Quantum chemical modeling of some biological mechanisms involving proton transfer

Dušan Hadži

Boris Kidrič Institute of Chemistry and Lek-Pharmaceutical and Chemical Works, 61115 Ljubljana, POB.30, Yugoslavia

Abstract - Quantum mechanical treatments of biochemical processes require a reduction of the real system to models of manageable dimensions and the choice of efficient computational techniques. The importance of proper basis set selection and inclusion of medium effects are illustrated by ab-initio calculations on the methylamine-formic acid system. Results of model calculations on possible proton transfers in rhodopsin and aspartic proteinases are presented.

INTRODUCTION

Translocation of charge concomitant to the proton transfer is essential for many biological mechanisms. In fact, there is hardly any biological process without involving in some way proton transfer (PT). The understanding at atomic level of such mechanisms in terms of energy profiles requires a possibly accurate production of proton potential functions. In view of the size and complexity of biological structures mediating PT this is a very demanding computational task and therefore the quantum mechanical treatments have to be carried out on appropriately reduced model systems using economical methods. Despite the ensuing reduction of accuracy the results of such treatments are useful in rationalizing and complementing experimental data. Often they can be instrumental in reaching decisions between mechanistic alternatives suggested by experiments. In the following some representative mechanisms involving PT will be examined first in order to illustrate the modellist's task and to set the stage for a brief discussion of possible quantum mechanical approaches. These will be exemplified by work in the author's group on rhodopsin and the proteolysis by aspartic proteinases.

EXAMPLES OF MECHANISMS INVOLVING PROTON TRANSFER

The mechanisms may be roughly divided in three groups. In one, several PTs occur in concert so that charge is translocated over large distance, for instance across the cell membrane or along the membrane surface. Such translocations occur in compensation for electronic charge translocation in chemical energy transforming processes as e.g. mediated by H⁺-ATPases (ref. 1) or to create an electrochemical gradient used in chemical energy storage, e.g. by bacteriorhodopsin (ref. 2). The carriers of long distance proton translocations are transmembrane helical proteins and, possibly, water channels. However, the nature of discrete groups such as ionic aminoacid side chains is not yet defined and this limits the treatments to rather general models such as chains of water, methanol or imidazole molecules (ref. 3). However, the possibility of identifying discrete aminoacids involved in proton transfers by site directed mutagenesis (ref. 4) is promising better defined model approaches.

PTs in shorter aminoacid chains are ubiquitous in enzyme mechanisms in which such charge relays provide for structural changes in the substrate, facilitating its conversion to the transition state. Perhaps the most intensely quantum mechanically elaborated enzyme group are the serine proteinases with their Asp-His-Ser charge relay to be discussed later. Charge relays also occur in the role of pH driven redox potential regulating devices in cytochromes (ref. 5). Besides the well established enzyme mechanisms, topical PTs may be involved in other highly important biological mechanisms.
such as induction of mutations by PT in nucleic base pairs (ref. 6), and drug receptor triggering (refs. 7,8) which also attracted quantum mechanical treatments. The third group of mechanisms is represented by overall changes in the conformation of functional proteins triggered by changes in the state of ionization by PT to or from amino acid side chains. This perturbs the electrostatic component of the potential energy of the molecule with concomitant conformational change and alteration of biological properties. A good example is the diphtheria toxin which undergoes an abrupt conformational change at low pH (<5 at 23°C). Burried tryptophans become exposed and a general increase in hydrophobicity occurs that facilitates the penetration of cell membranes by the toxin (ref. 9).

**COMPUTATIONAL ASPECTS**

PTs appear in biological mechanisms within large molecular complexes which as such are hardly accessible to any high level quantum methods even with the best available computers. They have therefore to be seriously reduced on one hand and, on the other, economical versions of quantum methods have to be used what anyway includes model approaches to the exact Hamiltonian. The situation is strongly aggravated by the necessity of considering the effects of the medium on the proton potential functions. For the selection of appropriate computational schemes high level computations on relatively small representative proton donating and accepting molecules such as water, ammonia, formic acid, etc. are needed for reference. On such systems it is possible to investigate the influence of the basis set, basis set superposition error (BSSE), electron correlation and other quality determining factors all of which strongly influence the depths of the energy minima and the potential barrier in the proton potential functions. Both type of test calculations have been done for several fundamentally important hydrogen bonded systems (ref. 10), at least "in vacuo". The results concerning the basis sets used in ab-initio calculations show that extended sets with corrections for BSSE and electron correlation are optimal for reproducing experimental protonation energies. However, these hardly can be applied to larger systems and particularly when geometry optimization is required. However, small split valence basis sets such as 4-31G, particularly if extended by some polarization functions (ref. 11) or even 3-21G are quite useful and, on the other hand, their deficiencies are well known and can be used in estimating the induced errors. Amongst the semiempirical methods the newly parameterized MNDO scheme AM1 (ref. 12) is claimed to be equivalent to ab-initio calculations with the 4-31G basis set, but this needs more testing (cf. Fig. 1). The obvious advantage of semiempirical methods is the possibility of extending the computations to larger numbers of atoms, which is particularly important in considering the effects of surrounding groups.

![Proton potential functions](image)
Systems in which PTs occur are usually directly connected by hydrogen bonds to neighbour aminoacids or such connections may be mediated by water molecules which influence the proton potential functions. Local electric fields originating in more or less distinct charged residues and, particularly, from dipoles created by $\alpha$-helices (ref. 13) are also highly important in shaping the proton potential function. Defined neighbour groups can be considered to a reasonable extent in the computation in the usual supermolecular approach and discrete charges can be included in the Hamiltonian of the main system (ref. 14). This also is the way to simulate the effects of the polarizable medium (e.g. 15).

Some of the points made above are illustrated by the calculations on the methylamine-formic acid dimer which is a good model for possible PT between side chains of glutamic and aspartic acids and lysine. The calculations were made with several basis sets, and monomer and intermolecular geometry optimization in the neutral form of the complex. Full geometry optimization was also carried out with the 4-31G basis set. In addition, the method of Tomasi and coworkers (ref. 15) was applied to simulate the effect of a polarizable medium (Fig. 1). With the minimal basis set the second minimum in the potential function is barely indicated and becomes distinct in the 4-31G calculation. The deeper minimum keeps, however, the proton still on the side of the carboxylic acid, but the switching on of the reaction field brings it over to the nitrogen side. This happens only with the larger basis sets. Thus we may expect that in real acid-base systems PT in one or the other way will be induced by a change in dielectric properties of the medium. Perhaps mechanistically even more interesting is the influence of a neighbour charge that strongly influences the shape of the potential function (refs. 16,17,18). Depending on the position of the charge on one or the other potential well becomes deeper. Thus the approach of a charged ligand to a hydrogen bonded pair of acidic and basic groups containing amino acid side chains (e.g. aspartate-lysine) may induce PT with ensuing charge shifts that may influence the conformational equilibrium of the protein. Moreover, PT in such a system may directly influence the protein force field by displacement of the side chain involved in PT. This follows from the computation in which we carried out a full geometry optimization on the methylamine-formic acid pair (ref. 19). Fig. 2 shows the relative displacements of the hydrogen bonded entities consequent with PT. This and related results will be used in developing models of ligand induced conformational changes in functional proteins, particularly drug receptors (ref. 20). Calculations with full geometry optimization are very demanding on computer time, but obviously necessary for realistic modeling.

**Some Examples of Modeling of Mechanisms with Proton Transfer**

Serine proteinases. A more complex example of quantum mechanical modeling of PT is that of the already mentioned charge relay in the active site of serine proteinases constituted by the triad Asp-His-Ser. Based on the X-ray structure of chymotrypsin (for a review see 21) is the suggestion (ref. 22) that the charge relay facilitates the nucleophilic attack of serine oxygen on the carbonyl carbon of the substrate by splitting off a proton from the serine OH group.

$$\text{(Asp)}\text{COO}^-...\text{NH(His)}...\text{HO(Ser)} \rightleftharpoons (\text{Asp})\text{COOH}...\text{N(His)}\text{NH}...\text{O(Ser)}.$$
The mechanistic proposals of Wang (ref. 23) and Polgar and Beveridge (ref. 24) differ in that the final PT between Asp and His is not occurring so that the leading charge system would be (Asp)COO\(^-\)...NH(His)NH\(^+\)...O(Ser).

The question whether one or two PTs are occurring stimulated extensive investigations based on both quantum mechanical modeling and experimental structural methods. For these calculations the triad aminoacids are usually reduced to formic acid (Asp), imidazole (His) and methanol (Ser), and arranged according to the actual geometry of the active site. In vacuo calculations in the ab-initio scheme on the formic acid-imidazole system yield lower energy for the neutral pair whereas gradual inclusion of the effect of discrete neighbours reverses the energy of the potential minima by about the same amount (refs. 25,26). Similar effects are obtained by the inclusion of the reaction field (ref. 27) and the protein electrostatic field (ref. 28). Thus the multicharged form of the triad is corroborated by quantum mechanical calculations.

Experimental approaches to the actual position of the proton in the Asp-His hydrogen bond of serine proteinases include \(^1\)H and \(^1\)N NMR spectroscopy and neutron diffractions (refs. 29-31). The results were variably interpreted and the conclusions contradictory. Most revealing appear to be the results yielded by hydrolysis rate measurements (ref. 32) in mixtures of H\(_2\)O and D\(_2\)O. The proton inventory technique, which allows the estimation of the number of protons participating in a reaction, was applied to the human leukocyte elastase with various oleopeptide substrates and the results show that the actual number of transferred protons may be one or two, depending upon the substrate. Obviously the small changes induced in the conformation of the active site of the enzyme by the binding of the substrate are sufficient to cause differences in the proton potential. This example clearly demonstrates the limitations of quantum mechanical modeling which are primarily due to the complexity of the systems. The simple models did not include the geometry changes induced by the mutual adaptation of the enzyme and substrate.

Rhodopsins. The conversion of light energy into chemical energy by the membrane holoprotein of Halobacterium halobium includes several PT events in the photoisomerization and the following thermal isomerization of the chromophoric retinal Schiff base (RSB) in which a proton is translocated between two sites within the protein opsin (ref. 2). This translocation is followed by the ejection of at least one proton to the surrounding medium. The photoisomerization of RSB was extensively studied and the intermediates are spectroscopically well characterized. The cycle is schematically shown in Fig. 3. RSB is covalently bound to Lys216 of the opsin in the trans conformation. Light absorption induces in essence a trans-cis isomerization of retinal whereby the hydrogen bond to an acidic side chain is broken and the proton of the protonated RSB is transferred to a different aminoacid residue. The intermediate stages are terminated by thermal isomerization back to the initial conformation and reprotonation of RSB.

Indications to the molecular nature of the proton acceptors are obtained from infrared and Raman spectroscopy. Several aspartate residues are changing during the isomerization cycle their state of protonation and a model (Fig. 3) was proposed that accounts for these changes (ref. 33). However, there is good spectroscopic evidence of deprotonations of tyrosines (ref. 34). Moreover, RSB seems to be initially hydrogen bonded to the carboxylic side chain of opsin residue via one water molecule (ref. 35). There were arguments about the initial state of protonation of RSB, i.e. whether it is anchored by a strong, neutral hydrogen bond or is it protonated. Considering the basicity of the Schiff base nitrogen relative to a carboxylic or even tyrosine phenolic group the former situation might be expected, but spectroscopic arguments favour the latter (ref. 36).

As far as the primary photocycle is concerned there are certain parallels between bacteriorhodopsin and the visual pigment, rhodopsin, but there also are differences. The RSB of the latter is in the dark adapted state in the cis conformation and isomerizes with photon absorption to the trans form (ref. 37). In the ensuing thermal isomerization of RSB in the visual rhodopsin the covalent link is hydrolysed and the Schiff base is enzymatically reconstituted in the last stage. The essential difference between bacteriorhodopsin and visual rhodopsin lays in the proton transfer. Whereas in the former protons are translocated across the membrane, in the latter an exchange only of the protons on the RSB seems to take place during the photoisomerization (ref. 37). However, this proton switching triggers a
Fig. 3. Photocycle of bacterial rhodopsin with deprotonation and reprotonation involving aspartates as proposed by Engelhard et al. (ref. 33).

conformational change of opsin. This is sensed by another protein, transducin, which couples the primary event to an enzymatic chain that turns it eventually into a nerve signal. The primary reactions of both rhodopsins are then rather similar in what PTs in the role of switching devices are concerned but in neither mechanisms is the microstructure of groups engaged in PT unambiguously resolved. Carboxylic and phenolic groups which have very different pK's appear to be involved and the question may be posed under what conditions PT may occur to and from the Schiff base nitrogen. Neighbour charges and medium differences are likely to be decisive in influencing the actual depth of the potential minima determining the proton position. The scope of quantum mechanical modeling is rather restricted by the lack of structural information, but calculations might nevertheless show whether at all and in which direction PT between candidate groups is possible.

Modeling of PT involved in rhodopsin functioning requires severe curtailing of the involved molecules lest the computation can be done at sufficiently high level. RSB is in our calculations represented by methylallylimine, asparagine by formic acid and tyrosine by phenol. The proton potential curves were calculated in the ab-initio scheme with the 4-31G basis set. The details of the calculations appear elsewhere (ref. 38) and only the main results will be considered here.

The first calculation concerns the methylallylimine-formic acid pair and the preferential position of the proton in this hydrogen bond. The respective curve in Fig. 4 shows that the lower energy minimum is at the carboxylic side will keep there the proton. However, the inclusion of a negative charge near the RSB nitrogen results in lowering the minimum nearer to this atom leading to PT (ref. 18). The existence of such charge was postulated from spectroscopic evidence (ref. 39). Similarly, the simulation of the medium
effect using the method of Tomasi (ref. 15) yields potential functions with deeper minima at the nitrogen side even with a small dielectric parameter (\( \varepsilon = 5 \)) (Fig. 4). Further, we wished to explore the effect of an intercalated water molecule according to the spectroscopic results of Alshuth et al. (ref. 35). Again it is the medium effect that brings about PT (Table 1).

TABLE 1. Ab initio (4-31G basis set) of the methylallylimine-water-formic acid system calculated at four points (minima in the STO-3G computed proton potential function) along the N...Owater (\( r_1 \)) and Owater...Oformiate (\( r_2 \)) hydrogen bonds

<table>
<thead>
<tr>
<th>( r_1 (\text{A}) )</th>
<th>( r_2 (\text{A}) )</th>
<th>Charge state</th>
<th>Relative energy (kcal/mol(^{-1}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vacuum</td>
<td>Solvent(^a)</td>
<td>--------------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>1.0</td>
<td>1.0</td>
<td>C(_2)H(_2)N(\cdot)H(_2)O(\cdot)HOOCH</td>
<td>0</td>
</tr>
<tr>
<td>1.9</td>
<td>1.0</td>
<td>C(_2)H(_2)N(\cdot)H(_2)O(\cdot)HOOCH</td>
<td>28.80</td>
</tr>
<tr>
<td>1.0</td>
<td>1.9</td>
<td>C(_2)H(_2)N(\cdot)H(_2)O(\cdot)HOOCH</td>
<td>41.56</td>
</tr>
<tr>
<td>1.9</td>
<td>1.9</td>
<td>C(_2)H(_2)N(\cdot)H(_2)O(\cdot)HOOCH</td>
<td>59.14</td>
</tr>
</tbody>
</table>

\(^a\)Continuum model, \( \varepsilon = 78.5 \) (ref. 15)

These results are in agreement with spectroscopic evidence on fully dried bacteriorhodopsin in which the RSB is nonprotonated. The last calculation was intended to show whether reversible PT may occur between RSB and tyrosine. The "in vacuo" computed curve (Fig. 5) has a deep minimum near the phenolic oxygen and the other minimum is barely indicated. However, the effect of the medium reverses the depth of the potential minima. These calculations demonstrate that PT is possible in the model hydrogen bonds and that the direction of the PT is governed by environmental effects whereby changes in conformation may approach or remove discrete charges or bring the proton transferring units to surroundings with different dielectric properties.
Proteolytic mechanism of aspartic proteinases. The structure of several enzymes of this group was determined by high resolution X-ray diffractions (refs. 40-42). The active site is constituted by two aspartic residues supposed to be connected on one side by a proton and on the other by a solvent molecule (Fig. 6a). X-ray diffractions cannot distinguish between this being water or a hydronium ion. The latter is possible in view of the low pH at which these enzymes are most active. Several mechanisms were proposed in which protonation of the peptide oxygen would facilitate the subsequent nucleophilic attack of water on the amide carbon atom. In the mechanism proposed by Pearl (ref. 43) the latter event should be facilitated by hydrogen bonding of the peptide carbonyl to an amino group. In the mechanisms proposing initial protonation of the carbonyl oxygen, PT would occur either from the solvent $H_2O^+$ (ref. 40) or the proton from the upper hydrogen bond would be transferred (ref. 42).

Our computational approach was aimed at (i) elucidating the energetics of the solvent molecule chelated by the aspartate residues and the possible positions of protons in a model of the active site (ref. 44), and (ii) setting the energetic frame for possible protonations of the peptide oxygen and, as an alternative, of nitrogen (ref. 45). The model consisted of two formiate molecules spatially related so that they simulate the geometry of aspartate residues 39 and 215 of endothiapeptidase (ref. 40). The solvent molecule and the proton positions were optimized. The results summarized in Fig. 6a show that the solvent is asymmetrically placed and that one proton migrates from $H_2O^+$ to the carboxylate group. Next, we replaced the solvent molecule by acetamide so that the carbonyl oxygen fits on to the site of the solvent oxygen (Fig. 6b). However, this arrangement is unstable the energy calculated in the STO-3G basis set being $+7$ kcal/mol. Energy minimization favours an attachment of the model substrate by NH...OCO hydrogen bonding to one carboxyl group (Fig. 6c) with the energy of $-19$ kcal/mol. Another stable complex (Fig. 6d) was obtained by placing the amide between the two
carboxylates so that it becomes hydrogen bonded to the upper proton. This arrangement is stabilized by 22 kcal/mol (9.6 kcal/mol in the 4-31G basis set). However, in view of the positive energy of the complex depicted in Fig. 6b a considerable barrier will oppose the approach of the peptide bond to this position.

PT energies were computed for the complex in Fig. 6d and the amide carbonyl protonation by H$_2$O$^+$. The former energy is 30 kcal/mol and thus the proposal of James and Sie?ecki (ref. 42) appears unlikely. Protonation energies by N-methylacetamide were calculated with the H$_2$O$^+$ ion bonded to hydrogen diformiate (Fig. 6e) and, alternatively to two additional waters, i.e. the trihydrated proton. The latter calculation was done in order to make the comparison between the energetics of carbonyl protonation by the enzyme and understanding of simple acid catalysed hydrolysis. The respective energies yielded by the 4-31G calculation and with optimized geometries are 29.8 kcal/mol and 4.75 kcal/mol, respectively. This means that protonation by H$_2$O$^+$ is energetically more favourable. It is worth mentioning that the protonation energy by H$_2$O$^+$ alone is negative (-15.27 kcal/mol). The details of these calculations will be published elsewhere (ref. 46), but we note another mechanistically important detail of these calculations which concerns the approach of the nucleophile to the scissile bond. The results obtained in the calculation of protonation of N-methylacetamide by the hydrated proton show that the electronic population of the carbonyl carbon is reduced by strong hydrogen bonding even more than in the full protonation in vacuo. This suggests that strong hydrogen bonding of the carbonyl would suffice to make its carbon susceptible for a nucleophilic attack by water. Obviously, the present results are also of consequence for the mechanism of amide hydrolysis in acidic media without enzyme catalysis.

The protonation energy calculations on the model of aspartic proteinases serve as a frame to which possible mechanisms have to be fitted. The details concerning the placement of a peptide in the enzyme active site which would allow for both appropriate hydrogen bonding and the approach of a water molecule have yet to be worked out. However, the quantum mechanical modeling appears to be a good guide even in such complex mechanisms despite all approximations imposed by reducing the real system to dimensions manageable by ab-initio calculations with at least a modest sized basis set.

Acknowledgement

This work and the quoted work from this laboratory were supported by the Research Community of Slovenia and the Lek Works.

REFERENCES

Quantum chemical modeling of some biological mechanisms

20. D. Hadži, to be published.