



<http://www.diva-portal.org>

## Postprint

This is the accepted version of a paper published in *Journal of Computational Chemistry*. This paper has been peer-reviewed but does not include the final publisher proof-corrections or journal pagination.

Citation for the original published paper (version of record):

Lundberg, M. (2012)  
Understanding Cross-Boundary Events in ONIOM QM:QM' Calculations.  
*Journal of Computational Chemistry*, 33(4): 406-412  
<http://dx.doi.org/10.1002/jcc.21982>

Access to the published version may require subscription.

N.B. When citing this work, cite the original published paper.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-181239>

# Understanding Cross-Boundary Events in ONIOM QM:QM' Calculations

Marcus Lundberg<sup>†,\*</sup>

*<sup>†</sup>Fukui Institute for Fundamental Chemistry, Kyoto University, 34-4 Takano Nishihiraki-cho,  
Sakyo-ku, Kyoto 606-8103, Japan*

<sup>#</sup> Present address: Department of Physical and Analytical Chemistry, Uppsala University, Box 518, 751 20 Uppsala, Sweden.

<sup>\*</sup> Corresponding Author: E-mail: [marcus.lundberg@kvac.uu.se](mailto:marcus.lundberg@kvac.uu.se), Telephone: +46-18-471-3708. Fax: +46-18-471-5830.

## Abstract

QM:QM' models, where QM' is a fast molecular orbital method, offers advantages over standard QM:MM models, especially in the description of charge transfer and mutual polarization between layers. The ONIOM QM:QM' scheme also allows for reactions across the layer boundary, but the understanding of these events is limited. To explain the factors that affect cross-boundary events, a set of proton transfer processes, including the acylation reaction in serine protease, have been investigated. For reactions *inside out*, i.e., when a group breaks a bond in the high layer and forms a new bond with a group in the low layer, QM' methods that are overbinding relative to the QM method, e.g., Hartree-Fock vs B3LYP, can severely overestimate the exothermicity of the reaction. This might lead to artificial reactivity across the QM:QM' boundary, a phenomenon called *model escape*. The accuracy for reactions that occur *outside in*, i.e., when a group in the low layer forms a new bond with the high layer, is mainly determined by the QM' calculation. Cross-boundary reactions should generally be avoided in the present ONIOM scheme. Fortunately, a better understanding of these events makes it easy to design stable ONIOM QM:QM' models, e.g., by choosing a proper *model* system. Importantly, an accurate description of cross-boundary reactions would open up possibilities to simulate chemical reactions without *a priori* limiting the reactivity in the design of the computational model. Challenges to implement a simulation scheme (ONIOM-XR) that can automatically handle chemical reactions between different layers are briefly discussed.

## I. Introduction

Multi-scale methods, e.g., quantum mechanics/molecular mechanics (QM/MM), can achieve a balanced combination of time and accuracy and have been successfully applied to a wide spectrum of applications.<sup>1,2</sup> The success of QM/MM has stimulated the development of the multi-scale ONIOM method, which is based on an extrapolation scheme rather than an additive scheme.<sup>3-7</sup> One of the major advantages of the ONIOM scheme is that it can combine any methods without previous parameterization or significant coding efforts. It is also possible to combine any number of layers, although the current implementation is limited to three layers.<sup>5,8</sup>

Although the ONIOM method is different in design and implementation compared to other multi-scale methods, there are also many similarities. In both QM/MM and ONIOM QM:MM models the combination of two fundamentally different descriptions of a molecular system leads to basic problems, e.g., in the description of charge transfer between MM and QM layers,<sup>9</sup> and the lack of polarization of the MM layer (with standard force fields). A possible alternative is to combine the expensive QM method with fast method based on molecular orbital theory (QM'). To achieve high accuracy, an expensive correlated wavefunction calculation can be embedded in the environment of a relatively fast DFT or wavefunction method.<sup>10,11</sup> For large systems an alternative is the combination of a density functional as QM method with a semiempirical method or density-functional tight-binding (DFTB) as QM' method.<sup>12-14</sup> The flexibility of the ONIOM QM:QM' scheme makes it possible to choose an optimum combination of QM and QM' for a wide range of different phenomena.

In almost all multi-layer designs, groups that are expected to be reactive are all placed in the QM layer. Consequently, the concept of cross-boundary reactivity has not been widely discussed. Here it is shown that the QM:QM' boundary distorts the shape of the potential energy surface and often causes significant errors in relative energies. In some cases the presence of the boundary itself can trigger, or prevent, cross-boundary reactions, and may make it impossible to find the expected minimum structure.

To explain when and why these events occur, a series of simple proton transfer reactions between amino-acid side chains are investigated, as well as a well-known catalytic reaction, peptide bond cleavage in serine proteases. Several QM:QM' combinations are included to illustrate how the relative behavior of the methods affect the description of cross-

boundary events. The primary goal of the investigation is to help modelers design stable ONIOM QM:QM' models where cross-boundary reactions can be controlled.

We also recognize that the possibility to describe reactions between layers is a big principal advantage, as the system's reactivity no longer is limited by the initial selection of the *model* system. In many enzymatic reactions, a primary chemical reaction leads to large changes in pK<sub>a</sub> values, which triggers secondary proton transfer reactions to, or from, the active site. Fascinating long-term goals are to describe proton pumping in cytochrome c oxidase and photosystem II, where redox reactions in transition metal systems (treated by QM) are coupled to proton pumping (treated by QM') across an electrochemical gradient. We will briefly discuss the possibilities to implement an ONIOM scheme that accurately handles chemical reactions between different layers. Such a method would significantly extend the range of multi-scale models applied to chemical reactions.

## II. Computational details

The basic concepts in the ONIOM approach are repeated to better explain its advantages and limitations. In the two-layer ONIOM approach, the *high* method is applied only to a selected *model* system, a selection that typically contains the reacting core region. The cheaper *low* method is applied to both the *real* system and the *model* system, and the expensive *real,high* calculation is approximated from an extrapolation scheme. The ONIOM energy in a two-layer calculation is:

$$E^{ONIOM} = E^{model,high} + E^{real,low} - E^{model,low} \quad (1)$$

The flexibility in the choice of methods comes from the fact that all calculations are made on complete systems.

In mechanical embedding (ME) the calculation of the *model* system is independent of the *real* system, while in electronic embedding (EE) the *model* system Hamiltonians include the charge distribution of the environment.<sup>6,15,16</sup> In mechanical embedding interactions between layers are included only in the QM' calculation of the full system, the *real,low* calculation. This calculation is performed over the entire (*real*) system where there are no artificial boundaries between layers, and therefore naturally describes mutual polarization and charge transfer between layers. In fact, the *real,low* calculation can describe any type of

interaction across the QM:QM' boundary, including chemical reactions between groups in different layers.

A convenient measure for evaluating ONIOM calculations is the  $\Delta S$  value that describes the environmental effect at the different computational levels:<sup>17</sup>

$$\Delta S^{high} = (E^{real,high} - E^{model,high})_{product} - (E^{real,high} - E^{model,high})_{reactant} \quad (2)$$

$$\Delta S^{low} = (E^{real,low} - E^{model,low})_{product} - (E^{real,low} - E^{model,low})_{reactant} \quad (3)$$

$\Delta S^{high}$  is the error introduced when the *real* system is approximated by the *model* system.  $\Delta S^{low}$  is the corresponding value for the low-level method, and should be seen as the correction introduced by adding the low layer. The ONIOM energy is exact, i.e., becomes the same as the target calculation,  $\Delta E^{real,high}$ , if the low-level method describes the environmental effect in the same way as the high-level method ( $\Delta S^{low} = \Delta S^{high}$ ).

Terms that have a specific meaning in the ONIOM scheme, e.g., *model* and *real* systems and *high* and *low* methods, are written in italic. In the discussion the terms high and low layer are also used to describe different parts of the system. The high layer consists of atoms in the *real* system that are treated by the high-level method (QM), while the low layer consists of atoms that are only treated by the low-level method (QM'). The difference between the high-layer and the *model* system is that the later includes also potential link atoms. The ONIOM scheme never calculates the low layer explicitly, it is only included in the calculations as part of the *real* system.

The hybrid density functional B3LYP, with the 6-31G(d) basis set, has been used as QM method in all ONIOM calculations. For brevity, this method is written as B3LYP in ONIOM combinations. As will be shown later, an important aspect of the QM:QM' selection is the relative description of bond dissociation, and from this perspective B3LYP is a convenient intermediate between HF and a GGA functional like BLYP. The limited accuracy of the *high* method is not important in this study because the main purpose is to analyze how well the approximate ONIOM scheme reproduces the target *real,high* calculation. Six different methods have been used for QM': HF/6-31G(d), HF/3-21G, BLYP/3-21G, SCC-DFTB, AM1, and PM3. These methods are realistic choices for the low-level methods in practical ONIOM QM:QM' combinations.

The selection of test systems is inspired by proton transfer reactions in enzymes. Histidine side chains are modeled using 4-methylimidazole, and aspartate side chains are modeled using acetic acid or acetate (depending on the protonation state).

Proton transfer reactions between aspartate and histidine models (Tests 2-5) were performed in a continuum solvent model. This significantly stabilizes the charge-separated state gives a more reasonable potential energy profile for the *real* system. The default PCM settings in Gaussian were used, including water as a solvent. ONIOM calculations were performed using the ONIOM-PCM/C scheme where the polarized continuum is applied only to the *real,low* calculation.<sup>18</sup> Calculations have been performed using a private development version of Gaussian, with the capability to perform SCC-DFTB and ONIOM-PCM calculations.<sup>14,19</sup>

Calculations where the PCM model is applied also to the *model* systems show limited effects (<1 kcal/mol) on the energy profiles for O–H bond dissociation in acetic acid and N–H bond dissociation in protonated 4-methylimidazole (bond lengths stretched to 2.0 Å). The reason for the small solvent effect is that the *model* calculations describe homolytic bond cleavage rather than charge separation. That is, the charge of the leaving group (hydrogen) is closer to 0 than +1. This is true also when the *model* system is solvated. Homolytic bond dissociation in the *model* system also explains the large difference between HF and BLYP. As in H<sub>2</sub> dissociation, HF significantly overestimates the energy of the dissociated state because it lacks left-right correlation.

Data are presented as potential energy profiles where markers show the actual data points and connecting lines are constructed using a quadratic Bezier algorithm. The reference point (zero of energy) is the fully optimized “reactant” at each respective level, without any constraints. The exception is the proton transfer from protonated 4-methylimidazole to acetate (Tests 3 and 5), for which no minimum exists in the “reactant” region. Instead the reference point is a structure with a frozen N–H distance of 1.05 Å. Reaction coordinates are explored by using a single reaction coordinate, the O–H bond for acetic acid and the N–H bond for protonated 4-methylimidazole. For some method combinations stable minima cannot be located in the product region, and sometimes no stable minima can be located at any part of the potential energy surface. Failures to locate minima are not due to the choice of reaction coordinate. Optimizations starting from the proper minima of the *real,high* calculations, give the same outcomes as stepwise optimizations along the assigned reaction coordinate.

The model design of the serine protease system was taken from reference <sup>20</sup>. Stationary points along the reaction coordinate were optimized using single-layer B3LYP/6-31G(d) calculations.

### III. Results and discussion

Cross-boundary reactions can occur in two directions. A reaction is classified as *inside out* when a group in the *model* system both breaks a bond in the high layer and forms a new bond with the low layer (a group that is only part of the *real* system). A reaction is classified as *outside in* if a group in the low layer forms a bond with an atom in the high layer. The latter classification can be applied even if no bond is broken in the low layer, see Figure 1.

Figure 1.

The ONIOM descriptions of the *inside-out* and *outside-in* reactions have fundamental differences. The analysis is therefore divided into different sections, starting with the more challenging *inside-out* reactions.

#### A. Crossing from inside to outside

The reaction of interest is  $AX + B \rightarrow A + XB$ . The analysis is simplified by first making A and B identical species, giving  $AX + A \rightarrow A + XA$ , a process that is thermoneutral at all levels of theory. To make this a cross-boundary reaction, the ONIOM selection is made so that AX is in the *model* system and A is only in the *real* system  $A_{model}X_{model} + A_{real} \rightarrow A_{model} + X_{model}A_{real}$ . This is no longer an identity reaction, and as all methods give the same reaction energy for the *real* system, deviations from thermoneutrality can be attributed to presence of the QM:QM' boundary.

A test is made for a proton transfer between two histidine side chains, Test 1 in Figure 2. In this reaction, A is 4-methylimidazole and X is a proton. The relative energy of the product state ( $A_{model} + A_{real}X_{model}$ ) compared to the reactant state ( $A_{model}X_{model} + A_{real}$ ), calculated with B3LYP/6-31G(d) as QM and with several different QM' methods, is given in Table 1.

Figure 2.

Table 1.

The most striking results are the large errors for the combination of B3LYP and HF. In a reaction where the two methods give exactly the same reaction energy in single-layer calculations, the errors in the B3LYP:HF ONIOM combinations are >10 kcal/mol. To better understand the behavior of the B3LYP:HF/3-21G model, a relaxed scan of the  $N_{\text{model}}\text{-H}$  distance is performed from 1.00 to 2.00 Å. The relative ONIOM energy along the scan, and the individual terms in the ONIOM expression are shown Figure 3.

Figure 3.

From Figure 3 it is clear that the proton transfer reaction is thermoneutral in the *real* system, and that the large contributions to the relative energy comes from the two *model* calculations that effectively describe N–H bond dissociation. The contribution to the ONIOM energy from the two model calculations ( $\Delta E^{\text{Model,High}} - \Delta E^{\text{Model,Low}}$ ) is highly negative at long distances because the QM' method HF is overbinding compared to the QM method B3LYP. Results for all method combinations broken down by ONIOM components are given in Table S1 in the Supporting information.

Further insights into the *inside-out* process can be obtained by a different view of the low-energy region of potential energy surface, see Figure 4. In this plot, the large *model,high* and *model,low* contributions have not been plotted separately, but rather their contribution to the relative ONIOM energy ( $\Delta E^{\text{Model,High}} - \Delta E^{\text{Model,Low}}$ ). This contribution mainly reflects the difference in bond dissociation energy between QM and QM'.

Figure 4.

Due to the choice of a single N–H bond as reaction variable, the energy profile for the *real* system is not symmetric with regard to the reactant and product minima, but still shows a thermoneutral reaction with two minima corresponding to reactant and product. As expected, the proton transfer barrier is higher with HF compared to B3LYP. At short  $N_{\text{model}}\text{-H}$  distances the increase of the *real,low* energy dominates over the contributions from the *model* calculations, and a local minimum can be found in the reactant region with a reasonable

structure. However, at distances  $>1.25 \text{ \AA}$  the increasingly negative contributions from  $\Delta E^{Model,High} - \Delta E^{Model,Low}$  makes the ONIOM potential fully dissociative. The incorrect preference for dissociation caused by the presence of the QM:QM' boundary is here named *model escape*.

Optimizations in the product region do not lead to any minimum, and results are in general unstable as the model allows for large changes in the alignment between donor and acceptor. In the present calculation a distortion of the system around  $N_{model}-H=1.9 \text{ \AA}$  leads to a further drop in the ONIOM energy.

Although the errors associated with this reaction are large in relation to the total reaction energy, this is not a spurious error of the ONIOM method. The error should always be evaluated in relation to the environmental effect, and in the *inside-out* reaction, the presence of the low layer changes the description of the N–H bond stretch from a pure bond dissociation to a normal proton transfer reaction. The environmental effect  $\Delta E^{Real,High} - \Delta E^{Model,High}$  is 62 kcal/mol at an  $N_{model}-H$  distance of 1.6  $\text{\AA}$ , and the ONIOM error of 11 kcal/mol is 18 % of that effect.

All the above results are obtained for closed shell singlet states in all calculations. However, as the cross-boundary reaction involves artificial bond breaking in the *model* system, results become radically different if the singlet-triplet instability is taken into account. When the open-shell singlet is converged for both *model,high* and *model,low* at an  $N_{model}-H$  distance of 2.0  $\text{\AA}$ , the relative ONIOM energy goes from -20.2 to +10.1 kcal/mol, because HF overestimates the stability of the spin-contaminated open-shell species. The reason mainly closed-shell results are discussed in the present investigation is that the open-shell solutions are not likely to appear, unless the *model* systems are treated by an unrestricted formalism and separately tested for instability.

Now that the behavior of the B3LYP:HF/3-21G combination has been understood, a similar analysis can be applied to the other QM:QM' combinations in Table 1. The results for HF/6-31G(d) as QM' method, see Figure S1 in the Supporting information, are similar to those of HF/3-21G. The critical factor is how bond dissociation is described, and this depends more on the choice of method than on the choice of basis set. Again the combination of a low barrier in the *real,low* calculation and the *model escape* mechanism cause the proton to spontaneously “leave” the model system and optimizations in the product region fail.

The results are very different with BLYP/3-21G as QM' method. As BLYP/3-21G is underbinding relative to B3LYP, the contribution  $\Delta E^{Model,High} - \Delta E^{Model,Low}$  becomes positive

as the  $N_{\text{model}}\text{-H}$  distance increases. The ONIOM B3LYP:BLYP combination thus disfavors proton transfer from the *model* system, even to the extent that the expected minimum in the product region disappears completely, see Figure 5. The lack of product minima is also due to the fact that the *real,low* BLYP/3-21G calculation places the proton symmetrically between the two histidine residues.

Figure 5.

Among the very fast molecular orbital methods, DFTB and AM1 gives apparently small errors (absolute values  $< 1$  kcal/mol) in the relative ONIOM energy, while the PM3 error is larger (7.0 kcal/mol). However, closer inspection of the potential energy surfaces, see Figure S1, shows that all these method combinations give distorted potential energy surfaces. Both AM1 and PM3 are dissociative at long  $N_{\text{model}}\text{-H}$  distances, while the DFTB “product” minimum is pushed very close to the reactant ( $N_{\text{model}}\text{-H}=1.34$  Å).

Several important conclusions about *inside-out* reactions can be drawn from this simple proton transfer. QM’ methods that are overbinding relative to the QM method artificially favor cross-boundary reactions, even to the extent that reactions can be spontaneously triggered by the presence of the QM:QM’ boundary. Underbinding QM’ methods artificially prevent groups in the *model* system from reacting with the surrounding. Even for method combinations that give small energetic errors along the reaction coordinate, e.g., B3LYP:AM1, the wavefunctions (electron densities) and geometries become unstable on the product side.

Considering the general failure to describe a proton transfer reaction, a relevant question is how well QM:QM’ describes non-bonded interactions,<sup>21,22</sup> especially low-barrier hydrogen bonds i.e., hydrogen bonds that are subject to proton transfer reactions. In the present study, interaction energies have been calculated for a water dimer that has no cross-boundary reactivity and the protonated 4-methylimidazole dimer (Test 1 in Figure 2). As expected, interaction energies are close to the low-level (QM’) results. Under the condition that the proton stays bonded to a group in the *model* system, the strength of the low-barrier hydrogen bonds are described with similar, or even slightly better, accuracy than normal hydrogen bonds. Detailed results and discussions are provided in the Supporting information (Table S2).

In most calculations, the main interest is a primary reaction within the *model* system. An *inside-out* reaction at the boundary would lead to an incorrect description of the primary reaction because the electronic structure of the *model* system is different than expected, e.g., what appears to be a neutral histidine residue is better described as a histidine cation radical. The reason for the discrepancy is that the proton transfer in the *real* system is artificially described as a homolytic bond cleavage of the N–H bond in the *model* system. *Inside-out* reactions should therefore always be avoided in the present ONIOM scheme. This can easily be done either by choosing a more appropriate combination of QM and QM', or by moving the QM:QM' boundary so that all possible cross-boundary reactions have such a high barrier in the *real,low* calculation that effects from the layer boundary can be overcome.

However, in many studies the detailed reaction mechanism is not known, and the presence of a cross-boundary reaction, can provide important information about the true reactivity of the system. To help interpret that information two prototypical *inside-out* reactions will be discussed below.

**Endothermal reactions in the *real,low* description.** In most multi-layer setups, the layer assignment is made so that all groups assumed to be reactive are placed in the high layer. Reactions between the high layer and the surrounding are thus expected to be endothermal or have very high barriers. To illustrate the result for an endothermal reaction in a simple test system, the protonated 4-methylimidazole in the high-layer of Test 1 is replaced by acetic acid (a model of aspartate), Test 2 in Figure 6. In this system a polarized continuum model (ONIOM-PCM/C) have been applied to give a more reasonable description of the charge separation process.<sup>18</sup> Two QM' methods have been tested, one overbinding (HF/3-21G) and one underbinding (BLYP/3-21G).

Figure 6.

Compared to the identity reaction (Test 1), the only thing that changes is that the reaction energy, and the difference in reaction energy between QM and QM', affect the ONIOM result. In the B3LYP PCM description, the proton transfer from acetic acid to 4-methylimidazole is endothermal, and the relative energy increases with increasing distance between acetate and the proton, see Figure 7. It should be noted that the proton transfer from acetic acid to 4-methylimidazole should really be exothermal (or rather exergonic) based on

their relative pKa values. However, this error has no effect on the evaluation of the ONIOM QM:QM' method. A single-layer HF/3-21G calculations gives a similar potential, but the B3LYP:HF ONIOM model favors the *inside-out* reaction to the point that it becomes exothermic instead of endothermic, see Figure 7. Again, the product region is dissociative and no stable product structure can be found. If BLYP is instead used as the low-layer method, the endothermicity of the reaction is overestimated, see Figure S2 in the supporting information.

Figure 7.

**Exothermic in the low,real description.** In most cases reactions *inside out* occur because they are exothermic in the QM' description. The initial layer selection is then inappropriate and needs to be modified so that the reacting groups are properly included in the *model* system. It is therefore important that the information about reactivity is not lost due to errors associated with the QM:QM' boundary. To illustrate the problem, the same system is used, although this time the proton is transferred from a high-layer protonated 4-methylimidazole to acetate (Test 3 in Figure 6), which makes it an exothermic process at the present level. B3LYP:HF further favors the *inside-out* reaction, although, as expected, with a large error in the reaction energy and no stable product structure, see Figure S3. If the process of cross-boundary reactivity is well understood, it is obvious that the *model* system needs to be modified.

The problem is instead the B3LYP:BLYP combination because it artificially prevents the *inside-out* reaction and predicts that the proton stays on 4-methylimidazole, see Figure 8. Once again, this error can be fixed by moving the QM:QM' boundary, but as the initial calculations will not show any cross-boundary reactivity, there would be no reason to modify the layer selection. An underbinding QM' method should therefore not be used to probe the reactivity of the system. Table 2 describes how different QM:QM' combinations affect the outcome of the ONIOM calculations.

Figure 8.

Table 2.

## B. Crossing from outside to inside – breaking bonds in the real system

Reactions where a group in the *real* system forms a bond with a group in the *model* system, can be classified as *outside-in* reactions, see Figure 1. These reactions are fundamentally different from the *inside-out* reactions, because they do not lead to any artificial bond dissociation in the *model* system. They therefore simpler to analyze, and the behavior can be properly understood from a single reaction.

The example is an endothermic reaction, proton transfer from acetic acid to 4-methylimidazole, where only the 4-methylimidazole appears in the *model* system (Test 4 in Figure 6). For this reaction, the B3LYP:HF/3-21G energy profile largely follows the *real,low* calculation, see Figure 9, with only minor improvements. In the *outside-in* reaction, the reactive group, i.e., the proton, is not included in the calculations of the *model* system, but there is a small indirect effect through the change in geometry. The accuracy of the ONIOM calculation is therefore mainly determined by the *real,low* calculation.

Figure 9.

There are no major differences in the description of endothermic and exothermic *outside-in* reactions. Results for the exothermic proton transfer from protonated 4-methylimidazole to acetate (Test 5) are shown in Figure S4. Again, the ONIOM energy is close to the *real,low* result.

In most applications, the primary interest is a process inside the *model* system. For this process, the effect of the secondary cross-boundary event is solely included in the *real,low* calculation. As the environmental effect of a secondary chemical reaction can be very big, the reactive low-layer group should be moved into the *model* system.

## C. QM:QM' model of the catalytic triad in serine protease

To illustrate the effect of the QM:QM' boundary in a more realistic situation, calculations are made on peptide cleavage in the serine protease trypsin, a textbook example of an enzymatic reaction.<sup>23</sup> An important catalytic motif is a conserved Ser-His-Asp triad, known as the catalytic triad. The first part of the reaction is shown in Figure 10. Ser195 performs a

nucleophilic attack on the substrate peptide and transfers a proton to His57 (INT1). This proton transfer is significantly stabilized by the polarization of His57 from the neighboring negative residue Asp102.<sup>24</sup> Another possibility is that the proton is actually transferred from His57 to Asp102 in the oxyanion intermediate (INT1). In the next step, the peptide C–N bond breaks and the newly formed N-terminal group accepts the proton from His57 (INT2), which completes the acylation part of the reaction.

Figure 10.

Following the textbook mechanism, a reasonable QM:QM' model of the catalytic triad is to put the peptide bond and the reacting side chains (Ser195 and His57) in the *model* system using a highly accurate QM method, while the polarizing residue Asp102 and other hydrogen bonding residues can be treated with QM'. To isolate the effect of the potential proton transfer from His57 to Asp102, the actual ONIOM test system was designed with all groups except Asp102 in the *model* system, see Figure 11.

Figure 11.

Geometries were initially optimized using single-layer B3LYP/6-31G(d),<sup>20</sup> and all ONIOM calculations were performed at the B3LYP geometries. Deviations between methods are thus not affected by spurious optimizations to different local minima. At the B3LYP level, a proton is transferred from His57 to Asp102 in INT1, which gives a good opportunity to test an *inside-out* reaction.

At the B3LYP/6-31G(d) level, the relative energy of INT1 and INT2 is 26.6 and 11.8 kcal/mol respectively, see Table 3. The deviation between B3LYP/6-31G(d) and the single-layer HF/6-31G(d) calculation is very small, 0.1 and 1.4 kcal/mol for INT1 and INT2 respectively. However, the B3LYP:HF combination with Asp102 outside the *model* system has an error of -24.2 kcal/mol for INT1 (but only -0.5 kcal/mol for INT2). As INT1 is the state where the *inside-out* reaction occurs, this is a clear illustration of the effects of cross-boundary reactions in ONIOM QM:QM' calculations.

Results for other QM' methods are given in Table 3, and they all show significant errors for INT1 due to the cross-boundary event, but perform well for INT2 where Asp102 just has a polarizing effect.

Table 3.

#### D. Prospects for a consistent treatment of cross-boundary events

The results show that the present ONIOM QM:QM' scheme is not designed to handle cross-boundary reactions. Although not discussed here in detail, similar issues are expected to appear also in QM:QM' schemes with electronic embedding, although the situation will be complicated by the presence of external charges in the *model* calculations.<sup>15,25,26</sup>

An accurate treatment of cross-boundary reactions would be a big principal advantage, as the reactivity of the system would no longer be limited by the initial setup of the computational model. A long-term goal is to be able to describe the mechanism of proton pumping in photosystem II and cytochrome c oxidase. For these reactions, the QM method should describe the redox reactions in the metal centers, while the QM' method should describe the long-range proton transfer reactions. However, as the proton flow is also involved in water oxidation/oxygen reduction, the model must also be able to accurately describe the transfer of groups (protons) between the QM and the QM' layer.

Several schemes have been designed to allow for transfer of groups between layers in multi-scale models,<sup>27</sup> e.g., the “hot spot” method,<sup>28</sup> adaptive partitioning (AP),<sup>29</sup> and the ONIOM-XS scheme.<sup>30</sup> However, the need to simulate bond breaking and formation between groups in different layers introduces additional challenges. Here some of the challenges to develop an ONIOM-XR (eXtension to Reactions) scheme are briefly discussed. Such a scheme must automatically move the boundary between QM and QM' when reactions occur, and must therefore be able to: 1) identify cross-boundary events, 2) select which atoms and how many electrons to move across the boundary, 3) ensure a continuous connection of potential energy and gradients.

Both *inside-out* and *outside-in* reactions include bond formation between layers. To identify a cross-boundary reaction, one possible solution is thus to evaluate distances between atoms in different layers, e.g., if a hydrogen in the QM' layer approaches within 1.4 Å of a nitrogen in the QM layer. Other options, e.g., based on gradients of different ONIOM components are not likely to be able to cover all types of reactions.

The movement of groups across the boundary requires an automated procedure for making ONIOM *model* selections based on chemical groups. In a protein, this requires a

library of partitions for each amino acid and an input format that uses information about residue type in the input. Significantly more problematic is probably to determine the identity of the group that is transferred between layers. As the number of electrons in QM and QM' layers are set in the input, the scheme must be able to recognize the difference between e.g., a hydrogen atom transfer and a proton transfer, This could possibly be solved by an analysis of the wavefunction, e.g., population analysis.

Finally, the connection of potential energy and gradients can be made in similar mode as proposed for the ONIOM-XS scheme, possibly modified to ensure that both energy and momentum is conserved.<sup>29</sup> One problem is that the present QM:QM' scheme also leads to error in reaction energies beyond the errors in the QM' method, and these errors should not propagate into the ONIOM-XR scheme. The design of ONIOM-XR is a therefore an elaborate project, and has not been attempted as part of the present study.

## Summary

We have investigated the properties of the boundary between layers in ONIOM QM:QM' models. The description of environmental effects from the *real, QM'* calculation leads to a simple description of charge transfer and mutual polarization between layers, but the scheme also allows for reactions between layers. An understanding of these processes are of great use when designing accurate and reliable QM:QM' models

For reactions *inside out*, i.e., when a group in the high layer breaks a bond in the high layer and forms a new bond in the low layer, the model system describes an artificial bond dissociation. This process can be very differently described by *high* and *low* methods, and the ONIOM results therefore heavily depend on the relative descriptions of bond dissociation in the two *model* calculations. QM' methods that are overbinding relative to the QM method, e.g., Hartree-Fock relative to B3LYP, can lead to artificial reactivity across the QM:QM' boundary, a phenomenon called *model escape*. This artificial reactivity, which leads to highly unstable models, is sometimes limited by the presence of reaction barriers, but can still be crossed, e.g., in molecular dynamics simulations.

On the other hand, QM' methods that are underbinding, e.g. pure DFT functionals, underestimate the cross-boundary reactivity, and may artificially prevent reactions from occurring. This is also problematic because it can lead to incorrect conclusions about the true reaction mechanism.

Reactions occurring *outside in*, i.e., when a group in the low layer forms a new bond in the high layer, are easier to describe. The difference is that unlike the *inside-out* reaction, there is no artificial bond dissociation in the *model* system. The accuracy of the *outside-in* reaction is mainly determined by the accuracy of the QM' calculation. However, if a reaction inside a different part of the *model* system is the primary focus of the study, the effect of the secondary cross-boundary reaction is only included in the *real,low* calculation. As the environmental effect is expected to be large this can have large effects on the accuracy of the ONIOM calculation.

In the present ONIOM scheme all cross-boundary reactions are expected to give significant energetic errors, and in many times even highly unstable models. These reactions should therefore be avoided. Fortunately, when the properties of the layer boundary are properly understood, it is relatively easy to design stable ONIOM QM:QM' models. A recommended procedure is to start the calculations with a QM' method that is not underbinding. Possible cross-boundary reactions can then be detected, out of which some may be artificial. These reactions are then addressed by moving the QM:QM' boundary until the reactivity occurs within a single layer. At present this can only be done manually. To develop a method that can properly handle reactivity between layers, here tentatively called ONIOM-XR (eXtension to Reactions), is believed to be an elaborate project, and has not been attempted.

## Supplementary material

Expanded section on the performance of QM:QM' for low-barrier hydrogen bonds. Additional tables and graphs that complement the discussion on cross-boundary reactivity.

## Acknowledgments

The author thanks Keiji Morokuma for useful discussion and the Fukui Institute for Fundamental Chemistry Fellowship for financial support. The present work was in part supported by a CREST (Core Research for Evolutional Science and Technology) grant in the Area of High Performance Computing for Multi-scale and Multi-physics Phenomena from the Japan Science and Technology Agency (JST).

## References

1. Lin, H.; Truhlar, D. G. *Theor Chem Acc* 2007, 117(2), 185-199.
2. Senn, H. M.; Thiel, W. *Top Curr Chem* 2007, 268, 173-290.
3. Dapprich, S.; Komaromi, I.; Byun, K. S.; Morokuma, K.; Frisch, M. J. *J Mol Struct Theochem* 1999, 462, 1-21.
4. Maseras, F.; Morokuma, K. *J Comput Chem* 1995, 16(9), 1170-1179.
5. Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. *J Phys Chem* 1996, 100(50), 19357-19363.
6. Vreven, T.; Byun, K. S.; Komaromi, I.; Dapprich, S.; Montgomery, J. A.; Morokuma, K.; Frisch, M. J. *J Chem Theory Comput* 2006, 2(3), 815-826.
7. Morokuma, K. *Proc Jap Acad Series B-Phys Biol Sci* 2009, 85(5), 167-182.
8. Morokuma, K.; Wang, Q. F.; Vreven, T. *J Chem Theory Comput* 2006, 2(5), 1317-1324.
9. Zhang, Y.; Lin, H. *Theor Chem Acc* 2010, 126(5-6), 315-322.
10. Henderson, T. M. *J Chem Phys* 2006, 125(1), 014105.
11. Huang, P.; Carter, E. A. *J Chem Phys* 2006, 125(8), 084102.
12. Cui, Q.; Guo, H.; Karplus, M. *J Chem Phys* 2002, 117(12), 5617-5631.
13. Gogonea, V.; Westerhoff, L. M.; Merz, K. M. *J Chem Phys* 2000, 113(14), 5604-5613.
14. Zheng, G. S.; Lundberg, M.; Jakowski, J.; Vreven, T.; Frisch, M. J.; Morokuma, K. *Int J Quantum Chem* 2009(109), 1841-1854.
15. Hratchian, H. P.; Parandekar, P. V.; Raghavachari, K.; Frisch, M. J.; Vreven, T. *J Chem Phys* 2008, 128(3), 034107.
16. Mayhall, N. J.; Raghavachari, K.; Hratchian, H. P. *J Chem Phys* 2010, 132(11), 114107.
17. Vreven, T.; Morokuma, K. *J Comput Chem* 2000, 21(16), 1419-1432.
18. Vreven, T.; Mennucci, B.; da Silva, C. O.; Morokuma, K.; Tomasi, J. *J Chem Phys* 2001, 115(1), 62-72.
19. Gaussian Development Version, G., Inc., Wallingford CT. 2009.
20. Lundberg, M.; Sasakura, Y.; Zheng, G. S.; Morokuma, K. *J Chem Theory Comput* 2010, 6(4), 1413-1427.
21. Tschumper, G. S.; Morokuma, K. *J Mol Struct Theochem* 2002, 592, 137-147.
22. Re, S.; Morokuma, K. *J Phys Chem A* 2001, 105(30), 7185-7197.
23. Hedstrom, L. *Chem Rev* 2002, 102(12), 4501-4523.
24. Ishida, T.; Kato, S. *J Am Chem Soc* 2004, 126(22), 7111-7118.
25. Parandekar, P. V.; Hratchian, H. P.; Raghavachari, K. *J Chem Phys* 2008, 129(14), 145101.
26. Mayhall, N. J.; Raghavachari, K. *J Chem Theory Comput* 2010, 6(10), 3131-3136.
27. Nielsen, S. O.; Buló, R. E.; Moore, P. B.; Ensing, B. *Phys Chem Chem Phys* 2010, 12(39), 12401-12414.
28. Rode, B. M.; Schwenk, C. F.; Hofer, T. S.; Randolf, B. R. *Coord Chem Rev* 2005, 249(24), 2993-3006.
29. Heyden, A.; Lin, H.; Truhlar, D. G. *J Phys Chem B* 2007, 111(9), 2231-2241.
30. Kerdcharoen, T.; Morokuma, K. *Chem Phys Lett* 2002, 355(3-4), 257-262.

Figure 1. Classification of cross-boundary reactions as *inside out* and *outside in*. Atoms labeled real appear only in the *real* system, not in the *model* system.

Figure 2. Cross-boundary reaction in an ONIOM QM:QM' model. Atoms in the *model* system are shown in ball&stick representation and atoms in the *real* system are shown in stick representation.

Figure 3. ONIOM(B3LYP/6-31G(d):HF/6-31G(d)) energies for a relaxed scan of the  $N_{\text{model}}\text{--H}$  distance from 1.00 Å to 2.00 Å in the system  $[His_{\text{model}}H_{\text{model}} + His_{\text{real}}]^+$  (*inside-out* reaction, Test 1 in Figure 2). The plot also contains the data from a separate scan at the B3LYP/6-31G(d) level. A different view of the low-energy region appears in Figure 4.

Figure 4. ONIOM(B3LYP/6-31G(d):HF/3-21G) potential energy profile for a relaxed scan of the  $N_{\text{model}}\text{--H}$  distance from 1.00 Å to 2.00 Å in the system  $[His_{\text{model}}H_{\text{model}} + His_{\text{real}}]^+$  (*inside-out* reaction, Test 1 in Figure 2). The plot also contains data from a separate scan at the B3LYP/6-31G(d) level.

Figure 5. ONIOM(B3LYP/6-31G(d):BLYP/3-21G) potential energy profiles for a relaxed scan of the  $N_{\text{model}}\text{--H}$  distance from 1.00 Å to 2.00 Å in the system  $[His_{\text{model}}H_{\text{model}} + His_{\text{real}}]^+$  (*inside-out* reaction, Test 1 in Figure 2). The plot also contains data from a separate scan at the B3LYP/6-31G(d) level.

Figure 6. Four test systems for cross-boundary reactions in ONIOM QM:QM' models. The reactions are proton transfers (protons highlighted) from the residue on the left to the residue

on the right. Atoms in the *model* system are shown in ball and stick representation and atoms in the *real* system are shown in stick representation.

Figure 7. Relaxed scan of the  $O_{\text{model}}\text{-H}$  distance from 0.90 Å to 2.00 Å for a proton transfer from a high-layer acetic acid to a low-layer 4-methylimidazole (*inside-out* reaction, Test 2 in Figure 6) in a PCM environment using ONIOM(B3LYP/6-31G(d):HF/3-21G). The plot also contains data from a separate scan at the B3LYP/6-31G(d) level.

Figure 8. Relaxed scan of the  $N_{\text{model}}\text{-H}$  distance from 1.00 Å for a proton transfer from a high-layer protonated 4-methylimidazole to a low-layer acetate (*inside-out* reaction, Test 3 in Figure 6) in a PCM environment using ONIOM(B3LYP/6-31G(d):BLYP/3-21G). SCF convergence problems limit the scan to 1.925 Å. The plot also contains data from a separate scan at the B3LYP/6-31G(d) level.

Figure 9. Relaxed scan of the  $O_{\text{model}}\text{-H}$  distance from 0.90 Å to 2.00 Å for a proton transfer from a low-layer acetic acid to a high-layer 4-methylimidazole (*outside-in* reaction, Test 4 in Figure 6) in a PCM environment using ONIOM(B3LYP/6-31G(d):HF/3-21G). The plot also contains data from a separate scan at the B3LYP/6-31G(d) level.

Figure 10. Textbook mechanism for the acylation step in serine proteases.

Figure 11. ONIOM QM:QM' model of the trypsin active site. Atoms in ball&stick representation are part of the *model* system, while Asp102 (in stick representation) is treated only with QM'. Atoms marked with x in the reactant structure are frozen during geometry optimization of all stationary points.

Table 1. Selected bond distances and reaction energies for the reaction  $[His_{model}H_{model} + His_{real}]^+ \rightarrow [His_{model} + H_{model}His_{real}]^+$  (Test 1 in Figure 2). ONIOM optimizations start from the two local minima of the B3LYP/6-31G(d) model. Calculations with HF/6-31G(d), HF/3-21G, and PM3 as low-level methods give purely dissociative potentials at large N-H distances. B3LYP:AM1 has a minimum in the product region, but at large N-H distances also this system becomes unstable. For B3LYP:BLYP there is no minimum in the product region.

Table 2. Classes of ONIOM QM:QM' cross-boundary reactions, how they are affected by the QM:QM' boundary, and how potential problems can be addressed. The reaction types are illustrated in Figure 6.

Table 3. Accuracy of ONIOM QM:QM' models for the acylation reaction in trypsin using the layer selection in Figure 11. In INT1 a proton is transferred from the high-layer His57 to the low-layer Asp102.