CMLS, Cell. Mol. Life Sci. 54 (1998) 378–382
1420-682X/98/040382-05 $ 1.50 + 0.20/0

CMLS Cellular and Molecular Life Sciences

Catalytic reaction pathways approached by quantum chemistry: a challenge

G. Dive*, D. Dehareng and L. Ghosez

[a]Centre d'Ingénierie des Protéines, Université de Liège, Institut de Chimie B6, Sart Tilman, B-4000 Liège (Belgium), Fax +32 4 366 3364, e-mail: gdive@ulg.ac.be
[b]Laboratoire de Chimie Organique de Synthèse, Université de Louvain, Bâtiment Lavoisier, Place L. Pasteur 1, B-1348 Louvain-la-Neuve (Belgium)

Abstract. This review explores the potential of quantum chemistry to help understand complex biochemical reactions such as enzyme catalysis. Starting from a historical background, the article introduces the reader to the great diversity of problems than can be dealt with in the framework of quantum chemistry.

Key words. Quantum chemistry; reaction pathway; energetic and electronic properties.

Introduction

This presentation is devoted to that part of Jean-Marie Ghuysen's interest in analysing with us how quantum chemistry can provide new insights into the complexity of biochemical processes, in particular, in catalytic reactions.

At the University of Liège, as early as 1969, E. Scoffeniels presented Huckel-type calculations in his course for undergraduate students in pharmaceutical science to describe base pairing in DNA and the K-region of carcinogenic drugs as proposed by Alberte and Bernard Pullman in the 1960s [1]. It was the beginning of a long story involving the development of quantum chemistry methods which can be used to study biochemical and pharmacological molecules. Little by little, this 'quantum biochemistry' became the centre of theoretical chemistry applications involving molecular mechanics and molecular dynamics. Today, this computational chemistry is very often called molecular modelling or, worse and also pretentious, rational drug design.

* Corresponding author.

This article aims to describe the quantum chemistry applications which were initiated in Jean-Marie Ghuysen's laboratory with the help of the first Data General computer installed in 1984, and their development from the study of isolated compounds to the exploration of catalytic reaction paths.

Historical background

Since the original work of Hartree and Fock in the 1930s and the 'linear combination of atomic orbitals' (LCAO) expansion proposed by Roothaan in 1951 [2], numerical calculations began to be tractable in the 1960s. The main problem is the huge number of bi-electronic integrals between two local charge distributions. As illustrated in a lecture by A. Veillard in 1976, substantial computer resources are needed to handle such a large number of integrals, which is the main reason so many methods to simplify the theoretical framework have been developed since 1965 by J. A. Pople [3] and many others. In 1978 Halgren and Lipscomb proposed a new approach to solving equations using approximate solutions to this bi-electronic problem. Their PRDDO [4] method was presented as a realistic compromise between relative errors and computer cost.

Roothaan's formulation solves the Schroedinger equation in terms of LCAO expanded molecular orbitals, and provides energetic and electronic information on the system under study. In the eigenvalue equation $FC = SCE$, the Fock matrix F takes into account all the nucleus-electron and electron-electron interactions. The S matrix results from the overlap between the basis functions, which are linearly combined in the C matrix to generate the orthogonal molecular orbitals. As the C matrix represents the weights of each of the basis functions combined in each of the molecular orbitals which give rise to the energy minimum at a given nuclear geometry, energetic and electronic properties can be analysed at one and the same time. This is a feature of any variational method.

In order to optimize the geometry of a molecule, all the relative atomic coordinates have to be displaced up to an equilibrium position corresponding to a local minimum energy. In many minimization procedures, the first and second derivatives of the function are needed. The analytic solutions of the electronic energy function derivatives were only published in the 1980s, and their algorithms have since been implemented in the computer programs. It was a significant turning point for quantum chemistry: calculations that involved molecules important in pharmacology and organic chemistry suddenly became tractable.

Today, with the parallel computers installed at the Centre for Protein Engineering, an ab initio energy calculation of a molecule containing 90 atoms: 45 atoms of the second period and 45 hydrogens takes 5 min for 300 basis functions. The determination of the first energy derivative needs 10 min and the second one more than 10 h.

Molecular interactions

Before studying the interaction between two molecules, we should consider the isolated entities, since, when the two partners are at infinite distance, examination of their electronic properties provides ideas about the ways they can interact.

Electrostatic potential

After the publication in 1971 by J. Tomasi's group [5] of the first electrostatic potential (EP) calculation on three-membered rings, this electronic property has been recognized as the fingerprint of a molecule. Extending the two-dimensional (2D) drawing of iso-contours, illustrated in the adenine molecule computed by Pullman et al. [6], the three-dimensional (3D) representation became more explicit and widely used to compare molecules belonging to different chemical series. This feature is well illustrated by the 'cache-oreilles' (earmuff) effect [7] discovered on platelet antiaggregation molecules bearing such different moieties as a tetrahydrofuran ring in Merck compounds, a benzofuranoid ring in kadsurenone, a triazolothienobenzodiazepine of Boehringer molecules, and a terpenoid structure built on a spiro-(4,4)-nonane for compounds extracted from *Ginkgo biloba*. On both sides of each compound, two impressive negative clouds appear at a common distance of 12–14 Å.

For large molecules, like enzyme-active sites, accurate EP calculations have often been replaced by rapid, crude drawings based on tabulated net charges which are not even derived from the self-consistent solution of Roothaan's equation. The case of an active-site model of α-chymotrypsin, constituted by 23 amino acids [8]. In contrast, the quantum chemistry calculation points out an impressive negative potential around the nucleophilic serine acting as a magnet towards a ligand. Today, EP can also be used in the estimation of local pK_a values often referred to by the biochemist. In the world of β-lactamases, the best example, which remains unsolved, is Lys 73, whose putative role in catalysis has been postulated [9].

EP is the first term of the interaction between a molecule and a bare proton. This charge induces polarization of the charge distribution [10]. Its contribution can be immediately interpreted in terms of proton affinity and also as an index of reactivity towards electrophilic and nucleophilic affinity. EP alone locates electrostatic attack above the aromatic ring, but the inclusion of the polarization term induces the formation of funnels which drive the reactant to the specific carbon [11]. In the case of aromatic compounds, a nice correlation and a physical meaning can be found by comparison with the inductive and resonance σ indices of Hammet [12].

Electrostatic energy

If the proton is replaced by a molecule, the electrostatic interaction between the two charge distributions becomes the electrostatic energy. It can be calculated following several schemes, such as numerical integration of the product between the charge distribution of one partner and the electrostatic potential of the other [13] or by analytical expression at an approximate level as complete neglect of differential overlap (CNDO) [14].

The very small model of serine peptidase's catalytic artillery involves a methanol interacting with an imidazole via a water molecule. The location of the equilibrium structures can be achieved by calculation of the electrostatic energy, which can then be compared with the result of the full self consistent field (SCF) procedure.

Energy decomposition scheme

In practice, electrostatic energy is the main component of the total interaction energy of complexes like water dimer and several geometrical arrangements of water and formamide. The energy decomposition scheme proposed by Morokuma [15] in the 1970s can help to interpret the relative contributions of the different terms to the total energy. By contrast, with complexes containing strong hydrogen bonds, like the guanine-cytosine pair, this scheme cannot be useful, and a significant residual term appears which is, in fact, the deviation with regard to Morokuma's model [16] implying a simple additive scheme.

Another topic frequently discussed in enzyme-ligand interactions is the $\pi-\pi$ stacking of aromatic amino acids like Phe or even Arg with the added problem of local charge. For the benzene dimer, three geometrical arrangements have been discussed: the T-shape, PD (parallel displaced) stacking and the S (stacked or sandwich) conformation [17]. Neutron diffraction experiments show the existence of PD and T conformations. In the light of the protein data bank (PDB) analysis carried out by B. Maigret over 404 nonredundant structures, it would seem that true S structures are scarce for steric reasons. What can be said is that the stabilization energy experimentally estimated as 1.6 ± 0.2 kcal/mol is due mainly to quadrupole-quadrupole $1/r^5$ interactions [18]. This stabilization effect can only be quantified by the use of large basis sets post-SCF calculations, which are very expensive in computer power.

In an enzyme-active site, the dense network of hydrogen bonds weaves a cobweb whose shape is modulated by the ligand. A model of the α-chymotrypsin enzymatic cleft containing 23 amino acids looks like a pair of tweezers capable of inducing several local conformational changes of the nucleophilic serine in a very small energetic range [19]. Again, Morokuma's scheme was applied to several complexes between the ligand and selected individual residues which induced stabilizing or destabilizing interactions, but the overall interaction cannot be reduced to one highly predominant component.

Short-range interactions of quantum origin (exchange and charge transfer), as well as classical long-range interactions such as the electrostatic and polarization terms, play an equivalent role in the stability of the complex. The optimization of geometry allows calibration of the deformation energy undergone by the ligand and the active site. It is very important to be aware of this deformation in the interpretation of classical 2D conformational analysis. The example of the RX821002, an α-2 antagonist, is a good example and demonstrates the contribution of complete reoptimization of all the 3N-8 degrees of freedom at each step of the grid [20].

Reaction path

The chemical reaction pathway study is a challenge for two main reasons. The first reason, well illustrated by the α-chymotrypsin model, is the extremely dense network of interactions between the amino acid side chains and the ligand. The second reason is the difficulty in locating the transition-state structure on the pathway itself [21]. For such a critical point, the energy is maximum in only one direction defined by one particular linear combination of all the degrees of freedom, all the other directions corresponding to energy minima. The reaction path between minima and transition state (TS) is not univocal and thus has no meaning by itself. The intrinsic reaction coordinate (IRC) definition proposed by Fukui [22] is related to the path of minimum energy, but several pathways could be taken to connect two minima as, for example, the Michaelis and the acylenzyme structures.

The model of the acylation reaction can be represented along two different reaction coordinates, r and r′, followed by the entities described either by molecular mechanics or quantum chemistry energy functions [23]. By definition, the first one can only describe energy minima as the equilibrium parameters of the energy function are built from small constitutive entities. The quantum chemistry view can also locate transition state structures in which covalent bonds can be created or broken.

In 1984, our first theoretical investigation of the C–N bond cleavage that could be catalysed by serine peptidases led to a model in which a water molecule could act as a transient carrier of the hydrogen, going from the nucleophilic serine to the leaving group [24]. For an enzyme active-site model, the main limitation of that type of investigation remains its size. Today, two approaches can be followed. Either the complete set of geometric variables is optimized, keeping the Cα's fixed at their positions in the X-ray structure [25] or a mixed model is applied in which only the active site is described by a quantum chemistry function, the perturbation of the rest of the protein being included by evaluating its energy terms at the molecular mechanics level [26].

This challenge becomes more and more accessible by the development of software tools and more powerful computers. Starting from a very simple model involving the dyad methanol-water as nucleophile and an ester as ligand, the location of equilibrium structures was searched on energy surfaces of increasing complexity by a stepwise introduction of the amino acids surrounding the nucleophilic serine [27].

Intrinsic reactivity

Since the first contract, which received a financial support of the Région Wallonne in the 1980s, efficient collaboration with the organic chemists of Louvain has

allowed the definition of new concepts about the reactivity of potential penicillin-binding protein (PBP) and β-lactamases ligands [28]. A significant contribution to emerge from the synthesis of numerous γ-lactam compounds pointed out the importance of heteroatoms in the cycle which were named 'atomes mous' (jelly atoms), able to absorb the geometric deformation occurring during the opening of the ring [29].

Prospects

By virtue of the complexity of their electronic, energetic and geometric features, biological systems show their contempt for theoreticians.

Intramolecular electronic transfer is likely to become a subject of intense research in the near future. This phenomenon has recently been discovered in organic compounds, generating the concept of mixed valency for radical cations able to 'transfer' one electron from one side to the other [30]. It seems that common features can be found for both saturated and unsaturated compounds. Everyone knows the importance of porphyrin in biological systems. By their pairwise action, they bring to mind the $\pi-\pi$ stacking problem but also this electronic signal transmission.

Another, simpler, example has also been found in α-2 agonist molecules developed by UCB in Belgium. Mivazerol is a benzylimidazole derivative containing three amino acid side chains but no peptide backbone: the phenol moiety bears, on both sides of the OH group, a carboxamide group and a CH_2-imidazole fragment like the substituent of a tyrosine and a histidine. Thus, one could easily imagine that its high interaction with receptors is related to its peptidic feature, while its specificity is linked to the rigidity of its skeleton. For the N imidazole protonated form, the search of equilibrium stuctures leads to several geometric arrangements which can be seen as a topological information transfer (TIT) from one side of the molecule to the other [31]. Moreover, the existence of quinoic structures, characterized by loss of benzenic aromaticity and a fundamental state wave-function instability points up the existence of complex processes such as spin-orbit coupling [32]. The interaction between the receptor and the ligand could also induce electronic modifications in the complex and lead to a signal transmission that could be denoted as electronic information transfer (EIT).

Conclusion

Nowadays, more and more accurate methods of quantum chemistry are being applied to enzyme active site models. The resulting structure, a sort of formal arrangement in which electrons saraband around the nuclei, compels the researcher's admiration.

Acknowledgements. This work was supported by the Belgian Government as part of the Pôles d'Attraction Interniversitaires (PAI no. P4/03).

1 Pullman B. and Pullman A. (1963) Quantum Biochemistry, Interscience, New York
2 Roothaan C. C. J. (1951) A study of two-centre integrals useful in calculations of molecular structure. J. Chem. Phys. **19:** 1445–1458
3 Pople J. A. and Segal G. A. (1965) A self-consistent molecular orbital theory. II. Calculations with complete neglect of differential overlap. J. Chem. Phys. **43:** S136–S151
4 Halgren T. A., Kleier D. A., Hall J. H., Brown L. D. and Lipscomb W. N. (1978) Speed and accuracy in molecular orbital calculations. A comparison of CNDO/2, INDO, STO-3G and other methods, including AAMOM, VRDDO and ESE MO. J. Am. Chem. Soc. **100:** 6595–6608
5 Bonaccorsi R., Scrocco E. and Tomasi J. (1971) Molecular SCF calculations for the ground state of some three-membered ring molecules: *cis* and *trans* diaziridine, oxaziridine, and the corresponding imminium ions. Theoret. Chim. Acta **21:** 17–27
6 Bonaccorsi R., Pullman A., Scrocco E. and Tomasi J. (1972) The molecular electrostatic potentials for the nucleic acid bases: adenine, thymine and cytosine. Theoret. Chim. Acta **24:** 51–60
7 Dive G., Godfroid J. J., Lamotte-Brasseur J., Batt J. P., Heymans F., Dupont L. et al. (1990) 'Cache-oreilles' effect of selected high potency PAF antagonists. J. Lipid Mediat. **1:** 201–215
8 Lamotte-Brasseur J., Dive G. and Ghuysen J. M. (1990) Electrostatic potential maps at the quantum chemistry level of the active sites of the serine peptidases, α-chymotrypsin and subtilisin. J. Theoret. Biol. **145:** 183–198
9 Damblon C., Raquet X., Lian L. Y., Lamotte-Brasseur J., Fonze E., Charlier P. et al. (1996) The catalytic mechanism of β-lactamases: NMR titration of an active-site lysine residue of the TEM-1 enzyme. Proc. Natl. Acad. Sci. USA **93:** 1747–1752
10 Dehareng D., Dive G. and Ghuysen J. M. (1991) Polarization correction to the electrostatic potential at the CNDO and the ab initio level. Influence of the basis set expansion. Theoret. Chim. Acta **79:** 141–152
11 Dive G. and Dehareng D. (1993) Polarization correction of the electrostatic potential for aromatic compounds: study of the nucleophilic attack. Int. J. Quant. Chem. **46:** 127–136
12 Dehareng D., Dive G. and Ghuysen J. M. (1993) Polarization-corrected electrostatic potentials of aromatic compounds. J. Am. Chem. Soc. **115:** 6877–6882
13 Dehareng D., Dive G., Lamotte-Brasseur J. and Ghuysen J. M. (1989) Numerical computation of the electrostatic interaction energy between methanol and the dyad water-imidazole. Theoret. Chim. Acta **76:** 85–94
14 Dehareng D., Dive G. and Ghuysen J. M. (1993) Analytical calculation of the electrostatic interaction energy within the CNDO framework. Int. J. Quant. Chem. **46:** 711–734
15 Kitaura K. and Morokuma K. (1976) A new energy decomposition scheme for molecular interactions within the Hartree-Fock approximation. Int. J. Quant. Chem. **10:** 325–340
16 Dive G., Dehareng D. and Ghuysen J. M. (1993) Energy analysis on small to medium-sized H-bonded complexes. Theoret. Chim. Acta **85:** 409–421
17 Hobza P., Selzle H. L. and Schlag E. W. (1994) Structure and properties of benzene-containing molecular clusters: nonempirical and ab initio calculations and experiments. Chem. Rev. **94:** 1767–1785
18 Chipot C., Jaffe R., Maigret B., Pearlman D. A. and Kollman P. A. (1996) Benzene dimer: a good model for $\pi-\pi$ interactions in proteins? A comparison between the benzene and the toluene dimers in the gas phase and in an aqueous solution. J. Am. Chem. Soc. **118:** 11217–11224
19 Dive G., Dehareng D. and Ghuysen J. M. (1994) A detailed study of a molecule into a molecule: the *N*-acetyl-L-tryptophanamide in an active site model of the α-chymotrypsin. J. Am. Chem. Soc. **116:** 2548–2556

20 Rondu F., Le Bihan G., Wang X., Lamouri A., Touboul E. G., Dive et al. (1997) Design and synthesis of imidazoline derivatives active on glucose homeostasis in a rat model of type II diabetes. 1. *N*-benzyl,*N*′-alkyl,2-(4′5′-dihydro-1′*H*-imidazol-2′yl)piperaz ines. J. Med. Chem. **40:** 3793–3803
21 Culot P., Dive G., Nguyen V. H. and Ghuysen J. M. (1992) A quasi-Newton algorithm for first-order saddle-point location. Theoret. Chim. Acta **82:** 189–205
22 Fukui K. (1970) A formulation of the reaction coordinate. J. Phys. Chem. **74:** 4161–4163
23 Ghuysen J. M. and Dive G. (1994) Biochemistry of the penicilloyl serine transferases. In: Bacterial Cell Wall, pp. 103–129, Ghuysen J. M. and Hakenbeck R. (eds), Elsevier, Amsterdam
24 Dive G., Peeters D., Leroy G. and Ghuysen J. M. (1984) Theoretical study of the C–N bond breakage catalysed by the serine peptidases. J. Mol. Struct. (THEOCHEM) **107:** 117–126
25 Wladkowski B. D., Chenoweth S. A., Sanders J. N., Krauss M. and Stevens W. J. (1997) Acylation of β-lactams by class A b-lactamase: an ab initio theoretical study on the effects of the oxyanion hole. J. Am. Chem. Soc. **119:** 6423–6431
26 Monard G., Loos M., Théry V., Baka K. and Rivail J. L. (1996) Hybrid classical quantum force field for modeling very large molecules. Int. J. Quant. Chem. **58:** 153–159
27 Dive G., Dehareng D. and Peeters D. (1996) Proposition for the acylation mechanism of the serine proteases: a one-step process? Int. J. Quant. Chem. **58:** 85–107
28 Ghosez L., Dive G., Dumas S., Génicot C., Kumli F., Love C. et al. (1993) Design and synthesis of D,D peptidase inhibitors. In: Antibiotics and Antiviral Compounds: Chemical Synthesis and Modification, pp. 177–186, Krohn K., Kirst H., Mass H. (eds), VCH, Weinheim
29 Marchand-Brynaert J., Couplet B., Dive G. and Ghosez L. (1993) New g-lactam homologs of penems. Bioorg. Med. Chem. Lett. **3:** 2303–2308
30 Lahlil K., Moradpour A., Bowlas C., Menou F., Cassoux P., Bonvoisin J. et al. (1995) Intervalence transitions in mixed valence bis(tetrathiafulvalene) compounds. J. Am. Chem. Soc. **117:** 9995–10002
31 Vancampenhout N., Dive G. and Dehareng D. (1996) Energetic and conformational study of four benzylimidazole compounds with α-2 agonist profile: the Mivazerol and three methylated derivatives. Int. J. Quant. Chem. **60:** 911–930
32 Remacle F., Petitjean S., Dehareng D. and Lorquet J. C. (1987) An ab initio study of the isomerization and fragmentation of the CHO_2^+ ions: an example of spin-controlled chemistry? Int. J. Mass Spectrom. Ion Process. **77:** 187–201