Theor Chem Acc (2000) 104:336–343
DOI 10.1007/s002140000153

Theoretical
Chemistry Accounts

Regular article

The generalized hybrid orbital method for combined quantum mechanical/molecular mechanical calculations: formulation and tests of the analytical derivatives

Patricia Amara[1], **Martin J. Field**[1], **Cristobal Alhambra**[2], **Jiali Gao**[2]

[1]Institut de Biologie Structurale – Jean-Pierre Ebel, 41 rue Jules Horowitz, 38027 Grenoble Cedex 1, France
[2]Department of Chemistry, State University of New York at Buffalo, Buffalo, NY 14260, USA

Received: 4 October 1999 / Accepted: 18 December 1999 / Published online: 5 June 2000

Abstract. Hybrid quantum mechanical (QM) and molecular mechanical (MM) potentials are becoming increasingly important for studying condensed-phase systems but one of the outstanding problems in the field has been how to treat covalent bonds between atoms of the QM and MM regions. Recently, we presented a generalized hybrid orbital (GHO) method that was designed to tackle this problem for hybrid potentials using semiempirical QM methods [Gao et al. (1998) *J Phys Chem A* 102: 4714–4721]. We tested the method on some small molecules and showed that it performed well when compared to the purely QM or MM potentials. In this article, we describe the formalism for the determination of the GHO energy derivatives and then present the results of more tests aimed at validating the model. These tests, involving the calculation of the proton affinities of some model compounds and a molecular dynamics simulation of a protein, indicate that the GHO method will prove useful for the application of hybrid potentials to solution-phase macromolecular systems.

Key words: Hybrid method – Analytic first derivatives – Geometry optimization – Molecular dynamics – Protein simulation

1 Introduction

Hybrid quantum mechanical (QM) and molecular mechanical (MM) potentials have proved to be a powerful tool for studying chemical reactions in solution and in enzymes [1–3]. In the QM/MM approach, a QM potential is used to study a small region of the system, whereas the rest of the atoms are treated with an MM potential. The QM region typically contains the atoms that are involved in the chemical reaction and for which a knowledge of the changes in their electron density is required, whereas the MM atoms act as the environment to the QM region.

For solution-phase systems, in which the solutes are the reacting species, the partitioning of the atoms between the QM and MM regions is straightforward and there will be no molecules with atoms in both regions. For enzymatic systems, however, some of the enzyme's amino acid residues typically take part directly in the reaction and so will need to be treated quantum mechanically. This means that the enzyme will have atoms in both the QM and MM regions and that there will be covalent bonds between QM and MM atoms.

A number of schemes have been developed to treat covalent bonds at the QM/MM boundary. An early and still probably the most widely used approach is the "dummy"- or "link"-atom method in which extra, unphysical atoms are added to the QM region, one for each QM/MM bond. These link atoms, usually hydrogens, are placed at an appropriate distance from the QM atoms of the broken bonds and replace the MM atoms of the broken bonds in the QM calculation [4, 5]. The link-atom method is simple to implement and works reasonably well but is not very elegant. More appealing schemes dispense with the unphysical atoms and employ hybrid orbitals. Warshel and Levitt [6] in their original work used such a method. More recently, Rivail and coworkers [7–10] developed a similar scheme, which they call a local self-consistent field (LSCF) algorithm. Other methods that have been developed include those by Maseras and Morokuma [11], by Bersuker et al. [12], by Bakowies and Thiel [13] and by Zhang et al. [14].

In a recent article we presented a generalized hybrid orbital (GHO) method to treat the link-atom problem for semiempirical QM potentials [15]. Our approach is similar in spirit to the method of Rivail and coworkers [7–10] in that atomic orbitals on a "boundary" atom are transformed into hybrid orbitals, some of which are optimized in the self-consistent field (SCF), but is different in that the parameters needed in the definition of the model are transferable. In our original study [15], we

Correspondence to: M. J. Field or J. Gao

tested the method on a series of hydrocarbon model compounds and showed that it yielded reasonable structural, energetic and electronic results in comparison with the results from the corresponding QM and MM methods. In this article, we describe the procedure for calculating analytically the first derivatives of the energy within the GHO formalism and present the results of computed proton affinities and molecular dynamics simulations of the SH2 domain in water. Although it would, in principle, be possible to calculate the derivatives numerically as is the case in some semiempirical QM programs, it is clear to us that analytic evaluation will be more precise and much less costly than numerical evaluation, particularly for cases in which there are large numbers of atoms in the MM region.

The outline of this article is as follows. The main features of the GHO model and the expressions for the derivatives within the GHO formalism are summarized in Sect. 2. The results of our calculations using the GHO model are presented in Sect. 3 and Sect. 4 concludes with a summary of the present work.

2 Methods

The GHO model was introduced and described in Ref. [15]. Here, the GHO model is briefly outlined before the derivatives of the GHO energy expression are presented.

2.1 The GHO model

In the GHO model the boundary atom in the MM region that is directly linked to the QM atom is represented by a set of four orthogonal hybrid orbitals $\{\eta_Q, \eta_A, \eta_B, \eta_C\}$, where the subscript Q denotes the orbital that forms a σ covalent bond with the QM atom (Fig. 1). The other three orbitals (η_A, η_B, η_C) are auxiliary orbitals that point roughly in the directions of the three MM atoms A, B and C. These orbitals do not participate in the SCF optimization procedure but they are included in the calculation of the effective Hamiltonian matrix.

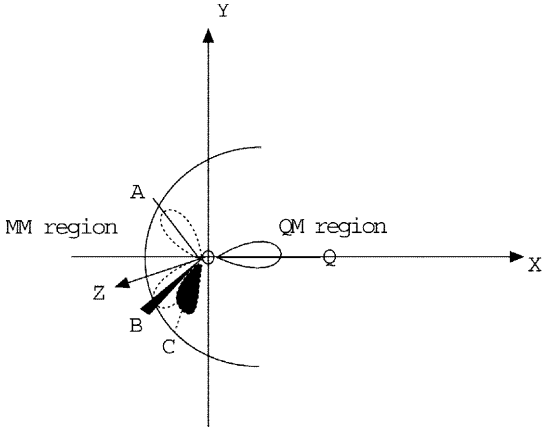

Fig. 1 A schematic diagram showing the generalized hybrid orbital (*GHO*) model applied to a covalent bond between an atom of the quantum mechanical (*QM*) region and an atom of the molecular mechanical (*MM*) region. *O* is the boundary atom to which the GHO model is applied. It is linked to the quantum atom *Q* and to the MM atoms *A*, *B* and *C*. The hybrid orbitals on the boundary atom are η_Q, η_A, η_B and η_C. η_Q points towards Q and is optimized in the self-consistent-held calculation. The remaining orbitals point towards the MM atoms, A, B and C, and are not optimized

In the following discussion, for clarity, we consider the case where there is only one boundary atom that is connected to the QM fragment. Generalization to systems containing more than one boundary atom is straightforward. The QM fragment is assumed to have N atomic orbital basis functions $\{\xi_\mu,\ \mu = 1, \ldots, N\}$, which together with the orbital η_Q of the boundary atom constitute the active orbital basis set for the SCF calculation. The molecular orbitals (MOs) of the system, ϕ_i^H, are linear combinations of these $N + 1$ functions:

$$\phi_i^H = \sum_{\mu=1}^{N} c_{\mu i}\xi_\mu + c_Q\eta_Q \ , \tag{1}$$

where the subscript i runs from 1 to $N + 1$ and the superscript H emphasizes the fact that the MOs are formed from a hybrid orbital basis set.

To construct the Fock matrix, the density matrix in the hybrid orbital basis set $\mathbf{P}^H$ is expanded so that the densities of the auxiliary orbitals are included as extra diagonal elements. This generates a new density matrix, $\mathbf{P}_t^H$, of dimension $(N + 4) \times (N + 4)$, where we have used the subscript t to denote the larger dimension. This density is then transformed into the density matrix in the standard atomic orbital basis, $\mathbf{P}_t^{AO}$, using the transformation matrix, $\mathbf{T}$:

$$\mathbf{P}_t^{AO} = \left(\mathbf{T}^{-1}\right)^\dagger \mathbf{P}_t^H \left(\mathbf{T}^{-1}\right) \ . \tag{2}$$

The Fock matrix, $\mathbf{F}_t^{AO}$, is obtained as usual from the atomic orbital density matrix, $\mathbf{P}_t^{AO}$, and an effective one-electron matrix, $\mathbf{H}_t^{eff}$, which contains contributions from the auxiliary orbitals [15].

The energy of the QM/MM system is expressed in terms of the atomic orbital basis as follows:

$$E = \sum_{\mu\nu}^{N+4} P_{\mu\nu}^{AO} H_{\mu\nu}^{eff} + \frac{1}{2}\sum_{\mu\nu}^{N+4}\sum_{\lambda\sigma}^{N+4} P_{\mu\nu}^{AO} P_{\lambda\sigma}^{AO}\left[(\mu\nu, \lambda\sigma) - \frac{1}{2}(\mu\sigma, \lambda\nu)\right] + E_{nuc}^{QM} + E_{nb}^{QM/MM} + E^{MM} \ , \tag{3}$$

where the first two terms are the QM/MM electronic energy, the third term is the nuclear repulsion energy for the QM nuclei, the fourth term is the QM nuclei/MM atom electrostatic and Lennard-Jones interaction energy and the last term is the energy of the atoms in the MM region.

2.2 Definition of the transformation matrix

It is convenient to write the transformation matrix as a product of a hybridization matrix, $\mathbf{H}$, in a local coordinate system and a rotation matrix, $\mathbf{B}$, that effects the transformation from the local coordinate system to the molecular coordinate system:

$$\mathbf{T} = \mathbf{BH} \ . \tag{4}$$

Because of the neglect of differential diatomic overlap (NDDO) approximation that is used in semiempirical calculations, the matrix $\mathbf{T}$ will be equal to the identity matrix, $\mathbf{I}$, except at the boundary atoms. For the case of one boundary atom $\mathbf{T}$ can be written as

$$\mathbf{T} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{T}_b \end{pmatrix} \ , \tag{5}$$

where $\mathbf{T}_b$ is a 4×4 matrix corresponding to the four orbitals of the boundary atom. $\mathbf{T}_b$ is used to obtain a set of hybrid orbitals on the boundary atom that point approximately towards the other atoms to which the boundary atom is bound:

$$\begin{pmatrix} \eta_Q \\ \eta_A \\ \eta_B \\ \eta_C \end{pmatrix} = (\mathbf{T}_b)^{-1} \begin{pmatrix} s \\ p_x \\ p_y \\ p_z \end{pmatrix} \ , \tag{6}$$

where the orbitals s, p_x, p_y and p_z are the valence orbitals of the boundary atom.

We begin by discussing the definition of the local coordinate system shown in Fig. 1, which determines the direction of the hybrid orbitals and the matrix $\mathbf{B}$. In what follows we use the subscripts 1, 2 and 3 to denote the three Cartesian coordinates in the

molecular frame and the vectors **x**, **y** and **z** to denote the three directional unit vectors in the local coordinate system. The vector **x** is taken to be orthogonal to the plane α that is defined by the three unit vectors **a**, **b** and **c** that point from the boundary atom, O, to the MM atoms A, B and C, respectively. The active hybrid orbital, which does not necessarily coincide with the covalent bond between the boundary atom and the QM atom Q is placed along the vector **x** towards Q. The vector **z** is chosen to be perpendicular to the plane containing the vectors **x** and **a**. The direction of **y** is defined by the vector cross product of **z** and **x**, i.e. $\mathbf{y} = \mathbf{z} \otimes \mathbf{x}$.

The rotation matrix $\mathbf{B}_b$ is defined by the directional cosines of the local coordinates as

$$\mathbf{B}_b = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & x_1 & y_1 & z_1 \\ 0 & x_2 & y_2 & z_2 \\ 0 & x_3 & y_3 & z_3 \end{pmatrix} . \tag{7}$$

The hybridization matrix, $\mathbf{H}_b$, should yield four orthonormal hybrid orbitals. This is conveniently done by using the local geometry of the three MM atoms attached to the boundary atom:

$$\mathbf{H}_b = \begin{pmatrix} C_s & C_p/\sqrt{3} & C_p/\sqrt{3} & C_p/\sqrt{3} \\ C_p & -C_s/\sqrt{3} & -C_s/\sqrt{3} & -C_s/\sqrt{3} \\ 0 & \sqrt{2/3} & -\sqrt{1/6} & -\sqrt{1/6} \\ 0 & 0 & \sqrt{1/2} & -\sqrt{1/2} \end{pmatrix} , \tag{8}$$

where the coefficients C_s and C_p represent the *s*- and *p*-orbital components in the hybrid orbitals. An equal partition of the *p* orbital is assumed in constructing the matrix in Eq. (8). For the hybrid orbital η_Q, C_s is proportional to the distance, *L*, between the atom O and the plane α and satisfies the relationship, $C_s^2/C_p^2 = L$. Along with the normalization condition $C_s^2 + C_p^2 = 1$, we obtain

$$C_s = \sqrt{\frac{L}{L+1}} , \tag{9}$$

$$C_p = \sqrt{\frac{1}{L+1}} . \tag{10}$$

The projection distance, *L*, between the boundary atom and the plane α is given by

$$L = -\mathbf{a}^\dagger \mathbf{x} = -\mathbf{b}^\dagger \mathbf{x} = -\mathbf{c}^\dagger \mathbf{x} \tag{11}$$

2.3 Derivatives of the GHO energy expression

The gradients of the energy are found by taking the partial derivatives of the energy expression of Eq. (3) with respect to the Cartesian coordinates of the atoms in the system. For atoms that are not directly connected to the boundary atom and for the QM atom Q which is connected to the boundary atom, the gradients can be obtained in the normal way because the partial derivatives of **T** with respect to the coordinates of these atoms are zero. For the three MM atoms A, B and C and the boundary atom O, there is an extra term involved due to the dependence of the density matrix on **T**. The derivatives for these atoms can be written as:

$$\frac{\partial E}{\partial q_i} = \frac{\partial E^{\mathrm{HF}}}{\partial q_i} + \sum_{\mu\nu}^{N+4} \frac{\partial P_{\mu\nu}^{\mathrm{AO}}}{\partial q_i} F_{\mu\nu}^{\mathrm{AO}} , \tag{12}$$

where q_i specifies one of the Cartesian coordinates of the atoms A, B, C and O and the superscript HF indicates the normal derivatives of the Hartree–Fock energy expression obtained in the atomic orbital basis.

The second term of Eq. (12) occurs because the SCF calculation is carried out in the GHO basis of $N + 1$ functions which depends on **T**. From Eq. (2), the derivatives of the density matrix are

$$\frac{\partial \mathbf{P}_t^{\mathrm{AO}}}{\partial q_i} = \frac{\partial \left(\mathbf{T}^{-1}\right)^\dagger}{\partial q_i} \mathbf{P}_t^{\mathrm{H}} \left(\mathbf{T}^{-1}\right) + \left(\mathbf{T}^{-1}\right)^\dagger \times \frac{\partial \mathbf{P}_t^{\mathrm{H}}}{\partial q_i} \left(\mathbf{T}^{-1}\right) + \left(\mathbf{T}^{-1}\right)^\dagger \mathbf{P}_t^{\mathrm{H}} \frac{\partial \left(\mathbf{T}^{-1}\right)}{\partial q_i} . \tag{13}$$

The first and third terms on the right-hand side of this equation involve the derivatives of the transformation matrix, explicit expressions for which can be found in the Appendix. The second term involves the partial derivatives of the density matrix in the GHO basis $\mathbf{P}_t^{\mathrm{H}}$, which is the matrix that is determined in the SCF procedure. The derivatives of this matrix do not contribute to the total derivatives for the same reason that the derivatives of the density matrix in a normal NDDO Hartree–Fock calculation do not contribute.

3 Results

We performed two sets of calculations to test the GHO/derivative method. The first concerned the proton affinities of a series of small molecules and were designed to test the perturbations induced by the GHO method on the electronic structure of the system. The second involved the comparison of two molecular dynamics simulations of a small protein in water, one using a purely MM potential and the second a QM(GHO)/MM hybrid model. Both sets of tests employed the GHO model as implemented in the CHARMM molecular modeling program [16]. The AM1 semiempirical method of Dewar et al. [17] was used to treat the quantum atoms and the CHARMM all-atom force field [18] for the MM atoms. We discuss each set of calculations separately.

3.1 Proton affinity

The calculation of the proton affinities of organic bases provides a stringent test of the accuracy of the QM(GHO)/MM method in comparison with the corresponding QM model. Following Dewar and Dieter [19], who provided an extensive list of experimental and AM1-derived proton affinities, the proton affinity of a basic compound, B, is defined as the negative value of the enthalpy of reaction at 25 °C for the protonation reaction:

$$\text{Proton affinity (B)} = \Delta H_f(\mathrm{H}^+) + \Delta H_f(\mathrm{B}) - \Delta H_f(\mathrm{HB}^+) . \tag{14}$$

Following the procedure in Ref. [19], we calculate all the quantities in Eq. (14) except for the heat of formation of the proton, $\Delta H_f(\mathrm{H}^+)$, which is poorly estimated by the AM1 method. Instead, we use the experimental value of 365.7 kcal mol^{-1}, which is also the value adopted by Bakowies and Thiel [13] for their proton affinity calculations. This value differs by a small amount from the value originally used by Dewar and Dieter (367.2 kcal mol^{-1}) [19].

We determined the proton affinities for three different sets of model compounds – carboxylate anions, alcoxide ions and primary amines and their anions – that were chosen because they contain functional groups occurring in the side chains of amino acids. In the cases where the molecules were large enough, we partitioned them into different QM and MM regions so that the influence of the size of the QM fragment could be systematically investigated as the boundary atom moved away from the protonation site. Almost all the compounds we chose

were in the set of Dewar and Dieter [19] and some of them were also studied by Field et al. [5] and by Bakowies and Thiel [13].

The results of our calculations are displayed in Tables 1–3 along with the experimental data [19], the results using a link-atom approach [5] and the results obtained with the AM1 method. We performed the last two sets of calculations using the current implementation of the link-atom and the AM1 semiempirical methods in the CHARMM program [16] and so, as a result, the numbers we obtain may differ slightly from those reported in Refs. [5, 19]. The structures of all deprotonated and protonated species used to calculate the proton affinities in the tables were fully geometry optimized at the relevant level of theory.

For the carboxylate and alcoxide anions, the GHO method gives results which are of equivalent accuracy to those of AM1, whereas the link-atom method gives values with larger errors. For the amine anions, the GHO method gives slightly poorer results than those of AM1, but it still performs better than the link-atom method. The only case where the link-atom method has an advantage is for the amines, but here the differences between the two methods become less important as the size of the QM region is increased. In general, the data in the tables show that the errors in the proton affinities are largest for those systems in which the boundary atom is either directly connected to the atom that is being protonated or is separated from it by only one other atom. When the boundary atom is separated by at least three bonds from the protonation site, the GHO results have a mean unsigned error of 0.9 kcal mol^{-1} (for the 13 such cases in the tables). The corresponding error for the link-atom results is 2.8 kcal mol^{-1}. This confirms that the QM region should be reasonably large to avoid the artefacts introduced by the boundary atoms and indicates that the choice of the GHO boundary atom should be at least two atoms away from the reactive center involving bond forming and breaking.

Table 1. Proton affinities for carboxylate anions. The experimental values are taken from Ref. [19]. The values for the AM1, generalized hybrid orbital (*GHO*) and link-atom orbital (*LA*) calculations were all obtained using the CHARMM program [16]. All energies are in kilocalories per mole. The *symbol* * indicates the boundary atom for the GHO calculations or the atom which is replaced by the LA in the LA quantum mechanical (*QM*) molecular mechanical (*MM*) calculations. Atoms on the right of C* are treated quantum mechanically

Acids	GHO	LA	AM1	*exp.*
$H_3C^*—CO_2^-$	349.9	355.3	353.3	348.5
$CH_3—H_2C^*—CO_2^-$	352.6	356.2	352.5	347.3
$H_3C^*—CH_2—CO_2^-$	351.9	355.1	352.5	347.3
$CH_3—H_2C^*—CH_2—CO_2^-$	353.3	355.7	352.2	346.6
$H_3C^*—CH_2—CH_2—CO_2^-$	351.9	353.5	352.2	346.6
$(CH_3)_2—HC^*—CO_2^-$	356.4	357.0	351.8	346.3
$H_3C^*—CH(CH_3)—CO_2^-$	351.0	354.5	351.8	346.3

Table 2. Proton affinities for alcoxide anions. The experimental values are taken from Ref. [19]. The values for the AM1, GHO and LA calculations were all obtained using the CHARMM program [16]. All energies are in kilocalories per mole. The *symbol* * indicates the boundary atom for the GHO calculations or the atom which is replaced by the LA in the LA QM/MM calculations. Atoms on the right of C* are treated quantum mechanically

Alcohols	GHO	LA	AM1	*exp.*
$H_3C^*—H_2C—O^-$	383.2	385.1	382.9	376.1
$CH_3—H_2C^*—CH_2—O^-$	385.8	385.4	381.5	374.7
$H_3C^*—CH_2—CH_2—O^-$	381.7	384.6	381.5	374.7
$H_3C^*—CH(CH_3)—O^-$	384.6	387.0	384.5	374.1
$(CH_3)_2—HC^*—O^-$	411.1	422.8	384.5	374.1
$(H_3C^*)—C(CH_3)_2—O^-$	383.1	385.9	383.1	373.3
$(CH_3)_3—C^*—O^-$	415.8	438.7	383.1	373.3
$(CH_3)_3—C^*—CH_2—O^-$	393.1	387.2	381.0	372.1
$H_3C^*—C(CH_3)_2—CH_2—O^-$	380.8	383.9	381.0	372.1
$H_3C^*—CH_2—CH_2—CH_2—O^-$	382.7	383.8	382.7	375.2
$CH_3—H_2C^*—CH_2—CH_2—O^-$	384.5	387.3	382.7	375.2
para $H_3C^*—C_6H_4—O^-$	345.1	349.5	346.3	350.9
para $CH_3—H_2C^*—C_6H_4—O^-$	346.7	349.8	346.0	349.9

Table 3. Proton affinities for amines and their anions. The experimental values are taken from Ref. [19]. The values for the AM1, GHO and LA calculations were all obtained using the CHARMM program [16]. All energies are in kilocalories per mole. The *symbol* * indicates the boundary atom for the GHO calculations or the atom which is replaced by the LA in the LA QM/MM calculations. Atoms on the right of C* are treated quantum mechanically

Amines	GHO	LA	AM1	*exp.*
$H_3C^*—NH^-$	416.8	433.3	404.9	403.2
$CH_3—H_2C^*—NH^-$	421.9	434.4	404.2	399.4
$H_3C^*—CH_2—NH^-$	414.5	407.5	404.2	399.4
$CH_3—H_2C^*—CH_2—NH^-$	407.5	409.3	402.8	398.4
$H_3C^*—CH_2—CH_2—NH^-$	403.0	407.0	402.8	398.4
$H_3C^*—NH_2$	202.9	214.4	209.5	214.1
$CH_3—H_2C^*—NH_2$	204.3	213.6	213.5	217.0
$H_3C^*—CH_2—NH_2$	212.6	212.3	213.5	217.0
$CH_3—H_2C^*—CH_2—NH_2$	211.4	214.5	214.9	217.9
$H_3C^*—CH_2—CH_2—NH_2$	212.6	212.7	214.9	217.9
$CH_3—CH_2—H_2C^*—CH_2—NH_2$	210.6	214.0	215.2	218.4
$CH_3—H_2C^*—CH_2—CH_2—NH_2$	213.1	214.1	215.2	218.4

3.2 Molecular dynamics simulations of an SH2 protein domain

To test the GHO method for applications to protein systems, we carried out molecular dynamics simulations of the SH2 domain of the human $p56^{lck}$ tyrosine kinase protein complexed to an 11-residue phosphotyrosyl peptide. The SH2 domain consists of about 100 amino acid residues, which binds phosphotyrosine-containing amino acid sequences with high affinity and specificity. These domains are of great interest because they are intimately involved in cellular signal transduction pathways.

For our simulations we used the SH2-domain structure of Eck et al. [20] (Protein Data Bank, PDB code 1LCJ [21]), which consists of two chains, the 104-residue SH2 domain (PDB numbering 123–226) and an 11-residue peptide that contains a phosphotyrosyl residue (pY). Starting from the PDB structure, we built the hydrogens using the HBUILD algorithm [22] in the

CHARMM program and solvated the protein by immersing it in a preequilibrated cubic box of water molecules of side 56.1 Å and deleting all water molecules whose oxygens were less than 2.8 Å from any non-hydrogen protein atom. The final system had about 16900 atoms – 1850 protein atoms and 5020 water molecules. The solvated system was then refined by minimizing its geometry with a conjugate-gradient algorithm and treating the entire system with the CHARMM22 MM force field.

We performed two molecular dynamics simulations at a temperature of 298 K. The first was a purely MM simulation of 1-ns duration. The second was a hybrid potential simulation of 500-ps duration in which the side chain of the phosphotyrosyl residue, which is bound at the SH2-domain bonding site, was chosen as the QM region and the Cβ atom of the side chain as the GHO boundary atom. The MM simulation was started from the minimized protein/water box structure described earlier, whereas the starting structure for the hybrid potential simulation was taken from the MM simulation at 600 ps after the MM simulation had equilibrated. For each simulation, the Nosé constant temperature algorithm was used [23] with an integration timestep of 1 fs. The non-bonding interactions were treated using a minimum-image group-switching approximation in which the switching function was applied between 8 and 12 Å.

To illustrate the general behavior of the dynamics with the MM potential, we show the root-mean-square coordinate deviations (RMSCDs) between the crystallographic structure and the structures produced during the simulation as a function of time in Fig. 2. The RMSCD slowly increases until a plateau value of about 2.5 Å is attained at 600 ps. These RMSCD values are similar to those found by Feng et al. [14], who performed a molecular dynamics simulation on a related SH2-domain protein.

That the structures obtained in the last 400 ps of the MM dynamics are reasonable can be seen by comparing the calculated and experimental isotropic B factors that are shown in Fig. 3. The agreement between the two is, in general, excellent. We highlight two regions for special attention. First, there is a small loop in the SH2 domain, comprised of residues 142–147, for which the MM simulation overestimates the B factors. This can be explained because the loop interacts with a neighboring protein molecule in the crystallographic structure but is free to move in solution in the simulation. Second, the MM simulation severely underestimates the fluctuations in the region immediately around the phosphotyrosyl residue.

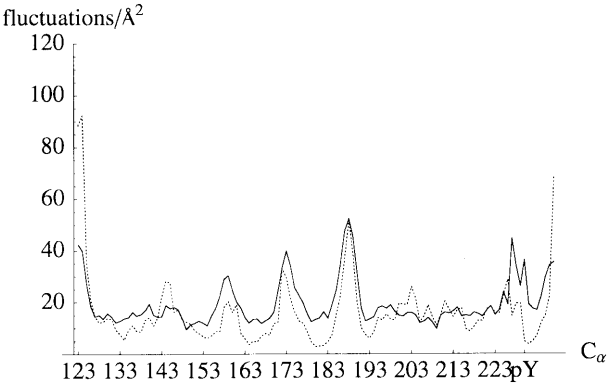

Fig. 3 The calculated (*dotted line*) and experimental (*solid line*) isotropic B factors for the Cα atoms of the SH2 protein. The calculated values were obtained from the structures of the last 400 ps of the 1-ns MM potential molecular dynamics simulation

Similar plots for the trajectory generated with the hybrid AM1/CHARMM22 potential are shown in Figs. 4 and 5. The RMSCD between the hybrid potential simulation and the crystallographic structures reaches a plateau value of about 2.5 Å, which is similar to that of Fig. 2, but this value is attained more rapidly than in the pure MM simulation because the starting structure was already well equilibrated. Similarly, the B factors are in excellent agreement with the experimental values, although the fluctuations in the SH2-domain loop region around residue 145 are much larger. In contrast, however, the fluctuations for the region pY – 2

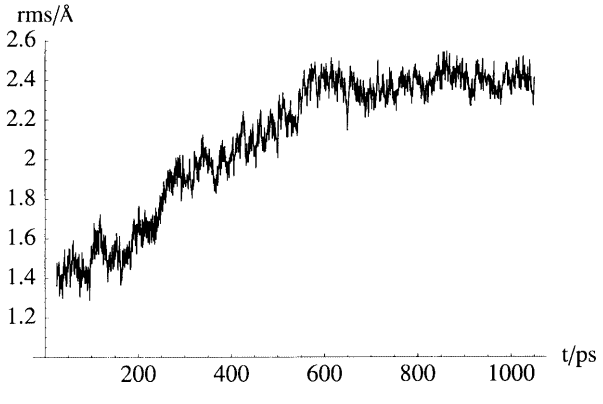

Fig. 2 A plot of the root-mean-square coordinate deviations (*RMSCDs*) as a function of time for the trajectory produced using the MM potential. The RMSCDs were calculated between the structures along the MM trajectory and the crystallographic structure using the backbone atoms only

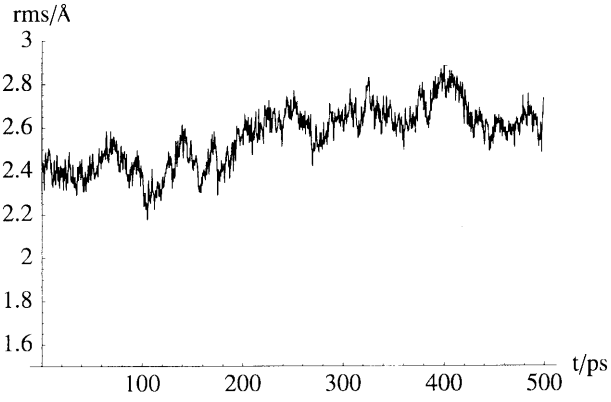

Fig. 4 A plot of the RMSCDs as a function of time for the trajectory produced using the hybrid potential. The RMSCDs were calculated between the structures along the QM/MM trajectory and the crystallographic structure using the backbone atoms only

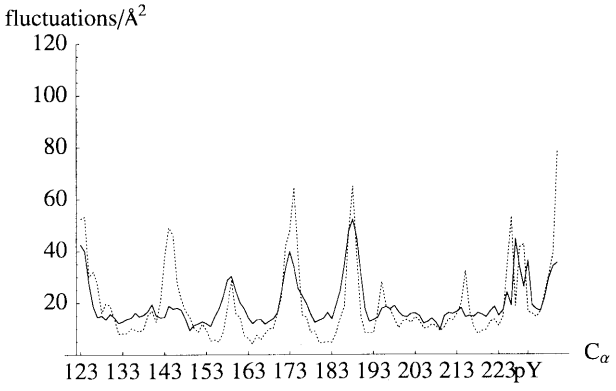

Fig. 5 The calculated (*dotted line*) and experimental (*solid line*) isotropic B factors for the Cα atoms of the SH2 protein. The calculated values were obtained from the structures of the last 400 ps of the 500-ps hybrid potential molecular dynamics simulation

to pY + 4 are much better reproduced than in the MM simulation.

Of particular importance in the binding of the peptide to the SH2 domain is a hydrogen-bond network around the phosphotyrosyl residue, which has been described in detail by Eck et al. [20]. Two arginines play critical roles – R134, which has one hydrogen bond with a phosphate oxygen, and R154, which has two. An analysis of the hydrogen-bond network around the phosphate ion during the hybrid potential simulation shows that there are hydrogen bonds between both arginines and the phosphate for about 85% of the time. In addition, there are five water molecules which have hydrogen bonds to the phosphate for over 65% of the time and a sixth water molecule which has a hydrogen bond with the phosphate for about 35% of the time. The positions of all these water molecules remain approximately constant during the course of the simulation and there is no exchange with the water molecules of the bulk solvent.

A similar analysis of the hydrogen-bond network in the MM simulation reveals that R154 has a hydrogen bond with the phosphate for over 95% of the trajectory structures, whereas R134 has no hydrogen bonds at all with the phosphate. Instead, the binding pocket has been deformed so that there are seven water molecules with hydrogen bonds to the phosphate, four for over 95% of the time and three for between 75 and 95% of the time. These observations help to explain that the B factors of the atoms around the phosphotyrosyl residue are so small in the MM simulation because the phosphate is locked into the binding pocket by an overstrong series of hydrogen bonds. In the QM/MM simulation, however, the hydrogen-bond network is less rigid and allows the pY residue more flexibility. The reason for this difference is no doubt due to the different potentials used to treat the highly charged phosphate group. The nonpolarizable MM potential overestimates the charge–charge interactions between the phosphateand the water molecules, whereas the hybrid potential, because it permits polarization of the phosphate group, provides a more realistic description of these interactions.

4 Conclusions

In this article we have extended the GHO method developed in our earlier work [15] by explaining how the GHO energy derivatives can be calculated and by performing more stringent tests of the method's precision. We have shown that the GHO method is applicable to the study of large macromolecular systems and is as accurate as the link-atom method that has been the most widely used method to date for hybrid potential studies. In contrast to the link-atom method, however, the GHO method does not introduce spurious atoms into the QM region and results in a well-defined potential-energy surface for the system. It also has the advantage over the LSCF method in that it does not need to be reparameterized for each new system. Although the GHO works well, it can be improved. In particular, we would like to refine the way in which the charges on the MM atoms at the interface are handled and we are working on a way of extending the method which is appropriate to ab initio QM wavefunctions. In summary, we feel that the GHO method will provide a useful complement to other methods that exist for tackling the "boundary-atom" problem and will find an important role when applying hybrid QM/MM potential techniques to the study of biomacromolecular systems.

Acknowledgements. This work has been supported in part by a joint grant from the National Science Foundation and the Centre National de la Recherche Scientifique. Gratitude is also expressed to the National Institutes of Health (J.G.) for partial support and to the Institut de Biologie Structurale – Jean-Pierre Ebel and the Commissariat à l'Energie Atomique (P.A., M.J.F.).

Appendix

In this appendix we give explicit expressions for the derivatives of the transformation matrix, **T**, and its inverse, $\mathbf{T}^{-1}$. From Eq. (4), it is clear that the derivatives of **T** can be expressed in terms of the derivatives of the matrices **B** and **H**:

$$\frac{\partial \mathbf{T}}{\partial q_i} = \frac{\partial \mathbf{B}}{\partial q_i}\mathbf{H} + \mathbf{B}\frac{\partial \mathbf{H}}{\partial q_i} \quad . \tag{A1}$$

The derivatives of $\mathbf{T}^{-1}$ can also be expressed in terms of the derivatives of **B** and **H** because the inverses of **B** and **H** are $\mathbf{B}^\dagger$ and $\mathbf{H}^\dagger$, respectively. Thus

$$\mathbf{T}^{-1} = (\mathbf{BH})^{-1} = \mathbf{H}^{-1}\mathbf{B}^{-1} = \mathbf{H}^\dagger\mathbf{B}^\dagger \quad . \tag{A2}$$

To determine the derivatives of the matrices **B** and **H** we suppose that the atom O is at the origin and that the unit vectors along the three boundary atom/MM atom bonds can be written as

$$\mathbf{a} = \frac{1}{R_\mathrm{A}}\begin{pmatrix} A_1 \\ A_2 \\ A_3 \end{pmatrix}, \quad \mathbf{b} = \frac{1}{R_\mathrm{B}}\begin{pmatrix} B_1 \\ B_2 \\ B_3 \end{pmatrix}, \quad \mathbf{c} = \frac{1}{R_\mathrm{C}}\begin{pmatrix} C_1 \\ C_2 \\ C_3 \end{pmatrix} \quad , \tag{A3}$$

where A_i, B_i and C_i are Cartesian coordinates for atoms A, B and C, respectively, and R_A, R_B and R_C refer to bond distances between these atoms and the boundary atom.

The three unit vectors defining the local coordinate system are

$$\mathbf{x} = \frac{1}{R_X}(\mathbf{b}\otimes\mathbf{c} + \mathbf{c}\otimes\mathbf{a} + \mathbf{a}\otimes\mathbf{b}) = \frac{1}{R_X}(X_1, X_2, X_3) \tag{A4}$$

$$\mathbf{z} = \frac{1}{R_Z}(\mathbf{x}\otimes\mathbf{a}) = \frac{1}{R_Z}(Z_1, Z_2, Z_3) \tag{A5}$$

$$\mathbf{y} = \mathbf{z}\otimes\mathbf{x}\ , \tag{A6}$$

where $R_X = |\mathbf{X}|$ and $R_Z = |\mathbf{Z}|$. Note that vectors in capital letters are unnormalized vectors.

The derivatives of the **B** matrix are

$$\frac{\partial \mathbf{B}_t}{\partial q_i} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{\partial x_1}{\partial q_i} & \frac{\partial y_1}{\partial q_i} & \frac{\partial z_1}{\partial q_i} \\ 0 & \frac{\partial x_2}{\partial q_i} & \frac{\partial y_2}{\partial q_i} & \frac{\partial z_2}{\partial q_i} \\ 0 & \frac{\partial x_3}{\partial q_i} & \frac{\partial y_3}{\partial q_i} & \frac{\partial z_3}{\partial q_i} \end{pmatrix}\ , \tag{A7}$$

where q_i represents one of the coordinates of the atoms A, B or C.

For implementation purposes, we found it convenient to express these partial derivatives as

$$\frac{\partial x_j}{\partial q_i} = \frac{\partial}{\partial q_i}\left(\frac{X_j}{R_X}\right) = \frac{1}{R_X}\frac{\partial X_j}{\partial q_i} - \frac{X_j}{R_X^3} \times \left(X_1\frac{\partial X_1}{\partial q_i} + X_2\frac{\partial X_2}{\partial q_i} + X_3\frac{\partial X_3}{\partial q_i}\right) \tag{A8}$$

$$\frac{\partial z_j}{\partial q_i} = \frac{\partial}{\partial q_i}\left(\frac{Z_j}{R_Z}\right) = \frac{1}{R_Z}\frac{\partial Z_j}{\partial q_i} - \frac{Z_j}{R_Z^3} \times \left(Z_1\frac{\partial Z_1}{\partial q_i} + Z_2\frac{\partial Z_2}{\partial q_i} + Z_3\frac{\partial Z_3}{\partial q_i}\right) \tag{A9}$$

$$\frac{\partial y_j}{\partial q_i} = \left[\frac{\partial \mathbf{z}}{\partial q_i}\otimes\mathbf{x}\right](j) + \left[\mathbf{z}\otimes\frac{\partial \mathbf{x}}{\partial q_i}\right](j)\ . \tag{A10}$$

In these equations, the lower-case variables correspond to unit vectors and the upper-case variables to the quantities defined in Eqs. (A4)–(A6). The indices i and j specify a Cartesian coordinate and an index in parenthesis (j) means the jth element of the cross product of two vectors.

The derivatives in Eq. (A8) can be expressed as follows:

$$\frac{\partial X_j}{\partial A_i} = -\frac{A_i}{R_A^2}\left[\frac{\mathbf{A}\otimes\mathbf{B}(j)}{R_A R_B} + \frac{\mathbf{C}\otimes\mathbf{A}(j)}{R_C R_A}\right] + \frac{\epsilon_{ijk}}{R_A}\left[\frac{C_k}{R_C} - \frac{B_k}{R_B}\right] \tag{A11}$$

$$\frac{\partial X_j}{\partial B_i} = -\frac{B_i}{R_B^2}\left[\frac{\mathbf{B}\otimes\mathbf{C}(j)}{R_B R_C} + \frac{\mathbf{A}\otimes\mathbf{B}(j)}{R_A R_B}\right] + \frac{\epsilon_{ijk}}{R_B}\left[\frac{A_k}{R_A} - \frac{C_k}{R_C}\right] \tag{A12}$$

$$\frac{\partial X_j}{\partial C_i} = -\frac{C_i}{R_C^2}\left[\frac{\mathbf{C}\otimes\mathbf{A}(j)}{R_C R_A} + \frac{\mathbf{B}\otimes\mathbf{C}(j)}{R_B R_C}\right] + \frac{\epsilon_{ijk}}{R_C}\left[\frac{B_k}{R_B} - \frac{A_k}{R_A}\right]\ , \tag{A13}$$

where ϵ_{ijk} is the standard permutation symbol with the definition

$$\begin{aligned} \epsilon_{ijk} &= +1 \text{ if } ijk \text{ is an even permutation of 1,2,3} \\ &= -1 \text{ if } ijk \text{ is an odd permutation of 1,2,3} \\ &= 0 \text{ otherwise .} \end{aligned} \tag{A14}$$

The gradient $\partial Z_j/\partial q_i$ can be expressed as

$$\frac{\partial Z_j}{\partial q_i} = \left[\frac{\partial \mathbf{x}}{\partial q_i}\otimes\mathbf{a}\right](j) + \left[\mathbf{x}\otimes\frac{\partial \mathbf{a}}{\partial q_i}\right](j)\ . \tag{A15}$$

The derivatives of the vector **x** occurring in this expression have already been defined in Eq. (A8), whereas the derivatives of the vector **a** are easy to determine from the definition of Eq. (A3).

The only derivatives that are required for the **H** matrix are those of L with respect to the coordinates of the atoms O, A, B and C and these may also be determined straightforwardly using the definition of Eq. (11) and the expressions for the derivatives of the vectors **x** and **a**.

This completes the formulation of the derivatives except to say that the presented formulae assume that the boundary atom is at the origin. To generalize to the case where the boundary atom is elsewhere requires a simple, linear transformation of coordinates and an application of the product rule. This results in identical derivative expressions for the atoms A, B and C and derivative expressions for the boundary atom that are the negative of the sum of the derivative expressions for the atoms A, B and C.

References

1. Warshel A (1991) Computer modeling of chemical reactions in enzymes and solutions. Wiley, New York
2. Gao J (1995) In: Lipkowitz KB, Boyd DB (eds) Reviews in computational chemistry 7. VCH, New York, pp 119–185
3. Amara P, Field MJ (1999) In: Leszczynski J (ed) Hybrid potentials for large molecular systems in computational molecular biology. Elsevier, Amsterdam, pp 1–33
4. Singh UC, Kollman PA (1986) J Comput Chem 7: 718–730
5. Field MJ, Bash PA, Karplus M (1990) J Comput Chem 11: 700–733
6. Warshel A, Levitt M (1976) J Mol Biol 103: 227–249
7. Théry V, Rinaldi D, Rivail J-L, Maigret B, Ferenczy GG (1994) 15: 269–282
8. Assfeld X, Rivail J-L (1996) Chem Phys Lett 263: 100–106
9. Gorb LG, Rivail J-L, Théry V, Rinaldi D (1996) Int J Quantum Chem 30: 1525–1536
10. Monard G, Loos M, Théry V, Baka K, Rivail J-L (1996) Int J Quantum Chem 58: 153–159
11. Maseras F, Morokuma K (1995) J Comput Chem 16: 1170–1180
12. Bersuker IB, Leong MK, Boggs JE, Pearlman RS (1997) Int J Quantum Chem 63: 1051–1063
13. Bakowies D, Thiel W (1996) J Phys Chem 100: 10580–10594
14. Zhang Y, Lee T-S, Yang W (1999) J Chem Phys 110: 46–54

15. Gao J, Amara PA, Alhambra C, Field MJ (1998) J Phys Chem A 102: 4714–4721
16. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminanathan S, Karplus M (1983) J Comput Chem 4: 187–217
17. Dewar MJS, Zoebisch EG, Healy EF, Stewart JJP (1985) J Am Chem Soc 107: 3902–3209
18. MacKerell Jr AD, Bashford D, Bellott M, Dunbrack Jr RL, Evanseck J, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE III, Roux B, Schlenkrich M, Smith J, Stote R, Straub JE, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M (1998) J Phys Chem B 102: 3586–3616
19. Dewar MJS, Dieter KM (1986) J Am Chem Soc 108: 8075–8086
20. Eck MJ, Shoelson SE, Harrison SC (1993) Nature 362: 87–91
21. Protein Data Bank. Brookhaven National Laboratory, Upton, NY
22. Brünger AT, Karplus M (1988) Proteins Struct Funct Genet 4: 148–156
23. Nose S (1984) Mol Phys 52: 255–268
24. Feng M-H, Philippopoulos M, MacKerell AD Jr, Lim C (1996) J Am Chem Soc 118: 11265–11277