Theor Chem Acc (1999) 101:241–245
DOI 10.1007/s00214980m108

Theoretical
Chemistry Accounts

Regular article

Modelling the active site of glyceraldehyde-3 phosphate dehydrogenase with the LSCF formalism*

Alain Cartier[1], David Brown[1], Bernard Maigret[1], Sandrine Boschi-Muller[2], Sophie Rahuel-Clermont[2], Guy Branlant[2]

[1] Laboratoire de Chime Théorique, UMR CNRS-UHP 7565, Université Henri Poincaré-Nancy I,
B.P. 239, F-54506 Vandoeuvre-lès-Nancy Cedex, France (Part of the Institut Nancéien de Chimie Moléculaire)
[2] UMR CNRS-UHP 7567, Université Henri Poincaré-Nancy I, B.P. 239, F-54506 Vandoeuvre-lès-Nancy Cedex, France

Received: 24 March 1998 / Accepted: 3 September 1998 / Published online: 23 November 1998

Abstract. In the framework of a theoretical approach to the relationship between structure and reactivity of the catalytic centers of enzymes, glyceraldehyde-3 phosphate dehydrogenase (GAPDH) has been chosen as a model enzyme. In GAPDH, the proximity of His_{176} increases the reactivity of Cys_{149} at neutral pH; however, its presence alone is not sufficient to explain the reactivity of the catalytic Cys. In order to determine which other interactions play an important role, a study of the geometric and electronic structure of the catalytic site has been made using a hybrid quantum mechanics/molecular mechanics local self-consistent field method. This allows the computation of the electronic properties of amino acid residues in subsystems influenced by other parts of the macromolecule. The quantum subsystem was centered on the Cys_{149} residue of GAPDH. The structures of GAPDH taken from the crystallographic database did not include hydrogen atoms and these had to be added taking into account the fact that, in the active site, His_{176} has three tautomeric forms: δ-His protonated, ϵ-His protonated and His^+. The results presented here suggest that the most stable His...Cys system in GAPDH is a strongly hydrogen-bonded Cys^-_{149}/His^+_{176} ion pair.

Key words: Enzymatic catalysis – Glyceraldehyde-3 phosphate dehydrogenase – Quantum mechanics/molecular mechanics methods

1 Introduction

A number of chemical reactions catalyzed by enzymes involve the formation of covalent intermediates with reactive residues in the active site. All amino acids bearing lateral chains which potentially display nucleophilic properties can be considered, i.e. Ser, Thr, Tyr, Lys, Cys, His, Asp, Glu and Selenocys. For each of them, covalent enzyme intermediates have already been biochemically characterized. However, the intrinsic nucleophilicity of these amino acids is so different that this implies a major role of their protein environment in modulating their chemical reactivity.

For instance, changing a catalytic Cys into a Ser and vice versa, generally leads to a dramatic decrease in enzymatic efficiency although both residues belong to the same group in the periodic table and thus are electrochemically the closest. The potential reasons are various:

1. The amino acid introduced could express a lower nucleophilic reactivity within its proteic environment resulting in the chemical attack step being rate-limiting.
2. The positioning of the substrate relative to that of the nucleophile and of the corresponding covalent intermediates formed during the catalytic process relative to those of residues which can act as acid/base catalysts or stabilize the intermediates, might not be optimal.
3. The chemical stability and reactivity of the new intermediates formed during the catalysis process could lead to a new limiting step associated with their decomposition.

To gain more insight into these possible causes, phosphorylating glyceraldehyde-3 phosphate dehydrogenases (GAPDHs) have been chosen as models to perform a theoretical approach to the relationship between the reactivity of the catalytic active site Cys_{149} and its protein environment.

The ability of the hybrid quantum mechanics/molecular mechanics (QM/MM) scheme to credibly model the biochemistry of large complex biomolecules has been shown [1–3], even for geometry optimization and transition state searching [4]. The method used is a local self-consistent field (LSCF) approach developed in our

* Contribution to the Proceedings of Computational Chemistry and the Living World, April 20–24, 1998, Chambery, France

Correspondence to: A. Cartier

laboratory [5]. This hybrid QM/MM method allows us to determine the electronic properties of amino acid residues included in subsystems experiencing the influence of the other atoms in the macromolecule as well as of a surrounding medium. We can compute the usual observables of quantum chemistry such as the distribution of charges, the electronic density, the reactivity (HOMO/LUMO), bond indices and electrostatic potentials.

2 Molecular systems studied

The GAPDH structures taken from the crystallographic database did not include hydrogen atoms. These had to be added taking into account the following dilemma: in the active site of the enzyme, there are two catalytic amino acid residues, Cys_{149} and His_{176}, which are close to each other, the latter having three tautomeric forms (protonated in δ or in ϵ, and His^+). In order to determine which is the most stable form of the system His...Cys, we studied the system twice, each time with one of the two following hypotheses: $His+ \ldots Cys-$ and His δ-protonated.

The quantum subsystem used is centered on the catalytic Cys_{149} residue of GAPDH. We will be able to study the modifications which occur in the environment of this residue due to a mutation of Cys_{149} into Ser or Selenocys as well as those resulting from modifications within the environment of this residue. Actually, when compared to the usual methods of free-energy simulation and pK_a calculation, which assume that the system studied is in a fixed ionization state, the LSCF method allows us to take into account the numerous changes of protonation during the simulation.

3 Calculations

3.1 Classical simulations using molecular dynamics and energy minimization

Before performing any of the LSCF calculations, we first performed completely classical simulations of the tetramer of GAPDH in a bath of explicit water molecules. The starting point for these simulations was the X-ray structure determined by Skarzynski et al. [6].

This preliminary classical simulation step is necessary because in the X-ray structure, the Cys group is oxidized and carries a bulky group which displaces the thiol group from the position it would have when engaged in the Cys^-/imidazole$^+$ ion pair. Minimization alone though runs the risk of falling into the nearest local minimum. For this reason, the following procedure was carried out.

The GAPDH was placed in a periodic simulation box of side length 100 Å. The box was then filled with water. The system thus contained the tetramer (4NAD (cofactor) molecules of 70 atoms each plus the 20392 atoms in the protein chain), 68 Na^+ ions, 52 Cl^- ions and 26146 water molecules, i.e. a total of 99230 atoms altogether at a density of 1029.928 kg m^{-3}.

Due to the large size of this system, we used a parallelized classical simulation package (*ddgmq*) developed in our laboratory [7]. Using the *ddgmq* program, this system was first energy minimized (EM1) to remove high energy overlaps resulting from the introduction of the water molecules. A molecular dynamics (MD) simulation was then performed in which the temperature was raised gradually to 300 K at constant volume. This was achieved by initially randomizing the velocities using a Gaussian distribution characteristic of 300 K and setting the temperature loose-coupling constant to 1 ps. After 20 ps of MD, the system had relaxed at the desired temperature. It was then quenched by setting the required temperature to 1 K using MD. After 10 ps, the system was energy minimized (EM2). A second period of MD at 300 K followed in the same way. This time, the total period of dynamics was extended to 100 ps before the system was quenched and energy minimized (EM3) in the same way as before. With the following Ewald summation parameters [8] ($\alpha = 0.2$ Å^{-1}, $R_c = 11.5$ Å, $K_{max} = 10$) and a tolerance of 10^{-5} used in the SHAKE routine maintaining all bond lengths rigid, the program, in MD mode, used 3.3 s of processing time per 1 fs time step on 27 processors of the SGi Origin 2000 in Nancy.

The force field used in *ddgmq* is based on the standard CVFF (Consistent Valence Force Field) one but differs from it slightly in two respects. First, the angle bending term is, for computational reasons, harmonic in the cosine of the angle, rather than in the angle itself. Second, the out-of-plane potential is harmonic in the distance of the central (trivalent) atom from the plane of the other three. In both the angle bending and out-of-plane bending cases we obtain the corresponding force constant for use in *ddgmq* from the CVFF form by equating the curvatures at the minimum of the potentials. The results of the minimization procedures are given in Table 1.

Table 1. Summary of the classical simulations for the Cys/imidazole hypothesis: the various energies given are from the final configurations of the three energy minimizations (EM) (see text for details) except for the first column which shows the energies after just the first 100 steps of EM1. NB these energies are quoted in joules per mole of atoms for the entire GAPDH plus water system, i.e. all 99230 atoms

Configuration	EM1(100 steps)	EM1	EM2	EM3
Total energy	−10260.48	−13244.16	−13890.47	−13912.12
Total angle bending energy	686.3351	663.6125	660.9465	656.5342
Total torsional energy	95.81995	74.84495	61.29700	56.30569
Total van der Waals energy	1648.106	2034.960	2259.575	2233.795
Total coulombic energy	−12699.73	−16020.72	−16874.72	−16860.99
Out-of-plane energy	6.438299	3.144966	2.426598	2.227787

3.2 Final optimization using classical mechanics

The structures obtained at the end of the three minimization procedures described above were then fully optimized using DISCOVER [9] with the CVFF. For this final optimization all but the water molecules of the X-ray structure were discarded and the bonds were treated harmonically. The results are given in Table 2 and compared with the X-ray structure.

We can see a significant decrease of the S...Nϵ distance in the Cys^-/imidazole$^+$ hypothesis related to both the X-ray value and the Cys/imidazole one. We note that optimization in a spherical water shell of radius 50 Å (limit of our version of DISCOVER [9]) of the two forms at the end of the EM3 stage leads to geometries very close to those obtained without a water box. For this reason we retain only the tetramer, the water molecules included in the X-ray structure and the counter-ions for the QM/MM computations. The size of the remaining system is 22796 atoms with 2016 belonging to 672 explicit water molecules.

3.3 Optimization with the LSCF method

The basic assumption of the LSCF method [5] is that the chemical bonds separating the quantum subsystem from the rest of the macromolecule (frontier bonds) have constant and well-defined properties, which can be determined on a well-chosen model system. One assumes that each of these bonds is a simple bond and can be described by a strictly localized bond orbital (SLBO) which is expressed as a linear combination, with constant coefficients, of two well-defined hybrid orbitals, one on each atom of the bond. The atom of the bond belonging to the quantum subsystem is called the frontier atom. On each frontier atom, the hybrid orbital which defines the SLBO of the frontier bond is called a frozen orbital (Fig. 1). The electrons of the quantum subsystem can be described by a set of molecular orbitals built with the atomic orbitals of the atoms of this subsystem, provided that the orbitals of the frontier atoms remain orthogonal to the SLBO describing the frontier bond. This condition is easily fulfilled in semiempirical methods provided that the orbitals of the frontier atoms entering this set of molecular orbitals are orthogonal to the hybrid orbitals defining the SLBOs.

In our system, we chose to build the quantum subsystem in one of the monomers as follows: Cys_{149}, His_{176}, Ser_{177}, Asn_{313}, C* and O from Ser_{148}, C* and O from Val_{175}, C* and O from Asp^-_{312}, N* and HN from Thr_{150}, N* and HN from Tyr_{178}, and N* and HN from Glu^-_{314}, (*designates the frontier atoms).

Table 2. Final optimization using classical mechanics

Config. (hypothesis)	S...Nϵ	O...Nδ	S...H	Nϵ...H
EM1 (Cys...His)	3.57	2.82	1.33	2.55
EM2 (Cys...His)	4.14	2.92	1.33	4.20
EM3 (Cys...His)	3.08	2.85	1.32	2.82
EM3 (Cys−...His+)	2.66	2.87	1.61	1.06
Experiment [6]	3.60	2.63		

The semiempirical method used for the quantum part is PM3 [10], which has been shown to give results closer to the ab initio ones than other semi empirical methods in the study of hydrogen bonds [11–13] and the classical force field is CVFF. The computer program used was GEOMOP [14] which is a modification of GEOMOS [15] which accounts for interactions between QM and MM subsystems within the LSCF approximation. The

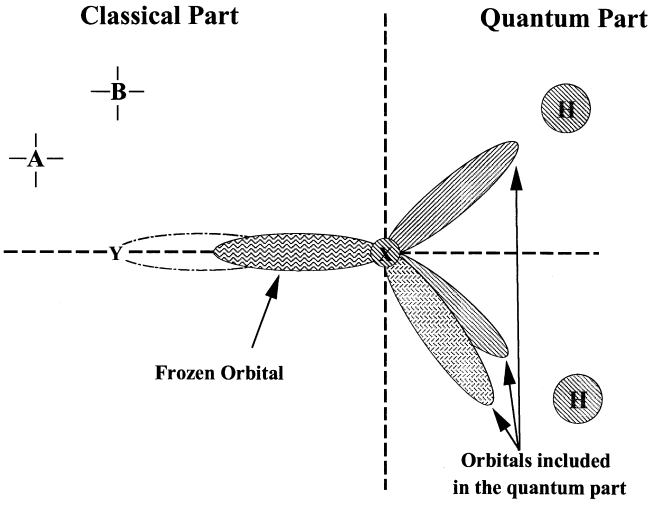

Fig. 1. Description of the system division in the hybrid method

Table 3. Proton transfer from Cys to His in the environment Cys/His

S...H (Å)	Nϵ...H (Å)	S...Nϵ (Å)	Energy (kcal mol^{-1})
1.40	1.753	3.146	−161383.679
1.50	1.685	3.180	−161384.337
1.60	1.617	3.213	−161382.615
1.70	1.530	3.227	−161378.343
1.80	1.157	2.955	−161374.819
1.90	1.104	3.001	−161372.346
2.00	1.049	3.065	−161369.622

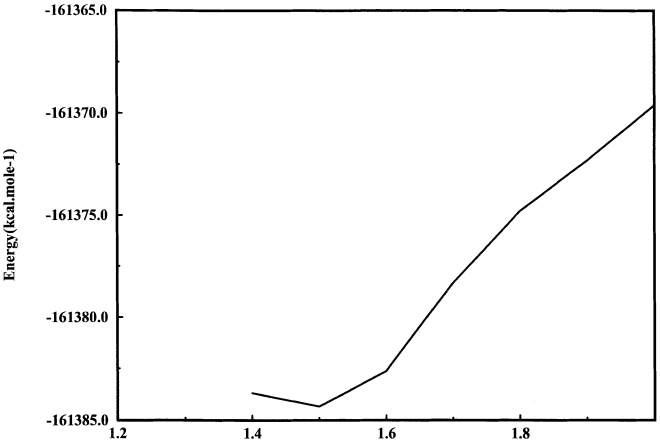

Fig. 2. Energy (kcal mol^{-1}) as a function of S...H bond length (Å) in the Cys/His environment

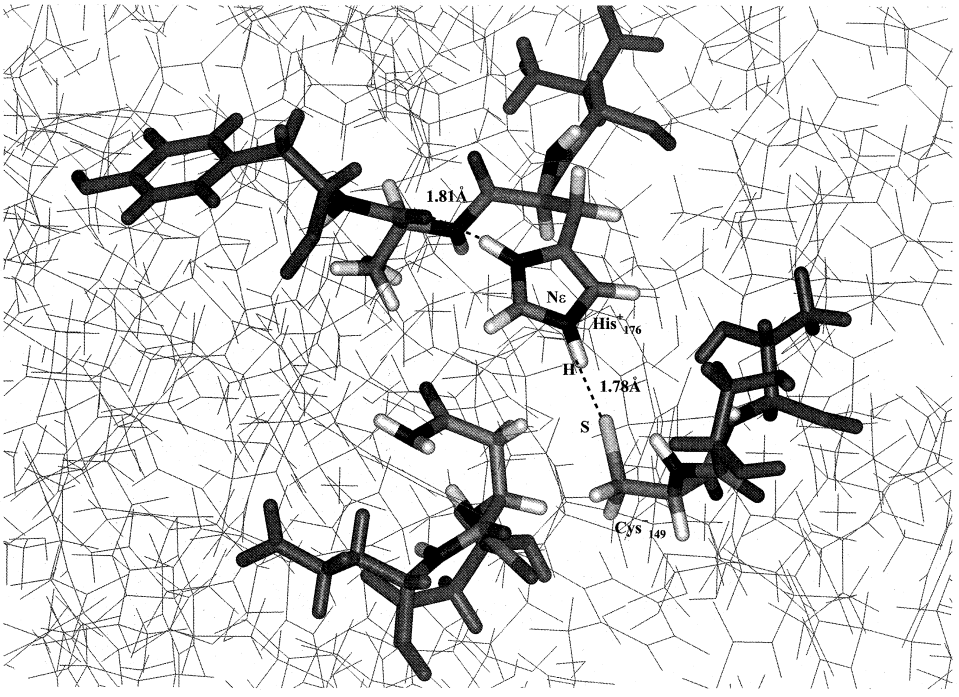

Fig. 3. Results in the Cys^-/imidazole$^+$ hypothesis

geometry of the quantum subsystem is fully optimized, the classical part remaining fixed and being that resulting from the final classical optimization after the EM3 step in each of the two hypotheses.

3.3.1 Results for the Cys/His hypothesis

Minimum:

S...Nϵ	S...H	Nϵ...H	Energy
3.175 Å	1.487 Å	1.693 Å	−161384.369 kcal mol^{-1}

If, in the environment computed for the Cys/imidazole hypothesis, we increase, with a step of 0.1 Å, the S—H bond length from 1.40 Å to 2.00 Å, while optimizing all the other geometric parameters in the quantum subsystem, we obtain the results given in Table 3.

These results lead to the unsatisfactory curve in Fig. 2 which shows a hydrogen atom neither bonded to S nor to N.

3.3.2 Results for the Cys^-/His^+ hypothesis

Minimum:

S...Nϵ	S...H	Nϵ...H	Energy
2.810 Å	1.767 Å	1.090 Å	−161394.018 kcal mol^{-1}

Results of the equivalent proton transfer calculation in the Cys^-/imidazole$^+$ case are shown in Table 4.

This time, an energy minimum is found for a structure (Fig. 3) corresponding to a Nϵ...H distance of 1.09 Å and exhibiting a hydrogen bond length of 1.78 Å confirming the existence of a Cys^-_{149}/His^+_{176} ion pair.

4 Conclusion

The existence of a Cys^-_{149}/His^+_{176} ion pair in GAPDH is consistent with the results of Harrison et al. [4] for Papain, using a hybrid QM/MM potential and a AM1 semiempirical method, and with the direct experimental observations of Talfournier et al. [16] on GAPDH itself in solution. The presence of this ion pair may explain the dramatic loss in activity found when Cys_{149} is replaced by Ser_{149} which is always in the neutral form.

Table 4. Proton transfer from Cys to His in the environment Cys^-/His^+

S...H (Å)	Nϵ...H (Å)	S...Nϵ (Å)	Energy (kcal mol^{-1})
1.467	1.790	2.997	−161390.890
1.567	1.673	3.012	−161391.198
1.667	1.129	2.754	−161391.184
1.767	1.090	2.810	−161394.018
1.867	1.062	2.877	−161394.008
1.967	1.042	2.948	−161393.794

References

1. Warshel A, Levitt M (1976) J Mol Biol 103:227
2. Singh UC, Kollman PA (1986) J Comput Chem 7:718
3. Antonczak S, Monard G, Ruiz-Lopez MF, Rivail JL (1998) J Am Chem Soc 120:8825
4. Harrison MJ, Burton NA, Hillier IH (1997) J Am Chem Soc 119:12285
5. Monard G, Loos M, Thery V, Baka K, Rivail JL (1996) Int J Quantum Chem 58:153
6. Skarzynski T, Moody PCE, Wonacott AJ (1987) J Mol Biol 193:171
7. Brown D, Minoux H, Maigret B (1997) Comput Phys Commun 1178:1
8. Brown D, (1998) The *gmq* User Manual; available at www.lctn.u-nancy.fr/Chercheurs/David.Brown
9. DISCOVER, Molecular Simulations Inc., 9685 Scranton Road, San Diego, CA 92121, USA

10. Stewart JJP (1989) J Comput Chem 10:209
11. Coitino EL, Irving K, Rama J, Iglesias A, Paulino N, Ventura ON (1990) J Mol Struct (Theochem) 210:405
12. Coitino EL, Ventura ON, Sosa RM (1992) J Mol Struct (Theochem) 254:315
13. Kallies B, Mitzner R. (1995) J Mol Model 1:68
14. Rinaldi D, Hoggan PE, Cartier A, Baka K, Monard G, Loos M, Mokrane A, Dillet V, Thery V (to be published)
15. (a) Rinaldi D, Hoggan PE, Cartier A (1989) QCPE 584a; (b) Rinaldi D, Hoggan PE, Cartier A (1989) QCPE Bull 1989:128
16. Talfournier F, Colloc'h N, Mornon JP, Branlant G (1998) Eur J Biochem 252:447