GTO exceeds the number of  $(2l + 1)$  physical functions of angular momentum  $l$ . For example among the six cartesian functions with  $l = 2$ , one is spherically symmetric and is therefore not a d-type, but an s-function. Similarly the ten cartesian  $l = 3$  functions include an unwanted set of three p-type functions. GTOs have a slope in the proximity of the nuclei and do not cover the proper cusp behavior. When  $r$  approaches infinity GTOs decay too rapidly in contrast STOs. The Cartesian-Gaussian normalization constant is:

$$N_{GTO} = \left( \frac{2\alpha}{\pi} \right)^{3/4} \left[ \frac{(8\alpha)^{i+j+k} i! j! k!}{(2i)!(2j)!(2k)!} \right]^{1/2} \quad (2.62)$$

If we have  $i+j+k = 0$  then the GTO is a s-type orbital. If  $i+j+k = 1$  then the GTO is a p-type orbital. If we have  $i+j+k = 2$  then the GTO is a d-type orbital. If  $i+j+k = 0$  then the Cartesian-Gaussian normalization constant is  $(2\alpha/\pi)^{3/4}$  and therefore the 1s hydrogen atom expressed as a GTO has the following form:

$$\phi(\vec{r} - \vec{R}) = \left(\frac{2\alpha}{\pi}\right)^{3/4} e^{-\alpha|\vec{r} - \vec{R}|^2} \quad (2.63)$$

To circumvent the difficulties with single GTO's the AOs are expressed as contracted Gaussian functions (CGF):

$$\phi_{\tau}^{CGF} = \sum_a^A d_{a\tau} \phi_a^{GTO} \quad (2.64)$$

The contraction coefficients  $d_{a\tau}$  can be chosen in that way that a linear combination of CGFs approximate a single STO. The current practice is to take a linear combination between three and six so-called primitive Gaussians in a single CGF. Anions, compounds with *lone pairs*, have significant electron density at large distances from the nuclei. To improve the accuracy of quantum chemical computations, basis sets with diffuse functions are constructed. A highly diffuse function is one with very small orbital exponents (see eq. 2.59-2.64). The 6-31G basis set, which was used in the present thesis adds four highly diffuse functions (s,p<sub>x</sub>,p<sub>y</sub>,p<sub>z</sub>) on each non-hydrogen atom. To invoke the diffuse function in a quantum chemical computation the following notation is used: 6-31+G. [91, 92, 94]

### 2.2.3 Density functional theory

**The Hohenberg-Kohn theorem.** As stated in eq. 2.32 the electronic Hamiltonian depends on the spatial coordinates of the electrons and the coordinates of the stationary nuclei. To completely describe electrons their spin was introduced. Since the Hamiltonian operator contains only one- and two-electron spatial terms the molecular energy can be written in terms of integrals involving only six spatial coordinates. As stated in I. N. Levine "Quantum Chemistry" the wave functions contains more information than needed but also lacks direct physical significance. One approach to solve this problem led to the discovery of the Hartree-Fock theory. The other approach, which is also a direct outcome of this problem is the establishment of **Density Functional Theory**. The probability density is obtained from the following multiple integral over spin coordinates of all electrons and over all but one of the spatial coordinates. [92]

$$\rho(\vec{r}) = N \int \int |\Psi(\bar{x}_1, \bar{x}_2, \bar{x}_3, \dots, \bar{x}_n)|^2 d\omega_1 d\bar{x}_2, d\bar{x}_3, \dots, d\bar{x}_n \quad (2.65)$$

$\rho(\vec{r})$  determines the probability of finding any of the N-electrons within the volume element  $d\vec{r}_1$  but with arbitrary spin while the other N-1 electrons have arbitrary positions and spins according to the wave function. However since the electrons are indistinguishable the probability of finding any electron at this position is just N times the probability for one particular electron. Clearly  $\rho(\vec{r})$  is a non-negative function of only three spatial variables, which vanishes at infinity and integrates to the total number of electrons. [94]

$$\rho(\vec{r} \rightarrow \infty) = 0 \quad (2.66)$$

$$\int_V \rho(\vec{r}) d\vec{r}_1 = N \quad (2.67)$$



The Hohenberg-Kohn theorem says that the molecular energy, the wave function and all other molecular properties can be computed from the ground state electron probability density  $\rho_0(x,y,z)$ . The Hohenberg-Kohn theorem starts with a reformulation of equation 2.32: <sup>[92]</sup>

$$H_{elec} = - \sum_{i=1}^N \frac{1}{2} \nabla_i^2 + \sum_j \sum_{i>j}^N \frac{1}{r_{ij}} + \sum_{i=0}^n v(\vec{r}_i) \quad (2.68)$$

$v(\vec{r}_i)$  in eq. 2.68 is the so called external potential. The potential energy of  $v(\vec{r}_i)$  depends on the coordinates  $x_i, y_i, z_i$  of the electron and on the nuclear coordinates.

$$v(\vec{r}_i) = - \sum_A \frac{Z_A}{r_{iA}} \quad (2.69)$$

Eq. 2.68 exhibits the identical form than the Hartree-Fock equation (eq 2.32). Whereas in the Hartree-Fock theory the electron-electron repulsion was exchanged by an external, it is now the electron nuclei attraction that is replaced by the external field  $v(\vec{r}_i)$ . From the external potential  $v(\vec{r}_i)$  and the number of electrons in the considered systems molecular wave functions and molecular energies can be determined as solutions of the Schrödinger equation. Hohenberg and Kohn showed in their seminal work *that the external potential and the number of electrons are a function of the electron density  $\rho_0$  only. Hence the molecular wave function and molecular energies, which are computed from the external potential are also a function of the electron density.* <sup>[92]</sup>

The first Hohenberg-Kohn theorem is not sufficient to serve as a practical tool for the computation of ground state energies of molecular systems and chemical reactions. It was shown that the ground state electronic energy is a functional of the ground state electronic density  $\rho_0$ . The electronic Hamiltonian (see eq. 2.32) itself is a sum of three terms: the kinetic energy, the energy of electron nuclear attractions and the electron-electron repulsion energy. The ground state electronic energy is therefore the sum of the average of these three energy terms. The averaged energy terms can be expressed as a functional of the ground state electronic density  $\rho_0$ .

$$E_0 = E_v [\rho_0] = \bar{T} [\rho_0] + \bar{V}_{Ne} [\rho_0] + \bar{V}_{ee} [\rho_0] \quad (2.70)$$

The subscript  $v$  in eq. 2.70 indicates the dependence of the ground state energy of the external potential  $v(\vec{r}_i)$ . In eq. 2.70  $\rho_0$  is the electron density of the ground state,  $\bar{T}$  the averaged kinetic energy,  $\bar{V}_{Ne}$  the averaged electron-nuclear attraction energy and  $\bar{V}_{ee}$  the averaged electron-electron repulsion energy. The potential energy of the interaction between the nuclei and the electrons can be determined from the electron density  $\rho_0$  and the external potential  $v(\vec{r}_i)$ . But the energy contribution of the additional two energy functionals of eq. 2.70 remains unknown. Hence, one can write the ground state electronic energy  $E_0$  as:

$$E_0 = E_v [\rho_0] = \int \rho_0(\vec{r}) v(\vec{r}) d\vec{r} + \bar{T} [\rho_0] + \bar{V}_{ee} [\rho_0] = \int \rho_0(\vec{r}) v(\vec{r}) d\vec{r} + \bar{F} [\rho_0] \quad (2.71)$$

The second Hohenberg-Kohn theorem is the **variational theorem**. The variational theorem states that for every trial density function  $\rho_{tr}(\vec{r})$  that satisfies  $\int \rho_{tr}(\vec{r}) d\vec{r} = N$  and  $\rho_{tr}(\vec{r}) \geq 0$  the following inequality holds:

$$E_0 \leq E_v [\rho_{tr}]. \quad (2.72)$$

Using the statement of the variational theorem and exchange of the ground state electron density  $\rho_0$  with a trial electron density  $\rho_{tr}$  in eq.2.71 leads to the following expression:

$$\bar{T} [\rho_0] + \bar{V}_{ee} [\rho_0] + \int \rho_{tr} v(\vec{r}) d\vec{r} \geq E_v [\rho_0] \quad (2.73)$$

**The Kohn-Sham approach: the reference system.** From the Hohenberg-Kohn theorem we know that in principle the wave function and all molecular properties can be derived from the electronic density of the ground state. The Hohenberg-Kohn theorem does not give practical advice as it cannot answer how to derive  $E_0$  without knowing the wave function of the ground state. It also does not provide an answer how to obtain the kinetic energies and the electron-electron repulsion. In 1965 Kohn and Sham offered a practical solution to these questions. Kohn and Sham considered a reference system of non-interacting electrons that each experience the same external field or external potential  $v(\vec{r}_i)$ . The external potential is chosen to make the ground state electronic density of the reference system equal to the exact ground state density. Hohenberg and Kohn showed that the ground state probability density matrix determines the external potential and one can derive from the electron density of the reference system the external potential of the reference system. Therefore one can write the electronic Hamiltonian of the non-interacting reference system as follows: <sup>[92, 94]</sup>

$$H_s = \sum \left[ -\frac{1}{2} \nabla_i^2 + v_s(\vec{r}_i) \right] \equiv \sum_{i=1}^n h_i^{KS}, \text{ where } h_i^{KS} \equiv -\frac{1}{2} \nabla_i^2 + v_s(\vec{r}_i) \quad (2.74)$$

In equation 2.74  $h_i^{KS}$  is the one-electron Kohn-Sham Hamiltonian.

The wave function of the reference system is the antisymmetrized product of the Slater determinant with the lowest Kohn-Sham spin orbitals of this reference system. The spatial part of each spin orbital is an eigenfunction of the one-electron Kohn-Sham Hamiltonian. The Kohn-Sham approach included a reformulation of the kinetic energy and the electron-electron repulsion that is given by the first Hohenberg-Kohn theorem. The kinetic energy  $\bar{T}$  is now expressed as follows:

$$\Delta \bar{T} [\rho_0] \equiv \bar{T} [\rho_0] - \bar{T}_{ref} [\rho_0] \quad (2.75)$$

In eq. 2.75  $\Delta \bar{T} [\rho_0]$  is the difference in the averaged ground-state electronic kinetic energy between the molecule and the reference system. In the Kohn-Sham approach the electron-electron repulsion is defined as stated by the next equation:

$$\Delta \bar{V}_{elec} [\rho_0] = \bar{V}_{elec} [\rho_0] - \frac{1}{2} \iint \frac{\rho_0(\vec{r}_1) \rho_0(\vec{r}_2)}{\bar{r}_{12}} d\vec{r}_1 d\vec{r}_2 \quad (2.76)$$

$\bar{r}_{12}$  is the distance between two interacting electrons and  $\bar{V}_{elec} [\rho_0]$  the electron-electron repulsion in the real system. The second term in equation 2.76 is the classical expression for the interelectronic repulsion energy. Using the definition of the kinetic energy and electron-electron repulsion a new expression of the ground-state electronic energy can be derived:

$$E_0 = E_v [\rho_0] = T_s [\rho_0] + \int \rho_0(\vec{r}) v(\vec{r}) + \frac{1}{2} \iint \frac{\rho_0(\vec{r}_1) \rho_0(\vec{r}_2)}{\bar{r}_{12}} d\vec{r}_1 d\vec{r}_2 + \Delta \bar{T} [\rho_0] + \Delta \bar{V}_{elec} [\rho_0] \quad (2.77)$$

$\rho_0(\vec{r}) v(\vec{r})$  is the potential energy of the interaction between the nuclei and the electrons (see eq. 2.71). The exchange-correlation energy is now defined from eq. 2.77 as:

$$E_{XC} = \Delta \bar{T} [\rho_0] + \Delta \bar{V}_{elec} [\rho_0] \quad (2.78)$$

From eq 2.77 the Kohn-Sham potential can be explicitly written as:

$$v_s(\vec{r}_1) = \rho_0(\vec{r}_1) v(\vec{r}_1) + \frac{1}{2} \iint \frac{\rho_0(\vec{r}_1) \rho_0(\vec{r}_2)}{\bar{r}_{12}} d\vec{r}_2 + E_{XC}(\vec{r}_1) \quad (2.79)$$

Using the above formulation we can now write the Kohn-Sham equation:

$$\left( -\frac{1}{2}\nabla^2 + v_s[\rho_0](\vec{r}) \right) \phi_i(\vec{r}) = \varepsilon_i \phi_i(\vec{r}) \quad (2.80)$$

Preliminary for the evaluation of equation 2.80 is the determination of the ground state electron density. As Kohn-Sham defined the electron density of the reference system to be identical to the electron density of the real system, the electron density of the real system can be derived from the one-electron Kohn-Sham spin-orbitals, which are the exact eigenfunctions of the reference system of non-interacting particles.

$$E_0 = -\sum_A Z_A \int \frac{\rho(\vec{r}_1)}{\bar{r}_{1A}} d\vec{r}_1 - \frac{1}{2} \sum_{i=1}^n \langle \theta_i^{KS}(1) | \nabla_1^2 | \theta_i^{KS}(1) \rangle + \frac{1}{2} \iint \frac{\rho(\vec{r}_1)\rho(\vec{r}_2)}{\bar{r}_{12}} d\vec{r}_1 d\vec{r}_2 + E_{xc}[\rho_0] \quad (2.81)$$

We have to transform the above equation into a matrix form, so that the energy of a molecular system using for  $\theta_i^{KS}$  linear combinations of STOs or CGTOs, can be derived from a Slater determinant. Eq. 2.81 will be transformed first into a matrix equation and the basis set within the DFT theory is introduced.

The one-electron Kohn-Sham orbitals are derived using the Hohenberg-Kohn variational theorem. The Kohn-Sham orbitals that minimize the molecular ground state energy satisfy the following eigenfunction: <sup>[92, 94]</sup>

$$\left[ -\frac{1}{2}\nabla_1^2 - \sum_A \frac{Z_A}{r_{1A}} + \int \frac{\rho(\vec{r}_2)}{\bar{r}_{12}} d\vec{r}_2 + v_{xc}(1) \right] \theta_i^{KS}(1) = \varepsilon_i^{KS} \theta_i^{KS}(1) \quad (2.82)$$

, which is written in a more compact form as:

$$\left[ -\frac{1}{2}\nabla_1^2 + v_s(1) \right] \theta_i^{KS}(1) = \varepsilon_i^{KS} \theta_i^{KS}(1) \quad (2.83)$$

with  $v_s(1)$  given in eq. 2.79 and even more compact as:

$$h_i^{KS}(1) \theta_i^{KS}(1) = \varepsilon_i^{KS} \theta_i^{KS}(1) \quad (2.84)$$

In Hartree-Fock theory a similar set of integro-differential equations was

It was shown that the integro-differential equation could be transformed into a set of algebraic equations using the set of known basis function and solved by standard matrix techniques. Within the DFT approach the one-electron Kohn-Sham orbitals instead of the wave functions are expanded into a set of known basis functions:

$$\theta_i^{KS}(\vec{r}) = \sum_{\mu=1}^K C_{\mu i} \eta_{\mu}(\vec{r}) \quad (2.85)$$

The expression for the Kohn-Sham orbitals is now pasted into eq. 2.84,

$$h_i^{KS}(1) \sum_{\nu} C_{\nu i} \eta_{\nu}(1) = \varepsilon_i \sum_{\nu} C_{\nu i} \eta_{\nu}(1) \quad (2.86)$$

In complete analogy to the Hartree-Fock scheme the left side is multiplied by a basis function  $\eta_{\mu}^*(\vec{r}_1)$  and then integrated over  $\vec{r}_1$  leading to the Kohn-Sham matrix  $F_{\mu\nu}$  and the overlap matrix  $S_{\mu\nu}$ .<sup>[94]</sup>

$$S_{\mu\nu} = \int \eta_{\mu}^*(\vec{r}_1) \eta_{\nu}(\vec{r}_1) d\vec{r}_1 \quad (2.87)$$

$$F_{\mu\nu} = \int \eta_{\mu}^*(\vec{r}_1) h_i^{KS}(\vec{r}_1) \eta_{\nu}(\vec{r}_1) d\vec{r}_1 \quad (2.88)$$

The matrix equation for the one-electron Kohn-Sham orbitals is then:

$$\mathbf{F}^{\text{KS}}\mathbf{C} = \mathbf{S}\mathbf{C}\boldsymbol{\epsilon} \quad (2.89)$$

**Basis set.** During the past decades many different basis sets were developed for wave function based approaches. In the Kohn-Sham approach the one-electron orbitals play an indirect role, since they are used as a tool to construct the charge density. Consequently the usage of particular basis sets should be less important in the framework of DFT than in the Hartree-Fock scheme. The studies of Bauschlicher<sup>[95]</sup> and Martin<sup>[96]</sup>, for example, support this assumption. This result would also explain why the electronic energies needed to evaluate PA (see eq. 2.10) values for closed shell organic molecules (see chapter 3.1) converged for the anionic states with DFT, although diffuse terms were not applied.<sup>[94]</sup>

Raymond and Wheeler<sup>[97]</sup> suggested in 1999 to use the correlation-consistent basis sets with triple zeta or better in connection with the B3LYP functional (see below). In the present thesis I computed accurate PA and optimal EA values with cc-pVXZ basis set developed by Dunning and coworkers.<sup>[79, 80]</sup> The abbreviation of cc means correlation-consistent. pV stands for polarized valence, the XZ for the size of the basis, which can be double zeta, triple zeta, quadruple zeta, etc. To compute EA values quantitatively (see 2.24) diffuse functions were added to all atoms (the difference between PA and EA is explained in chapter 2.1). For the computation of PA values I used the quadruple cc-pVXZ version, without additional diffuse functions. It turned out that the use of diffuse functions significantly improved the quality of the DFT results for one-electron reduction potentials. The applied basis set used GTOs to describe the spatial orbitals. Polarization functions were included into the basis set, which is indicated by the label pV.

The major problem to be solved to make DFT a practical and accurate tool in computational chemistry is to find computable and sufficiently exact approximations for  $E_{xc}$ . To derive reasonable expressions for the exchange-correlation term of eq. 2.82 and 2.77 we need the adiabatic-connection formula, which connects the real system with the non-interacting reference system through an interelectronic coupling parameter:

$$E_{xc} = \int_0^1 \frac{\partial F[\rho_{tr}]}{\partial \lambda} d\lambda - J[\rho_{tr}] \quad (2.90)$$

where F is given as

$$F[\rho_0] = \bar{T}[\rho_0] + \bar{V}_{el}[\rho_0] \quad (2.91a)$$

and  $J[\rho_{tr}]$  as:

$$J[\rho_{tr}] = \frac{1}{2} \int \int \frac{\rho(\vec{r}_1)\rho(\vec{r}_2)}{r_{12}} d\vec{r}_1 d\vec{r}_2 \quad (2.91b)$$

where  $\lambda$  is the interelectronic coupling strength parameter that switches on the  $1/r^{12}$  Coloumb repulsion between electrons and  $E_{xc}^\lambda$  is the potential energy of exchange-correlation at intermediate coupling strength. This formula connects the non-interacting Kohn-Sham reference system (defined by  $\lambda = 0$ ) with the fully interacting real system (defined by  $\lambda = 1$ ) through a continuum of partially interacting systems ( $0 \leq \lambda \leq 1$ ), all of which share a common density  $\rho$ , which is the density of the real and fully interacting system. Although the integrand of the “adiabatic connection formula (eq. 2.90) refers explicitly to potential energy only, the kinetic part of the exchange-correlation part is generated also by integration over  $\lambda$ .

**The uniform electron gas:** Hohenberg-Kohn derived an expression for the exchange correlation that is called the local-density approximation (LDA). In this approach the electron density of a

fictitious electron system  $\rho$  is varied extremely slowly. Then the exchange correlation can be expressed as: <sup>[92]</sup>

$$E_{XC}^{LDA}[\rho_0] = \int \rho(\vec{r}) \varepsilon_{XC}(\rho) d\vec{r} \quad (2.92)$$

$\varepsilon_{xc}$  is the exchange plus correlation energy per electron in a homogenous electron gas with electron density  $\rho$ . The electron gas is called Jellium, which is an infinite volume-system consisting of an infinite number of electrons moving in a space where the positive charge is continuously and uniformly distributed. The number of electrons per unit volume has a non-vanishing constant value  $\rho$ . The electrons inside Jellium constitute a homogenous (or uniform) electron gas. Taking the functional derivative of  $E_{XC}^{LDA}$ , one finds the following relations:

$$v_{XC}^{LDA} = \frac{\delta E_{XC}^{LDA}}{\delta \rho} = \varepsilon(\rho(\vec{r})) + \rho(\vec{r}) \frac{\partial \varepsilon_{XC}(\rho)}{\partial \rho} \quad (2.93)$$

Kohn and Sham suggested the above relation as a good approximation to  $E_{xc}$  in eq. 2.83. In the LDA approach  $\varepsilon_{xc}$  can be expressed as a sum of the exchange and the correlation part.

$$\varepsilon_{XC} = \varepsilon_x(\rho) + \varepsilon_c(\rho) \quad (2.94)$$

In the LDA approach of Kohn and Sham the exchange part has the form:

$$\varepsilon_x(\rho) = -\frac{3}{4} \left( \frac{3}{\pi} \right)^{\frac{1}{3}} (\rho(\vec{r}))^{\frac{1}{3}} \quad (2.95)$$

The  $\varepsilon_c(\rho)$  has been explicitly calculated for example by Vosko, Wilk and Nusair (VWN).<sup>[98]</sup>

## 2.2.4 Hybrid functionals

The above introduction on the basics of DFT with the Kohn-Sham extensions revealed that improved expressions for electron exchange and correlation lead to exacter wave functions and electronic energies. Therefore, a complete expression for electron exchange and correlation would lead to exact wave functions and electronic energies. Therefore major attempts were carried out in the past to find new and better electron exchange and correlation expressions. Since their creation in the early nineties *hybrid functionals, with explicit exchange* and correlation expressions were applied with success.<sup>[94]</sup>

**Correlation functionals** are part of modern hybrid functionals. Frequently used in QC calculations are the correlation functionals of Colle and Salvetti (CS),<sup>[99]</sup> which comprise the basis for the construction of the LYP functional (see below).<sup>[77]</sup>, the Perdew-Wang correlation functional PW91<sup>[100]</sup> and Perdew's own correlation functional P86.<sup>[101]</sup> A. Becke derived a dynamical correlation functional in 1988,<sup>[102]</sup> which will be marked as Bc88 and must not be confused with the exchange functional,<sup>[76]</sup> which is referred as B in the B3LYP<sup>[76, 77]</sup> functional. Becke introduced a new correlation functional, referred to as Bc95,<sup>[103]</sup> which fulfills all four criteria in 1995. Each of the above mentioned functionals is based on its own model. In 1995 Becke<sup>[103]</sup> analyzed the above five correlation functionals on the basis of four criteria:

- (1) attainment of the exact uniform electron gas limit
- (2) distinct treatment of opposite-spin and parallel-spin correlations
- (3) perfectly self-interacting free (i.e. exactly zero correlation energy in any one-electron system),
- (4) good fit to exact correlation energies of atomic systems

**Table 2.1.** The following table is taken from Becke's work, where he introduced a new functional Bc95:<sup>[103]</sup> which fullfills all of the above 4 criteria.

Criteria	CS	LYP	Bc88	P86	PW91	Bc95
(1)	no	no	no	yes	yes	yes
(2)	no	no	no	yes	Yes	yes
(3)	yes	yes	yes	no	No	yes
(4)	yes	yes	yes	yes	yes	yes

Hohenberg and Kohn proposed the electron gas model so-called Jellium when they introduced LDA as an exchange-correlation functional (see section above). In contradiction to Becke many chemists argue that the electron gas assumption is of little relevance if atomic and molecular systems are treated. Consequently many do not agree that a good correlation functional has to obey criteria one. Table 2.1 reveals that the CS, LYP and Bc88 do not meet criteria one.

The correct treatment of parallel and antiparallel spin correlation is physically relevant. Therefore dynamical correlation functionals gauged on the Helium, which has zero parallel spin correlation energy cannot be expected to correctly reproduce the correlation energy of systems with parallel spin contributions. Therefore the LYP functional gives incorrectly zero correlation energy in any ferromagnetic system, which means that all spins are aligned.<sup>[103]</sup> In the Hartree-Fock the self-interaction of an electron cancels, because the local – non-local Fock operator equals zero when  $a = b$ . It was shown in many works that DFT lacks this elegant treatment of the self-interaction.<sup>[94]</sup> Within the Kohn-Sham scheme (see eq. 2.76) the electrostatic repulsion term is:

$$V_{el} [p_0] = \frac{1}{2} \iint \frac{\rho(\vec{r}_1)\rho(\vec{r}_2)}{r_{12}} d\vec{r}_1 d\vec{r}_2 \quad (2.96)$$

This term does not exactly vanishes for a one electron system, where there must not be any self-interaction. Correction of eq. 2.69 so that  $E_{xc}[p] - E_{xc}[p] = 0$  hardens the development of correlation functionals. Table 2.1 shows the CS, LYP and Bc88 to be deficient of criteria two.<sup>[94, 103]</sup>

Requirement three is of utmost importance if chemical reactions that include hydrogen atoms are considered. The PW91<sup>[100]</sup> correlation functional gives small but nonzero correlation energy for the hydrogen. These errors are too large if the goal of density functional theory is to predict thermochemical data, PA or EA values that are accurate up to a few kcal/mol.

Criteria four recommends to include a good fit to exact correlation energies of atomic systems. Dynamical correlation is an implicitly short-range phenomenon. Hence dynamical correlation functionals should be calibrated only on atomic and never on molecular systems.<sup>[103]</sup>

**The Becke<sup>(1/2)</sup> functional**, which was applied in computation of pK<sub>a</sub> values for closed shell organic molecules expresses the exchange–correlation behavior as follows:<sup>[75, 94]</sup>

$$E_{XC} = \frac{1}{2} E_{XC}^{\lambda=0} + \frac{1}{2} E_{XC}^{\lambda=1} \quad (2.97)$$

Becke<sup>(1/2)</sup> uses the LDA exchange-correlation functional to express the adiabatic connection formula. It is a combination of exact exchange and density functional exchange-correlation as introduced by Becke in 1993.

**The Becke 3 Lee, Yang and Parr (B3LYP) hybrid-functional**<sup>[76, 77, 94]</sup> is frequently used in quantum chemical computations. The DFT functional B3LYP comprises three different units: B stands for the exchange functional developed by A. Becke. The 3 indicates the usage of a hybrid-functional, which includes a mixture of Hartree-Fock exchange with DFT exchange-correlation. LYP is the correlation functional that was developed by Lee, Yang and Parr in the beginning of the nineties. The structure of the hybrid functional B3LYP was originally suggested by Stephens et al. in 1994.<sup>[94, 104]</sup>

$$E_{XC}^{B3LYP} = (1-a)E_X^{LSD} + aE_{XC}^{\lambda=0} + bE_X^{B88} + cE_C^{LYP} + (1-c)E_C^{LSD} \quad (2.98)$$

In the above expression  $E_X^{LSD}$  and  $E_C^{LSD}$  are the hybrid functionals with the local exchange and correlation terms, respectively.  $E_{XC}^{\lambda=0}$  is the exchange-correlation energy of the non-interacting reference system. The remaining two terms are the exchange functional  $bE_X^{B88}$  and the correlation functional  $cE_C^{LYP}$ . The relative weight difference between the exchange, correlation and hybridization is given by the parameters a,b and c. Becke derived from 56 atomization energies, 42 ionizations potentials and 10 first-row total atomic energies the following values for the three parameters: a = 0.20; b = 0.72 and c = 0.81. The next section will convey the general ingredients needed to construct an exchange functional. Then the essential parts of the correlation functional LYP will be explained. Finally I will explain briefly Becke's attempts to construct an additional dynamical exchange-correlation functional.

**The exchange functional proposed by Becke in 1988**<sup>[76]</sup> is a gradient corrected exchange functional. The LDA functional suggested by Kohn and Sham was insufficient to model chemical reactions. The situation changed significantly with the outcome of gradient-corrected LDA functionals. One of the first gradient-corrected functionals was the lowest-order gradient correction to the LDA by Herman et al.<sup>[105] [102]</sup>

$$E_X^{LGC} = E_X^{LDA} - \beta \sum_{\sigma} \int \frac{(\nabla \rho_{\sigma})^2}{\rho_{\sigma}^{4/3}} d^3r \quad (2.99)$$

$E_X^{LDA}$  denotes the Kohn-Sham LDA exchange-correlation functional.  $\sigma$  denotes either spin up or spin down. The value of the constant  $\beta$  was determined empirically. To understand the deficiencies of this equation let us follow Becke's analysis of it. A widely accepted DFT exchange functional must obey the following three constraints. First, the exact asymptotic behavior of the exchange-energy is given by:

$$\lim_{(r \rightarrow \infty)} U_X^{\sigma} = -\frac{1}{r} \quad (2.100)$$

where  $U_X^{\sigma}$  is the Coulomb potential of the exchange charge. The Coulomb potential  $U_X^{\sigma}$  is connected to the total exchange energy  $E_X$  by:

$$E_X = \frac{1}{2} \sum_{\sigma} \int \rho_{\sigma} U_X^{\sigma} d^3r \quad (2.101)$$

The asymptotic behavior of the spin density is given by

$$\lim_{(r \rightarrow \infty)} \rho_{\sigma} = e^{-\alpha_{\sigma} r} \quad (2.102)$$

In contradiction to previous gradient-corrected exchange functionals the following equation by Becke in 1988 fulfills the above three requirements:

$$E_X = E_X^{LDA} - b \sum \int \rho_\sigma^{4/3} \frac{x_\sigma^2}{(1 + 6bx_\sigma \sinh^{-1} x_\sigma)} d^3\vec{r} \quad (2.103)$$

It is also important to notice that the exchange functional is expressible as Taylor series in even powers of  $X_\sigma$ , which is due to the antisymmetry of the function  $\sinh^{-1} x_\sigma$ .  $x_\sigma$  is a dimensionless non-uniformity parameter defined by  $x_\sigma = \frac{|\nabla \rho_\sigma|}{\rho_\sigma^{4/3}}$ . Eq. 2.100 is an improvement of the local spin-

density approximation (LSDA), which is a simple exchange-correlation approximation. The major advantage of the  $E_{XC}$  correlation term given by eq. 2.101 is the exactness of the asymptotical behavior of the exponential tails of atomic and molecular charge distribution. Incorporation of this  $E_{XC}$  term into the Kohn-Sham equation (2.77) significantly improved the quality of computed atomization energies for atoms of the G1 data set. In the next step Becke included correlation corrections to the LSDA. For this purpose Becke used the PW91 correlation functional.

Eq. 2.103 introduces an exchange-only gradient corrected exchange correlation term. This application yielded a drastically improvement of the quality of the computed atomization energies. While DFT computation with non-corrected LSDA yielded atomization energies of 36.2 kcal/mol the modified exchange corrected LSDA yielded an average value of 3.7 kcal/mol. The uncorrected non-correlated LSDA were able to reproduce atomization energies but did to reproduce experimental ionization energies accurately. The next logical step was to introduce correlation correction to the exchange correlation term.

**The correlation-functional by Lee, Yang and Parr**<sup>[77]</sup> was derived from the correlation energy formula of Colle and Salvetti. The latter two derived the correlation energy in terms of electron density and a Laplacian of the second-order Hartree-Fock density matrix. The authors of the LYP functional converted the Hartree-Fock second order density matrix into a “true” density-functional formula. The expression is based on the assumption (not mathematical proven) that the individual density orbitals in the Colle Salvetti<sup>[99]</sup> formula are Kohn-Sham orbitals. For closed shell systems the correlation functional LYP has the following form:

$$E_C = -a \int \frac{1}{1 + d\rho^{-1/3}} \left\{ \rho + b\rho^{-2/3} \left[ C_F \rho^{5/3} - 2t_w + \left( \frac{1}{9} t_w + \frac{1}{18} \nabla^2 \rho \right) \right] e^{-c\rho^{-1/3}} \right\} d\vec{r} \quad (2.104)$$

for open shell systems the correlation functional becomes:

$$E_C = -a \int \frac{\gamma(r)}{1 + d\rho^{-1/3}} \left\{ \rho + b\rho^{-5/3} \left[ 2^{2/3} C_F \rho_\alpha^{8/3} + 2^{2/3} C_F \rho_\beta^{8/3} - \rho t_w + \frac{1}{9} (\rho_\alpha t_w^\alpha + \rho_\beta t_w^\beta) + \frac{1}{18} (\rho_\alpha \nabla^2 \rho_\alpha + \rho_\beta \nabla^2 \rho_\alpha) \right] e^{-c\rho^{-1/3}} \right\} d\vec{r} \quad (2.105)$$

The correlation potential is the functional derivative of  $E_C$ . In the above equation  $C_F$  is defined as

$C_F = \frac{3}{10} (3\pi^2)^{\frac{2}{3}}$  and  $t_w$  is a local Weizsaecker kinetic-energy density.  $\rho(\vec{r})$  is the Hartree-Fock

electron density. The parameters  $a = 0.049$ ,  $b = 0.132$ ,  $c = 0.2533$  and  $d = 0.349$  were already defined in the originally Colle and Salvetti formula.  $\alpha$  and  $\beta$  denote spin down and spin up states of electrons.



**Non-B3LYP functionals.** An additional exchange-correlation functional was introduced by Becke. Becke merged the exchange-only gradient corrected part, which is part of the B3LYP functional and the correlation corrected generalized gradient approximation with the exact-exchange-local DFT approach. This procedure yielded results for the G1 data set that were superior to previous results on all three fields: thermochemical data, atomic energies, ionization energies. An additional feature was the achievement of the *uniform electron gas limit*. The following formula expresses this hybrid functional:

$$E_{XC} = E_{XC}^{LSDA} + a_0(E_X^{exact} - E_X^{LSDA}) + a_x \Delta E_X^{B88} + a_c \Delta E_C^{PW91} \quad (2.106)$$

where  $a_0$ ,  $a_x$ ,  $a_c$  are semiempirical coefficients, which were determined by a fitting to experimental data. These coefficients are identical to three parameters of the B3LYP functional. Becke determined the semiempirical coefficients for 56 atomization energies, 42 ionization potentials and 8 proton affinities and 10 first-row total atomic energies. The values he obtained are :  $a_0 = 0.20$ ,  $a_x = 0.72$  and  $a_c = 0.81$ .  $E_X^{exact}$  is the exact exchange energy,  $\Delta E_X^{B88}$  is Becke's 1988 gradient correction to the LSDA for exchange and  $\Delta E_C^{PW91}$  is the gradient corrections. An exact *ab initio* expression of the exchange-correlation energy can be derived in the framework of the Kohn-Sham DFT approach.

The exchange correlation functional of Perdew and Wang is a *generalized gradient approximation* (GGA). Local exchange-correlation functionals depend explicitly on local spin densities and local spin-density gradients as follows:

$$E_{XC}^{GGA} = \int f \left[ \rho_\alpha(\vec{r}), \rho_\beta(\vec{r}), \nabla \rho_\alpha(\vec{r}), \nabla \rho_\beta(\vec{r}) \right] d^3 \vec{r} \quad (2.107)$$

In equation 2.107  $\alpha$  and  $\beta$  denote up and down electron spins. This exchange correlation functional of Perdew and Wang is referred to in the literature as PW91.<sup>[94, 100]</sup>

Let us now inspect Becke's new correlation functional of 1995, which obeys the four minimal requirements (see paragraph above). The total correlation energy is expressed as a sum of the correlation energy of electrons of opposite spins and of the correlation of electrons with  $\alpha$  spin and with electrons with  $\beta$  spin.

$$E_C = E_C^{opp} + E_C^{\alpha\alpha} + E_C^{\beta\beta} \quad (2.108)$$

The opposite spin correlation energy is expressed as :

$$E_C^{opp} = \left[ 1 + c_{opp} (\chi_\alpha^2 + \chi_\beta^2) \right]^{-1} E_{opp}^{UEG} \quad (2.109)$$

The component  $E_{opp}^{UEG}$  is the correlation energy of opposite spin electrons in the framework of the uniform electron gas assumption, which was presented above. The term  $E_{opp}^{UEG}$  was originally derived by Stoll, Pavlidou and Preuss. The correlation energy for electrons of parallel spin is expressed with equation 2.110:

$$E_C^{\sigma\sigma} = \left[ 1 + c_{\sigma\sigma} \chi_\sigma^2 \right]^{-2} \frac{D_\sigma}{D_\sigma^{UEG}} E_{C\sigma\sigma}^{UEG} \quad (2.110)$$

The correlation energy  $E_C^{\sigma\sigma}$  expresses the correlation energy between  $\alpha\alpha$  and  $\beta\beta$  spin electrons.

The terms  $D_\sigma$  and  $D_C^{UEG}$  are taken from the correlation model derived by Becke in 1988. Each of two equations (eq. 2.109 and eq. 2.110) comprise one fitting parameter  $c_{opp}$ . These parameters were fitted to Helium ( $c_{opp}$ ) and Neon ( $c_{\sigma\sigma}$ ). Becke derived values of  $c_{opp} = 0.0031$  and  $c_{\sigma\sigma} = 0.038$ .

### 2.3.5 Gaussian-3(MP2)

The **G3MP2** method was applied to compute EA values for the considered redox-active compounds. G3MP2 is a variation of Gaussian-3 (G3) theory,<sup>[106]</sup> which itself is an improvement of the Gaussian-2 (G2) theory.<sup>[107]</sup>

G2 and G3 are composite techniques in which sequences of *ab initio* molecular orbital calculations are performed.<sup>[106, 107]</sup> In G2 and G3 calculations optimized geometries are first obtained at the Hartree-Fock level with a 6-31G\* basis set. The optimized coordinates are then used to compute the harmonic frequencies at the same level of theory. Obtained zero-point energies are finally scaled by a factor of 0.89. Both approaches refine the Hartree-Fock optimized geometries at the MP2 (full)/6-31G(*d*) level of theory, using all electrons for the calculation. A series of high-level QC calculations is then carried out differently in both theories. In step 4 of G3 theory a complete computation at the fourth-order Moller–Plesset perturbation theory with the 6-31G(*d*) basis set (MP4/6-31G(*d*)) is carried out. In G2 theory the complete MP4 calculations are performed at the with the 6-311G(*d*) basis set. In both theories a series of corrections is then performed after the MP4 calculations. First, G3 corrects the energy to account for the influence of diffuse functions:

$$\Delta E (\text{diffuse}) = E [\text{MP4/6-31+G} (d)] - E [\text{MP4/6-31G} (d)] \quad (2.111)$$

In G2 theory this  $\Delta E$  (diffuse) is computed at the MP4/6-311+(*d*). Subsequently G3 corrects the energy for the influence of higher polarization functions at the MP4 level with the 6-31+G (2df, p) basis set, so that  $\Delta E$  (2df, p):

$$\Delta E (2df, p) = E [\text{MP4/6-31G} (2df, p)] - E [\text{MP4/6-31G} (d)] \quad (2.112)$$

The 2*d* assigns two sets of uncontracted *d*-primitives with exponents twice and half the standard values. Again G2 theory employs the 6-311 variant of the 6-31G basis set. Therefore G2 calculates  $\Delta E$  (2df, p) with MP4/6-311G (2df, p). A correction of correlation effects beyond fourth order perturbation theory is performed in G3 calculations using the method of quadratic con-figuration interaction,  $\Delta E$  (QCI), according to:

$$\Delta E (\text{QCI}) = E [\text{QCISD(T)/6-31G}(d)] - E [\text{MP4/6-31G} (d)] \quad (2.113)$$

A correction for larger basis set effects and for the nonadditivity caused by the assumption of separate basis set extensions for diffuse functions and higher polarization functions,  $\Delta E$ (G3large):

$$\begin{aligned} \Delta E (\text{G3 large}) = & E [\text{MP2}(\text{full})/\text{G3large}] - E [\text{MP2}/6-31\text{G} (2df,p)] \\ & - E [\text{MP2}/6-31+\text{G} (d)] + - E [\text{MP2}/6-31\text{G} (d)] \quad (2.114) \end{aligned}$$

The MP2(full) calculation is calculated in G2 theory with the 6-311G (3df, 2p) basis set instead of the G3large that is applied in G3 computations. The 6-311G(3d f, 2p) assigns 3d f polarization functions on first and second row atoms and 2p polarization functions on hydrogen atoms. Because of limitations of this basis the G3large basis set was developed. The G3large includes 2df polarization functions on second row atoms and additional core polarization functions for all atoms. The total electronic energy is finally calculated according to eq. 2.115:

$$E_0 = E [\text{MP4/6-31G (d)}] + \Delta E (\text{diffuse}) + \Delta E (6\text{-}311\text{G}/2\text{df, p}) + \Delta E (\text{QCI}) + \Delta E (\text{G3 large}) \\ + \Delta E(\text{SO}) \quad (2.115)$$

The spin-orbit correction  $\Delta E (\text{SO})$  is taken from experiments when available or from accurate theoretical calculations in other cases. According to explanation given in the above paragraph one can formulate the electronic energy in G2 theory:

$$E_0 = E [\text{MP4/6-31G (d, p)}] + \Delta E (\text{diffuse}) + \Delta E (2\text{df, p}) + \Delta E (\text{QCI}) + \Delta E (6\text{-} \\ 311\text{G}(3\text{df},2\text{p}) + \Delta E(\text{SO}) \quad (2.116)$$

G3MP2<sup>[78]</sup> is a reduced variant of G3 theory that replaces the QC MP4 method with MP2 for the calculations of the basis set extensions. Therefore in a QC computation with G3MP2 the  $\Delta E (\text{QCISD(T)})$  is the most demanding part. The G3MP2large basis set is the same as the G3large basis set used in G3 theory, except that the core polarization functions are not included. In addition, the MP2/G3MP2large calculation in G3(MP2) theory is done with a frozen core (fc) approximation, whereas the MP2/G3large calculation in G3 theory includes all electrons in the correlation treatment. In a G3MP2 calculation the final energies are computed according to:

$$E_0 [\text{G3(MP2)}] = \text{QCISD(T)}/6\text{-}31\text{G(d)} + \Delta E_{\text{MP2}} + \Delta E (\text{SO}) + E(\text{HLC}) + E(\text{ZPE}) \quad (2.117)$$

The correction to the second-order MP2 level is given in G3MP2 theory as:

$$\Delta E_{\text{MP2}} = E [\text{MP2/G3MP2large}] - E [\text{MP2/6-31G}] \quad (2.118)$$

## 2.3 Atomic partial charges

The present work evaluates solvation energies within a *two-step* procedure. In the *two-step* procedure first a QC computation in vacuum is performed to compute the electron density distribution and the wave function. Subsequently, the electron density is transformed into atomic partial charges. In the second step the electrostatic free energies of solvation are evaluated from this point charge distribution separately without inclusion of any QC method by solving the Poisson-Boltzmann equation. Unfortunately, atomic charges are not observables; i.e., they cannot be determined by experiments or quantum chemical calculations. A variety of methods exist to compute atomic partial charges from the molecular wave function.

In the Mulliken population analysis, atomic partial charges are calculated according to the electron distribution inside atomic orbitals. Electrons in overlap regions of different orbitals are distributed evenly to each atom, independently of the characteristics of the participating atom types, which build up that region. Hence, this method does not separate electron density in overlap regions according to the electronegativity of atoms. The starting point of the Mulliken analysis is eq. 2.119, which relates the total number of electrons to the density matrix and the overlap integrals.<sup>[108]</sup>

$$q_A = Z_A - \sum_{\mu=1; \mu \text{ on } A}^K P_{\mu\mu} - \sum_{\mu=1; \mu \text{ on } A}^K \sum_{\nu=1, \nu \neq \mu}^K P_{\mu\nu} S_{\mu\nu} \quad (2.119)$$

In eq. 2.119  $P_{\mu\mu}$  is the electron density in an orbital that can be allocated to the atom on which the considered orbital  $\phi_\mu$  is centered. The remaining electron density  $P_{\mu\nu}$  is associated with the overlap region  $\phi_\mu \phi_\nu$  which is expressed by the overlap matrix  $S_{\mu\nu}$ . The advantage of the Mulliken analysis is that the computation is trivial once the elements of the density matrix are known. It is so simple that most QC programs output Mulliken charges by default at the end of the computation. The method proposed by Mulliken is crucially dependent on a balanced basis set on each atom. The quality of Mulliken charges suffers from the inexact assignment of p, d and f functions to different atoms. These orbitals extend considerably into space and overlap with orbitals of other atoms. The electron density of p, d and f orbitals is now sometimes assigned to other atoms that build up the overlap region. In the Mulliken approach the charges are therefore dependent on the applied basis set. According to Wiberg and Rablen the charges on the central carbon of isobutene increase from +0.1 to +1.0 when the 6-311++G\*\* is used instead of the 6-31G\*\* basis set. Other orbital-based methods are the Löwdin population analysis or the natural population analysis. In the Löwdin approach the atomic orbitals are transformed into an orthogonal set, along with the molecular coefficients. Thereby reduces the basis set dependence of the calculated charges. Calculated Löwdin charges are deprived of the problem of negative population or population greater than two.<sup>[108]</sup>

The theory *atom-in-molecule* was originally formulated by Baader in 1985.<sup>[108]</sup> In the *atom-in-molecule* method each point in the molecular space is assigned to one of the atoms that build up the molecule. After the molecular space was divided into volumes occupied by atoms, the electron density in each volume is assigned to the corresponding atom. Wiber and Reblan showed that this method is relatively invariant to the applied basis set. For methane and ethyne the computed *atom-*

*in-molecule* charges reproduced measured dipoles accurately. The disadvantage of the method is that the partition of the electron density into the volumes is CPU expensive, because the electron population that is assigned to each atom is calculated by numerical integration.<sup>[108]</sup>

### 2.3.1 Potential based methods

Widely distributed methods to generate atomic partial charges from QC computations are so called potential-based methods as CHELP,<sup>[72]</sup> CHELPG,<sup>[109]</sup> the Merz-Kollmann procedure<sup>[71]</sup> and the restrained electrostatic potential method (RESP).<sup>[69, 70]</sup> Charges are calculated by matching the electrostatic potential (ESP)<sup>[109]</sup> from the instantaneous atomic partial charge distribution with the QC ESP, computed on optimized geometries in the vacuum. These methods have in common that they use values of the electrostatic potential on spatial positions that are distributed around the atoms of the molecule. All potential-based methods take into account only potential points outside the vdW radii of the atoms to generate charges. As they cannot be determined undoubtedly effective vdW radii are a matter of debate in the literature and are treated differently in the potential-based methods. Different schemes are currently in use to select the points outside the vdW radii of the atoms of the molecule. The CHELP method selects points spherically around each atom, with 14 points per shell. Spheres are considered only, when they are 2.5 Å, 3.5 Å, 4.5 Å, 5.5 Å and 6.5 Å away from the center of the atom. CHELPG picks the potential points from a regular cubic grid placed around the atoms of the substance. Points that lie between 0 and 2.8 Å plus the vdW radius from the atomic center are included into the computation. The Merz-Kollmann method takes into account only potential points, which lie on nested Connolly surfaces and are 1.4, 1.6, 1.8 and 2.0 times the vdW radius of the atom. It is clear from the above explanations that a certain grade of arbitrariness remains in each potential-based method. Four methods for deriving atomic partial charges from the QC electrostatic potential CHELP, CHELPG, Merz-Kollman and RESP have been compared and critically evaluated. It was shown the charges strongly depend on the selection scheme of potential points around the atoms.<sup>[74, 110]</sup>

Modified version of the potential based methods have been developed. Whereas the conventional potential methods as CHELP and Merz-Kollman fit by means of the sum of atomic partial charges directly to the QC ESP, modified versions carry out the fitting process with constrains or restrains. In the matrix form of the Merz-Kollmann approach the atomic partial charge are given as

$$\bar{q} = A^{-1}B \quad (2.120)$$

In the modified version the constraints or restrains can be included into eq. 2.120, which leads to:

$$wB\bar{q} = wc \quad (2.121)$$

where  $q_i$  are the partial charges,  $c_k$  are the constraints or restrains, and  $w_k$  are weight factors. All restraints or constraints that are linear in the charges can be expressed in this way, e.g., the total charge, electrostatic moments, electrostatic potential, or equality of certain charges resulting from symmetry. A modified version of CHELP the CHELP-BOW method was introduced recently by Sigfridsson and Ryde.<sup>[74, 110]</sup> Description of long-range electrostatic interaction properly forces to include potential points even far away from the atom. Naturally, all points up to infinity cannot be included; at some finite distance, the point selection has to stop. A solution to the problem would be to make a more sparse sampling of points at larger distances from the molecule and compensate for this by a higher weight in the fit. Potential points are picked randomly outside the vdW radii of the atoms with a density of 2.500 points per atom and are weighted with Boltzmann

factor ( $w = e^{-\frac{E}{RT}}$ ). Potential points are picked up to a distance of 8 Å away from the center of the atom. The potential points are Boltzmann-weighted after their occurrence in actual simulations using the energy function of the program in which the charges will be used. E is part of the Boltzmann factor in the Lennard-Jones potential.

$$E = \sum_{atom,i} \begin{cases} A_i r_i^{-6} + B_i r_i^{-12} & \text{if } r_i < r_{i \min} \\ E_{i \min} & \text{if } r_i > r_{i \min} \end{cases} \quad (2.122)$$

Here, r is the distance between the electrostatic potential point i and atom i, and  $A_i$  and  $B_i$  are the Lennard-Jones constants obtained from the force field of the simulation package to be used in the simulations.  $E_{i \min}$  is the minimum energy of the Lennard-Jones potential, and it is the energy assumed at the distance  $r_i = r_{i \min}$ . Thus, only the repulsive part of the Lennard-Jones potential is considered.<sup>[74, 110]</sup>

### 2.3.2 The RESP method

In this work I used the Merz–Kollman RESP<sup>[69, 70]</sup> procedure to fit atomic partial charges from the molecular ESP<sup>[109]</sup> generated by QC computations at different level of QC theory. It will be shown in chapter 3 that the most valuable charges were computed at the Becke(1/2)/6-31G\*\* and B3LYP/6-31G\*\* level. The RESP procedure determines the atomic partial charges of a molecule by matching the ESP calculated on grid points in the neighborhood of the considered molecule with the ESP generated by the atomic partial charges, while constraining the net charge of the molecule. The ESP was defined on grid points outside of the vdW spheres (with radius  $R_{vdW}$ ) of each atom up to a maximum radius  $R_{max}$  placed equidistantly on a system of rays emerging from each atom center. The corresponding  $R_{vdW}$  values are 1.95 Å, 1.6 Å, 1.7 Å, 1.8 Å, 1.96 Å, the  $R_{max}$  values are 9.0 Å, 8.3 Å, 8.5 Å, 8.7 Å, 9.0 Å for C, H, O, N, Cl atoms, respectively. With a direct least square fit of atomic partial charges molecular dipole and quadrupole moments are often not well reproduced. This is due to charges of buried atoms whose influence on ESP may be poorly defined, since all ESP grid points lie outside of the vdW volume of all atoms of the molecule. A *two-step* RESP fitting procedure with different constraints can overcome this caveat without significant influence to reproduce the quantum chemically defined ESP. In the first step a weak hyperbolic penalty function (strength parameter  $a = 0.0005$  au) was applied on all atoms except hydrogens. In the second step a stronger hyperbolic penalty function (strength parameter  $a = 0.001$  au) was applied to all atoms except hydrogens, methyl and methylene carbons.



**all computations are performed without involmment of the solvent**

**Figure 2.1.** Flowchart to generate atomic partial charges.

The following paragraph shows how the RESP<sup>[69, 70]</sup> method has incorporated the penalty function into the fitting process of atomic partial charges to the QC ESP. The ESP of the computed atomic partial charges is computed by:

$$\hat{V}_i = \sum_{ij} \frac{q_j}{r_{ij}} \quad (2.123)$$

In eq. 2.4.1  $\hat{V}_i$  is the molecular mechanic ESP,  $q_j$  the charge point distributed on a cubic grid in distinct distant  $r_{ij}$  to the atomic center. The charge point  $q_j$  is found between the vdW volume of the atom and the maximum distance given above for each atom. In a least-square fit the function  $\chi_{esp}^2$  has to be minimized to maximize the agreement between  $V_i$  and  $\hat{V}_i$  for each charge point.

$$\chi_{esp}^2 = \sum_i (V_i - \hat{V}_i)^2 \quad (2.124)$$

To minimize the function the first derivative has to be found:

$$\frac{\partial (\chi_{esp}^2)}{\partial q_i} = 0 \quad (2.125)$$

where :

$$\frac{\partial (\chi_{esp}^2)}{\partial q_i} = -2 \sum_i \frac{V_i - \hat{V}_i}{r_{ij}} = 0 \quad (2.126)$$

to arrive to eq. 2.126 the expression  $\sum_{ij} \frac{q_j}{r_{ij}}$  was substituted into eq.2.4.2; then the first derivative was formed and finally  $\hat{V}_i$  could be substituted back into the equation. With the addition of a penalty function to the charge fitting procedure an additional term is added to  $\chi_{esp}^2$  so that the function to be minimized becomes:

$$\chi^2 = \chi_{esp}^2 + \chi_{rstr}^2 \quad (2.127)$$

and the least square minimum is now defined as :

$$\frac{\partial (\chi^2)}{\partial q_i} = \frac{\partial \chi_{esp}^2}{\partial q_i} + \frac{\partial \chi_{rstr}^2}{\partial q_i} \quad (2.128)$$

Different penalty function were tried to optimize solvation energies for model compounds. Finally Kollmann and coworkers have chosen a hyperbolic function as the one of choice:

$$\chi_{rstr}^2 = a \sum_j (q_j^2 + b^2)^{1/2} - b \quad (2.129)$$

where a is a scale factor, which defines the asymptotic limits of the strength of the restraint and b determines the tightness of the hyperbola around its minimum. The second term in eq. 2.4.6 becomes now:

$$\frac{\partial(\chi_{rstr}^2)}{\partial q_i} = a q_i (q_j^2 + b^2)^{-1/2} \quad (2.130)$$

The system of equations is now transformed into a matrix form, with the following matrix elements:

$$A_{jk} = \sum_i \frac{1}{r_i r_{ik}}; \quad A_{ij} = \sum_i \frac{1}{r_{ij}^2} + \frac{\partial \chi_{rstr}^2}{\partial q_i} \quad \text{and} \quad B_j = \sum_i \frac{V_i}{r_{ij}} + q_{0j} \frac{\partial(\chi_{rstr}^2)}{\partial q_i} \quad (2.131)$$

yielding the matrix equation:

$$\begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{22} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ A_{nm} & A_{n2} & \cdots & A_{nm} \\ 1 & 1 & \cdots & 1 \end{pmatrix} \begin{pmatrix} q_1 \\ q_2 \\ \vdots \\ q_n \\ \lambda \end{pmatrix} = \begin{pmatrix} B_1 \\ B_2 \\ \vdots \\ B_n \\ q_{tot} \end{pmatrix} \quad (2.132)$$

the matrix can be formulated as  $Aq=B$  which leads to  $q = A^{-1}B$ . The quality of the fit is analyzed, computing the relative root-mean-square value of  $\hat{V}_i$  to  $V_i$

$$RRMS = \left( \frac{\chi_{esp}^2}{\sum V_i^2} \right)^{-1/2} \quad (2.133)$$



## 2.4 The Poisson and Poisson Boltzmann equation

### 2.4.1 Derivation

The Poisson-Boltzmann equation is fundamental to compute electrostatic interaction energies between a titratable or a redox-active group and the environment. If the ionic strength of the medium is zero the Poisson-Boltzmann equation is truncated to the Poisson equation. Because of its importance in the present thesis I will give the derivation of this equation in following paragraph.

$$\phi_i(\vec{r}) = \sum_i \frac{q_i}{\epsilon |\vec{r} - \vec{r}_i|} \quad (2.134)$$

The derivation of the Poisson-Boltzmann equation starts with the Coulomb potential (eq. 2.134). The sum runs over all point charges  $q_i$  at the position  $r_i$ .  $\epsilon$  is the dielectric constant. The Coulomb potential can be reformulated to express the electrostatic potential with respect to charge density instead of point charges  $q_i$ .

$$\phi(\vec{r}) = \int_V \frac{\rho(\vec{r}') d\vec{r}'}{\epsilon |\vec{r} - \vec{r}'|} \quad (2.135)$$

Applying the Laplace operator to both sides of this equation results after a few transformation in the Poisson equation. In cases when the ionic strength equals zero, the solvation process is described correctly using the Poisson equation.

$$\Delta \phi(\vec{r}) = -4\pi \frac{\rho(\vec{r})}{\epsilon} \quad (2.136)$$

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

$$\nabla \left[ \epsilon(\vec{r}) \nabla \phi(\vec{r}) \right] = -4\pi \rho(\vec{r}) \quad (2.137)$$

To describe the effect of mobile ions in the solvent the Debye-Hückel theory is needed. The derivation of it continues under the assumption that mobile ions that are distributed in the solvent behave as the atoms of an ideal gas. The charge density of ions  $\rho_{ion}(\vec{r})$  is added to the charge density of eq. 2.137, which yields eq. 2.138:

$$\rho_{ion}(\vec{r}) = \sum_s c_s(\vec{r}) q_s e^{-\beta q_s \phi(\vec{r})} \quad (2.138)$$

The sum runs over all mobile ions  $s$ .  $c_s$  is the original concentration of ions  $s$  and  $q$  is the charge of the ions  $s$ .  $\beta$  is equal to  $(k_B T)^{-1}$ . Adding the ionic strength to the Poisson equation leads to the Poisson-Boltzmann equation (PBE).

$$\nabla \left[ \epsilon(\vec{r}) \nabla \phi(\vec{r}) \right] + 4\pi \sum_s c_s(\vec{r}) q_s e^{-\beta q_s \phi(\vec{r})} = -4\pi \rho(\vec{r}) \quad (2.139)$$

The Debye-Hückel theory requires overall electroneutrality:

$$\sum_s c_s(\vec{r}) q_s = 0 \quad (2.140)$$

In the special case of only monovalent ions and overall electroneutrality, the PBE can be written elegantly as:

$$\nabla[\varepsilon(\vec{r}) \nabla \phi(\vec{r})] - 8\pi \sum_s c_s(\vec{r}) \sinh[\beta e \phi(\vec{r})] = -4\pi \rho(\vec{r}) \quad (2.141)$$

In this equation  $e$  is the unit elementary charge. To ease the solution of the PBE, it can be linearized by expanding the exponentials.

$$\sum_s c_s(\vec{r}) q_s e^{-\beta q_s \phi(\vec{r})} \approx \sum_s c_s(\vec{r}) q_s - \beta \sum_s c_s(\vec{r}) q_s^2 \phi(\vec{r}) \quad (2.142)$$

$\sum_s c_s(\vec{r}) q_s$  is zero due to the requirement of the 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. 2.142 the linearized PBE (LPBE) can be written as:

$$\nabla[\varepsilon(\vec{r}) \nabla \phi(\vec{r})] - 8\pi \beta I(\vec{r}) \phi(\vec{r}) = -4\pi \rho(\vec{r}) \quad (2.143)$$

For a more detailed description of the derivation of the linearized PBE and the Debye-Hückel theory, see chapter 15-1 (pp-328-340) of McQuarrie (1976)<sup>[111]</sup> and chapter 18 (pp-321-339) of Hill (1986).<sup>[112]</sup>

## 2.4.2 Numerical solution of the linearized PBE and finite difference method

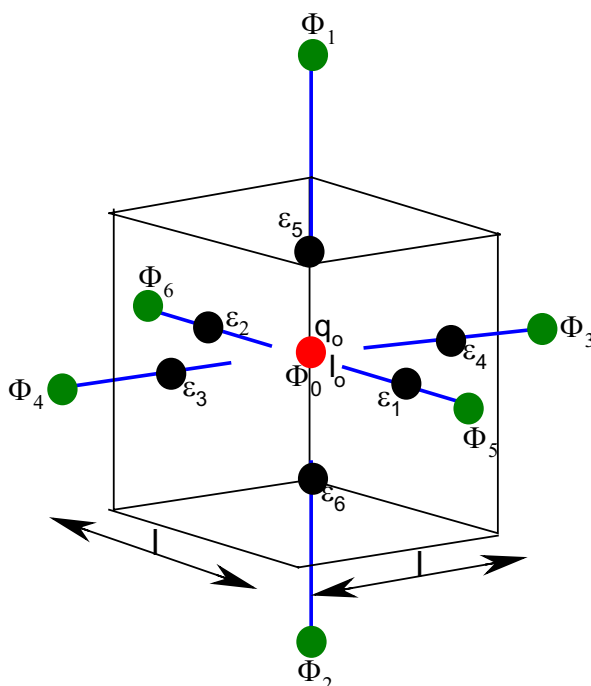
Analytical solutions of the LPBE can be achieved only for simple geometries. For complex geometries a numerical solution of the LPBE must be carried out. In the present thesis the finite difference method was applied as the method of choice for numerical solutions of the LPBE. A widely distributed program to compute electrostatic solvation free energies is SOLVATE from the Macroscopic electrostatic with atomic details (MEAD) suite.<sup>[58, 113, 114]</sup> This program solves the LPBE and is accessible free of charge. It was applied in the current thesis to compute solvation free energies.

The linearized PBE is integrated over a small volume  $V = l^3$  of one cubic grid element:

$$\int_V \nabla[\varepsilon(\vec{r}) \nabla \phi(\vec{r})] d\vec{r} - \int_V 8\pi \beta I(\vec{r}) \phi(\vec{r}) d\vec{r} = \int_V -4\pi \rho(\vec{r}) d\vec{r} \quad (2.144)$$

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

$$\int_A \varepsilon(\vec{r}) \nabla \phi(\vec{r}) dA - 8\pi \beta I_0 \phi_0 l^3 = -4\pi q_0 \quad (2.145)$$



**Figure 2.2:** It shows the physical relevant quantities of LPBE mapped on a cubic grid

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:

$$\sum_{i=1}^6 \frac{\varepsilon_i (\phi_i - \phi_0) l^2}{l} - 8\pi \beta I_0 \phi_0 l^3 = -4\pi q_0 \quad (2.146)$$

Eq. 2.146 is simplified and rearranged to yield a finite difference expression for  $\phi_0$  :

$$\phi_0 = \frac{\left(\sum_{i=1}^6 \varepsilon_i \phi_i\right) + \frac{4\pi q_0}{l}}{\left(\sum_{i=1}^6 \varepsilon_i\right) + 8\pi \beta I_0 l^2} \quad (2.147)$$

Starting from arbitrary values, the electrostatic potential is iteratively calculated for each grid point according to eq. 2.147 until a convergence criterion is met. For details of this procedure see Nicholis and Honig (1991)<sup>[115]</sup> and chapter (pp 762-768) of Press et al. (1992).<sup>[116]</sup>

Grid points at the edges of the cubic grid have less than six neighbor points for the cubic grid elements. To overcome the problem one starts the calculation with a grid that is much larger than the molecule to be solvated. Then  $\phi_0$  outside the grid is set to zero or to a value according to the Debye-Hückel sum. To reduce the CPU time the grid resolution in the initial step is low. In a second step, a focussing step, the grid resolution is enhanced but it is only applied in a small volume around the molecule. In the application, where small organic molecules are transferred from the gas phase into the condensed phase a *two-step* focusing procedure was sufficient.

### 2.4.3 Grid artefact and electrostatic self energy

The electrostatic energy is calculated from the electrostatic energy potential  $\phi$ :

$$G_{el} = \frac{1}{2} \sum_i q_i \phi_i \quad (2.148)$$

The sums run over all atoms  $i$ .  $q_i$  is the atomic partial charge of atom  $i$  and  $\phi_i$  the electrostatic potential at the position of atom  $i$ .

The electrostatic energy of a classical point charge of atom  $i$  in its own electrostatic potential is the self-energy. For point charges these self-energies are infinity. 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 on grid resolution and position, remain from the infinite self-energy of the singularities. This grid artefact is called grid energy. There are several possibilities to get rid of the grid energy. The easiest way is to compute only relative electrostatic energies between systems with equal grid energy. This requires that the coordinates of the molecule are placed identically relative to the grid points in both phases. To make the Poisson-Boltzmann electrostatics suitable for larger systems mathematical concepts to approximate the potential of the LPBE had to be developed. One such method is the analytical continuum electrostatic method (ACS) founded by Schaefer and Karplus in second half of the nineties.<sup>[117, 118]</sup> Consistent with other approaches ACS the self-energy part of ACE is based on the integrated field concept. It combines the self-energy potential with the generalized Born equation for charge-charge interaction. The basic ideas of this concept date back to the theory founded by Born in the twentieth of the last century. Adding non-polar free energy to the solvation energy, the method of Born was generalized to the ACS. The following is an introduction of the basic ideas of ACS according to Schaefer and Karplus (1996,1998).<sup>[117,118]</sup>

In the ACS approach the electrostatic energy  $G_{el}$  is expressed in terms of the energy density  $u(\vec{r})$  instead of the electrostatic potential formulated by the Poisson equation.  $u(\vec{r})$  of an electrostatic field generated by a charge distribution  $\rho(\vec{r})$  is expressed in terms of the electric displacement vector.

$$D(\vec{r}) = \varepsilon(\vec{r}) E(\vec{r}) \quad (2.149)$$

and therefore:

$$u(\vec{r}) = \frac{1}{8\pi\varepsilon(\vec{r})} \vec{D}^2(\vec{r}) \quad (2.150)$$

By integrating the energy density over the full space, Born (1920) obtained the electrostatic energy  $G_{el}$ .

$$G_{el} = \int_V u(\vec{r}) d\vec{r} = \frac{1}{8\pi} \int_V \frac{1}{\varepsilon(\vec{r})} \vec{D}^2(\vec{r}) d\vec{r} \quad (2.151)$$

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  $\phi(\vec{r})$ :

$$G_{el} = \frac{1}{2} \int_V \rho(\vec{r}) \phi(\vec{r}) d\vec{r} \quad (2.152)$$

The discrete form of eq. 2.152 is eq. 2.148. With the Poisson equation 2.137 the charge density  $\rho(\vec{r})$  is substituted by  $-\frac{1}{4\pi} \nabla[\varepsilon(\vec{r})\nabla\phi(\vec{r})]$ , so that  $G_{el}$  can be expressed as:

$$G_{el} = \frac{1}{8\pi} \int_V \phi(\vec{r}) \nabla [\varepsilon(\vec{r}) \phi(\vec{r})] d\vec{r} \quad (2.153)$$

Integration leads to:

$$G_{el} = \frac{1}{8\pi} [\phi(\vec{r}) \varepsilon(\vec{r}) \nabla \phi(\vec{r})]_V + \frac{1}{8\pi} \int_V \varepsilon(\vec{r}) \nabla \phi(\vec{r}) d\vec{r} \quad (2.154)$$

The first term vanishes due to the vanishing electrostatic potential  $\phi(\vec{r})$  at the boundaries of the volume  $V$ . Using eq. 2.149 and the relationship  $\vec{E}(\vec{r}) = -\nabla\phi(\vec{r})$  the following expression is obtained:

$$G_{el} = -\frac{1}{8\pi} \int_V \varepsilon(\vec{r}) \vec{E}(\vec{r}) \cdot \vec{E}(\vec{r}) d\vec{r} = \frac{1}{8\pi} \int_V \frac{1}{\varepsilon(\vec{r})} D(\vec{r}) \cdot D(\vec{r}) d\vec{r} \quad (2.155)$$

$$G_{el} = \frac{1}{8\pi\varepsilon_s} \int_{V_s} \vec{D}^2(\vec{r}) d\vec{r} + \frac{1}{8\pi\varepsilon_p} \int_{V_p} \vec{D}^2(\vec{r}) d\vec{r} \quad (2.156)$$

with  $\varepsilon_s$  and  $\varepsilon_p$ , which define the reduced dielectric constant  $\Delta\varepsilon$  :

$$\frac{1}{\Delta\varepsilon} = \frac{1}{\Delta\varepsilon_p} + \frac{1}{\Delta\varepsilon_s} \quad (2.157)$$

Since  $\varepsilon_p < \varepsilon_s$ , the reduced dielectric constant  $\Delta\varepsilon$  is always positive. Using the reduced dielectric constant  $\Delta\varepsilon$ , eq. 2.157 can be written extending the integral over the full space  $V$ :

$$G_{el} = \frac{1}{8\pi\varepsilon_s} \int_V \vec{D}^2(\vec{r}) d\vec{r} + \frac{1}{8\pi\Delta\varepsilon} \int_{V_p} \vec{D}^2(\vec{r}) d\vec{r} \quad (2.158)$$

It is now assumed that the first term in eq. 2.158 corresponds to the situation in a homogeneous dielectric medium, where the dielectric displacement vector is described, by a simple Coulomb field. The field lines at the dielectric boundary, which is the interface between solute and solvent are somewhat smeared and that is not properly accounted for with the Coulomb field approximation. According to Schaefer and Frömmel 1990,<sup>[119]</sup> the error introduced by this assumption is not more than a few percent. Such a simplifying approximation is not possible for the second integral, which is harder to evaluate. Because we have discrete atoms with atomic partial charges, the interaction energy of the atoms with themselves  $G_i^{self}$  can be distinguished from the interaction energy between different atoms  $G_{ji}^{self}$  :

$$G_{el} = \sum_i (G_i^{self} + \sum_{j>i} G_{ij}^{int}) \quad (2.159)$$

Both sums run over all atoms.  $G_i^{self}$  and  $G_{ji}^{int}$  can be calculated according to eq. 2.159 with the approximation that the displacement of the first integral is expressed by the Coulomb field:

$$G_{ij}^{int} = \frac{q_i q_j}{\epsilon_s |\vec{r}_{ij}|} + \frac{1}{8\pi \Delta \epsilon} \int_{V_p} \vec{D}_i(\vec{r}) \vec{D}_j(\vec{r}) d\vec{r} \quad (2.160)$$

$$G_i^{self} = \frac{q_i^2}{2\epsilon_s R_i} + \frac{1}{8\pi \Delta \epsilon} \int_{V_p} \vec{D}_i^2(\vec{r}) d\vec{r} \quad (2.161)$$

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. 2.161: The atom is no longer considered as a point, but as a sphere with radius  $R_i$ , which is the van-der Waals radius. The point charge  $q_i$  is distributed over the surface of this sphere. The first term in eq. 2.161 is obtained for this special distribution of charge  $q_i$  and is called the Born energy term.

In the presented continuum approach the solvation energy is computed as the electrostatic free energy that arises, when a molecule is placed from a homogeneous medium, where solute and solvent have identical dielectric constants, into a heterogeneous medium, where solute and solvent are represented by different dielectric constants. This energy is identical to the electrostatic part of the solvation energy, if  $\epsilon$  for the solute is unity.  $\Delta G_{el}$  is calculated by taking the difference between the electrostatic energy in the heterogeneous dielectric medium  $G_{el}$  and the electrostatic energy in the homogeneous medium  $G_{hom}$ :

$$\Delta G_{el} = G_{el} - G_{hom} \quad (2.162)$$

The Born formula is applied for the calculation of the self-energy in the homogeneous dielectric field:

$$G_{hom} = \sum_i \frac{q_i^2}{2\epsilon_p R_i} + \sum_{j>i} \frac{q_i q_j}{\epsilon_p |\vec{r}_{ij}|} \quad (2.163)$$

The energy is splitted into a self energy term and an interaction term:

$$\Delta G_{el} = \Delta G^{self} + \Delta G^{int} = \sum_i (\Delta G_i^{self} + \sum_{j>i} \Delta G_{ij}^{int}) \quad (2.164)$$

Taking into account eq. 2.161 and 2.162,  $G_i^{self}$  and  $G_{ji}^{int}$  can be expressed as follows:

$$\Delta G_{ij}^{int} = \frac{q_i q_j}{\epsilon_s |\vec{r}_{ij}|} + \frac{1}{8\pi \Delta \epsilon} \int_{V_p} \vec{D}_i(\vec{r}) \vec{D}_j(\vec{r}) d\vec{r} - \frac{q_i q_j}{\epsilon_p |\vec{r}_{ij}|} \quad (2.165)$$

$$\Delta G_i^{self} = \frac{q_i^2}{2\epsilon_s R_i} + \frac{1}{8\pi \Delta \epsilon} \int_{V_p} \vec{D}_i^2(\vec{r}) d\vec{r} - \frac{q_i^2}{2\epsilon_p R_i} \quad (2.166)$$

With the expression for the reduced dielectric constant eq. 2.157, these expressions can be truncated:

$$\Delta G_{ij}^{\text{int}} = -\frac{q_i q_j}{\Delta \varepsilon |\vec{r}_{ij}|} + \frac{1}{8\pi \Delta \varepsilon} \int_{V_p} \vec{D}_i(\vec{r}) \vec{D}_j(\vec{r}) d\vec{r} \quad (2.167)$$

$$\Delta G_i^{\text{self}} = -\frac{q_i^2}{2\Delta \varepsilon R_i} + \frac{1}{8\pi \Delta \varepsilon} \int_{V_p} \vec{D}_i^2(\vec{r}) d\vec{r} \quad (2.168)$$

#### 2.4.4 Generalized Born approximation

The Born term  $\Delta G_i^{\text{self}}$  can be formally written as an analogous Born energy term with an unknown effective Born radius  $b_i$ , which accounts for the volume covered by the solute surrounding atom  $i$ :

$$\Delta G_i^{\text{self}} = \frac{q_i^2}{2\Delta \varepsilon b_i} \quad (2.169)$$

If  $\Delta G_i^{\text{self}}$  is already calculated (e.g. by solving the PBE or by applying the analytical approximation), the effective Born radius  $b_i$  is given as:

$$b_i = -\frac{q_i^2}{2\Delta \varepsilon \Delta G_i^{\text{self}}} \quad (2.170)$$

In analogy  $\Delta G_{ij}^{\text{int}}$  can be formulated with an effective interaction distance  $R_{ij}^{\text{solv}}$  accounting for the influence of the solute volume:

$$\Delta G_{ij}^{\text{int}} = -\frac{q_i q_j}{\Delta \varepsilon R_{ij}^{\text{solv}}} \quad (2.171)$$

The generalized Born equation can now be formulated as from eq.2.170, eq.2.171 and eq.2.164. The generalized Born equation was first formulated by Still et al. 1990:<sup>[68]</sup>

$$\Delta G_{el} = -\sum \left( \frac{q_i^2}{2\Delta \varepsilon b_i} + \sum_{j>i} \frac{q_i q_j}{\Delta \varepsilon R_{ij}^{\text{solv}}} \right) \quad (2.172)$$

To evaluate the integral in eq.2.172. Schaefer and Karplus introduced a molecular density function  $P_s(\vec{r})$ , with  $P_s(\vec{r}) = 1$  (if  $\vec{r}$  is inside the solute volume) or  $P_s(\vec{r}) = 0$  (if  $\vec{r}$  is not inside the solute volume). Then they expressed eq.2.173:

$$\Delta G_i^{\text{self}} = -\frac{q_i^2}{2\Delta \varepsilon R_i} + \frac{1}{8\pi \Delta \varepsilon} \int_{V_p} \vec{D}_i^2(\vec{r}) P_s(\vec{r}) d\vec{r} \quad (2.173)$$

Eq.2.173 is now divided into two parts  $P_s(\vec{r})$  is expressed as the sum of atomic density functions. For each atom  $k$ , there is one atomic density function  $P_k(\vec{r})$ , which describes the volume distribution of the atom:

$$P_s(\vec{r}) = \sum_k P_k(\vec{r}) \quad (2.174)$$

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

$$\Delta G_{ik}^{self} = \frac{1}{8\pi \Delta \epsilon} \int_V \vec{D}_i^2(\vec{r}) P_k(\vec{r}) d\vec{r} \quad (2.175)$$

The sum of  $\Delta G_{ik}^{self}$  for all atoms k leads to the second term in eq.2.175, so that this equation can be reformulated as:

$$\Delta G_i^{self} = -\frac{q_i^2}{2\Delta \epsilon R_i} + \sum_{k \neq i} G_{ik}^{self} \quad (2.176)$$



## 2.5 Proton solvation free energy

Computation of relative  $pK_a$  values does not depend on the proton solvation free energy ( $\Delta G_{\text{solv}}(\text{H}^+)$ ). The following equation documents that  $\Delta G_{\text{solv}}(\text{H}^+)$  vanishes if the relative  $pK_a$  value between compound A and B is considered:<sup>[51]</sup>

$$\Delta pK_a (A/B) = \frac{1}{2.303RT} \Delta \Delta G_R (A/B) = \frac{1}{2.303RT} \left[ (\Delta G_{\text{gas}}(A) - \Delta G_{\text{gas}}(B)) + (\Delta \Delta G_{\text{solv}}(A) - \Delta \Delta G_{\text{solv}}(B)) - (\Delta G_{\text{solv}}(\text{H}_A^+) - \Delta G_{\text{solv}}(\text{H}_B^+)) \right] \quad (2.177)$$

In eq. 2.177  $\Delta \Delta G_R (A/B)$  is the reaction free energy difference between two deprotonation reactions and  $\Delta G_{\text{gas}}$  is the free energy of a deprotonation reaction in the vacuum for compound A and B.  $\Delta \Delta G_{\text{solv}}$  describes the solvation free energy difference between the deprotonated and protonated states for A and B. Eq 2.177 was employed by MacKerell Jr. and coworkers<sup>[51]</sup> to evaluate relative  $pK_a$  values of substituted pyridines. In this study the experimental  $pK_a$  value of unsubstituted pyridine served as the basis for all computed relative  $pK_a$  values. To avoid the dependence on a reference compound absolute instead of relative  $pK_a$  values were computed in the present study. Let us recall the definition of the absolute  $pK_a$  value, which reveals its dependence on the proton solvation free energy ( $\Delta G_{\text{solv}}(\text{H}^+)$ ):

$$pK_a (A) = \frac{1}{2.303RT} \Delta G_R (A) = \frac{1}{2.303RT} \left[ \Delta G_{\text{gas}}(A) + \Delta \Delta G_{\text{solv}}(A) + \Delta G_{\text{solv}}(\text{H}^+) \right] \quad (2.178)$$

$\Delta G_R$  is the reaction free energy in solution and  $\Delta G_{\text{gas}}$  is the reaction free energy of the deprotonation reaction in the vacuum.  $\Delta \Delta G_{\text{solv}}$  is the solvation free energy between the deprotonated and protonated states of compound A. The spectrum of experimental data on the free proton solvation energy comprises an energy range between  $-252.5$  kcal/mol and  $-266.00$  kcal/mol.<sup>[47, 54, 84-90, 120, 121]</sup> Therefore, computed  $pK_a$  values can be influenced by the choice of  $\Delta G_{\text{solv}}(\text{H}^+)$  by more than 11  $pK_a$  units. As indicated by these numbers principle experimental difficulties to determine  $\Delta G_{\text{solv}}(\text{H}^+)$  are expectable. In the following paragraph I will give an overview of the usage of  $\Delta G_{\text{solv}}$  values in recent studies on absolute  $pK_a$  values. Subsequently I will explain more elaborately three different methods to determine  $\Delta G_{\text{solv}}(\text{H}^+)$ .

### 2.5.1 General overview on the usage of $\Delta G_{\text{solv}}(\text{H}^+)$

Noyes determined  $\Delta G_{\text{solv}}(\text{H}^+)$  to be equal to  $-260.5$  kcal/mol.<sup>[84]</sup> This proton solvation free energy was employed for instance by Richardson et al.<sup>[46]</sup> in a study on absolute acidities of organic molecules and by Quennville et al.<sup>[42]</sup> to determine the  $pK_a$  value of a histidine in the catalytic center of cytochrome c oxidase. Wilson et al.<sup>[122]</sup> measured enthalpic solvation energies of aliphatic carboxylic acids using an enthalpic proton solvation energy ( $\Delta H_{\text{solv}}(\text{H}^+)$ ) equal to –

271.6 kcal/mol. The corresponding free energy measured by Klots in the beginning of the eighties equals  $-262.4$  kcal/mol.<sup>[87]</sup> This value was actually a refinement of the experimental studies performed by Randless and Trasatti.[Randless, 1956 #49][Klots, 1981 #51] From electrochemical experiments of the potassium ion Randles derived a proton solvation free energy of  $-259.4$  kcal/mol.<sup>[123]</sup> The refined value of Trasatti accounted for the surface potential of the surrounding water.<sup>[88]</sup> The average value of  $-259.5$  kcal/mol derived from five independent measurements of the standard hydrogen electrode was applied in  $pK_a$  computations of Lim et al.<sup>[47]</sup> The  $4.44$  V measured by Reiss and Heller,<sup>[90]</sup> the  $4.43$  V published by Gurevich and Pieskov,<sup>[124]</sup> the  $4.48$  V proposed by Lohman<sup>[125]</sup> and the  $4.44$  V recommended by Trasatti<sup>[88]</sup> are in close proximity, whereas the  $4.73$  V suggested by Gomer and Tryson<sup>[121]</sup> deviates significantly from the other four measurements. Reiss and Heller explained the deviation of the latter value as they showed that additional potentials of materials inside the electrode were included into the measurement. Measured and calculated solvation energies combined with measured PA values of organic molecules enabled Florian and Warshel<sup>[86]</sup> to deduce  $\Delta G_{solv}(H^+)$  equal to  $-259.5 \pm 2.5$  kcal/mol. The cluster-base-pair approximation by Tissandier et al.<sup>[85]</sup> yielded  $\Delta G_{solv}(H^+) = -263.96$  kcal/mol.<sup>[85]</sup> This procedure is based on the evaluation of the energy balance of a large number of protonation reactions, whose energetics involve uncertainties of about  $-2$  kcal/mol, while the procedure itself is connected with a negligible statistical error of only  $-0.07$  kcal/mol. This  $\Delta G_{solv}(H^+)$  value is closed to the  $-264.64$  kcal/mol employed by Liptak et al.<sup>[40, 43]</sup> to estimate absolute  $pK_a$  values for aliphatic carboxylic acids and substituted phenols. Recent computations on water clusters by Tawa et al.<sup>[89]</sup> suggested  $-262.23$  kcal/mol for  $\Delta G_{solv}(H^+)$ , which is consistent with similar computations from Zhan and Dixon<sup>[126]</sup> who estimated  $\Delta G_{solv}(H^+)$  to be  $-262.40$  kcal/mol.<sup>[126]</sup> These values are nearly identical to the one determined by Klots (see above). To compute the acidities of substituted imidazoles Topol et al.<sup>[45]</sup> employed a  $\Delta G_{solv}(H^+)$  value of  $-262.50$  kcal/mol.

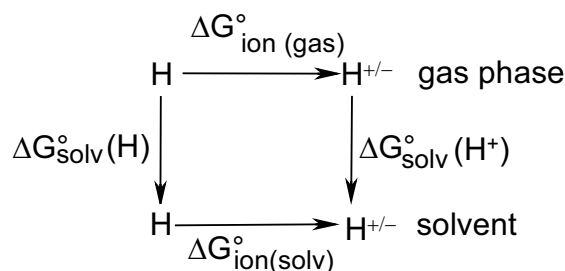
The experimental spectrum of  $\Delta G_{solv}(H^+)$  values is considerably narrowed if the measurement of Gomer and Tryson would be dismissed.<sup>[121]</sup> It would assume values between  $-258.5$  kcal/mol proposed by Lohman<sup>[125]</sup> and  $-266.00$  kcal/mol suggested by Kallies and Mitzner.<sup>[54]</sup> All studies in the last 25 years except the one by Florian and Warshel<sup>[86, 127]</sup> and Kallies and Mitzner determined  $\Delta G_{solv}(H^+)$  in the range between  $-262.00$  kcal/mol and  $264.00$  kcal/mol.

## 2.5.2 Thermodynamic measurements of ion hydration

Thermodynamic measurements by Noyes<sup>[84]</sup> in the sixties of the last century determined an electrostatic contribution of  $-260.5$  kcal/mol to the total free proton solvation energy. The central equation of that study describes the solvation of any cation or anion as a three-step process: First the considered ion is discharged in the vacuum, then the obtained neutral form is hydrated and finally recharged in solution. Eq. 2.179 describes this process quantitatively:

$$\Delta\Delta G_{solv}^\circ(H/H^+) = \Delta G_{solv}^\circ(H) + \Delta G_{solv}^\circ(H^+) \quad (2.179)$$

Inspection of the thermodynamic cycle depicted below shows that the sum of  $\Delta G_{solv}^\circ(H)$  and  $\Delta G_{solv}^\circ(H^+)$  accounts for  $\Delta\Delta G_{solv}^\circ(H/H^+)$ . In the scaffold of the cycle the corresponding chemical reactions are represented by  $\Delta G_{ion}^\circ(gas)$  and  $\Delta G_{ion}^\circ(solv)$ .

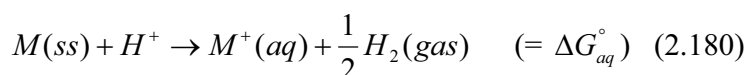


**Scheme 2.3:** Thermodynamic cycle describes the solvation process of any cation or anion as contributions of  $\Delta G_{\text{neut}}^{\circ}$  and  $\Delta G_{\text{el}}^{\circ}$ . The corresponding chemical reactions in the vacuum and in solution are covered by  $\Delta G_{\text{ion}}^{\circ}(\text{gas})$  and  $\Delta G_{\text{ion}}^{\circ}(\text{solv})$ .

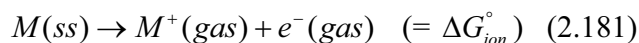
To solve eq. 2.179 for  $\Delta G_{\text{solv}}^{\circ}(\text{H}^{+})$  we need alternative expressions for  $\Delta \Delta G_{\text{solv}}^{\circ}(\text{H} / \text{H}^{+})$  and  $\Delta G_{\text{solv}}^{\circ}(\text{H})$ . The derivation of these expressions are explained in the following paragraphs.

The quantity  $\Delta G_{\text{solv}}^{\circ}(\text{H})$  refers to the process when 1 mole of neutral gaseous species at 1 atmosphere fugacity is dissolved in water to an ideal 1 molal solution under a pressure of 1 atmosphere. The zero point assumption applies if there is no energy of hydration so that the enthalpy and entropy changes are just those calculable from the change in volume associated with the change of state. The ideal gas law gives the volume of 24.47 liters for 1 molar gas. The above two assumptions yield a final value of 1.325 kcal/mol for  $\Delta G_{\text{neut}}^{\circ}$ .

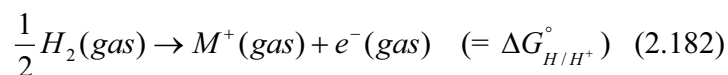
The following paragraph shows the derivation of an alternative expression for  $\Delta \Delta G_{\text{solv}}^{\circ}(\text{H} / \text{H}^{+})$ . The free energy change ( $\Delta G_{\text{aq}}^{\circ}$ ) that refers to the formation of an aqueous ion (cation or anion) from the corresponding element in the standard state can be expressed as:



M can be any cationic molecule in the standard state (ss) and  $\Delta G_{\text{aq}}^{\circ}$  can easily be formulated for anions.  $\Delta G_{\text{ion}}^{\circ}$  describes the free energy change for the formation of an gaseous ion (cation or anion) from the corresponding element in the standard state.:



$\Delta G_{\text{H}/\text{H}^{+}}^{\circ}$  accounts quantitatively for the free energy change of the charge donating reaction:



From the above three equation we can now formulate the alternative expression of:

$$\Delta \Delta G_{\text{solv}}^{\circ}(\text{H} / \text{H}^{+}) = \Delta G_{\text{aq}}^{\circ} - \Delta G_{\text{ion}}^{\circ} + \Delta G_{\text{H}/\text{H}^{+}}^{\circ} \quad (2.183)$$

Values for  $\Delta G_{aq}^\circ$ , are tabulated in various standard reference books.  $\Delta G_{ion}^\circ$  was determined according to the following equation:

$$\Delta G_{ion}^\circ = \Delta G_{atom}^\circ + \Delta H_{ion}^\circ - \Delta H_{atom}^\circ - RT \ln\left(\frac{Q_{ion}}{Q_{atom}}\right) \quad (2.184)$$

Values for  $\Delta G_{atom}^\circ$ ,  $\Delta H_{ion}^\circ$ ,  $\Delta H_{atom}^\circ$  are also provided by standard reference books. The last term in eq. 2.184 is the electronic partition function for the considered ion. A final value of  $-363.985$  kcal/mol is obtained for  $\Delta G_{aq}^\circ - \Delta G_{ion}^\circ$ , when the hydrogen proton is considered. If reliable values for  $\Delta G_{aq}^\circ$  and  $\Delta G_{ion}^\circ$  can be derived only  $\Delta G_{H/H^+}^\circ$  needs to be determined.

Here and in the following the subscript -th indicates that the quantity is derived from theoretical considerations. According to theoretical considerations the free energy change to transfer reversibly a charge  $q$  from a conducting sphere of radius  $r$  in a vacuum to an identical sphere in a medium of dielectric constant  $\epsilon$ , is expressed by eq. 2.185:

$$\Delta G_{el-th}^\circ = -\frac{q^2}{2r}\left(1 - \frac{1}{\epsilon}\right) \quad (2.185)$$

The corresponding entropy change to the above equation is expressed as follows:

$$\Delta S_{el}^\circ = -\left(\frac{\partial \Delta G_{el-th}^\circ}{\partial T}\right)_P \quad (2.186)$$

From the theoretically derived equations above the following two values for  $\Delta G_{el-th}^\circ$  and  $\Delta S_{el-th}^\circ$  can be deduced:

$$\Delta G_{th}^\circ = -163.89 \frac{z^2}{rA} \text{ kcal / mol} \quad (2.187)$$

$$\Delta S_{th}^\circ = -9.649 \frac{z^2}{rA} \text{ cal / mol deg} \quad (2.188)$$

The observed discrepancy between the theoretical derived and the measured values for  $\Delta G_{el-th}^\circ$  and  $\Delta S_{el-th}^\circ$  should vanish if an additional term is added to eq. 2.187 and eq.2.188

$$\Delta G_{th/modi}^\circ = \left(\frac{-163.89}{r} + \frac{B}{r^2}\right) \frac{z^2}{A} \text{ kcal / mol(cation)} \quad (2.189)$$

$$\Delta S_{th/modi}^\circ = \left(\frac{-9.649}{r} + \frac{D}{r^2}\right) \frac{z^2}{A} \text{ kcal / mol(cation)} \quad (2.190)$$

The above procedure was also applied to derive modified expressions for  $\Delta G_{el-th}^\circ$  (anion) and  $\Delta S_{el-th}^\circ$  (anion), so that finally four equations were obtained. Using the modified expressions together with eq. 2.179, 2.183 and 2.184 a system of four equations, which depends on three parameters is derived.  $\Delta G_{H/H^+}^\circ$  can now be calculated with the mathematical procedure of a three

parameter least square technique. The described procedure yields  $\Delta G_{H/H^+}^\circ$  equal to  $-104.81$  kcal/mol. According to eq 2.184  $\Delta\Delta G_{sol}^\circ(H/H^+)$  equals then  $-259.175$  kcal/mol.

Eq. 2.179 can now be solved for  $\Delta G_{sol}^\circ(H^+)$ :

$$-259.175 \text{ kcal/mol} = 1.325 \text{ kcal/mol} + \Delta G_{sol}^\circ(H^+) \quad (2.191)$$

$$\Delta G_{sol}^\circ(H^+) = -260.5 \text{ kcal/mol} \quad (2.192)$$

To study absolute  $pK_a$  values in solution Lim et al. used a proton free solvation energy that was the average value derived from five independent measurements of the standard (normal) hydrogen electrode. Instead of using the least square fit technique suggested by Noyes one can use the experimental value for  $\Delta G_{H/H^+}^\circ$  derived from measurements of the absolute potential of the standard hydrogen electrode (SHE). Table 2.2 explicitly converts measured potentials of the SHE into the proton solvation free energy.

**Table 2.2:** measured potentials of the standard hydrogen electrode are converted into the proton solvation free energy using eq. 2.197:

Experiment:	SHE (eV)	$\Delta G_{H/H^+}^\circ$ (kcal/mol)	$\Delta\Delta G_{sol}^\circ(H/H^+)$ (kcal/mol)	$\Delta G_{sol}^\circ(H^+)$ (kcal/mol)
Reis-Heller	4.43 V	102.156	-261.829	-263.15
Gurevich- Pieskov	4.43 V	102.156	-261.829	-263.15
Trasatti	4.44 V	102.386	-261.60	-2.62.93
Lohman	4.48 V	103.309	-260.676	-262.00
Gomer-Tryson	4.73 V	109.074	-254.912	-256.24

### 2.5.3 The cluster-pair-based approximation

It was shown above that the derivation of the proton solvation free energy from measurements of the standard hydrogen electrode still depends on the thermodynamical framework established by Noyes in the sixties of the last century. The next paragraphs will explain a method called the cluster-pair-base approximation, which was used by Tissandier et al.<sup>[85]</sup> to deduce the solvation free energy and enthalpy of the proton without additional thermodynamic assumptions. Within the cluster-pair-base assumption the values of  $-263.98$  kcal/mol and  $-274.88$  kcal/mol were determined for  $\Delta G_{sol}^\circ(H^+)$  and  $\Delta H_{sol}^\circ(H^+)$ , respectively.

Tissandier et al. derived  $\Delta G_{sol}^\circ(H^+)$  (2.193) and  $\Delta H_{sol}^\circ(H^+)$  (2.194) directly from the next two equations:

$$\Delta G_{sol}^{\circ}(H^{+}) \approx \frac{1}{2} \left[ \frac{\Delta G_n^{\circ}(A^{+}) + \Delta G_n^{\circ}(B^{-})}{c_{n,n}(A^{+}, B^{-})} - k(A^{+}) - k(B^{-}) \right] \quad (2.193)$$

$$\Delta H_{sol}^{\circ}(H^{+}) \approx \frac{1}{2} \left[ \frac{\Delta H_n^{\circ}(A^{+}) + \Delta H_n^{\circ}(B^{-})}{c_{n,n}(A^{+}, B^{-})} - k(A^{+}) - k(B^{-}) \right] \quad (2.194)$$

where  $c_{n,n}$  of eq. 2.193 is defined according to:

$$c_{n,n}(A^{+}, B^{-}) = \frac{\Delta G_n^{\circ}(A^{+}) + \Delta G_n^{\circ}(B^{-})}{\Delta G_{aq}^{\circ}(A^{+}) + \Delta G_{aq}^{\circ}(B^{-})} \quad (2.195)$$

Proton solvation free energy and enthalpy were estimated for the case that  $c_{n,n}(A^{+}, B^{-}) = c_n(A^{+}) = c_n(B^{-})$ , because then the above two expressions become exact relations.  $\Delta G_{aq}^{\circ}$  and  $\Delta H_{aq}^{\circ}$  are the conventional solvation free energy and enthalpy of an ion, respectively, which are defined for  $Z^{*}$  which can be  $A^{+}$  or  $B^{-}$  as:

$$\Delta G_{aq}^{\circ, con}[Z^{*}] = \Delta G_{aq}^{\circ, con}[Z^{*}] - \Delta G_{formation}^{\circ}(H^{+}) + \Delta G_{solv}^{\circ}(H^{+}) \quad (2.196)$$

$$\Delta H_{aq}^{\circ, con}[Z^{*}] = \Delta H_{aq}^{\circ, con}[Z^{*}] - \Delta H_{formation}^{\circ}(H^{+}) + \Delta H_{solv}^{\circ}(H^{+}) \quad (2.197)$$

In eq. 2.198  $\Delta G_n(Z^{\pm})$  accounts for the free energy change to place any arbitrary ion ( $(Z^{\pm})$  ( $A^{+}$  or  $B^{-}$ )) into a water cluster of  $n$  water molecules. This free energy change is quantitatively expressed as follows and can also be formulated for  $\Delta H_n(Z^{\pm})$ :

$$\Delta G_n^{\circ}(Z^{\pm}) = \sum_{i=1}^n \Delta G_{i-1,i}^{\circ}(Z^{\pm}) - \sum_{i=1}^n \Delta G_{i-1,i}^{\circ}(H_2O) \quad (2.198)$$

The three energy terms in the above expression are explained below:

$$Z^{\pm} + (H_2O)_n \rightarrow Z^{\pm}(H_2O)_n \quad (= \Delta G_n^{\circ}(Z^{\pm})) \quad (2.199)$$

$$Z^{\pm}(H_2O)_{i-1} + H_2O \rightarrow Z^{\pm}(H_2O)_i \quad (= \Delta G_{i-1,i}^{\circ}(Z^{\pm})) \quad (2.200)$$

$$(H_2O)_{i-1} + H_2O \rightarrow (H_2O)_i \quad (= \Delta G_{i-1,i}^{\circ}(H_2O)) \quad (2.201)$$

The summations of stepwise free energies and enthalpies,  $\Delta G_n^{\circ}(Z^{\pm})$ , which enter the final expression needed to derive  $\Delta G_{sol}^{\circ}(H^{+})$  and  $\Delta H_{sol}^{\circ}(H^{+})$  are obtained from averaging over literature data. Inspection of the literature reveals that these publications were made over a period of more than twenty years by several independent groups. In the cluster-pair-based approximation the constant  $k(Z^{\pm})$  expresses the absolute bulk single-ion free energies of various ions in terms of the unknown absolute solvation free energy or enthalpy of the proton.

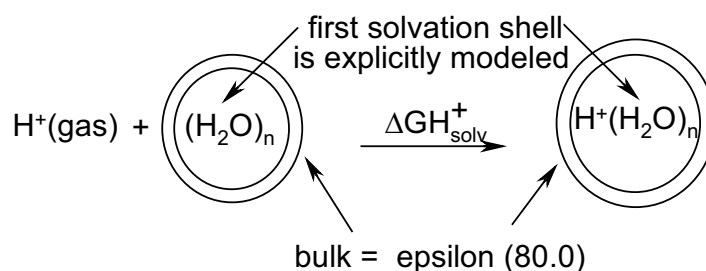
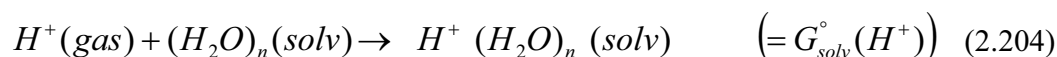
$$\Delta G_{sol}^{\circ}(Z^{\pm}) = \pm \left[ \Delta G_{sol}^{\circ}(H^{+}) + k(Z^{\pm}) \right] \quad (2.203)$$

$k(Z^{\pm})$  between two arbitrary ions ( $A^{-}$  and  $B^{+}$ ) can be computed from gas phase hydration data performed on water clusters and ion-water clusters. Because  $k(Z^{\pm})$  can be determined experimentally the only unknown quantity in eqs. 2.193 and 2.194 are the free proton solvation energy ( $\Delta G_{sol}^{\circ}(H^{+})$ ) and enthalpy ( $\Delta H_{sol}^{\circ}(H^{+})$ ).

### 2.5.4 A computational method to derive the proton solvation free energy

The theoretical approach<sup>[89]</sup> derived  $\Delta G_{solv}^{\circ}(H^+)$  without additional experimental information, but with thermodynamic assumptions. It is clear from the above paragraphs that being totally independent of such assumptions would have been a major improvement in the evaluation of the real proton solvation energy. To estimate  $\Delta G_{sol}^{\circ}(H^+)$  the free energy change that arises when the proton is transferred from the vacuum into a water-cluster embedded in bulk water was calculated. The bulk water was described through a continuum model and a dielectric constant equal to 80.0. The computational approach, which was performed by Tawa et al.<sup>[89]</sup> determined a proton solvation free energy equal to  $-262.23$  kcal/mol.

As explicit water molecules only accounted for the first solvation shell, four to six molecules were assumed to be enough to derive  $\Delta G_{sol}^{\circ}(H^+)$ . The subsequent equation (eq. 2.204) and Figure 2.3 describe the experimental approach:



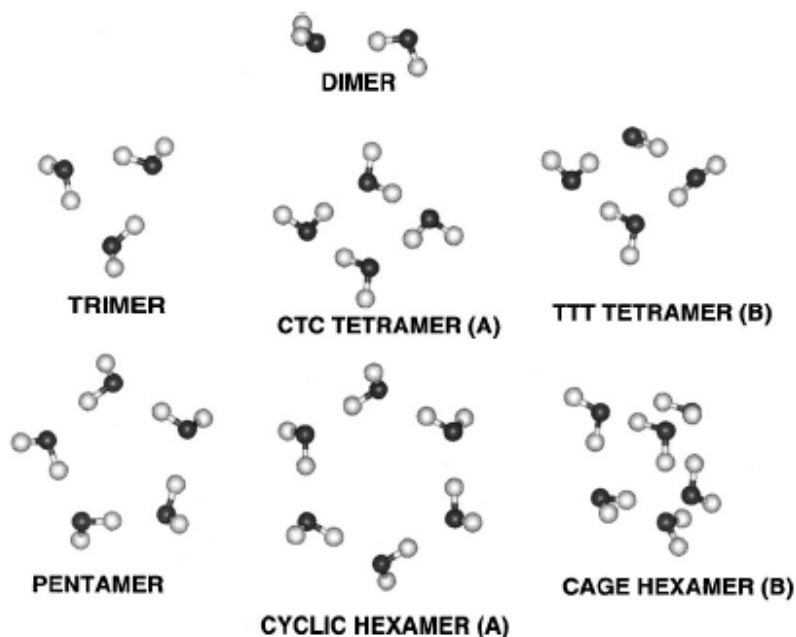
**Figure 2.3** Illustrates the theoretical approach to derive the proton solvation free energy.

In the study of Tawa et al. solvation free energies were obtained using a self-consistent reaction field (SCRf) cycle that couples the dielectric continuum with the supermolecule, which is treated at several correlated electronic structure levels. The authors employed DFT as well as second and fourth order Møller perturbation theory and couple-cluster with single, double and triple excitation theory [CCSD(T)]. The calculation of the hydration free energy of the proton, was derived as the free energy difference between the right and left side of eq. 2.204, so that the subsequent expression is justified:

$$\Delta G_{sol}(H^+) = G_{sol}[H^+(H_2O)_n] - G_{sol}((H_2O)_{n-1}) - G_{gas}(H^+) \quad (2.205)$$

The total free energies of the protonated water cluster and the neutral water cluster, which are displayed in the above equation were then separated based on thermodynamic assumptions into gas phase and solvent phase contribution as shown in the subsequent expression:

$$\begin{aligned} \Delta G_{solv}(H^+) = \\ \Delta G_{gas}(H^+(H_2O)_n) - \Delta G_{gas}((H_2O)_n) - \Delta G_{gas}(H^+) + \Delta G_{solv}(H^+(H_2O)_n) - \Delta G_{solv}(H_2O)_n \end{aligned} \quad (2.206)$$



**Figure 2.4** Water cluster considered in the described study to estimate the proton solvation free energy.

Having separated the free energy expressions for the protonated and neutral water cluster  $\Delta G_{gas}$  and  $\Delta G_{sol}$  were expressed in computable terms:

$$\Delta G_{gas}(M) = E(M) + E_V(M) + E_R(M) + E_T(M) - TS(M) + pV(M) \quad (2.207)$$

In the above equation (M) indicates the protonated and neutral water cluster. The translational  $E_T(M)$  and rotational corrections  $E_R(M)$  to the electronic energy were approximated in the ideal gas assumption with  $1.5 RT$  for each term.  $TS$  and  $pV$  are the entropic and translational energy contribution of the proton. The entropic energy contribution was derived from statistical thermodynamics. In fact, authors were using exactly the same expressions as the one explained in chapter 2.2.1. Thus, the free energy of the proton in the vacuum was taken to be  $-6.28$  kcal/mol.  $E(M)$ , the electronic energy of the protonated or neutral water clusters and the vibrational corrections  $E_V(M)$  were computed explicitly.

The solvation energy in the SCRF approach is dissected into three components: the electrostatic solvation part ( $\Delta G_{el}$ ); the solute-solvent dispersion interaction ( $\Delta G_{disp}$ ) and the entropy change due the ordering of the solute ( $T\Delta S_{hydr}$ ).

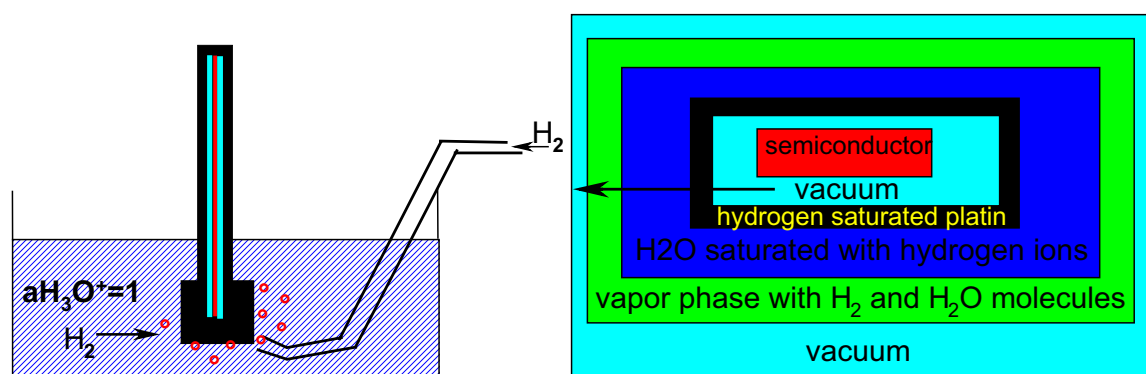
$$\Delta G_{sol} = \Delta G_{el} + \Delta G_{disp} - T\Delta S_{hydr} \quad (2.208)$$

Here another adjustable parameter of the SCRF approach came into play. The adjustable size of the cavity radii.



## 2.6 Standard hydrogen electrode

The dissociation of a molecule of molecular hydrogen  $H_2$  to  $2H^+$  and  $2e^-$  is the half-cell reaction of the normal (or standard) hydrogen electrode (NHE or SHE). Five independent measurements of the absolute half cell potential yielded values between 4.48 eV and 4.73 eV.<sup>[87, 88, 90, 124, 125]</sup> Latest measurements processed in the mid 80<sup>th</sup> by Reiss and Heller<sup>[90]</sup> and Trasatti<sup>[88]</sup> converged to 4.43 eV and 4.44 eV, respectively. In the present work the value determined by Reiss and Heller was taken when computed redox -potentials were referred relative to the SHE (see chapter 3.2 and 3.3). This value is commonly used in theoretical studies on redox-potentials. Latest examples are the computational works of Friesner and coworkers<sup>[5]</sup> and Truhlar and coworkers.<sup>[50]</sup>



**Figure:2.5.** A schematic drawing of the standard hydrogen electrode and of the open thermodynamic cycle used by Reiss and Heller to estimate the absolute half-cell potential (for explanations see text). The arrow in the right picture indicates the phases that the electron transfers in the open thermodynamic cycle.

When redox potentials are measured against the potential of a standard electrode, the potential of the standard electrode is set to zero, so that measured redox potentials are indeed relative potentials. The precision to which  $\Delta G^{\circ}_{\text{solv}}(H^+)$  can be known is limited as the potential cannot be determined from measurements alone, rather additional theoretical knowledge is required to derive the absolute half-cell potential of the standard hydrogen electrode. The following chapter summarizes how Reiss and Heller<sup>[90]</sup> approximated the potential with the help of theoretical knowledge.

Figure 2.5 gives a schematic drawing of the SHE and the different phases that are involved. The aqueous solution filled with hydronium ions is drawn in blue in the right sketch of Figure 2.5. The vapor phase next to the aqueous phase is filled with molecular hydrogen and  $H_2O$  molecules (depicted in green). Then the second vacuum of the SHE (drawn in cyan) occurs. The absolute half-cell potential of the SHE is defined as the *difference between the electrochemical potential*

of the electron, in the aqueous solution of the SHE, and the vacuum level just outside of the vapor phase (H, and H<sub>2</sub>O) of the SHE. Clearly two interface potentials are bridged by this difference.

To estimate the absolute half-cell potential the “work function” was employed to describe the potential of an electron between two phases.

$$\phi = -\phi_a - \frac{\bar{\mu}}{e^-} \quad (2.209)$$

In eq. 209  $e^-$  is the electronic charge and  $\phi_a$  is the electrostatic potential in the vacuum just outside of an electronic conductor.  $\bar{\mu}$  is the electrochemical potential of the electron in the bulk of the conductor. Both  $-\phi_a$  and  $\bar{\mu}$  must be measured with respect to the same reference level of energy. This means that the energy component of  $\bar{\mu}$ , that differs from the entropic component, must be measured relative to this reference energy level.

The choice of the reference point is the improvement in the work of Reiss and Heller and distinguishes this study from previous ones. Here, the electrochemical potential of the electron,  $\bar{\mu}$ , just outside of the semiconductor is considered as the reference point. Just outside of the semiconductor  $\phi_a$  is equal to zero and therefore eq. 2.209 can be reformulated, so that the redox potential of the half-cell is expressed through  $\bar{\mu}$ , which is not anymore dependent on  $\phi_a$ .

$$\bar{\mu} = -\frac{\phi}{e} \quad (2.210)$$

The authors name this method the *vacuum scale method*, because the potential of the electron, which is equivalent to the fermi level of the electron is referred to this vacuum reference point. Albeit the work function and therefore  $\phi_a$  is a free energy it can only be measured under none reversible conditions, whereas generally thermodynamic quantities are measured under reversible conditions. In fact, this means that the work function is measured in an open thermodynamic cycle, which prevents the electron from being in a true vacuum. The contact potential between two contacting phases is known as the Calvani potential. An open thermodynamic cycle is defined, which allows the measurement of the potential of the electron between the two phases that comprise the absolute half-cell potential. In an open *thermodynamic cycle* the electron is moved from the semiconductor vacuum interface (“vacuum just outside the semiconductor” = reference point) through a layer of platinum, encompassed with hydrogen and the aqueous solution, which contains hydrogen ions (H<sup>+</sup>) into the neighboring vapor phase that contains H<sub>2</sub> and H<sub>2</sub>O molecules and finally into the vacuum. The electron is not brought back through the semiconductor into the initial vacuum and therefore it is named open cycle. From the above thermodynamic considerations  $\bar{\mu}$  is now expressed as:

$$\bar{\mu} = V_{FB} + V_{DP} - V_{ZC} - \phi_{110} \quad (2.211)$$

$\phi_{110}$  is the potential at the semiconductor vacuum interface. In the present study the experimentalists considered indium phosphide as the semiconductor of choice. Its potential was measured from an isolated single crystal of p-indium phosphide with 110 points in contact with the surface. A Schottky diode is constructed by a layer of platinum, saturated with hydrogen at 1 atm, which is deposited onto a 110 surface of the same p-type indium phosphide. At the Schottky diode a flat-band potential ( $V_{FB}$ ) and a dipole potential ( $V_{DP}$ ) is measured.  $V_{FB}$  is determined by a Mott–Schottky plot. This plot is a *considerable source of error* as it is based on an approximated theory.  $V_{ZC}$  is the level bending in solution at the platinum water interface. A diffuse space

charge layer and a fixed dipole layer contribute to  $V_{ZC}$ . Measurements of this quantity cause problems and the necessary values for the hydrogen electrode had to be derived by extrapolation.

Measurements yielded  $\phi_{110}=5.78$  V,  $V_{FB} = 1.05$  V,  $V_{DP} \approx 0$  V and  $V_{ZC} = -0.3$  V. Substitution of these values into eq.2.211 yields  $\bar{\mu}$ :

$$\bar{\mu}_{redox} = \mu_{H^+/H_2} = 4.43 \text{ V} \quad (2.212)$$

The 4.43 V can also be considered as absolute half-cell potential of the standard hydrogen electrode.

## 2.7 Computational procedure to compute $pK_a$ values and one-electron reduction potentials for organic molecules:

In the following three chapters I give the computational details to estimate  $pK_a$  values and one-electron reduction potentials in solution. Computation of absolute  $pK_a$  values and one-electron reduction potentials for organic molecules requires the determination of the reaction free energy in the condensed phase ( $\Delta G_R$ ) (see definition in chapter 2.1). Using suitable thermodynamic cycles (see 2.1.1 and 2.1.2) I computed the reaction free energy from contributions of  $\Delta G_{gas}$  and  $\Delta \Delta G_{solv}$ .

$\Delta G_{gas}$  was determined using adequate QC methods. For the determination of solvation energies, a *one-step* and a *two-step* procedure are in use. In the one-step procedure the solvation energy is evaluated directly within a suitable QC method using a reaction field or the Poisson equation selfconsistently. In the *two-step* procedure first a QC computation in vacuum is performed to determine the atomic partial charges. These charges are calculated by matching the electrostatic potential (ESP) based on the QC wave function with the ESP generated by the atomic point charges using the RESP procedure (see below and chapter 2.3). In the second step the electrostatic energies of solvation are evaluated from this point charge distribution separately by solving the Poisson equation. This *two-step* procedure is the method of choice to calculate the  $pK_a$  shift between solvent and protein environment, since modeling the detailed charge distribution of a protein in a QC computation is a formidable task (see introduction). Interestingly, while the two-step procedure was successfully applied to compute  $pK_a$  values in a protein the one-step procedure was generally used to evaluate solution  $pK_a$ . However, a two-step method was also used to evaluate solvation energies of different organic molecules.

### 2.7.1 Computation of $pK_a$ values for closed-shell organic molecules:

#### Computation of gas phase free energies and enthalpic energies (PA values)

All QC calculations, which were necessary to calculate absolute  $pK_a$  values for closed-shell systems (for results see chapter 3.1) were performed with the quantum chemistry software JAGUAR V4.2.<sup>[128]</sup> According to eq. 2.8  $\Delta G_{gas}$  was estimated from contributions of  $G_{gas}(A^*)$  and  $\Delta G_{gas}(H^+)$ .  $G_{gas}(A^*)$  is expressed through the electronic energy  $E_0$ , the zero point vibrational energy ZPVE and the thermal correction  $\Delta G_{0K \rightarrow 298K}^\circ$ . To obtain  $pK_a$  values for the compounds

considered in chapter 3.1  $G_{gas}^{298K}(A^*)$  was computed for all protonated ( $A^* = AH$  or  $AH^+$ ) and all deprotonated states ( $A^* = A^-$  or  $A$ ). All energy computations are based on individual *geometry optimizations*, which were performed using the DFT functional B3LYP with the 6-31G\*\* basis sets. The  $E_0$  contributions to  $G_{gas}^{298K}(A^*)$  were estimated either with the DFT functional B3LYP or Becke<sup>1/2</sup> using the basis set cc-pVQZ. ZPVE and  $\Delta G_{0K \rightarrow 298K}^\circ$  values were estimated with B3LYP or Becke<sup>1/2</sup> using the smaller basis set 6-31G\*\*.

The free energy contribution of the proton to  $\Delta G_{gas}$  was described using the following relation:  $\Delta G_{gas}(H^+) = E_{gas}(H^+) - 298K\Delta S(H^+)$ , here  $E_{gas}(H^+) = 2.5 RT$  (with RM298  $K=1.48$  kcal/mol), where  $1.5 RT$  is the enthalpy contribution of the three translational degrees of freedom of  $H^+$  and  $1 RT = pV$  is the excess mechanical energy required to generate two particles ( $A^-, H^+$ ) out of one ( $AH$ ) as estimated from the ideal gas model.  $-\Delta S$  is the entropic contribution, which is estimated from the ideal gas model.

PA values are defined as the enthalpic energy differences between deprotonated and protonated forms and the contribution of the enthalpic energy of the proton in the gas phase (2.10). For computational details of  $E_0$  and ZPVE see the above paragraph. For  $E_{gas}(H^+)$  I used  $E_{gas}(H^+) = 2.5 RT$  (with RM298  $K=1.48$  kcal/mol (see explanation in the previous paragraph). Note that in some publications instead of PA, which are enthalpic quantities, gas-phase acidities/basicities are given that refer to Gibbs free energies. The corresponding PA value is then lower by 7.76 kcal/mol.

### Computation of electrostatic solvation free energies:

**Determination of atomic partial charges.** Unless otherwise stated partial atomic charges were derived from the molecular ESP generated by QC computations on optimized geometries (B3LYP/6-31G\*\*) at the Becke<sup>(1/2)</sup>/6-31G\*\* or B3LYP/6-31G\*\*. I used the RESP<sup>[69, 70]</sup> procedure to determine the atomic partial charges of a solute molecule. RESP matches the solute ESP<sup>[109]</sup> calculated with a QC method on grid points in the neighborhood of the considered molecule with the ESP generated by the atomic partial charges to be determined, while constraining the total charge of the molecule at the appropriate value. Thereby the ESP is defined on grid points outside of the vdW spheres (defined by radius  $R_{vdW}$ ) of each atom up to a maximum radius  $R_{max}$  placed equidistantly on a system of rays emerging from each atom center. The corresponding  $R_{vdW}$  values are 1.95 Å, 1.6 Å, 1.7 Å, 1.8 Å, 1.96 Å, the  $R_{max}$  values are 9.0 Å, 8.3 Å, 8.5 Å, 8.7 Å, 9.0 Å for C, H, O, N, Cl atoms, respectively. With a direct least square fit of atomic partial charges molecular dipole and quadrupole moments are often not well reproduced. This is due to charges of buried atoms whose influence on ESP may be poorly defined, since all ESP grid points lie outside of the vdW volume of all atoms of the molecule. To circumvent these effects I applied the RESP procedure with a hyperbolic penalty function applied to all atoms except hydrogens with strength parameter  $a = 0.001$  au.

**Solvation energies.** Electrostatic solvation free energies were computed according to the eq 2.13. Note that for titratable compounds that are neutral (carboxylic acids, benzoic acids, imides, phenols) or cationic (imidazoles and pyridines) in the protonated state we have  $AH^* \equiv AH$  and  $A^\# \equiv A^-$  or  $AH^* \equiv AH^+$  and  $A^\# \equiv A$ , respectively.

From solvation energies of a titratable compound I determined the electrostatic energy contributions only and ignored contributions from van-der-Waals (vdW) interactions and reorganization of the solvent in the presence of a solute molecule, i.e. the hydrophobic effect,

which is particularly difficult to evaluate. Unless there are charge specific solvent structures the influence of these interactions cancels in the difference between protonated and deprotonated species.

The program SOLVATE from the electrostatic energy program suite MEAD (macroscopic electrostatics with atomic detail) was employed to calculate the electrostatic energies generated if a solute molecule with its atomic partial charges is placed from vacuum ( $\epsilon = 1$ ) in a dielectric continuum of vanishing ionic strength with dielectric constant  $\epsilon = 80$  while  $\epsilon = 1$  is kept in the vdW volume of the solute. Unless explicitly specified (see chapter 3.1.5) all computed  $pK_a$  and  $\Delta\Delta G_{\text{solv}}$  values were computed with a solute dielectric constant  $\epsilon = 1$  in vacuum and solvent. The solute/solvent boundary was defined through a solvent probe radius of 1.4 Å. A focusing procedure was used to solve the Poisson equation with an initial low resolution grid of  $101^3$  points and 1.0 Å step size and a second high resolution grid of  $101^3$  points and 0.25 Å step size centered at the solute molecule. To estimate free energies of solvation we used the following vdW radii: for carbon in  $\text{CH}_3$  and  $\text{CH}_2$  groups  $R_{\text{CHn}} = 2.0$  Å, else  $R_{\text{C}} = 1.5$  Å,  $R_{\text{Cl}} = 1.9$  Å,  $R_{\text{O}} = 1.4$  Å,  $R_{\text{N}} = 1.4$  Å and  $R_{\text{H}} = 1.2$  Å, except for the vdW radii of the titratable proton, where we used  $R_{\text{p}} = 1.0$  Å. The vdW radii are close to the values of Richardson et al.,<sup>[46]</sup> who recommended the usage of small vdW radii according to suggestions from Bondi<sup>[129]</sup> to estimate solvation energies of organic molecules. As in the study of Richardson I also used an oxygen vdW radius of 1.4 Å, which is the Pauling radius also applied by Tunon et al.<sup>[130]</sup> These radii are also used within the quantum chemical software package Gaussian<sup>[131]</sup> to compute Merz and Kollman charges, except for  $\text{CH}_n$ , N, and Cl, where they used 1.5 Å, 1.5 Å, and 1.7 Å, respectively.

A balanced combination of the QC method applied to compute the ESP and the vdW radii is crucial to obtain accurate free solvation energies. The distinction between  $R_{\text{C}} = 1.5$  Å and  $R_{\text{CH}_2} = R_{\text{CH}_3} = 2.0$  Å has only a moderate impact on the computed solvation energies. Note that the vdW radii used for the RESP fitting procedure have an influence on the atomic partial charges and thus indirectly also on the electrostatic energies, while the choice of vdW radii used to define the solute solvent boundary have a direct impact on the electrostatic energies. For the protonated imidazole and its substituents we have two equivalent deprotonation sides, but only one leaving proton. For imidazoles we assigned a vdW radius of 1.0 Å to the titratable proton and 1.2 Å to the hydrogen atom bound to the second nitrogen atom, which is consistent with the values used above.

### 2.7.2 Computation of one-electron reduction potentials in protic and aprotic solvents:

All redox-active substances considered in chapter 3.2 refer to single electron-transfer equilibria between a neutral and anionic charge state. Except for *p*-benzoquinone and *p*-benzodithiyl, which are radicals in the anionic charge state, all redox-active substances are radicals in the neutral charge state. Radicals are open shell systems that bear an unpaired electron. The accurateness of the post Hartree-Fock method G3MP2 and the DFT functional B3LYP was evaluated in terms of EA values, which are the enthalpic contribution to  $\Delta G_{\text{gas}}$  for electron-transfer equilibria. The contributions of the free solvation energies were calculated using the two-step procedure. To compute free solvation energies in different solvent types (protic and aprotic) two sets of vdW radii of solvation were used.

#### Computation of gas phase free energies and enthalpic energies (EA values):

$G_{\text{gas}}^{T=298\text{K}}(A^*)$  (see eq. 2.26) was computed for all radicals ( $A^* = A^\bullet$  or  $A^{\cdot-}$ ) and non-radicals ( $A^* = A$  or  $A^-$ ). The enthalpic quantity,  $H_{\text{gas}}^{T=298\text{K}}(A^*)$ , was used to derive adiabatic EA values, according to the conventional definition (see eq.2.24).  $H_{\text{gas}}^{T=298\text{K}}$  contains only term one and two of eq. 2.26. The QC method G3MP2<sup>[78]</sup> computes electronic energies at the MP2(full)/6-31G(d) level of theory. A series of single point calculation at different level of theory is carried out to obtain the electronic energy (see chapter 2.2.5 for more details). The zero point corrections are obtained from scaled HF frequencies.

Calculated  $E_{\text{redox}}^\circ$  values based on DFT theory were obtained using the B3LYP functional with the aug-cc-pvTZ basis set, which assigns diffuse function to all atoms. These computations employed optimized geometries from the B3LYP/6-31G\*\* of theory for each redox state. The thermal corrections and the zero point vibrational energies were computed at the same level of theory than the geometry optimization.

### Computation of electrostatic solvation free energies:

To compute the electrostatic solvation free energies I used the atomic partial charges obtained with the Merz–Kollman RESP<sup>[69, 70]</sup> procedure for molecular geometries optimized with G3MP2 or B3LYP/6-31G\*\*, respectively (for more details see chapter 2.7.1 “computation of free electrostatic solvation energies”). The Poisson equations for these molecular charge distributions were solved with the program SOLVATE from the electrostatic energy program suite MEAD (macroscopic electrostatics with atomic detail).<sup>[58, 113, 114]</sup> In the present work, I estimated one-electron reduction potentials in water, AcN and DMAc using  $\epsilon_{\text{H}_2\text{O}} = 80$ ,  $\epsilon_{\text{AcN}} = 37.5$  and  $\epsilon_{\text{DMAc}} = 37.8$  to compute solvation from electrostatic energies. Note that the values of the dielectric constants for AcN and DMAc found in the “Dielectric Constant Reference Guide”<sup>[132]</sup> are virtually identical. The solute/solvent boundary was defined through a solvent probe radius of 1.4 Å for water and 2.0 Å for AcN and DMAc. Although dependencies on solvent probe radii are weak and therefore not critical. A focusing procedure was used to solve the Poisson equation with an initial low resolution grid of  $101^3$  points and 1.0 Å step size and a second high resolution grid of  $101^3$  points and 0.25 Å step size centered at the solute molecule.

In combination with the G3MP2 QC method a single set of vdW radii for solute atoms was sufficient to estimate the electrostatic solvation free energies in water for all considered redox-active compounds. These are the same vdW radii used for pK<sub>a</sub> computations,<sup>[Busch, 2004 #34]</sup> with the exception that the somehow artificial discrimination between aliphatic and non-aliphatic carbons was dropped in the present work. The solute vdW radii used for water are: 1.5 Å, 1.9 Å, 1.4 Å, 1.4 Å, 1.2 Å, 2.025 Å for C, Cl, O, N, H and S, respectively. The adiabatic EA values computed with B3LYP for sulfur-centered radicals were by more than 200 meV lower than the corresponding experimental values (see Table 3.15). To regain agreement between measured computed  $E_{\text{redox}}^\circ$  values for B3LYP of sulfur-centered radicals in water the deviations in EA was compensated by using a smaller sulfur radius of 1.80 Å for the evaluation of the solvation energies.

In AcN and DMAc redox-active compounds with G3MP2 geometries were solvated with the following set of vdW radii: 1.725 Å, 2.1 Å, 1.7 Å, 1.7 Å 1.2 Å and 2.3 Å for C, Cl, O, N, H, and S, respectively. Except for the hydrogen atom these vdW radii for solvation were enhanced by about 15% for the aprotic solvents relative to the values used for protic solvents like water. Again, to obtain agreement between computed and measured  $E_{\text{redox}}^\circ$  values for sulfur-centered

radicals in AcN/DMAc we decreased the sulfur vdW radius from 2.3 Å used for G3MP2 to 2.0 Å for B3LYP geometries.

### 2.7.3 Computation of $pK_a$ values for open-shell organic molecules:

Estimated  $pK_a$  values of semiquinones ( $Q^{\cdot-}$  and  $QH^{\cdot}$ ) and hydroquinones ( $Q^{2-}$ ,  $QH^-$ ,  $QH_2$ ) are based on  $\Delta G_{\text{gas}}$  values, which were obtained with the QC method G3MP2 (see chapter 3.3). The QC method G3MP2 was applied because semiquinones are open-shell systems. Hence, I calculated the individual free energies for  $Q^{\cdot-}$ ,  $QH^{\cdot}$ ,  $Q^{2-}$ ,  $QH^-$ ,  $QH_2$  and in the same way than the individual free energies for redox equilibria (see chapter 2.7.2). Free electrostatic solvation energies were computed as described in chapter 2.7.1, with the only exception that instead of B3LYP/6-31G\*\* optimized geometries were taken from the G3MP2 computation.