Development of Novel Computational Strategies to Match the Challenges of Supramolecular Chemistry, Biochemistry, and Materials Science

Tomasz A. Wesolowski*

Abstract: Recent formal developments and applications of the 'freeze-and-conquer' strategy proposed by Wesolowski and Warshel in 1993 to study large systems at quantum mechanical level are reviewed. This universal approach based on density functional theory allows one to link, *via* the orbital-free embedding potential, two parts of a larger system described at different levels of accuracy leading thus to significant savings in computational costs. As a result, applicability of conventional methods of quantum chemistry can be extended to even larger systems. It is shown that the 'freeze-and-thaw' approach applying the first-principles based approximation to the orbital-free embedding potential recently developed in our group provides a powerful and universal technique to study such embedded molecules (or molecular complexes), which are not linked with their microscopic environment by covalent bonds.

Keywords: Density functional theory · Embedding · Non-covalent interactions · Physisorption · Solvation

1. Introduction

Computer modeling at the quantum mechanical level has long been recognized as a useful technique to study molecules and chemical processes. In principle, it can provide a complementary insight into details of the system or process under investigation that are not available from experimental studies. The most universal are *ab initio* methods, which do not involve any empirical parameters. For small chemical molecules or molecular complexes, the results derived from high-level *ab initio* calculations are of 'chemical accuracy' quality [1]. Unfortunately, the high-level *ab initio* methods are not applicable for most of the systems relevant to the challenges of modern chemistry dealing usually with very large systems. In the Hartree-Fock theory, which is the simplest *ab initio* method and which does not describe correlation effects, the computational cost of solving relevant equations grows as fast as N^4 where N is proportional to the number of electrons in the system. In the Configuration Interaction method, which leads to virtually exact results, the computational cost grows exponentially with N [2]. Despite recent progress in the numerical implementation of *ab initio* methods reducing significantly the above scaling laws, the routine applications of the high-level *ab initio* methods are currently limited to systems of the size of the benzene dimer. Quantum Monte Carlo methods [3] provide another route to solve Schrödinger's equation with chemical accuracy, but the applicability is currently limited to even smaller systems. A computational chemist applying *ab initio* methods to model large systems faces, therefore, a dilemma of making trade-offs between the accuracy of the applied method and the available computer resources.

Computational schemes which, unlike *ab initio* methods, apply certain system-dependent empirical parameters provide, therefore, a useful alternative. *Semi-empirical methods*, in which the potential energy surface and expected values of observables are obtained by means of empirically parameterized approximation to the exact Hamiltonian (Hückel method, Tight-Binding method, ZINDO, AM1, *etc.*) are generally recognized as useful tools in interpretation and even quantitative prediction of the structure and reactivity of large molecules, supramolecular complexes, extended systems of interest in materials science, biomolecules, clusters, *etc.* [4]. Even simpler methods using parameterized Born-Oppenheimer potential energy surface, labeled frequently as *Molecular Mechanics*, provide a valuable tool to study static and dynamic properties of liquids, interfaces, solvent effects, and bio-molecules [5].

In the last decade, the formal developments in density functional theory (DFT) and the emergence of efficient computer implementations of the Kohn-Sham method (KS) [6] have enriched the set of tools available to a computational chemist [7]. The DFT label covers a wide range of methods including those based on first-principles and those which use empirical parameters. Numerous reports available in the literature show that the currently devel-

Correspondence: Dr. T.A. Wesolowski
Department of Physical Chemistry
University of Geneva
30, Quai Ernest-Ansermet
CH–1211 Genève 4
Tel.: +41 22 7026957
Fax: +41 22 702 6518
E-Mail: tomasz.wesolowski@chiphy.unige.ch
http://lcta.unige.ch/~tomek/

oped DFT methods applying Kohn-Sham formalism lead to results with an accuracy significantly better than that of the Hartree-Fock theory which is the lowest level *ab initio* method not describing electron correlation. They are usually of the quality of the simplest *ab initio* correlated methods. The practical benefits from this 'price-to-performance' ratio can be hardly overestimated. Computer modeling using Kohn-Sham formalism of large systems comprising thousands of atoms is currently possible [8]. Owing to efficient computer implementations of the Kohn-Sham formalism, systems such as proteins, membranes, nucleic acids, supramolecular complexes, large organometallic complexes, or nanostructures entered the domain of applicability of computer modeling at the quantum mechanical level (see *e.g.* [9] or the recent applications of Kohn-Sham DFT methods in our Geneva group [10]). Moreover, the Kohn-Sham equations provide also one of the key elements of the Car-Parrinello simulation technique, which is becoming one of the most commonly used method in the studies of large periodic systems [11].
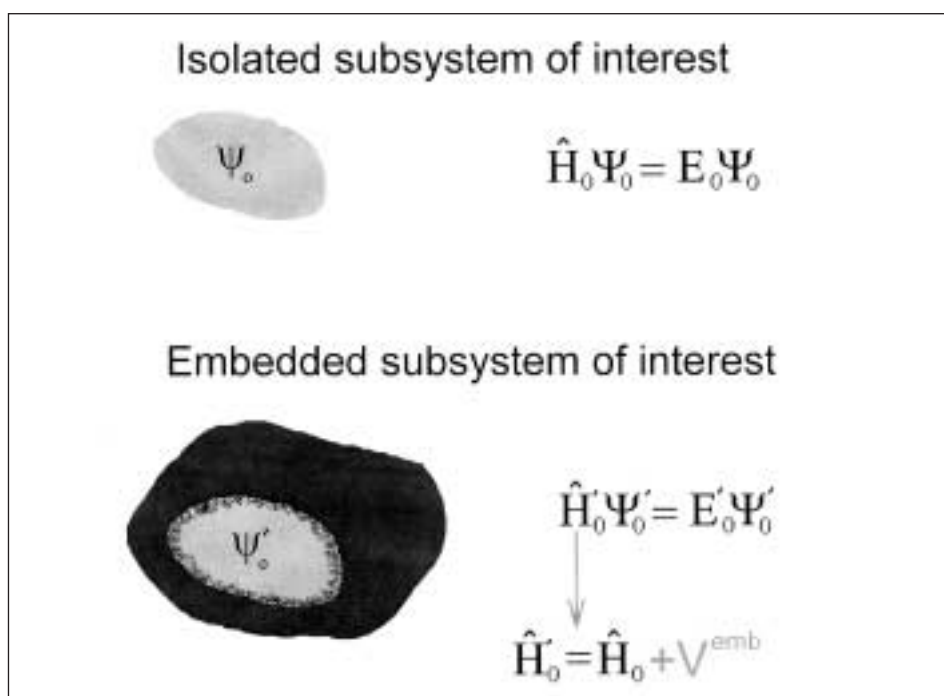

Fig. 1. Schematic division in real space of a larger system into subsystems in modeling by means of embedding techniques. In the orbital-free embedding formalism [12], the subsystem of interest (shaded) is represented as ρ_1 whereas its microscopic environment (dark) as ρ_2. The overlap between ρ_1 and ρ_2 in real space is indicated by the intermediate shade of the interface region.

2. Embedding Strategy for Studies of Large Systems

The strategy of using a high-level theoretical method of a controllable accuracy to the whole system of interest is not practical for sufficiently large systems. An alternative, which we will call here the *embedding strategy*, is based on reducing the quantum mechanical description to the selected part of a larger system and describing the effect of the remaining part by means of the *embedding potential* (V^{emb}) (see Fig. 1). The particular form of V^{emb} might originate from various theoretical considerations. It appears in models known under such labels as *hybrid methods*, QM/MD or QM/MM, *Reaction Field Method, Polarizable Continuum Method, COSMO, etc.* (see the reviews by Gao, Sauer *et al.*, or Wesolowski and Weber [13]). Studies of the solvent effect on the electronic structure and other properties of a) a solvated molecule or b) a solvated molecular complex undergoing chemical reaction in condensed phase are typical examples of applicability of the embedding strategy [14]. Embedding methods are also widely used in material science where computer models of extended surfaces or bulk material comprise an embedded cluster [15]. It is important to underline that the precondition for applicability of embedding methods is the *a priori* knowledge of the localization of the electronic states needed to define the subsystems.

3. Orbital-Free Embedding

Wesolowski and Warshel showed that the exact embedding potential takes the following universal system-independent form [15]:

$$V^{emb} = V^{emb(KSCED)}[\vec{r}, \rho_1, \rho_2] =$$
$$= -\sum_{\alpha}^{N_2} \frac{Z_\alpha^2}{|\vec{R}_\alpha - \vec{r}|} + \int \frac{\rho_2(\vec{r}')}{|\vec{r} - \vec{r}'|}dr' + \left.\frac{\delta E_{xc}[\rho]}{\delta\rho}\right|_{\rho=\rho_1+\rho_2} - \left.\frac{\delta E_{xc}[\rho]}{\delta\rho}\right|_{\rho=\rho_1} + \frac{\delta T_s^{nad}[\rho_1,\rho_2]}{\delta\rho_1} \quad (1)$$

where:

N_2 is the number of the nuclei in the subsystem 2; Z_α^2 is the nuclear charge of the nucleus α in the subsystem 2;

$$T_s^{nad}[\rho_1,\rho_2]=T_s[\rho_1+\rho_2]-T_s[\rho_1]-T_s[\rho_2] \quad (2)$$

and where the functional of the kinetic energy in the non-interaction electrons reference system ($T_s[\rho]$) and the exchange-correlation functional ($E_{xc}[\rho]$) are defined as in the Kohn-Sham formalism [6]. The effective potential of Eqn. (1) is a key element in the 'freeze-and-conquer' strategy introduced in [12]. We denote it with $V^{emb(KSCED)}$ to indicate that it results from the constrained minimization of the total energy bi-functional ($E[\rho_1,\rho_2]$) in which ρ_2 is kept frozen at a given initial value (Kohn-Sham Equations with Constrained Electron Density). The total energy bi-functional $E[\rho_1,\rho_2]$, is a particular case of the multi-

functional introduced by Cortona [16] in his formulation of DFT. Note that $V^{emb(KSCED)}$ includes *all* the information about the subsystem 2 *via* the electron density ρ_2 and the positions of nuclear charges. Orbital representation of neither ρ_2 nor $\rho_1+\rho_2$ is needed. Therefore, $V^{emb(KSCED)}$ is orbital-free. Obviously, as in the conventional Kohn-Sham model, ρ_1 is constructed using one-electron functions (Kohn-Sham-like) orbitals ($\rho_1 = \sum^{N_1}\phi_i^*\phi_i$).

For practical applications of Eqn. (1), good approximations for $T_s^{nad}[\rho_1,\rho_2]$ and $E_{xc}[\rho]$ are needed. The development of the approximations to the exchange-correlation potential as explicit functional of ρ has been an object of intensive studies in various theoretical groups (for review, see [7]). Less is known about approximations to $T_s^{nad}[\rho_1,\rho_2]$. In our earliest applications of $V^{emb(KSCED)}$ to study solvent effects [17],

we used the simplest approximations for $T_s^{nad}[\rho_1,\rho_2]$ derived from gradient expansion approximation for $T_s[\rho]$. Our subsequent dedicated studies of the accuracy of $T_s^{nad}[\rho_1,\rho_2]$ led us to a very accurate approximation applicable for such cases where the ρ_1–ρ_2 overlap is small. In this approximation, the analytic form of $T_s^{nad}[\rho_1,\rho_2]$ defined in Eqn. (2), depends explicitly on ρ_1 and ρ_2 as well as on their gradients (see [18] and reference therein). For stronger overlaps, however, the semilocal approximations to $T_s^{nad}[\rho_1,\rho_2]$ in which the functional depends explicitly on the local value of electron density and its gradient are too rude. In practice, such partitioning of a larger system into ρ_1 and ρ_2, which corresponds to cutting through covalent bonds, is beyond reach of such approximations.

4. Properties of Embedded Molecules Studied Using Orbital-Free Embedding Potential $V^{emb(KSCED)}$

Various computer implementations of the 'freeze-and-conquer' approach applying the orbital-free embedding potential of Eqn. (1) are possible. They can differ in the approximation used for $T_s^{nad}[\rho_1,\rho_2]$ and in the origin of the embedding electron density ρ_2. Several authors followed the 'freeze-and-conquer' approach of [12]. Stefanovitch and Truong used it in studies of molecules adsorbed on metal oxide surfaces [19]. Trail and Bird applied Eqn. (1) in calculations where partitioning of the total electron density into ρ_1 and ρ_2 was used to separate valence and core electrons in a solid [20]. Carter and coworkers combined orbital-free embedding potential of Eqn. (1) with the *ab initio* description of the isolated embedded subsystem [21]. In the following part, we will review our recent applications of the 'freeze-and-conquer' strategy using the gradient-dependent approximation to $T_s^{nad}[\rho_1,\rho_2]$ developed in our group in 1997 [18].

One of its first applications was the study of the effect of noble gas matrix on the hyperfine structure of the Mg^+ cation [22]. This rather academic problem provides a challenging test for the approximation to the kinetic energy dependent part in the orbital-free embedding potential of Eqn. (1). The effect of the noble gas atoms surrounding the Mg^+ cation (either Ne or Ar) on the spin distribution originates from Pauli repulsion arising from the overlap between the electron density of Ne (or Ar) and the highest occupied orbital of Mg^+. Our calculations, in which the matrix was represented by either Ar_8 or Ne_8 cage with frozen electron density (ρ_2) whereas the cation was represented as ρ_1, predicted very accurately the Fermi contact term of Mg^+ in Ne and Ar matrices ($A_{iso}^{calc(Ne)} = 222.7$ gauss *vs.* $A_{iso}^{exp(Ne)} = 222.4$ gauss; $A_{iso}^{calc(Ar)} = 209.8$ gauss *vs.* $A_{iso}^{exp(Ar)} = 211.6$ gauss). Other areas of applicability of the orbital-free embedding formalism include the studies of molecules physisorbed on metal oxide surfaces. A straightforward choice for subsystems is to represent the physisorbed molecule as ρ_1, whereas the surface as the frozen electron density ρ_2. The effect of physisorption on the properties of the adsorbed molecules results mainly from the electrostatic interactions and from Pauli repulsion [23]. Both effects are described by our orbital-free embedding potential: (a) electrostatic interactions are represented by the exact functionals (two first terms in Eqn. (1)) whereas (b) the Pauli repulsion is represented through the approximate non-additive kinetic energy functional (last term in Eqn. (1)). Our theoretical studies of the effect of physisorption on the infrared spectra of CO physisorbed on MgO or ZnO surfaces led to very accurate predictions (agreement within few cm^{-1}) of the surface-induced shifts of the CO stretching band [24]. The good numerical results of the above studies provide a justification of the use of the gradient-dependent approximations of [18] to the relevant approximate component of the total energy bi-functional for more complicated systems such as those reviewed in the next section.

4.1. *Molecules Trapped in the Zeolite*

Small molecules such as CO are frequently used as probes to obtain insight into the local structure of metal binding sites in zeolites [25]. The sign of the shifts and the shape of IR bands provide the details concerning the topology of the Si/Al replacement sites [26]. In our recent work on the influence of the interaction with the ZSM-5 zeolite framework on the stretching frequency of CO adsorbed on alkali metal cations, we demonstrated that the theoretical calculations could provide quantitative predictions [27]. In these studies, the complex Me^+..CO (where Me = Li, Ne, or K) was considered as ρ_1 whereas the zeolite framework represented as a cluster was considered as ρ_2 (see Fig. 2). Calculations using a series of clusters of variable size to represent the zeolite led us to the identification of microscopic interactions contributing to the shifts of the stretching frequency of the CO molecule trapped in a zeolite. It was found that the interactions between the bound CO and the wall of the channel opposite to the Si/Al replacement site as well as long-range electrostatic interactions affect the frequency of CO stretching by as much as 30 cm^{-1}.
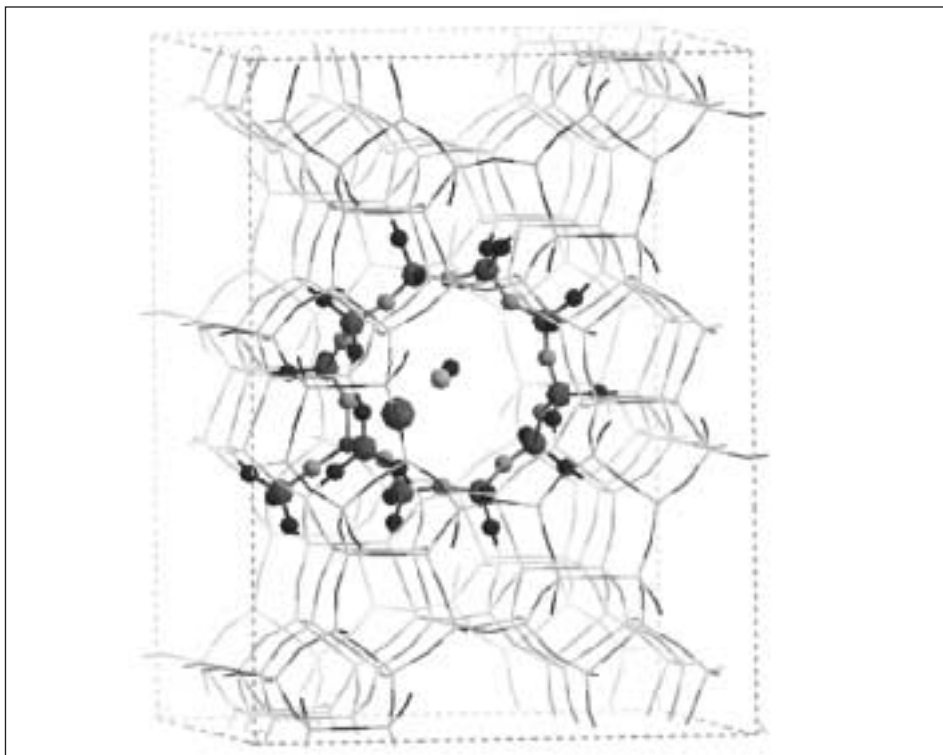

Fig. 2. The CO molecule adsorbed at the cationic site in the Me-ZSM5 zeolite. The Me+..CO complex is considered as ρ_1 whereas the other highlighted atoms represent ρ_2. The calculated shifts of the CO stretching frequencies amount to 46 cm^{-1}, 38 cm^{-1}, and 17 cm^{-1} for Me = Li, Na, and K, respectively. The corresponding shifts measured in the zeolite amount to 45 cm^{-1}, 29–35 cm^{-1}, 16–23 cm^{-1}. More details can be found in [27].

4.2. *Physisorption of Molecular Hydrogen on Graphitic Surfaces*

Adsorption of molecular hydrogen on carbonaceous surfaces is a key element in astrophysical theories aimed at understanding the abundance of hydrogen in interstellar space [28] and in development of hydrogen storage devices based on carbon [29]. We studied the adsorption of molecular hydrogen on various polycyclic aromatic hydrocarbons of increasing size used as models of the infinite graphite surface (see Fig. 3). It was found that the hydrogen molecule is adsorbed above the ring center and can rotate rather freely. The calculated binding energy of 1.27 kcal/mol is in excellent agreement with available experimental measurements (1.1–1.2 kcal/mol) [28].
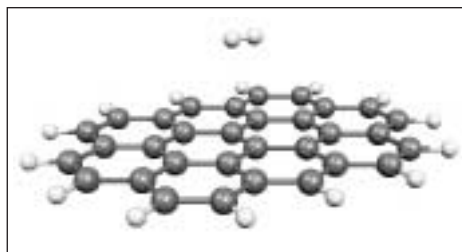

Fig. 3. The cluster model of the molecular hydrogen adsorbed on graphite surface. The H_2 molecule is considered as one subsystem and the cluster representing the surface as the other one in the total energy bi-functional $E[\rho_1,\rho_2]$. More details can be found in [31].

4.3. *Other Weak Intermolecular Complexes*

For weakly interacting complexes close to their equilibrium, the minimization of the total energy bi-functional depending on two variables ρ_1 and ρ_2 leads to better interaction energies than the corresponding Kohn-Sham calculations [30][31]. Fig. 4 compares the experimental dissociation energies in complexes involving carbazole with the binding energies derived from KSCED calculations [30]. Results obtained using two levels of approximation in KSCED are shown: (a) the gradient-free approximation for $T_s^{nad}[\rho_1,\rho_2]$ which was proposed as the simplest parameter-free approximation in our original work in 1993 [12] and the conventional LDA approximation for $E_{xc}[\rho]$ and (b) the gradient-dependent approximation for $T_s^{nad}[\rho_1,\rho_2]$ developed in 1997 [18] and the corresponding generalized gradient

approximation for $E_{xc}[\rho]$. Fig. 4 illustrates the recent progress in the development of first-principles based approximations to the relevant functionals. Currently, we are applying the KSCED method in studies of the structure of larger complexes formed by larger organic molecules used in the construction of nanocolumns [32].

4.4. *Proton Transfer in the Enzyme's Active Center*

Contrary to the examples discussed so far, partitioning the total electron density of a protein molecule into its ρ_1 and ρ_2 components cannot avoid cutting through covalent bonds. Therefore, a straightforward application of the approximation to $T_s^{nad}[\rho_1,\rho_2]$ of [18] for the arbitrarily chosen ρ_1 and ρ_2 is not appropriate. One of the possible ways to avoid high ρ_1–ρ_2 overlaps arising from cutting the covalent bonds is to use artificial link atoms. Such computer implementation of KSCED is in progress in our group [33]. Another option is to use an artificial cluster comprising only the amino acids in the enzyme's active center as a model of the whole enzyme. Fig. 5 shows the model used in our studies of the proton transfer in carbonic anhydrase [34]. Compared to full Kohn-Sham calculations, limiting the quantum description to the $Zn^{2+}...H_2O...H...H_2O$ complex and representing the three histidine molecules as frozen electron density ρ_2 has a negligible effect on the potential energy curve corresponding to the proton transfer [34].
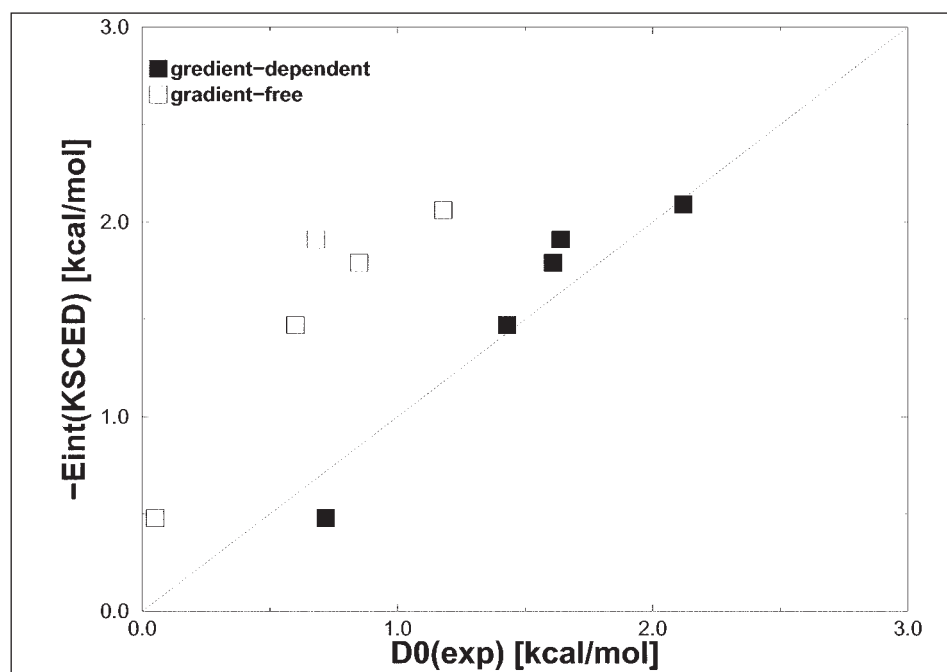

Fig. 4. Comparison between the experimental dissociation energies and the calculated interaction energies in the complexes of the X...carbazole type (X = Ne, Ar, N_2, CO, CH_4). The KSCED interaction energies were calculated using two levels of approximation to the relevant functionals in the total energy bi-functional $E[\rho_1,\rho_2]$: (a) gradient-free functionals used in our original paper [16] (empty squares), (b) recently developed [18] gradient-dependent functionals (full squares). More details can be found in [30].
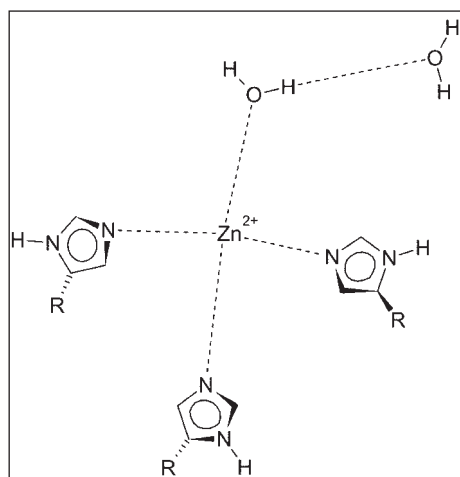

Fig. 5. The model of the active center in the carbonic anhydrase. The three histidine molecules are considered as one subsystem and the $Zn^{2+}...H_2O...H...H_2O$ complex, which is involved in the proton transfer reaction, as the other one in the total energy bi-functional $E[\rho_1,\rho_2]$. R indicates the rest of the polypeptide chain of the enzyme. In the calculations reported in [34], R represents hydrogen.

5. Conclusions and Future Outlook

We have reviewed the recent formal developments and applications of the 'freeze-and-conquer' approach to study large systems. It was shown that the first-principles based approximation to the kinetic energy part of the embedding potential developed in our group [18] is sufficiently accurate for such systems where the overlap between the electron density of the embedded subsystem (ρ_1) and the electron density of the surrounding molecules (ρ_2) is small. Therefore, our method can be expected to be an adequate and universal tool in such studies where the primary interest concerns *selected molecules in larger complexes bound by non-covalent interactions*. Currently, we are concentrating our efforts on a) theoretical progress in developing better approximations to $T_s^{nad}[\rho_1,\rho_2]$ to extend the range of applicability of the orbital-free embedding [35], b) improving the computer implementation of the orbital-free embedding (better CPU efficiency and the evaluation of other observables such as: NMR spectra, Raman spectra, polarizabilities and hyperpolarizabilities, *etc.*).

[1] R.J. Bartlett, in 'Modern Electronic Structure Theory', Ed. D.R. Yarkony, World Scientific, New York, **1995**, Part 2, Chapter 6.

[2] C.J. Cramer, 'Essentials of Computational Chemistry. Theories and Models', John Wiley & Sons, New York, **2002**.

[3] J.B. Anderson, in 'Quantum Mechanical Electronic Structure Calculations with Chemical Accuracy', Ed. S.R. Langhoff, Kluwer Academic, The Netherlands, **1995**, p. 1.

[4] T. Clark, *J. Mol. Structure (Theochem)* **2000**, *530*, 1.

[5] J.P. Bowen, N.L. Allinger, in 'Reviews in Computational Chemistry', vol. 2, Eds. K.B. Lipkowitz, D.B. Boyd, VCH, New York, **1991**, p. 81.

[6] W. Kohn, L.J. Sham, *Phys. Rev.* **1965**, *140*, A1133.

[7] W. Kohn, A.D. Becke, R.G. Parr, *J. Phys. Chem.* **1996**, *100*, 12974.

[8] G.E. Scuseria, *J. Phys. Chem.* **1999**, *103*, 4782.

[9] P. Ordejon, D. Sanchez-Portal, A. Garzia, E, Artacho, J. Junquera, J.M. Soler, *RIKEN Review* **2000**, *No. 29*(June), 42.

[10] J. Weber, P. Boulet, F. Gilardoni, A. Goursot, M.J. Mayor-Lopez, H. Sidorenkova, T.A. Wesolowski, *Abstracts of Papers of The American Chemical Society,* Part 1, **2002**, *221*, 32.

[11] R. Car, M. Parrinello, *Phys. Rev. Lett.* **1985**, *55*, 2471.

[12] T.A. Wesolowski, A. Warshel, *J. Phys. Chem.* **1993**, *97*, 8050.

[13] a) J. Gao, in 'Reviews in Computational Chemistry', vol. 7, Eds. K.B. Lipkowitz, B. Boyd, VCH, New York, **1995**, p. 119; b) J. Sauer, P. Ugliengo, E. Garrone, V.R. Sounders, *Chem. Rev.* **1994**, *94*, 2095; c) T.A. Wesolowski, J. Weber, in 'Molecular Orbital Calculations Applied to Biochemical Systems', Ed. A.-M. Sapse, Oxford University Press, **1998**, p. 85.

[14] L.R. Pratt, A. Pohorille, *Chem. Rev.* **2002**, *102*, 2671.

[15] J. Sauer, M. Sierka, *J. Comput. Chem.* **2000**, *21*, 1470.

[16] P. Cortona, *Phys. Rev. B* **1991**, *44*, 8454.

[17] a) T.A. Wesolowski, A. Warshel, *J. Phys. Chem.* **1994**, *98*, 5183; b) T.A. Wesolowski, R.P. Muller, A. Warshel, *J. Phys. Chem.* **1995**, *100*, 15444.

[18] T.A. Wesolowski, *J. Chem. Phys.* **1997**, *106*, 8516.

[19] E.V. Stefanovich, T.N. Truong, *J. Chem. Phys.* **1996**, *104*, 2946.

[20] J. Trail, D. Bird, *Phys. Rev. B* **2001**, *62*, 16402.

[21] a) N. Govind, Y.A. Wang, E.A. Carter, *J. Chem. Phys.* **1999**, *110*, 7677; b) T.A. Wesolowski, *Phys. Rev. Lett.* **2002**, *88*, 209701.

[22] T.A. Wesolowski, *Chem. Phys. Lett.* **1999**, *311*, 87.

[23] G. Pacchioni, G. Cogliandro, P.S. Bagus, *Intl. J. Quant. Chem.* **1992**, *42*, 1115.

[24] N. Vulliermet, T.A. Wesolowski, J. Weber, *Collect. Czech. Chem. Commun.* **1998**, *63*, 1447.

[25] A. Zecchina, S. Coluccia, C. Morterra, *Applied Spectroscopy Reviews,* **1985**, *21*, 259.

[26] H. Knözinger, S. Huber, *J. Chem. Soc. Faraday Trans.* **1998**, *94*, 2047.

[27] T.A. Wesolowski, A. Goursot, J. Weber, *J. Chem. Phys.* **2001**, *115*, 4791.

[28] F. Tran, J. Weber, T.A. Wesolowski, F. Cheikh, Y. Ellinger, F. Pauzat, *J. Phys. Chem. B* **2002**, *106*, 8689.

[29] a) A. C. Dillon, K.M. Jones, T.A. Bekkedahl, C.H. Kiang, D.S. Bethune, M.J. Heben, *Nature* **1997**, *386*, 377; b) P. Chen, X. Wu, J. Lin, K.L. Tan, *Science* **1999**, *285*, 91.

[30] T.A. Wesolowski, P.-Y. Morgantini, J. Weber, *J. Chem. Phys.* **2002**, *116*, 6411.

[31] a) T.A. Wesolowski, Y. Ellinger, J. Weber, *J. Chem. Phys.* **1998**, *108*, 6078; b) F. Tran, J. Weber, T.A. Wesolowski, *Helv. Chim. Acta* **2001**, *84*, 1489.

[32] T. Jenny, *VISION Le magasiné suisse de la science et de l'innovation* **2001**, *octobre*, 32.

[33] M. Casida, T.A. Wesolowski, in preparation.

[34] G. Hong, M. Strajbl, T. A. Wesolowski and A. Warshel, *J. Comput. Chem.* **2000**, *21*, 1554.

[35] a) F. Tran, T.A. Wesolowski, *Int. J. Quant. Chem.* **2002**, *89*, 441; b) F. Tran, T.A. Wesolowski, *Chem. Phys. Lett.* **2002**, *360*, 209.