

Computation of protein pK 's values by an integrated density functional theory/Polarizable Continuum Model approach

Vincenzo Barone, Roberto Improta, Nadia Rega

Dipartimento di Chimica, Università Federico II, Complesso Monte S. Angelo, via Cintia, 80126 Naples, Italy
Permanent address: R. Improta, Istituto Biostrutture e Bioimmagini-CNR, V. Mezzocannone 680134 Naples, Italy

Received: 20 March 2003 / Accepted: 30 April 2003 / Published online: 4 February 2004
© Springer-Verlag 2004

Abstract. This paper describes the extension of our computational strategy for pK predictions of small molecules to large solutes. The basic computational tool results from the coupling of quantum mechanical methods rooted in the density functional theory with the most recent version of the Polarizable Continuum Model. However, a third level is introduced, which includes solute regions far from the reactive center, which are described at a simplified level. This partition, together with the recent implementation of fast cavity generation, powerful iterative solvers, and fast multipole technology, allows us to tackle solutes of the dimension of a small protein. The problems and perspectives of this methodology are analyzed with special reference to the behavior of different Polarizable Continuum Model versions on the challenging playground represented by the pK 's of the different histidine residues occurring in the human prion protein.

Keywords: pK calculation – Polarizable Continuum Model – Human prion protein

1 Introduction

Several processes of biological relevance involve a proton transfer between ionizable groups belonging to protein residues, enzyme cofactors, or solvent molecules [1]. Reliable predictions of the pK of the different groups possibly involved in biological reactions are thus of paramount importance for a deeper understanding of their mechanism and a rational modification of key steps (as could be necessary, for instance, in the case of pathological behavior). Unfortunately, the experimental

determination of pK 's within a protein is not an easy task, and it is seldom possible to measure with the necessary accuracy the pK_a of a given residue [2]. A suitable computational approach would thus be very useful [3].

However, the accurate computation of acid dissociation constants is far from being a trivial task even for small molecules. As a matter of fact, reproduction of the experimental pK within 1 pK unit requires calculations able to provide for the condensed phase protonation energies with an error of 1 kcal/mol or less. On these grounds, a number of computational procedures have recently been devised with variable success. In general terms, trends within classes of related systems are correctly reproduced by most of the methods proposed [4, 5], whereas reliable absolute values are just starting to be produced [6]. Calculation of the pK of ionizable groups in biomacromolecules involves additional difficulties, since the dimensions of the systems do not allow straightforward quantum mechanical (QM) computations. In these cases the most effective solution is probably represented by combined approaches, in which a QM description of the ionizable groups is coupled to a molecular mechanics (MM) description of more distant regions of the solute and to a continuum model for the proper treatment of bulk solvent effects. Some initial attempts along these lines have recently been performed [3, 7], but they do not benefit from the most recent developments of the Polarizable Continuum Model (PCM) leading to linear scaling procedures for all the most computer intensive steps (building of the cavity including geometric derivatives and computation of polarization charges) [8]. We thus think it interesting to test different PCM versions including all the most recent improvements for the demanding problem represented by the computation of reliable pK 's for specific residues in proteins. It must be recalled in this connection that a proper computation of this quantity for systems containing several acid/base groups requires the consideration of different overall ionization states. Effective models have recently been proposed in the context of MM models [9] and their extension to QM/MM

Contribution to the Jacopo Tomasi Honorary Issue

Correspondence to: N. Rega
e-mail: nadia@lsdm.dichi.unina.it

techniques is straightforward provided that the computation time for a single step can be controlled. As a consequence, in this first feasibility study we consider a single ionization state.

We chose to study the pK_a of a histidine residue (His140) of the human prion protein (hereafter HuPr^P) as a “test case”. The prion protein has recently received increasing attention owing to its involvement in several neurodegenerative diseases, including scrapie in sheep, bovine spongiform encephalopathy in cattle, and Kuru, Creutzfeldt–Jakob disease, Gerstmann–Sträussler syndrome and fatal familial insomnia in humans [10, 11]. As a matter of fact, the transition from the normal cellular prion protein (PrP^C) to a misfolded isoform (PrP^{Sc}) has been recognized as the most significant pathogenic event [12, 13].

Recent experimental [14, 15, 16] and computational [17, 18, 19] studies suggest that the conformational behavior of HuPr^P is remarkably influenced by the pH of the embedding medium. In particular, it has been shown that the conformational rearrangements induced by low-pH conditions bear significant similarity with the transition leading to the misfolded isoform.

However, the role of the different ionizable groups in the pH-driven transition has not been clearly assessed. Protonation of the four histidine residues present in the C-terminal region of HuPr^P, the structured core (125–228 aa) of the protein, should play a significant role in this transition, since that conformational rearrangement starts even under mildly acidic conditions [14]. A reliable computation of the pK of each of the four histidine residues can thus be relevant for a better understanding of the chemical effects leading to the transition from the normal to the pathogenic form of the prion.

2 Methods

2.1 pK_a calculation

For an acid species AH the pK_a , defined as the negative logarithm of the dissociation constant of the reaction



is given by the well known thermodynamic relation

$$pK_a = \Delta G_{aq,AH} / 2.303RT. \quad (2)$$

The Gibbs energy variation of the deprotonation reaction in aqueous solution is calculated in our procedure by adding a solvation contribution to the gas-phase value as follows:

$$\Delta G_{aq,AH} = \Delta G_{gas,AH} + \Delta \Delta G_{solv,AH}, \quad (3)$$

$$\Delta G_{gas,AH} = G_{gas,A^-} + G_{gas,H^+} - G_{gas,AH}, \quad (4)$$

$$\Delta \Delta G_{solv,AH} = \Delta G_{solv,A^-} + \Delta G_{solv,H^+} - \Delta G_{solv,AH}. \quad (5)$$

The proton free energy in the gas phase at 298K and 1 atm is: [20] $G_{gas,H^+} = 2.5RT - T\Delta S^\circ = 1.48 - 7.76 = -6.28$ kcal/mol.

For the corresponding value in aqueous solution, G_{solv,H^+} , we use the most recent experimental value (–263.98 kcal/mol) [21], obtained using the cluster-pair-based approximation without extra thermodynamic assumptions. This value is more negative than the lower end of the range of values based on measurements of the general hydrogen electrode potentials [22], but is consistent with the value obtained from the experimental thermodynamic cycle of acetic acid [5].

2.2 The PCM picture

Solvation energies were computed by our most recent implementation of the PCM [23]. The United-Atom Topological Model was used for building the molecular cavity both in its standard version (United-Atom Hartree–Fock, UAHF) [24] and in a new simplified implementation (hereafter UA0). The PCM picture of isotropic solutions was used in the improved dielectric version, taking into proper account escaped-charge effects [25, 26] (hereafter referred to simply as PCM), in the conductor-like model (CPCM) [27], and in the original dielectric implementation (DPCM) [28, 29].

In the most recent PCM picture, the solvent reaction field is expressed in terms of a polarization charge density, $\sigma(\mathbf{s})$, spread on the cavity surface. The polarization charge density depends on the electrostatic potential, $V(\mathbf{s})$, generated by the solute on the cavity according to

$$\left(\frac{\epsilon + 1}{\epsilon - 1} \hat{S} - \frac{1}{2\pi} \hat{S} \hat{D}^* \right) \sigma(\mathbf{s}) = \left(-1 + \frac{1}{2\pi} \hat{D} \right) V(\mathbf{s}), \quad (6)$$

where ϵ is the solvent dielectric constant and $V(\mathbf{s})$ is the (electronic + nuclear) solute potential at point \mathbf{s} on the cavity surface. The \hat{S} and \hat{D}^* operators are related, respectively, to the electrostatic potential, $V^\sigma(\mathbf{s})$, and to the normal component of the electric field, $E_\perp^\sigma(\mathbf{s})$, generated by the surface charge density, $\sigma(\mathbf{s})$. It is noteworthy that in this PCM formulation $\sigma(\mathbf{s})$ is designed to take into account implicitly the effects of the fraction of solute electronic density lying outside the cavity (“outlying charge”).

Taking the limit for $\epsilon \rightarrow \infty$ one obtains the CPCM [30]:

$$\hat{S} \sigma(\mathbf{s}) = -V(\mathbf{s}). \quad (7)$$

Note that in the literature one finds several techniques to correct the outlying electronic density also for this kind of model: in the present case, however, the CPCM is considered as an approximation of the PCM version, which already takes escaped-charge effects into account, so we shall not correct CPCM charges further.

The original DPCM formulation can be formally recovered by taking the limit for vanishing outlying charge [26, 30]:

$$\left(\frac{\epsilon + 1}{\epsilon - 1} - \frac{1}{2\pi} \hat{D}^* \right) \sigma(\mathbf{s}) = \left(-\frac{1}{2\pi} \hat{D} \right) E_\perp(\mathbf{s}), \quad (8)$$

where $E_\perp(\mathbf{s})$ is the normal component of the electric field generated by the solute on the cavity.

In practice, $\sigma(\mathbf{s})$ is expressed in terms of a set of point charges, q_i , placed at the center of each of the N_{TS} finite elements (called tesserae) in which the cavity surface is partitioned. As a consequence, operators are replaced by the corresponding $N_{TS} \cdot N_{TS}$ matrices: the number of tesserae, N_{TS} , determines the dimension of the PCM electrostatic problem and can reach values of tens of thousands for large molecules like proteins.

In this connection the efficiency of different PCM versions is quite different since the CPCM involves a single symmetric matrix \mathbf{S} , DPCM a single nonsymmetric matrix \mathbf{D} , and PCM both \mathbf{S} and \mathbf{D} . As a consequence, use of the CPCM or DPCM versions could be advantageous for very large systems.

In particular, the DPCM formulation is exact when the whole solute charge is contained in the cavity. This happens in classical calculations, where solute atoms are assigned point partial charges or dipoles according to the force field parameterization. The numerical equivalence between both methods has recently been verified in a MM context [8].

On the other hand, when using the DPCM in QM calculations, one has to consider some correction (based, in general, on Gauss’ law), to adjust *a posteriori* the surface charges, in the approximation that the whole electron cloud was confined in the cavity [28]. In this approximation the total solvation density on the cavity surface should obey Gauss’ law:

$$\int_{\Gamma} \sigma(\mathbf{s}) d\mathbf{s} = \sum_i^{N_{TS}} q_i = -\frac{\epsilon - 1}{\epsilon} Q_{sol}, \quad (9)$$

where Q_{sol} is the solute net charge. Among the several techniques proposed in the literature, we consider here two of the most reliable

[28]. In the first procedure the difference is evaluated between the total solvation charge and its theoretical value obtained from the Gauss theorem:

$$Q_{\text{exc}} = -\frac{\epsilon - 1}{\epsilon} Q_{\text{sol}} - \sum_i^{N_{\text{TS}}} q_i . \quad (10)$$

The charge amount, Q_{exc} , is then distributed to each tessera, weighted by the surface area, a_i , and the solute electronic density, $\rho_{\text{el}}(\mathbf{s}_i)$:

$$q_i^{\text{corr}} = q_i + \delta q_i , \quad (11)$$

where

$$\delta q_i = \frac{a_i \rho_{\text{el}}(\mathbf{s}_i)}{\sum_i^{N_{\text{TS}}} a_i \rho_{\text{el}}(\mathbf{s}_i)} Q_{\text{exc}} . \quad (12)$$

This is called density-weighted charge compensation and we indicate it as CompIII.

The second approach considers the solute electronic charge lying outside the cavity as a further source of apparent charge. The contribution to the reaction field coming from the outlying electronic density is represented by an effective charge density, σ_{eff} , spread on the surface and depending on both Q_{exc} and $\rho_{\text{el}}(\mathbf{s})$ [31].

$$\sigma_{\text{eff}}(\mathbf{s}) = \frac{Q_{\text{exc}} \rho_{\text{el}}(\mathbf{s})}{\int_{\Gamma} \rho_{\text{el}}(\mathbf{s}) d\mathbf{s}} . \quad (13)$$

In practice σ_{eff} is expressed in terms of N_{TS} effective charges \mathbf{q}^{eff} distributed on the tesserae

$$q_i^{\text{corr}} = q_i + q_i^{\text{eff}} . \quad (14)$$

The normal component of the electric field generated by the \mathbf{q}^{eff} also adds to the solute contribution $E_{\perp}(\mathbf{s})$ in the right hand side of Eq. (8). We call this procedure the effective volume charge compensation and indicate it as CompIV.

Once the solvation charges (\mathbf{q} or \mathbf{q}^{corr}) have been determined, they can be used to compute energies and properties in solution: in MM, the calculation of solvent effects on energies is straightforward, while in QM calculations the molecular Hamiltonian must be corrected by a suitable PCM operator. The interaction energy between the solute and the solvation charges can be written

$$E_{\text{int}} = \mathbf{V}^{\dagger} \mathbf{q} = \sum_i^{N_{\text{TS}}} V_i q_i \quad (15)$$

or, if a compensation scheme has been adopted,

$$E_{\text{int}} = \mathbf{V}^{\dagger} \mathbf{q}^{\text{corr}} = \sum_i^{N_{\text{TS}}} V_i q_i^{\text{corr}} , \quad (16)$$

where V_i is the solute potential calculated at the representative point of tessera i . The charges act as perturbations on the solute electron density ρ : since the charges depend in turn on ρ through the potential or the field, the solute density and the charges must be adjusted until self-consistency. It can be shown [29] that for any self-consistent field (SCF) procedure including a perturbation linearly depending on the electron density (either HF or Kohn-Sham), the quantity that is variationally minimized corresponds to a free energy (i.e. E_{int} minus the work spent to polarize the dielectric and to create the charges). If $E^0 = E[\rho^0] + V_{\text{NN}}$ is the solute energy in vacuo, the free energy minimized in solution is

$$\mathcal{G} = E[\rho] + V_{\text{NN}} + \frac{1}{2} E_{\text{int}} , \quad (17)$$

where V_{NN} is the solute nuclear repulsion energy, ρ^0 is the solute electronic density for the isolated molecule, and ρ is the density perturbed by the solvent.

Note that all the terms of the solvation energy can be dissected into contributions arising from the different spheres forming the cavity. Since each sphere corresponds to a well-defined atom (or chemical group), this procedure allows a detailed analysis of the origin of differential solvation effects. However, the electrostatic contribution of each sphere originates from the electronic density of the whole solute, so the analysis should be considered only qualitative.

For small solutes described at the QM level, the computational cost of cavity generation and evaluation of polarization charges is negligible with respect to other steps (e.g. evaluation of the two-electron integrals also needed for the corresponding computation for a vacuum). This could be no longer true for MM methods or, even, for QM/MM approaches involving evaluation of the reaction field on a cavity containing thousands of tesserae even for relatively small QM subsystems. In order to treat these situations in an effective way, the most computer intensive parts of the PCM were completely rewritten, leading to an essentially linear scaling code both for energy and analytical gradient evaluations. Full details of this implementation are given elsewhere [8]. Here we concentrate on testing different issues involved in the resulting model for the specific problem of pK evaluation.

2.3 Computational details

For the QM method, our approach is based on the density functional theory (DFT), which, because of its favorable scaling with the number of active electrons, opens the possibility to treat very large molecules.

We chose, in particular, the hybrid PBE0 functional, which consistently provides satisfactory results for several structural and thermodynamic properties [32].

All the calculations were performed using a development version of the Gaussian package [33], by using internal 6-31 + G(d,p) and 6-311 + G(2d,2p) basis sets.

3 Results

The solvation energies calculated for four molecules tailored in order to model the side chain of the acid/base peptide residues, namely histidine, lysine, arginine and glutamic/aspartic acid (see Fig. 1), are compared in Table 1. We checked the performance of the different models for two different sets of radii (UAHF and UA0). The first set of radii was introduced in order to obtain very accurate solvation energies for both neutral and charged species formed by H, C, N, O, F, P, Cl, Br, and I

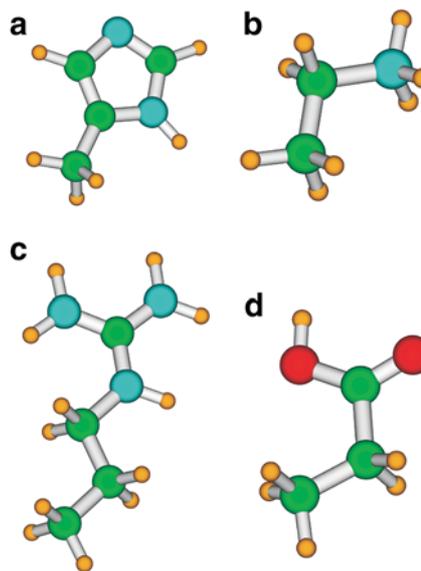


Fig. 1. Side chain models used in the pK_a computations: **a** histidine (δ isomer, neutral); **b** protonated lysine; **c** protonated arginine; **d** neutral aspartic acid

Table 1. Electrostatic contributions to the solvation free energies of acid/base residue side chains (see Fig. 1) obtained by different solvation models (polarizable continuum model, *PCM*, conductor-like PCM, *CPCM*, dielectric implementation of the PCM, *DPCM*) and escaped charge at geometries optimized in the gas phase. All the computations were performed at the PBE0/6-31+G(d,p) level

	PCM	DPCM ^a		CPCM ^a	Escaped charge	
		CompIII	CompIV			
UA0						
His	11.69	0.35	0.35	0.17	-0.11	0.22
His ⁺	63.95	-8.53	-8.38	0.88	-0.10	0.15
Lys ⁺	67.44	-4.61	-4.55	0.50	-0.07	0.08
Arg ⁺	63.74	-11.2	-11.7	1.4	0.15	0.23
Glu	9.65	-0.16	-0.16	0.07	-0.07	0.17
Glu ⁻	68.08	16.2	17.7	-2.5	-0.09	0.29
UAHF						
His	10.67	0.29	0.29	0.21	-0.09	0.21
His ⁺	60.45	-6.66	-6.55	0.75	-0.08	0.13
Glu	8.42	-0.61	-0.61	0.19	-0.06	0.18
Glu ⁻	71.29	23.54	23.48	-3.04	-0.12	0.38

^aDifferences with respect to PCM results

atoms [24]. However, the generation of radii is quite involved and depends on several contributions (effective charge, hybridization, first neighbors). As a consequence extension of the model to other systems is quite difficult taking also into account that reliable experimental solvation energies are not always available. This situation prompted us to develop a simplified set of radii (UA0) based on slight modifications of the unified force field radii [34], available for the whole periodic table. While some degradation of numerical accuracy is unavoidable, it would be more than compensated by the generality and ease of use if at least general trends are correctly reproduced at the UA0 level. Thus, one of the objectives of the present paper is to validate the UA0 model in the field of bioactive molecules.

As a first point our calculations confirm that, for aqueous solutions, the CPCM results are extremely close to their PCM counterparts. Furthermore, the CPCM solvation energies seem to be slightly overestimated (by approximately 0.1 kcal/mol) for both the neutral and the charged form of the amino acids (see the results concerning histidine and glutamic acid side chains). As a consequence the pK_a 's provided by CPCM and PCM methods are practically equivalent (since they depend on the difference between the free energy of the acid/base pair). On the other hand, the performances of the DPCM calculations dramatically depend on the compensation scheme adopted. As a matter of fact, when no compensation scheme or CompIII is used the DPCM calculations do not reach an acceptable agreement with the PCM results for charged species. The solvation energy of the positively charged residues is indeed remarkably overestimated, whereas that of the glutamate ion is largely underestimated. It is not surprising that the larger is the value of the charge escaped from the cavity the larger is the discrepancy between DPCM and PCM results. However, inspection of Table 1 shows that the accuracy of the DPCM method can be improved significantly by the use of CompIV. Interestingly, the

Table 2. pK and energy difference (in kilocalories per mole) between neutral and protonated forms of 4-methylimidazole, for both δ and ϵ isomers. Geometry optimizations were performed for a vacuum at the PBE0/6-31+G(d,p) level

	δ		ϵ	
	ΔE	pK	ΔE	pK^b
Gas phase	238.91		238.46(238.38 ^a)	
ZPE corr.	-8.76		-8.86	
UA0				
PCM	290.8		290.7	
CPCM	290.8	10.0	290.7	9.86
DPCM	290.0	9.4	290.0	9.3
UAHF				
PCM	288.7		288.7	
CPCM	288.7	8.5	288.6	8.3
DPCM	288.1	8.1	288.0	7.9

^a Single-point with the 6-311+G(2d,2p) basis set

^b The experimental value is 7.35

behavior is opposite to that found for the other compensation schemes: the solvation energy of the positively charged residues is overestimated by approximately 1 kcal/mol, i.e. only around 1.5%. The results are worse for glutamic acid, whose solvation energy is 2.5 kcal/mol larger than the PCM reference value. The results obtained by using UAHF radii are similar to the UA0 ones, showing that the previous conclusion does not depend significantly on the the cavity radii, at least for reasonable choices. Of course, very small cavity radii (increasing the value of the escaped charge) would increase the discrepancy between DPCM and CPCM/PCM results, while just the opposite would occur for very large cavities.

On the whole, the results obtained on model compounds show that, when using a suitable compensation scheme, DPCM calculations can reach very good accuracy for neutral compounds. For the ions, the accuracy is lower but still acceptable, provided that the escaped charge is not extremely high.

We then calculated the pK of 4-methylimidazole resorting to the computational procedure we recently developed [6] (Table 2). The only difference with respect to the original procedure is that we used the gas-phase equilibrium geometry also for the calculations for aqueous solution, without resorting to PCM geometry optimizations. This is not very significant in the present context, since we are not pursuing quantitative agreement with experiment, but, rather, validation of the cheap CPCM version for pK calculation. In any case, the results obtained by the UAHF radii show good agreement with experiment, with a discrepancy of less than 1 pK unit with respect to the experimental pK_a , i.e. an error approximately 1 kcal/mol in the differential free energy in solution. As expected, the results obtained by using UA0 radii are less accurate, but, since general trends are correctly reproduced, we prefer to discuss in detail in the following the results obtained by this model, which is easier and of more general use especially for QM/MM computations, which are, anyway,

approximate and can be expected to provide general trends rather than quantitative accuracy.

With those considerations in mind, let us now analyze the pK_a of a histidine residue within a protein, i.e. His140 in the structured part of the human prion protein (residues 125–228).

As a first step, we optimized the geometry of the His140 side chain in the gas phase at the ONIOM: PBE0/6-31+G(d,p)/AMBER level, keeping frozen all the remaining degrees of freedom of the protein. In the ONIOM [35] procedure we used the His140 side chain treated at the DFT level (cut between $C\beta$ and $C\alpha$ atoms of His140, Fig. 2). The same procedure was applied for protonated and neutral forms of His140, starting from the experimental NMR structure of HuPr^P [36]. The main optimized geometrical parameters of the histidine side chain are collected in Table 3. A comparison with the results of PBE0/6-31+G(d,p) geometry optimization of an isolated 4-methylimidazole molecule shows that the geometry changes induced by the presence of the protein are not significant, but for a slight lengthening of the N3–H bond owing to the presence of the salt bridge with Asp147. Owing to the similarity of the equilibrium geometry, we can safely assume that the effect of the

protein on the methylimidazole pK is not related, at least to first order, to structural modifications.

Finally, the histidine pK_a was calculated by single-point PCM calculations, using two different solvation models (DPCM and CPCM) on three different model systems 1, 2, 3, differing in the size of the part treated at the QM level. In model system 1 only the side chain of His140 was treated at the QM level. Model system 2 is constituted by the side chain of His140 and Asp147. These two residues are indeed hydrogen-bonded in the experimental structure of HuPr^P, and the pK_a of His140 is expected to be significantly influenced by the possible formation of a strong salt bridge with Asp147 in a pK range in which the aspartic acid side chain is not protonated (i.e. for $pH \geq 4$). Finally, model system 3 contains together with His140 and Asp147, also Asp144, i.e. the acid/base residue closest to the His140/Asp147 pair (Fig. 2).

The cavity used in the QM calculations is always that containing the whole protein. From a computational point of view, this model has quite modest requests: all the computations reported in this study were performed on a single Pentium IV 1800 Mhz node with 200 MB of memory and 1 GB of disk allocation. The computation time was about 1 h per SCF cycle for the most involved PCM calculation, which exploited about 100000 tesserae.

We also checked the convergence of the PCM computations with respect to the average dimension of the tesserae (keyword *tsare* in the Gaussian package) (Table 4). It is quite gratifying that converged CPCM results are obtained even using average areas of 0.4\AA^2 (the default in the Gaussian98 package), whereas the DPCM requires values as small as 0.2\AA^2 (which is indeed the default in Gaussian03).

We used the difference of the QM energy in solution of the protonated and the neutral form of the model system in solution to estimate the pK of His140 within the protein. From a physical point of view, this assumption corresponds to treating the remaining part of the protein as a dielectric medium with $\epsilon = 1$, whose influence on the histidine pK is mainly to reduce the solvent-accessible surface of His140 side chain.

The results of the computation on systems 1, 2 and 3 are collected in Table 5. The passage from the isolated

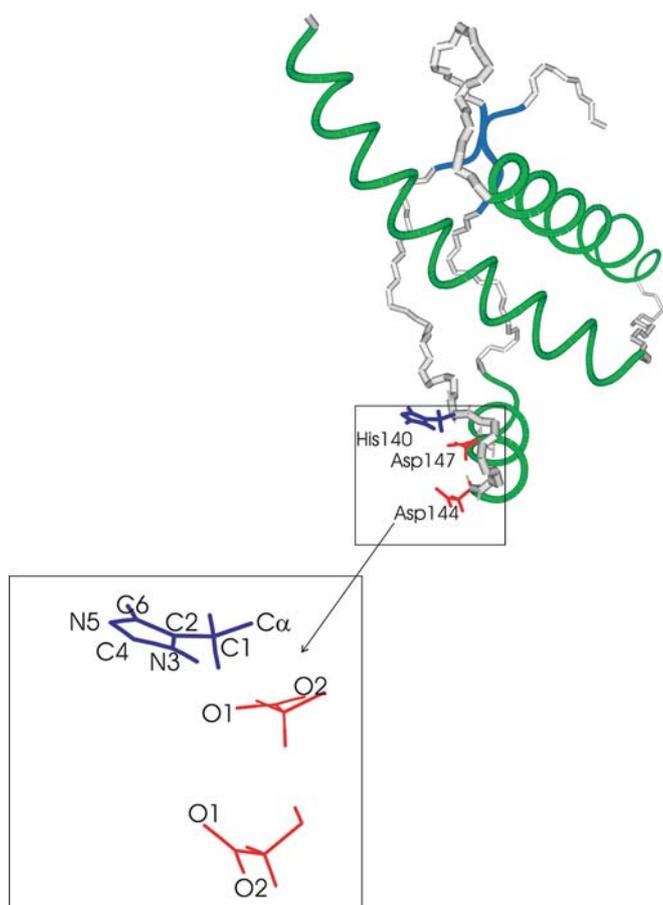


Fig. 2. Sketch of human prion protein(125–228). Residues treated at the quantum mechanical level are shown, together with the adopted atom labeling. The C1 atom of His140 corresponds to $C\beta$ in the standard protein nomenclature

Table 3. Optimized geometrical parameters of protonated and neutral forms of methylimidazole [δ isomer, PBE0/6-31+G(d,p) calculations] and of the His140 side chain [ONIOM:PBE0/6-31+G(d,p)/AMBER calculations]

	Neutral		Protonated	
	Isolated	Protein	Isolated	Protein
C1–C2	1.487	1.478	1.484	1.466
C2–N3	1.377	1.375	1.384	1.383
N3–C4	1.363	1.361	1.332	1.329
C4–N5	1.310	1.316	1.328	1.331
N5–C6	1.373	1.377	1.377	1.383
C6–C2	1.373	1.371	1.366	1.365
N3–H	1.010	1.029	1.012	1.035
N5–H			1.012	1.001

Table 4. Total energy (in atomic units) in solution of the His140 (system 1) side chain by PBE0/6-31+G(d,p) calculations. Energy differences are given in kilocalories per mole. 265 au has been added to all energies

Average tessera surface (\AA^2)	N_{TS}^a	Neutral		Protonated		Δ	
		CPCM	DPCM	CPCM	DPCM	CPCM	DPCM
0.4	36139	-0.26047	-0.26016	-0.71826	-0.71464	287.3	285.2
0.3	43780	-0.26052	-0.26018	-0.71826	-0.714097	287.2	284.8
0.2	57395	-0.26061	-0.26039	-0.71836	-0.71347	287.2	284.3
0.1	96734	-0.26057	-0.26055	-0.71837	-0.71307	287.3	284.0

^aFor the protonated species

Table 5. Differential energy in aqueous solution for the protonated and the neutral form of methylimidazole (δ isomer). PBE06-31+G(d,p) calculations with $T_{\text{sare}}=0.3$ (see text for details)

	Isolated His	Model 1		Model 2		Model 3	
	ΔG	ΔG	ΔpK_a	ΔG	ΔpK_a	ΔG	ΔpK_a
DPCM	290.0	284.8(272.2 ^a)	-3.7	296.3	4.6	299.5	7.0
CPCM	290.8	287.2(272.6 ^a)	-2.6	295.0	3.1	295.6	3.5

^aIncluding only the contribution of the spheres associated with methylimidazole

Table 6. Solvent accessible surfaces (SAS in angstroms squares) and solvation free energies (in kilocalories per mole) for the different groups of 4-methylimidazole [CPCM/PBE0 6-31+G(d,p) calculations] and of the His140 side chain [ONIOM:CPCM/PBE0 6-31+G(d,p)/AMBER calculations]

	SAS				ΔG	
	Isolated	Protein	Isolated		Protein	
			CPCM	DPCM	CPCM	DPCM
	Neutral					
C1	59.16	19.40	-0.89	-0.94	-0.56	-0.53
C2	0.67	0.75	-0.02	-0.02	-0.01	-0.01
N3-H	13.32	2.56	-5.85	-5.48	-0.06	-0.04
C4	26.51	17.20	-1.70	-1.60	-1.51	-1.43
N5	11.48	11.42	-5.51	-5.66	-5.10	-5.06
C6	23.04	21.14	-1.28	-1.20	-1.10	-1.04
Total	135.745	72.47	-15.33	-14.98	-8.34(-9.88 ^a)	-8.11(-9.57 ^a)
	Protonated					
C1	59.02	19.28	-7.75	-8.19	-5.25	-4.79
C2	0.77	0.72	-0.13	-0.08	-0.18	-0.11
N3-H	12.89	2.79	-17.19	-16.38	-1.05	-0.78
C4	26.23	16.24	-14.68	-14.89	-13.02	-12.45
N5-H	14.04	13.10	-18.41	-17.70	-19.21	-18.46
C6	22.65	20.48	-7.30	-7.16	-7.16	-8.48
Total	136.72	72.61	-65.79	-64.71	-45.71(-58.79 ^a)	-44.97(-56.07 ^a)

^a Including also the contribution of spheres not associated with the His140 side chain

molecule to system 1 induces a significant shift of the His140 pK_a toward lower pH values. This result is not surprising taking into account that the solvation energy of the protonated imidazole is obviously much larger than that of its neutral counterpart. Notwithstanding that the His140 side chain is on the protein surface, its solvent-exposed surface is obviously smaller (by around 50%, Table 6) than that of an isolated methylimidazole molecule. As a consequence, within the protein the differential solvation energy between the protonated and the neutral imidazole ring is reduced and the pK_a is lower. The decomposition of the total solvation energy into the contribution of the different spheres associated with each non-hydrogen atom allows us to highlight some interesting features. It is noteworthy that the decrease in the solvation energy of the two species is much smaller than the corresponding decrease in the

solvent-exposed surface. When going from the aqueous solution to the protein, the group experiencing the larger decrease in the solvent-exposed surface (around 60% of the overall decrease) is the methyl substituent ($C\beta H_2$ of the His140 side chain), whose contribution to the total solvation energy is very small. As a matter of fact, the decrease in the solvation energy is due mainly to the burial of the N3H group, whose solvent-exposed surface is vanishing owing to the formation of a salt bridge with the Asp147 side chain. Interestingly, the contribution coming from the spheres directly associated with the imidazole moiety is only 80% of the total solvation energy for the protonated species, while for the neutral side chain it is almost 90%. This is not surprising, since the presence of a positive net charge is expected to make nonlocal solvation effects much more significant. Inspection of Fig. 3 shows indeed that a quite large

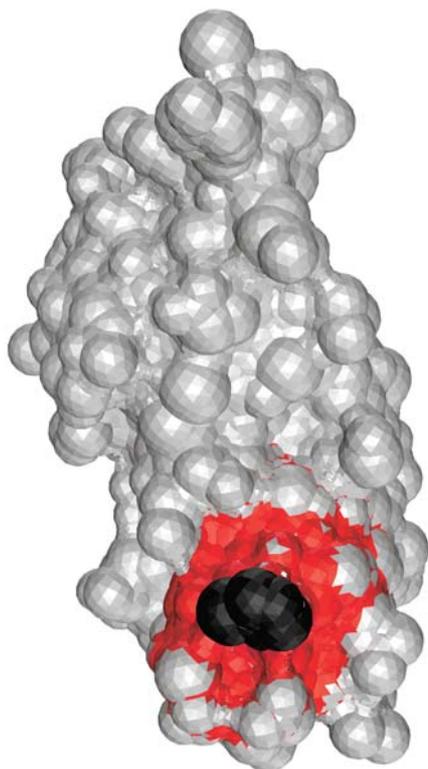


Fig. 3. Human prion protein (125–228) cavity. Tesserae associated with His140 atoms are shown in *black*. Tesserae whose polarization charge is zero and nonzero are shown in *white* and *red*, respectively

patch of the solute cavity around the His140 side chain exhibits nonnegligible solvation charges.

The presence of nonlocal solvation effects can also explain why the differences between the solvation energies predicted by CPCM and DPCM methods are larger within the protein than for the isolated molecule. As a matter of fact, when looking at the contribution of the spheres associated with the imidazole moiety, the discrepancy between DPCM and CPCM results is similar to those found for the calculations on the isolated molecules. Owing to the slower decay with distance ($\frac{1}{r}$ versus $\frac{1}{r^2}$), CPCM calculations predict a significantly larger contribution to the solvation energy of the spheres more distant from His140 (around 2 kcal/mol) when compared to the DPCM counterparts.

When the Asp147 side chain is also included in the QM calculations (system 2) the situation is reversed with respect to the prediction obtained on system 1. The His140 pK_a is indeed predicted to shift toward higher pH values with respect to the isolated molecule. The positive charge on the imidazole ring is indeed stabilized by the formation of a salt bridge with the Asp147 side chain. The extra stabilization would obviously be particularly large in the gas phase: the imidazole protonation energy increases by around 16 kcal/mol owing to the presence of the Asp147 side chain. The importance of this effect is strongly reduced by the solvent: for example, when the His140 side chain is protonated the contribution to solvation energy of the propionate group modeling the aspartic acid side chain is reduced by around 3 kcal/mol.

Table 7. Solvation free energies (in kilocalories per mole) obtained by different solvation methods (with respect to the PCM reference) for the system formed by the His140, Asp147 and Asp140 side chains. The partial contributions to G_{sol} of the spheres associated with polar groups are also reported. All the computations were performed at the PBE0/6-31 + G(d,p) level

	PCM	CPCM	DPCM
His140 Asp147			
Total	0.0	-0.21	-0.02
N3(His140)	-0.02	-0.02	-0.01
N5(His140)	-11.09	-11.16	-10.53
O1(Asp147)	-3.40	-3.42	-3.64
O2(Asp147)	-12.57	-12.62	-13.02
His140 Asp147 Asp144			
Total	0.0	-0.46	-8.83
N3(His140)	0.10	0.09	-0.06
N5(His140)	-6.92	-6.99	-6.08
O1(Asp147)	-2.86	-2.90	-3.93
O2(Asp147)	-20.04	-20.11	-23.67
O1(Asp144)	-23.09	-23.15	-27.88
O2(Asp144)	-31.83	-31.88	-37.15

This result is probably due to the onset of repulsive electrostatic interactions between the positive electron density appearing around the carboxylate group for the solvent reaction field and the positive charge on His140. The same effect is obviously present for the negative solvation density around His140. In other words, we can also say that when His140 is protonated the aspartic acid–histidine system becomes, as a whole, a neutral solute, whereas when histidine is not protonated the system is negatively charged.

The results obtained for system 3 show that residues not extremely close to the acid/base species under investigation can also influence its pK_a . The His140 pK_a is indeed increased by 0.6 units when Asp143 is included in the QM calculations. From a methodological point of view, it is important to highlight that DPCM and CPCM calculations, while providing the same qualitative indication, give remarkably different results from the quantitative point of view. The pK_a shift predicted by DPCM calculations is 3.5 pK units larger than that obtained at the CPCM level. The two models thus confirm their differences when treating the long-range solvation effects of charged systems, mostly when containing both positive and negative charges. In order to ascertain which is the most reliable model in those cases we studied a model system containing only His140 and Asp147 side chains and compared the performance of the DPCM, the CPCM and the PCM (Table 7). Although the three models give very similar solvation energies, the analysis of the decomposition of the electrostatic solvation energy in the contributions of the spheres associated with different atomic groups suggests that the similarity between the PCM and DPCM results is mainly due to error cancellation between the “positive” and the “negative” part of the solute cavity. As a matter of fact, when the Asp144 side chain is also included in the calculations, the DPCM gives results remarkably different from those obtained by the PCM and the CPCM. The discrepancy is much larger than that found for an isolated propionate group, confirming

that DPCM calculations are not able to handle long-range solvent effects in strongly charged systems.

Finally, it is worth highlighting that the use of the PCM method does not introduce any limit in the choice of the procedure used to compute the p*K*. For example, PCM computations can be used in a pure classical procedure computing the p*K* shift due to the presence of protein (ΔpK) according to the following relationship:

$$\Delta pK = (\Delta G_{MM}^{\text{Protein}} - \Delta G_{MM}^{\text{Model}})/1.3644, \quad (18)$$

where $\Delta G_{MM}^{\text{Protein}}$ is the MM free-energy difference in solution due to the protonation of a given residue of a protein, whereas $\Delta G_{MM}^{\text{Model}}$ is the MM free-energy difference in solution between the protonated and the neutral form of a suitable isolated model of that residue. For example, the total AMBER free energy in solution of HuPr^P(125–228) is –4314.66 kcal/mol when His140 is protonated and –4314.36 kcal/mol when it is neutral. On the other hand, it is 13.09 and 6.01 kcal/mol for the protonated and the neutral form, respectively, of a methylimidazole molecule in aqueous solution. The resulting ΔpK is 5.4, in reasonable agreement with the QM calculations. It is also worth noting that in the present case a PCM calculation involving about 1000 atoms and 100000 tesseræ can be performed using the Pentium IV 1800 Mhz system mentioned earlier in about 45 min.

4 Discussion and conclusions

From a methodological point of view, the most important outcome of this study is that, thanks to our latest methodological and implementative developments, the PCM approach can be used for studying by integrated QM/MM models the behavior in solution of systems containing thousands of atoms. For polar solvents, the CPCM version provides results very close to those delivered by the best PCM implementation taking escaped-charge effects into proper account. This is quite significant since the CPCM is significantly simpler and more effective from a computational point of view. At the same time the DPCM provides reasonable results only when employing the CompIV compensation procedure, but suffers from serious problems in describing in a balanced way long-range interactions. In any case, the remaining problems in QM/MM models concern the partitioning between different regions and are essentially equivalent for a vacuum and for a solution, the PCM embedding not involving any significant additional burden.

For the p*K*_a of His140 in HuPr^P, our calculations suggest that it is shifted toward higher pH values. In other words, at pH7 this residue is expected to be protonated. This result is due to the formation of a strong salt bridge with the side chain of Asp147, which should be negatively charged at least up to pH 4. The results of these preliminary QM calculations are in agreement with the comparison between the experimental NMR structures and molecular dynamics simulations in which

different protonation states are assigned to histidine residues [19]. Analogously, Mead computations [9], which use a combination of electrostatic calculations with a suitable Monte Carlo procedure averaging over all the possible protonation states of the titrable residues [37], predict that His140 should have p*K*_a ≥ 7. His140 should thus not be involved in the conformational transition exhibited by HuPr^P upon lowering of the pH.

Acknowledgement. This paper is a tribute to the scientific achievements and to the career of Jacopo Tomasi.

References

- (a) Hansch C, Sammes PG, Taylor JB (eds) (1990) Comprehensive medicinal chemistry Vol 4. Quantitative drug design. Pergamon, Oxford; (b) Shuurmann G, (1998) In: Ecotoxicology. Shuurmann G, Markert B (eds) John Wiley, New York, pp 665–749
- (a) Reeve W, Erikson CM, Aluotto PF (1979) Can J Chem 57: 2747; (b) Bordwell FG (1988) Acc Chem Res 21: 456
- (a) Li H, Hains AW, Everts JE, Robertson AD, Jensen JH (2002) J Phys Chem B 106: 3486; (b) Jang YH, Goddard WA III, Noyes KT, Sowers LC, Hwang S, Chung DS (2003) J Phys Chem B 107: 344
- (a) Shuurmann G (1997) In: Chen F, Shuurmann G (eds) Quantitative structure – activity relationship in environmental sciences Vol VII. SETAC, Pensacola, FL, p 225; (b) Shuurmann G, Cossi M, Barone V, Tomasi J (1998) J Phys Chem A 102: 6706; (c) Adzelm J, Kulmel C, Klant A (1995) J Chem Phys 103: 9312; (d) Cramer CJ, Truhlar D (1992) Science 256: 213; (e) da Silva CO, da Silva EC, Nascimento MAC (2000) J Phys Chem A 104: 2402; (f) Liptak MD, Shields GC (2001) J Am Chem Soc 123: 7314; (g) Liptak MD, Gross KC, Seybold PG, Feldgus S, Shields GC (2002) J Am Chem Soc 124: 6421; (h) Chipman DM (2002) J Phys Chem A 106: 7413
- Piego JR Jr (2003) Chem Phys Lett 367: 145
- Saracino GAA, Improta R, Barone V (2003) Chem Phys Lett 373: 411
- Li H, Pomelli CS, Jensen JH (2003) Theor Chem Acc 109: 71
- Scalmani G, Barone V, Kudin K, Pomelli C, Scuseria GE, Frisch M Theor Chem Acc (in press)
- Bashford D, Gerwert K (1992) J Mol Biol 224: 473–486
- Prusiner SB, DeArmond SJ (1994) Annu Rev Neurosci 17: 311
- Weissmann C (1996) FEBS Lett 21: 482
- Prusiner SB (1991) Science 252: 1515
- Prusiner SB, McKinley MP, Bowman KA, Bolton DC, Bendheim PE, Groth DF, Glenner GG (1983) Cell 35: 349
- Swietnicki W, Petersen R, Gambetti P, Surewicz WK (1997) J Biol Chem 272: 27517
- Jackson GS, Hill AF, Joseph C, Hosszu L, Power A, Waltho JP, Clarke AR (1999) Collinge J Biochim Biophys Acta 1431: 1
- Hornemann S, Glockshuber R (1998) Proc Natl Acad Sci USA 95: 6010
- Alonso DOV, DeArmond SJ, Cohen FE, Daggett V (2001) Proc Natl Acad Sci USA 98: 2985
- Alonso DOV, An C, Daggett V (2002) Philos Trans R Soc London Ser A 360:1165
- Langella E, Improta R, Barone V (submitted)
- Topol IA, Tawa GJ, Burt SK, Rashin AA (1999) J Chem Phys 111: 10998
- Tissandier MD, Cowen KA, Feng WY, Gundluach E, Cohen MH, Earhart AD, Coe JV, Tuttle TR (1998) J Phys Chem A 102: 7787
- Lim C, Bashford D, Karplus M (1991) J Phys Chem 95: 5610
- Cossi M, Rega N, Scalmani G, Barone V (2002) J Chem Phys 117: 43
- Barone V, Cossi M, Tomasi J (1997) J Chem Phys 107: 3210

25. Cancès E, Mennucci B, Tomasi J (1997) *J Chem Phys* 107: 3031
26. Cossi M, Rega N, Scalmani G, Barone V (2001) *J Chem Phys* 114: 5691
27. Cossi M, Rega N, Scalmani G, Barone V (2003) *J Comput Chem* 24: 669
28. Amovilli C, Barone V, Cammi R, Cancès E, Cossi M, Mennucci B, Pomelli CS, Tomasi J (1998) *Adv Quantum Chem* 32: 227
29. Tomasi J, Persico M (1994) *Chem Rev* 94: 2027
30. Chipman DM (2000) *J Chem Phys* 112: 5558
31. Mennucci B, Tomasi J (1997) *J Chem Phys* 106: 5151
32. Adamo C, Barone V (1999) *J Chem Phys* 110: 6158
33. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomery JA Jr, Vreven T, Kudin KN, Burant JC, Millam JM, Iyengar SS, Tomasi J, Barone V, Mennucci B, Cossi M, Scalmani G, Rega N, Petersson GA, Nakatsuji H, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Klene M, Li X, Knox JE, Hratchian HP, Cross JB, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Cammi R, Pomelli C, Ochterski J, Ayala PY, Morokuma K, Voth G, Salvador P, Dannenberg JJ, Zakrzewski VG, Dapprich S, Daniels AD, Strain MC, Farkas O, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Ortiz JV, Cui Q, Baboul AG, Clifford S, Cioslowski J, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Gonzalez C, Pople JA (2003) *Gaussian*, Pittsburgh, PA
34. Rappé AK, Casewit CJ, Colwell KS, Goddard WA III, Skiff WM (1992) *J Am Chem Soc* 114: 10024
35. Dapprich S, Komáromi I, Byun KS, Morokhuma K, Frisch MJ (1999) *J Mol Struct (Theochem)* 461: 1
36. Zahn R, Liu A, Luhrs T, Riek R, von Schroetter C, Garcia F, Billeter M, Calzolari L, Wider G, Wüthrich K (2000) *Proc Natl Acad Sci USA* 97: 145–150
37. Beroza P, Fredkin DR, Okamura MY, Feher G (1991) *J Mol Biol* 224: 473