APPENDIX

Appendix A

Theoretical Background

1. Calculation of NMR Chemical Shifts

NMR spectroscopy is a very important analytical tool for the identification and characterization of molecules. However, since there is no simple correlation between the measured chemical shifts and structural parameters, the interpretation of experimental NMR spectra is not trivial and can be in many cases quite involved. The ability to calculate NMR chemical shifts *ab initio* is therefore a very important advancement in quantum chemistry. The calculation of chemical shifts can provide in many cases the necessary information for the correct interpretation of experimental NMR spectra.

In recent years, calculations of NMR shieldings have included the effects of relativity non-perturbationally. Schreckenbach and Ziegler implemented a scalar-relativistic DFT routine within the framework of the ADF package for calculation NMR shielding tensors. In their calculations they used: gauge-including atomic orbitals (GIAO), a frozen core approximation, and the scalar relativistic draw in and mass-velocity terms.

1.1 Defining the NMR shielding tensor

In the absence of electrons, the magnetic field at the nucleus is the same as the external magnetic field. In the presence of electrons the magnetic field induces electronic currents that create their own magnetic fields. As a result the magnetic field at the nucleus may be different to the external magnetic field. Paramagnetic currents reinforce the external field and thus de-shield the nucleus. Diamagnetic currents counter the external field and thus shield the nucleus.

The magnetic field at the nucleus and the external magnetic field can be related by a 3X3 NMR shielding tensor:

$$B_{@\,nucl} = \delta B_{external}$$

$$\begin{bmatrix} B_1 \\ B_2 \\ B_3 \end{bmatrix}_{\textbf{@nucl}} = \begin{bmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{yx} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{zx} & \sigma_{zy} & \sigma_{zz} \end{bmatrix} \begin{bmatrix} B_1 \\ B_2 \\ B_3 \end{bmatrix}_{\textbf{external}}$$

In a crystal, the three components: $\sigma_{xx}, \sigma_{yy}, \sigma_{zz}$ can be distinguished. In solution, because the molecules are rapidly tumbling, only the trace of the tensor can be determined. One third of the trace of the tensor gives the NMR isotropic shielding constant:

$$\sigma = \frac{1}{3}\left(\sigma_{xx} + \sigma_{yy} + \sigma_{zz}\right) \tag{1}$$

In most experiments, the isotropic shielding constant is not measured directly, but rather its relation to a standard is measured. This gives the NMR shift:

$$\delta^{sample} = \sigma^{reference} - \sigma^{sample} \tag{2}$$

1.2 Calculating the NMR shielding tensor

If we have an expression for the energy of a system in terms of the nuclear magnetic moment μ and the external magnetic field B, then the NMR shielding tensor can be determined from:

$$\sigma_{kt} = \frac{\partial^2 E}{\partial \mu_t \partial B_k} \tag{3}$$

Where k and t are the components of the external magnetic field and induced magnetic moment, respectively. Considering SCF theory in general, the SCF energy and Fock matrix are

$$E = \langle hP \rangle + \tfrac{1}{2}\langle PG_{2e}(P) \rangle + E_{xc} + V \tag{4}$$

$$F = \frac{\partial E}{\partial P} = h + G_{2e}(P) + G_{xc} = h + G \tag{5}$$

Where P is the density matrix, h is the one-electron Hamiltonian, and V is the nuclear repulsion energy and

$$G_{2e}(P)_{\mu\nu} = \sum_{\lambda\sigma} P_{\lambda\sigma} \langle \mu\lambda \| \nu\sigma \rangle \tag{6}$$

where $\langle \mu\lambda \| \nu\sigma \rangle$ is the antisymmetrized two-electron integral over spin orbitals $\chi_\mu, \chi_\nu, \chi_\lambda$ and χ_σ which includes a coefficient for Hartree-Fock exchange, C_{HFX} ,as follows

$$\langle \mu\lambda \| \nu\sigma \rangle = \int \chi_\mu^*(1)\chi_\lambda^*(2)\frac{1}{r_{12}}[\chi_\nu(1)\chi_\sigma(2) - C_{HFX}\,\chi_\sigma(1)\chi_\nu(2)]d\tau_1\tau_2 \tag{7}$$

The exchange-correlation energy is

$$E_{XC} = \int f\left(\rho_\alpha, \rho_\beta, \gamma_{\alpha\alpha}, \gamma_{\alpha\beta}, \gamma_{\beta\beta}\right) d\vec{r} \tag{8}$$

where α and β refer to the spin components and f is a general first-order exchange-correlation functional and does not include an explicit magnetic field dependent term. The spin densities and the density gradient invariant are given as follows:

$$\rho_\alpha = \sum_{\mu\nu} P_{\mu\nu}^{\alpha} \chi_\mu \chi_\nu, \tag{9}$$

$$\gamma_{\alpha\alpha} = \nabla\rho_\alpha \cdot \nabla\rho_\alpha, \quad \gamma_{\alpha\beta} = \nabla\rho_\alpha \cdot \nabla\rho_\beta, \tag{10}$$

$$\nabla\rho_\alpha = \sum_{\mu\nu} P^\alpha \nabla\left(\chi_\mu \chi_\nu\right) \tag{11}$$

and G_{xc}, the exchange-correlation piece of F is

$$(G^\alpha_{xc})_{\mu\nu} = \int \left[\frac{\partial f}{\partial\rho_\alpha} \chi^*_\mu \chi^*_\nu + \left(2\frac{\partial f}{\partial\gamma_{\alpha\alpha}} \nabla\rho_\alpha + \frac{\partial}{\partial\gamma_{\alpha\beta}} \nabla\rho_\beta \right) \nabla\left(\chi^*_\mu \chi^*_\nu\right) \right] d\vec{r} \tag{12}$$

with a similar expression for G^β_{xc}. The coefficient for Hartree-Fock exchange, C_{HFX}, in equation (7) is one of Hartree-Fock, zero for pure DFT and non-zero for hybrid methods. Similarly, f is 0 for Hartree-Fock theory.

Using the notation, where first and higher derivatives are denoted by superscripts specifying the variable(s) of differentiation, the expression for the shielding tensor for nucleus *N* becomes:

$$\sigma^N_{kt} = \frac{\partial^2 \zeta}{\partial\mu_{Nt}\partial B_k} = \left\langle h^{(B_i, m_{Nt})} P \right\rangle + \left\langle h^{\mu_{Nt}} P^{B_k} \right\rangle \tag{13}$$

where the derivatives of the Hamiltonian are given by

$$h^{m_{Nt}}_{\mu\nu} = \left\langle \chi_\mu \middle| \hat{h}^{(B_k, m_{Nt})} \middle| \chi_\nu \right\rangle; \tag{14}$$

$$\hat{h}^{m_{Nt}} = -\frac{i}{c} \frac{\left|(\vec{r} - \vec{R}_N) \times \vec{\nabla}\right|_t}{\left|\vec{r} - \vec{R}_N\right|^3} \tag{15}$$

$$h_{\mu\nu}^{(B_k,m_{Nt})} = \left\langle \chi_\mu \middle| \hat{h}^{(B_k,m_{Nt}} \middle| \chi_\nu \right\rangle; \tag{16}$$

$$\hat{h}^{(B_k,m_{Nt})} = \frac{1}{2c^2} \frac{\vec{r}(\vec{r}-\vec{R}_N)\partial_{kt} - \vec{r}_k(\vec{r}-\vec{R}_N)_t}{\left|\vec{r}-\vec{R}_N\right|^3} \tag{17}$$

In addition to the derivatives of the one-electron Hamiltonian, the calculation of the nuclear magnetic shielding tensor also requires the derivative of the density matrix with respect to the magnetic field, P^{B_k}. This is obtained via solution of the coupled-perturbed (CP) equations for the appropriate perturbation.

Separating P into its occupied-occupied and virtual-occupied blocks, the CP equations for an external magnetic field perturbation B_k are

$$FP_{ov}^{B_k} - P_{ov}^{B_k}F - G(P_{ov}^{B_k} + P_{vo}^{B_k})_{ov} = h_{ov}^{B_k} + G_{ov}^{B_k}(P)_{ov} - FS_{ov}^{B_k} + G(S_{oo}^{B_k})_{ov}$$

$$P_{oo}^{B_k} = -S_{oo}^{B_k}, \qquad P_{vv}^{B_k} = 0 \tag{18}$$

where the subscripts *oo* and *ov* refer to the occupied-occupied and virtual-occupied blocks of the matrix, respectively. S is the overlap matrix and *G(X)* is define as

$$G(X)_{\mu\nu} = \sum_{\lambda\sigma} X_{\lambda\sigma} \langle \mu\lambda \| \nu\sigma \rangle \tag{19}$$

where $\langle \mu\lambda \| \nu\sigma \rangle$ was defined as the same. The term $G^{B_k}(P)_{\mu\nu}$ which results from the derivative of the basis function with respect to the field is

$$G^{B_k}(P)_{\mu\nu} = \sum_{\lambda\sigma} P_{\lambda\sigma} \langle \mu\lambda \| \nu\sigma \rangle^{B_k} + G_{xc}^{B_k} \tag{20}$$

Note that there is no exchange-correlation contribution to these equations because the standard functional considered here depends only on $\rho(\vec{r})$ and its derivative and not on the magnetic field explicitly. If C_{HFX} is taken to be one in equation (7) with $f = 0$ in equation (12), then the CPHF equation result from equation (16). For an imaginary perturbation, the coulomb contributions in both $G(P_{ov}^{B_k} + P_{vo}^{B_k})_{ov}$ and $G(S_{oo}^{B_k})_{ov}$ equation (18) vanish leaving only an explicit dependence upon the field (via the basis functions), the DFT exchange-correlation contribution to these two terms is zero. For hybrid methods, which include a mixture of Hartree-Fock exchange and DFT exchange correlation, only Hartree-Fock exchange is present in these two terms. As gauge-invariance is achieved in different ways, the gauge including atomic orbital (GIAO) method differs at this point in the formation of the right-hand side of equation (18).

1.3 GIAO Method

The GIAO method (Wolinski et al., 1990: Gauss 1993) for calculating magnetic properties uses the following explicit field-dependent basis functions.

$$\chi_\mu(\vec{B}) = \exp[-\frac{i}{2c}(\vec{B}\times\vec{R}_\mu)\cdot\vec{r}]\chi_\mu(\vec{0}) \tag{21}$$

where $\vec{R}_\mu$ is the position vector of basis function χ_μ and $\chi_\mu(\vec{0})$ denotes the usual field independent basis functions. The derivative of a field dependent basis function with respect to the magnetic field direction i is

$$\chi_\mu^{B_k} = -\frac{i}{2c}(\vec{R}_\mu\times\vec{r})_k\,\chi_\mu(\vec{0}) \tag{22}$$

Defining $\vec{R}_{\mu\nu} = \vec{R}_\mu - \vec{R}_\nu$, the derivative of the overlap S and Hamiltonian h matrices with respect to the external magnetic field in the AO basis are

$$S_{\mu\nu}^{B_k} = \frac{i}{2c}\left(\vec{R}\times\vec{r}\right)\left\langle\chi_\mu|\chi_\nu\right\rangle \tag{23}$$

$$h_{\mu\nu}^{B_k} = \left\langle \chi_\mu^{B_k} \middle| \hat{h} \middle| \chi_\nu \right\rangle + \left\langle \chi_\mu \middle| \hat{h}^{B_k} \middle| \chi_\nu \right\rangle + \left\langle \chi_\mu \middle| \hat{h} \middle| \chi_\nu^{B_k} \right\rangle \tag{24}$$

$$= \frac{i}{2c}\left(\vec{R}_{\mu\nu} \times \left\langle \chi_\mu \vec{r} \middle| \vec{h} \middle| \chi_\nu \right\rangle - \left\langle \chi_\mu \middle| \vec{r} \times \nabla \middle| \chi_\nu \right\rangle + \vec{R}_\nu \left\langle \chi_\mu \middle| \nabla \middle| \chi_\nu \right\rangle \right) \tag{25}$$

where the derivative of the Hamiltonian operator is

$$\hat{h}^{B_k} = -\frac{i}{2c}\left(\vec{r} \times \nabla \right)_k \tag{26}$$

The term $G_{xc}^{B_k}$ in equation (20), which results from the derivative of the field-dependent basis functions with respect to the field is

$$\left(G_{xc}^{\alpha}\right)_{\mu\nu}^{(B_k)} = \int \left[\frac{\partial}{\partial \rho_\alpha}\left(\chi_\mu^{*} \chi_\nu \right)^{B_k} + \left(2\frac{\partial f}{\partial \gamma_{\alpha\alpha}} \nabla \rho_\alpha + \frac{\partial f}{\partial \gamma_{\alpha\beta}} \nabla \rho_\beta \right) \cdot \nabla (\chi_\mu^{*} \chi_\nu)^{B_k} \right] d\vec{r} \tag{27}$$

where
$$\left(\chi_\mu^{*} \chi_\nu \right)^{B_k} = \frac{i}{2c}\left(\vec{R}_{\mu\nu} \times \vec{r} \right)_k \chi_\mu^{*} \chi_\nu \tag{28}$$

and χ for the component of the gradient,

$$\nabla_x (\chi_\mu^{*} \chi_\nu)^{B_k} = \frac{i}{2c}\left(\vec{R}_{\mu\nu} \times \vec{r} \right)_k \left[\left(\nabla_x \chi_\mu^{*} \right) \chi_\nu + \chi_\mu^{*} \left(\nabla_x \chi_\nu \right) \right] + \frac{i}{2c} \chi_\mu^{*} \chi_\mu \left[\left(\vec{B} \times \vec{R}_{\mu\nu} \right)_x^{B_k} \right]_{\vec{B}=0} \tag{29}$$

Three sets of equation (18) are then solved, one for each magnetic field direction. How well can we actually calculate theoretically NMR shielding? If we expect to make reasonable predictions and to perform meaningful analyses, the theoretical calculations must agree, at least semiquantitatively, with the experiment. Theoretical predictions are especially useful in NMR spectroscopy, because the interpretation of experimental information is not always straightforward. In particular, the spectral analysis for unusual molecules is difficult because there is no clear relationship between qualitative molecular and electronics structure and the NMR parameters.

2. Integral Equation Formalism Polarized Continuum (IEF-PCM) Solvation Continuum Model

In the IEF-PCM model, the solvent is represented by a homogeneous continuum medium which is polarized by the solute placed in a cavity built in the bulk of the dielectric. The solute-solvent interactions are described in terms of a solvent reaction potential. The basic hypothesis is that one can always define a new energetic functional, the free energy *G*, depending on the solute electronic wave function.

$$G(\psi) = \langle \psi | \hat{H}^0 | \psi \rangle + \left\langle \psi \left| \frac{1}{2} \hat{V}^R \right| \psi \right\rangle \qquad (30)$$

where $\hat{H}^0$ is the Halmiltonian describing the isolated molecule and $\hat{V}^R$ represents the solvent reaction operator. By applying the variational principle to this functional, we can derive the nonlinear Schrödinger equation specific for the solvated system.

In general, the computational strategy formulated to define the reaction potential is based on a modelization of the solvent interactions according to the theory of intermolecular forces. Within this framework, the energetic quantity *G* and the corresponding reaction operator $\hat{V}^R$ are written as a sum of contributions of different physical origin related to dispersion, repulsion, and electrostatic forces between solute and solvent molecules. However, we shall consider the electrostatic part of the interactions only.

The electrostatic problem of a charge distribution, ρ_M, embedded in a cavity, C, (within which the permittivity is assumed to be equal to 1) surrounded by an isotropic continuum dielectric with a given permittivity, ε, can be expressed as follows:

$$\begin{cases} -\Delta V = 4\pi\rho_M & \text{in C} \\ -\varepsilon\Delta V = 0 & \text{outside C} \\ [V] = 0 & \text{on } \Sigma \\ [\partial_x V] = 0 & \text{on } \Sigma \end{cases} \tag{31}$$

where V indicates the electrostatic potential and Σ is the cavity surface. The jump condition, $[V] = 0$, means that the potential V is continuous across the interface Σ, i.e., $V_e - V_i = 0$ on Σ where the subscripts e and i indicate the exterior and the interior of the molecular cavity, respectively. The equality $[\partial_x V] = 0$ is a formal expression of the jump condition of the gradient of the potential; for a homogeneous isotropic dielectric, it takes the well-known form

$$\left(\frac{\partial V}{\partial n}\right)_n - \varepsilon\left(\frac{\partial V}{\partial n}\right)_e = 0 \tag{32}$$

where n is the outward pointing unit vector perpendicular to the cavity.

Within the integral equation formalism (IEF), one can transform the first two equations in system in equation (31) into integral equations on the surface Σ that can be solved with standard numerical methods. The solution of system in equation (31) is thus reduced to a sum of two electrostatic potentials, one produced by ρ_M in vacuo and the other due to a surface charge distribution σ placed on the interface which arises from the polarization of the dielectric medium:

$$V(x) = V_M(x) + V_\sigma(x) = \int_{R^3} \frac{\rho_M(y)}{|x-y|} dy + \int_{\Sigma} \frac{\sigma(s)}{|x-s|} ds \tag{33}$$

where the integral in the first term is taken over the entire three-dimensional space. The problem is then shifted to the definition of the proper apparent surface charge (ASC), σ. In computational practice, use is made of a partition of the cavity surface into small regions, called tesserae, with known area, a_k. In the limit of a sufficiently accurate mapping, one can always approximate the continuum distribution σ on each tessera with a single-value quantify to define the equivalent sets of pointlike charges as $q(s_k) = \sigma(s_k)\, a_k$ where s_k indicates the representative point of tessera k (i.e., the point at which we compute σ).

In this scheme, the reaction potential, $\hat{V}^R$, to be introduced in the effective Hamiltonian is reduced to one-electron operators depending on $q(s_k)$, and thus, the IEF-PCM method can be straightforwardly applied to different levels of the quantum mechanical description and modeled to include various concepts and approaches provided by the general quantum mechanical theory. The important new aspect to be taken into account is the introduction of an additional nonlinear character not present in isolated systems; the apparent charges, $q(s_k)$, depend on the solute charge distribution they contribute to modify.

2.1 Nuclear Shielding for an IEF solute

For a molecular solute, the nuclear magnetic shielding tensor σ^X of a nucleus, X, is expressed as mixed second derivatives of the free energy functional, G, with respect to the external magnetic field, B, and the nuclear magnetic moment, μ^X,

$$\sigma_{ij}^X = \frac{\partial^2 G}{\partial B_i \partial \mu_j^X} \tag{34}$$

Where Bi and μ_j^X ($i, j = x, y, z$) are the Cartesian components of the external magnetic field, B, and of the nuclear magnetic moment, μ^X, respectively.

The presence of the magnetic field introduces the problem of the definition of the origin of the corresponding vector potential. However, because σ is a molecular property, it must be invariant with respect to changes of the gauge origin. To obtain this gauge invariance in the *ab initio* calculations, one can introduce gauge factors into the atomic orbitals of the basis set in such a manner that the results are independent of the gauge origin even though the calculation is approximate. Inclusion of gauge factors in the atomic orbitals may be accomplished by using gauge invariant atomic orbitals (GIAO)

$$X_v(B) = X_v(0)\exp\left[-\frac{i}{2c}(B \times R_v)\cdot r\right] \tag{35}$$

Where R_v is the position vector of the basis function, and $X_v(0)$ denots the usual field-independent basis function.

The GIAO method is used in conjunction with analytical derivative theory; in this approach, the magnetic field perturbation is treated in an analogous way to the perturbation produced by changes in the nuclear coordinates. For a solute described at Hartee-Fock or DFT level with expansion of the molecular orbitals over the previously defined field-dependent basis set, the components of the nuclear magnetic shielding tensor are obtained as

$$\sigma_{ij}^{X} = tr\left\lfloor \mathrm{P}h^{B_i\mu_j^X} + \mathrm{P}^{B_i}h^{\mu_j^X} \right\rfloor \tag{36}$$

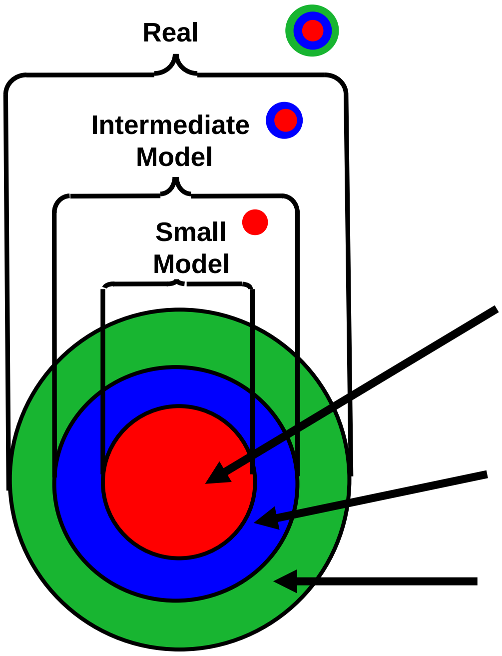
where P^{B_i} is the derivative of the density matrix with respect to the magnetic field. Matrices $h^{\mu_j^X}$ and $h^{B_i\mu_j^X}$ contain the first derivative of the standard one-electron Hamiltonian with respect to the nuclear magnetic moment and the second derivative with respect the magnetic field and the nuclear magnetic moment, respectively. Both terms do not contain explicit solvent-induced contributions as these contributions do not depend on the nuclear magnetic moment of the solute and thus the corresponding derivatives are zero. On the contrary, explicit solvent effects act on the first derivative of the density matrix P^B which can be obtained as solution of the corresponding first-order coupled-perturbed HF (or Kohn-Sham, KS) equation.

3. ONIOM calculations

The main disadvantage of accurate quantum chemical methods is that the computational cost scales extremely unfavorably with the size of system. One approach to improve the computational efficiency is to different methods for the different parts of system. Most of these so-called hybrid methods combined quantum chemical methods with molecular mechanics methods, often referred to as QM/MM, and proven very successful.

For systems normally treated with molecular mechanics (MM), the integrated molecular orbital and molecular mechanics method (IMOMM), has been proposed and tested for several different systems. For systems and reactions that require a more accurate description of the geometries and energies, the integrated MO+MO (IMOMO) method has been shown to reproduce high-level *ab initio* results. Consequently, it would be attractive to combine these approaches into an onion-like mutilayered scheme named the ONIOM method (our own-layered integrated molecular orbital + molecular mechanics method) (Svensson *et al*., 1996)

Recently, ONIOM method has been proven to be powerful tools for the theoretical treatment of large molecular systems where different levels of theory are applied to different parts of molecule (Appendix Figure 1) (Morokuma *et al*., 1995; Dapprich *et al*., 1999; Morokuma *et al*., 2000).



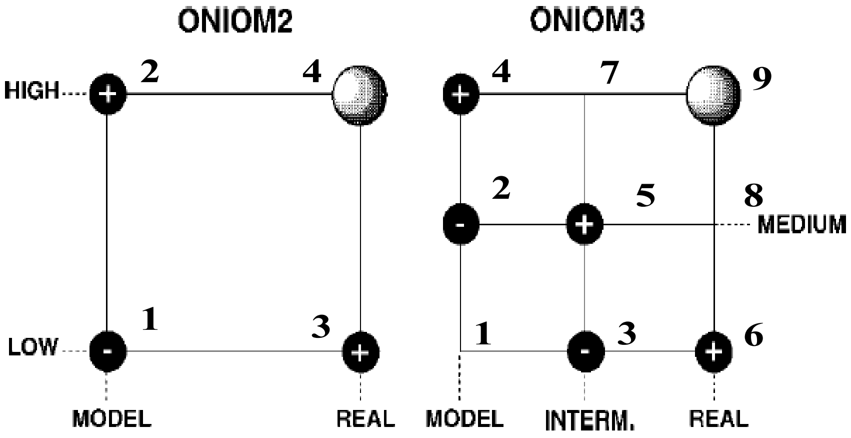
First Layer
Bond-formation/breaking takes place.
Use the 'High level' method.

Second Layer
Electronic effect on the first layer.
Use the 'Medium level' method.

Third Layer
Environmental effects on the first layer.
Use the 'Low level' method.

Appendix Figure 1 Schematic concept of ONIOM method.

Source: Morokuma, *et al.*, (2000)



Appendix Figure 2 ONIOM extrapolation scheme for a molecular system partitioned into two (left) and three (right) layers.

Source: Morokuma, *et al.*, (2000)

3.1 ONIOM energy definition

The basic idea behind the ONIOM approach can be explained most easily when it is considered as an extrapolation scheme in the two-dimensional space, spanned by the size of the system on one axis and the level of theory on the other axis. Figure 2 shows the extrapolation procedure schematically. The goal is E_4 describe the real system at the higher level of the theory, i.e., the approximation of the target E_4 (point 4) in the system partitioned into the two-layer ONIOM or E_9 (point 9) in the system consisting of the three layers. In the two-layered (ONIOM2) method, the extrapolated energy E_{ONIOM2} is then defined as

$$E_{ONIOM2} = E_3 - E_1 + E_2 \tag{37}$$

Where E_3 is the energy of the entire (real) system calculated at the low level of theory and E_1 and E_2 are the energies of the model system determined at the low and high level of theory, respectively. E_{ONIOM} is an approximation to the true energy of the real system E_4 :

$$E_4 = E_{ONIOM2} + D \tag{38}$$

Therefore, if the error D of the extrapolation procedure is constant for two different structures (e.g. between reactant and transition state), their relative ΔE_4 will be evaluated correctly by using the ONIOM energy ΔE_{ONIOM2}.

For a system partitioned into three different layers, the expression for the total energy E_{ONIOM3} as an approximation for E_9 is:

$$E_{ONIOM3} = E_6 - E_3 + E_5 - E_2 + E_4 \tag{39}$$

Since the evaluation of E_1 (the smallest model system at the lowest level of theory) does not require much computational effort, this value is used to determine the effect of the three-layer approach as compared to the two-layer partitioning with point

1, 4 and 6. If the energy difference between the two– and three–layer extrapolation is constant, a two layer partitioning with the intermediate layer omitted would give comparably accurate results.

It should be noted that the layers need not be inclusive or contiguous. The so-called “inner layer” does not have to be physically inside the “outer layer”. The layers can be any part of the system. Each layer does not have to be contiguous: it can consist of several separate regions of system.

4. The Theoretical Background of Molecualr Dynamics Simulations

The molecular dynamics simulations method is based on Newton’s second law or the equation of motion, F= ma, where F is the force exerted on the particle, m is its mass and a is its acceleration. From knowledge of the force on each atom, it is possible to determine the acceleration of each atom in the system. Integration of the equations of motion then yields a trajectory that describes the positions, velocities and accelerations of the particles. From this trajectory, the average values of properties can be determined. The method is deterministic; once the positions and velocities of each atom are known, the state of the system can be predicted at any time in the future or the past.

Newton’s equation of motion is given by

$$\mathbf{F}_\mathbf{i} = \mathbf{m}_\mathbf{i}\mathbf{a}_\mathbf{i} \qquad (40)$$

where $\mathbf{F}_\mathbf{i}$ is the force exerted on the particle **i**, $\mathbf{m}_\mathbf{i}$ is the mass on the particle **i**, and $\mathbf{a}_\mathbf{i}$ is its acceleration. The force can also be expressed as the gradient of the potential energy.

$$\mathbf{F}_\mathbf{i} = \nabla_\mathbf{i}\mathbf{V} \qquad (41)$$

Combining these two equations yields

$$-\frac{dV}{dr_i} = m_i \frac{d^2 r_i}{dt^2} \tag{42}$$

where V is the potential energy of the system. Newton's equation of motion can be related the derivative of the potential energy to the changes in position as a function of time.

In case of property calculations of the system,

$$\mathbf{F} = \mathbf{ma} = \mathbf{m}\frac{dv}{dt} = \mathbf{m}\frac{d^2 x}{dt^2} \tag{43}$$

and the acceleration is constant,

$$\mathbf{a} = \frac{dv}{dt} \tag{44}$$

Expression for the velocity after integration is obtained,

$$\mathbf{v} = \mathbf{at} + \mathbf{v_0} \tag{45}$$

and since

$$\mathbf{v} = \frac{dx}{dt} \tag{46}$$

therefore,

$$\mathbf{x} = \mathbf{vt} + \mathbf{x_0} \tag{47}$$

Combining this equation for the velocity, it can be obtained the following relation which gives the value of **x** at time **t** as a function of the acceleration, **a**, the initial position, $\mathbf{x_0}$ and the initial velocity, $\mathbf{v_0}$.

The acceralation is given as the derivative of the potential energy with respect to the position, **r**,

$$\mathbf{a} = -\frac{1}{m}\frac{dE}{dr} \qquad (48)$$

Therefore, to calculate a trajectory, one only needs the initial positions of the atoms, an initial distribution of velocities and the acceleration, which is determined by the gradient of the potential energy function. The equations of motion are deterministic, e.g., the position and the velocities at time zero determine the positions and velocities at all other times, t. The initial positions can be obtained from experimental structures, such as the x-ray crystal structure of the protein.

4.1 Initial condition

The initial condition in a molecular simulation is the initial positions and velocities of all atoms. In this study, the corrected structure was exerted to be the initial positions. Generally, the initial distribution of velocities is usually determined from a random distribution with the magnitudes conforming to the required temperature. Therefore, there is no overall momentum, for instance

$$\mathbf{P} = \sum_{i=1}^{N} m_i v_i = \mathbf{0} \qquad (49)$$

The velocities, $\mathbf{v_i}$, are often chosen randomly from a Maxwell-Boltzmann or Gaussian distribution at a given temperature, which gives the probability that an atom i has a velocity $\mathbf{v_x}$ in the x direction at a temperature **T**.

$$\mathbf{P(v_{ix})} = \left(\frac{m_i}{2\pi k_b T}\right)^{1/2} \exp\left[-\frac{1}{2}\frac{m_i v_{ix}^2}{k_b T}\right] \tag{50}$$

And the temperature can be calculated from the velocities using the relation

$$\mathbf{T} = \frac{1}{(3N)}\sum_{i=1}^{N}\frac{|P_i|}{2m_i} \tag{51}$$

where **i** is the number of atoms in the system.

4.2 An integration algorithm

The potential energy is a function of the atomic position (3N) of all the atoms in the system. Due to the complicated nature of this function, it must be solved numerically. Numerous numerical algorithms have been developed for integrating the equations of motion.

All the integration algorithms assume the positions, velocities and accelerations can be approximated by a Taylor series expansion as shown in equation (52) – (54),

$$\mathbf{r(t+\Delta t) = r(t) + v(t)\Delta t + (1/2)a(t)\Delta t^2 + ...} \tag{52}$$

$$\mathbf{v(t+\Delta t) = v(t) + a(t)\Delta t + (1/2)b(t)\Delta t^2 + ...} \tag{53}$$

$$\mathbf{a(t+\Delta t) = a(t) + b(t)\Delta t + ...} \tag{54}$$

where **r** is the position, **v** is the velocity (the first derivative with respect to time), **a** is the acceleration (the second derivative with respect to time), etc. The popular integration methods for MD calculations are the Verlet algorithm and predictor-corrector algorithms.

4.3 Periodic boundary conditions

It is the most common boundary condition for mimicking bulk behavior; the parent system is surrounded as infinite with copies of itself in all dimensions, a trick borrowed from solid state theory. Appendix Figure 3 illustrates the idea of periodic boundary condition. Periodicity applies not only to the interactions between atoms but also to the motion of atoms. If the atom leaves the parent cell its periodic image appears at the opposite side. In the cases of molecules, periodicity can also apply to the chemical connectivity. An interesting consequence of periodic boundary conditions (which need not be cubic as in the example but may be based on any space-filling polyhedron) is the artificial periodicity (superlattice) imposed on the system, even if the system is inherently non-periodic like a liquid. It is clear that in order to rule out artifacts, the system has to be “large enough” for effects of periodicity to disappear in the noise of other influences.

4.4 An ensemble

An ensemble is a collection of points in phase space satisfying the connditions of a particular thermodynamic state. A molecular dynamics simulations generates a sequence of points in phase space as a function of time, these points belong to the same ensemble, and it corresponds to the different conformations of the system and their respective momenta. Several different ensembles are described below.

An ensemble is a collection of all possible systems which have different microscopic states but have an identical macroscopic or thermodynamic state. There are different ensembles with different characteristics.

4.4.1 Microcanonical Ensemble (NVE)

The thermodynamic state characterized by a fixed number of atoms, **N**, a fixed volume, **V**, and a fixed energy, **E**. This corresponds to an isolated system.

4.4.2 Canonical Ensemble (NVT)

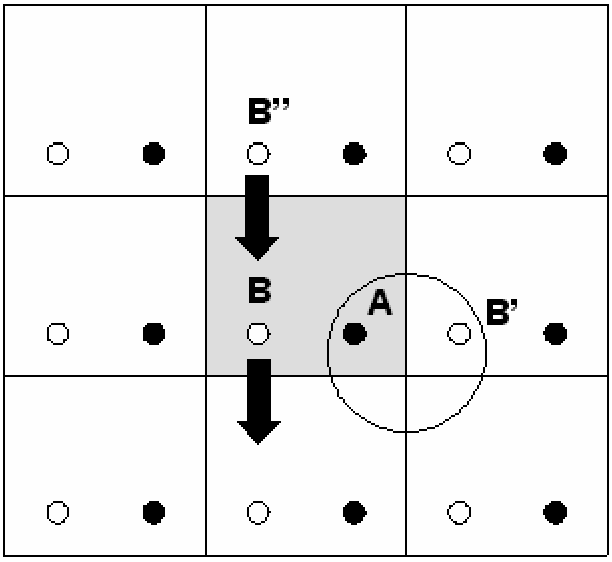
This is a collection of all systems whose thermodynamic state is characterized by a fixed number of atoms, **N**, a fixed volume,**V**, and a fixed temperature, **T**.

4.4.3 Isobaric-Isothermal Ensemble (NPT)

This ensemble is characterized by a fixed number of atoms, **N**, a fixed pressure, **P**, and a fixed temperature, **T**.

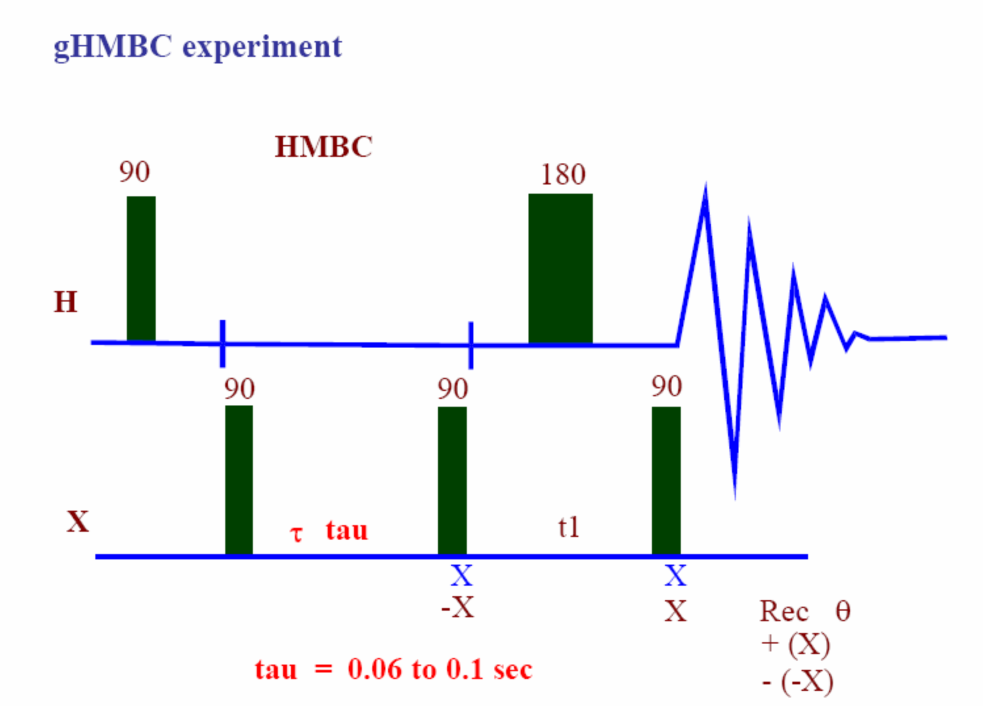
4.4.4 Grand Canonical Ensemble (μVT)

The thermodynamic state for this ensemble is characterized by a fixed chemical potential, μ, a fixed volume, **V**, and a fixed temperature, **T**.



Appendix Figure 3 Illustration of periodic boundary conditions (for clarity, in 2 dimensions): (i) Atoms A interacts with the nearest periodic image of atom B, B' ("minimum image convention") but not with atom B itself which is outside the interaction cut-off distance. (ii) When atom B moves out of the parent simulation cell (shaded), its periodic image B'' enter through the opposite face.

5. Gradient-selected versions Heteronuclear Multiple Bond Coherence (gHMBC) NMR



Appendix Figure 4 gHMBC experiment

Source: http://www.chem.uky.edu/research/miller/

Heteronuclear Multiple Bond Coherence (HMBC) is 2-dimensional inverse H, C correlation techniques that allow for the determination of carbon (or other heteroatom such as ^{15}N) to hydrogen connectivity. Gradient-selected versions HMBC (gHMBC) improves the acquired spectra by significantly reducing unwanted signal artifacts.

The HMBC experiment detects long range coupling between proton and hetereoatom (two or three bonds away) with great sensitivity. The length of the tau delay can be adjusted to detect relatively large coupling constants (4-10 Hz) tau = 0.06 s or snakker ciyokubgs (2-7 Hz) tau = 0.1 s.

In this sequence, the first 90° -pulse on ^{13}C serves as a low-pass filter that suppresses one-bond correlation and passes the smaller coupling. This pulse creates multiple quantum coherence for the one-bond coupling, which is removed from the spectra by alternating the phase of the ^{13}C pulse. The second 90º -pulse on ^{13}C creates multiple quantum coherence for the long-range couplings. After the evolution time t1, the magnetization is converted back into detectable single quantum proton magnetization. The carbon decoupler is never used in this sequence. Therefore the protons display homonuclear as well as heteromuclear couplings.

This technique is very valuable to detect indirectly quaternary carbons coupled to protons. Especially useful if direct ^{13}C is impossible to obtain due to low amount of material available. This very useful sequence provides information about the skeleton of a molecule. It could be an alternative to the 2D-INADEQUATE experiment (which is so insensitive). It is also very useful in carbohydrate area as a sequence analysis tool that provides unique information concerning connectivities across glycosidic linkages. Another area of interest for using HMBC is in the peptide-protein area. Especially when applied to a ^{15}N labeled protein. It is possible with this technique to get connectivities between the Nitrogen and the CHα proten of the amino acid of the next residue.