Towards a Generic Model of Catalysis

Martin Grayson,[1] Stefan J. Janusz[1]

[1] Department of Chemistry, the University of Shefield, Sheffield S3 7HF.

Internet Electronic Conference of Molecular Design **2003**, November 23 – December 6

Abstract

A *generic* method of modelling catalysis using fundamental experimental bond based data is described. An attempt has been made to put forward this model, to be used in conjunction with visualization tools, which uses tried and trusted chemical concepts without direct calculation, though the capabilities of other precise calculations at reactant and transition state geometries are discussed. Related to the electrically distorted bond model however is the *geometrical* distortion which forms the basis of the entasis effect. The current state of modelling entasis is reviewed. However model calculations of the *electrical* strain induced in a complex at the transition state are presented. Studies of entasis have concentrated on the energetics of geometrical distortion but by considering polarizabiities and hardness / softness parameters one can see how local polarizations of the electron density may also be responsible for activation of a localised area of a large molecule.

Keywords. Geometrical strain, environment strain, entasis, polarizability, hardness and softness.

Introduction

The methodology presented here is simple, almost to the extreme, but contains within it tried and trusted experimental data, which in itself contains the quantum mechanical values by the inclusion of polarizabilities, and some values from density functional theory by the implicit use of electronegativity and hardness / softness parameters.

The basic idea is that to assist the breaking of a bond you must create a field in the catalytic transition state which opposes the bond dipole of that bond leading to a local increase in energy. (If a single molecule were to be fixed in a homogeneous electric field the energy induced in the molecule by that field would not be evenly distributed. Apolar bonds would interact only at second order in electric field *via* the polarizability wheras polar bonds would be enhanced in strength or have energy *pumped* into them by the field.)

It is apparant in this model that a high polarizability will allow a bond to become more reactive regardless of the direction of local fields it experiences. This also accords with the criteria for a *good leaving group*, which requires a high polarizability and high value of the hardness / softness parameter *b*. However we have to consider the balance of polarization energies across the whole reaction profile. It is likely that the polarization energy of reactants and products is similar. However for the reactive bonds polarization will increase at the transition state as the bondlength is stretched or alternatively more excited states are mixed in a quantum mechanical sum over states picture. The effect of polarizability is to *lower* the transition state. A desciption of catalysis using essentially a sum over states and perturbation theory methodology has been expounded by Tchougréeff [1]. Though this kind of approach is potentially exact it cannot be generic, because so much complexity is folded into the *reaction coordinate* which describes the trajectory over the potential surface. A great deal of information about the energetics of the system, if not the whole reaction must be known before one knows this trajectory and so the solution is to some extent tautological.

Inorganic Biological Systems

A particularly interesting area of chemical catalysis which is largely unexplored by computational chemistry is the study of inorganic biomolecules. That inorganic elements are ubiquitous in living things is often overlooked, but they are actually instrumental in many vital processes [2]. Light-harvesting chlorophyll contains magnesium. The iron in haemoglobin carries a dioxygen ligand through mammalian blood. Copper-containing haemocyanin does the same thing in crustaceans. Ca^{2+} also has many roles in the soft tissues as well as the skeletons of living organisms.
Metals and their ions are also crucial in many *catalytic* processes, at the active site of enzymes. Because of the manner in which certain amino acid residues along the polypeptide ligate to the metal, it was originally thought that the metal's role was to provide a geometrical template to which the polypeptide could conform. However, crystallographic data have revealed that the geometry at metal centres in many metalloenzymes is 'non-standard'. Since 'standard' geometries are optimized for the best electronic interaction (and subsequent energy minimization through mutual stabilization) between the metal and its ligands it follows that 'non-standard' geometries imply that the system is strained. At the metal centre at least, the enzyme is in a non-equilibrium state. In these so-called 'entatic' complexes (from the Greek entasis, meaning strain), the tertiary structure of the polypeptide influences the geometry at the metal centre, *rather than* (or perhaps, *as well as*) the metal influencing the conformation of the protein. Geometrical effects can compress or stretch bonds, *entasis*, but electrical effects can also weaken bonds by moving electron density away from the bonding region.

The Entatic State and Catalysis

Since the 1950s, it has been proposed that the entatic state of enzymes might account for their tremendous power as catalysts [3]. Any system in a configuration that lies away from that at equilibrium must have an energy greater than the equilibrium, so it possible that certain catalytic pathways could be opened up or made more favourable by an intrinsic strain. Many non-enzymatic catalysts involving d-block metals require certain geometric or electronic strains to be formed at some intermediate stage of the catalysis. An example is the Ring-Opening Metathesis Polymerization (ROMP) reaction [4], whereby a thermodynamically strained hydrocarbon chelate is briefly formed, causing the complex to fall apart. The strain gives the system an incentive to exit the intermediate chelate stage and so propagate the catalysis. Other systems, which attempt to mimic the strains inside enzymes with working 'small molecule' models (i.e. non-enzymatic), have been developed and investigated [5].

The above complexes are several orders of magnitude more proficient catalysts than anything seen in the traditional realm of d-block coordination complexes. Enzymes can catalyze chemical reactions whose natural (*i.e.* non-catalyzed) half-lives are of the order of the age of the universe, and effect these changes in time frames of less than a second [6]. A major difference between enzymes and 'traditional' (that is, artificial; non-biological) catalysts is that the protein component imparts a non-equilibrium geometry at the active site, and on the coordination sphere of the metal it contains. Only very exotic and unstable artificial complexes display anything approaching this amount of strain (note that enzymes, though technically quite strained, are also remarkably stable). So it seems plausible that the novel, 'strained' geometries seen at the active sites in metalloenzymes could be a factor in producing this remarkable catalytic proficiency.

Investigation of these systems with computer codes is desirable because attempts to synthesize working models of the active sites of enzymes have usually failed. The metal centres in these small molecule mimics, lacking the 'rigid' backbone of a polypeptide, tend to adopt the standard equilibrium (*i.e*. non-entatic) geometries described earlier [7]. This can be rationalized in several ways. Most likely is that it is far more energetically favourable for the ligands in an artificial catalyst to rearrange to positions in the coordination sphere that maximize overlap with the *d* orbitals of the metal, so promoting a more thermodynamically stable structure. In the case of the enzyme, optimization of the geometry at the metal centre is a consideration of far less importance than the energetically unfavourable rearrangement of the bulk of the protein that such an action would necessitate. Generally, a protein's conformation is one that maximizes intramolecular interactions between domains within the structure. The number of thermodynamically favourable, so-called 'weak interactions', vastly outweigh the disfavoured mis-fitting at the metal centre, and it is conceivable that the folding of the protein is used to pay for the raising of the potential energy of the active site.
The active site in an enzyme can be thought of as existing in isolation. The polypeptide effectively buffers it from the external environment, and so it can exist in a strained state. In a calculation the model molecule is in isolation, unless the program is explicitly told to model the system as though there is a solvent, or an array of similar molecules, surrounding it. While real small molecule models of active sites, through energetic exchanges with their environment, break down - virtual models, effectively in isolation, do not. Computational theoretical work therefore offers a good way to study the entatic state.

A Brief Review of Existing Work on Entasis

There exists a wealth of literature that has relevance to this problem, that all attack it from a unique perspective. Chemists, biochemists, biologists and biophysicists all have their own particular views, many contradictory to others; it seems that the contention of the matter has inspired much work, but papers offering a real insight into the phenomenon are rare because a simple solution is so elusive.

An early attempt to rationalize the problem was outlined in a paper by Vallee and Williams (1968)[7]. It was here that the term 'entatic' was first invoked to describe an enzymatic system "in a stretched state or under tension... [implying] a catalytically poised state *intrinsic to the active site*". When bonds are stretched they are more polarizable and therefore more reactive. Because of the crudity of crystallographic analysis at the time, the nature of the active site was not known directly, but inferred from spectroscopic data. Metalloenzymes were deemed "exceptionally well suited for the examination of the physicochemical basis of [enzymes]...", because the physical (i.e. spectroscopic) properties of metals, "...constitute *intrinsic* probes." It was understood from the analyses of these data, that the natures of the metals in their coordination spheres were quite unlike those seen in model complexes (*i.e. d*-block coordination compounds). What was proposed, in light of the known proficiency of enzymes as catalysts, was that the active site has a configuration "...closer to that of a unimolecular transition state than to that of a conventional, stable molecule, thereby constituting an energetically poised domain."
Interest in a correlation between the structure and function of enzymes can be dated back to 1894 when Emil Fischer introduced his now widely-known 'lock and key' analogy. It was known then that enzymes catalyze reactions very well and are very specific to one particular substrate (or type of substrate), but due to the infancy of the field a

rationalization for the enzyme-substrate specificity in terms of binding could not be given. In 1930, J.B.S. Haldane proposed his theory of enzyme/transition state complimentarity to explain the kinetics of enzyme catalysis [8]. This was later expanded upon and popularized by Linus Pauling in two papers, written in 1946 and 1948 [9]. Here it was suggested that the active site is uniquely suited to substrate transition-state binding, and that a ground state substrate, on approaching the active site, "will be subjected to an environment that forces it into the transition state, thereby effecting catalysis."

In computational terms, much of the work on metalloenzymes has been tentative. New techniques are required for solving the problem of the QM/MM boundary, (how to mix quantum mechanics and molecular mechanics, as mixed QM/MM must be used to calculate properties of the active site and polypeptide, respectively) [10]. There have been some advances in terms of modelling proton tunneling in enzymes [11]. Fariselli *et al.* [12] have modelled the electron correlation in haemocyanin. Most of the work in the field is of this nature: a thorough study of one particular system, and often of one particular phenomenon to which that system gives rise. Much of the time, answers have already been determined experimentally, so that theoretical work, especially that of a computational nature, often confirms facts rather than predicting new data.

The geometrical strain in the activated complex is best modelled by quantum mechanical calculation or QM/MM methods. However when thinking about the characteristics required of an active site, and as a back of the envelope thinking estimation, suitable for display in a molecular graphics system, bond polarization considerations can be a help. There is suggested a model with some predictive power, extrapolated from the nature of the bonds in the active site. The general idea is this: that a chemical bond's strength is a function of its bond dipole, and by applying an electrostatic field, particularly, in opposition to the direction of that dipole, the bond is weakened. The model draws upon a dataset of bond based polarizabilities [13] to determine the likelihood of bond breaking (the more polar or polarizable the 'easier' energetically that it should be). The stability of the leaving group (more polarizable species tend to be better leaving groups) is also partially predicted by this model. The charge transfer component of the polarizability can also be influenced in the course of the reaction by the Pearson hardness factor b. In particular how much greater negative charge and how much increase in polarizability an atom or functional group undergoes is determined by the value of *b*. The value of *a*, the traditional electronegativity determines the unperturbed bond dipole.

2 MATERIALS AND METHODS

Electrical Strain in a Simple Model

We will now make a digression to looking at one of the simplest reaction systems: how the methyl halides undergo S_N2 hydrolysis and we will consider how the CH_3-X bond becomes broken.

From the modelling of protein NMR it is known that the sort of fields experienced in the equilibrium environment are on average 0.006 au rising to 0.008 au for the more perturbed atoms [14,15] (1 au = 5.14220×10^{11} V m^{-1}). It is clear that the model here requires a field

of about a half to one orders of magnitude greater than these equilibrium fields to break a bond. This would seem to be logistically sensible that the catalytic site must be capable of generating a change of electrical environment one order of magnitude greater than the *normal* perturbed environment. Therefore fields of 0.04 and 0.08 aus have been used in the model calculations. (Even greater fields than these might be possible in zeolites [16].)

It is hoped that this model can be combined with a model of solvation effects in enzyme reactions, such as described by Warshel [17]. Since these reactions take place in aqueous conditions, the Sheffield solvation model, recently developed by the Pickup group specifically for high throughput computations, is potentially useful [18].

3 Model Calculations

Table 1 – Contributions to the Induced Energy at Applied Electric Fields 0.04 and 0.08 au

Bond	Energy from μ /kJ mol^{-1}	Energy from α /kJ mol^{-1}	Total energy /kJ mol^{-1}	Bond energy /kJ mol^{-1}
C-H(0.04)	53.72	8.10	61.81	414.22
C-H(0.08)	107.43	33.76	139.82	414.22
C-F	187.60	8.49	196.08	485.34
C-F	375.19	33.94	409.13	485.34
C-Cl	202.47	29.48	254.49	338.90
C-Cl	404.94	208.08	613.02	338.90
C-Br	195.86	43.66	261.77	284.51
C-Br	391.72	263.64	655.36	284.51
C-I	172.31	66.05	268.83	217.57
C-I	344.62	386.11	730.72	217.57

Table 2 – Bond Dipole Moments x 10^{30} / Cm

carbon sp^3	hydrogen	-1.300
carbon sp^3	fluorine	4.540
carbon sp^3	chlorine	4.900
carbon sp^3	bromine	4.740
carbon sp^3	iodine	4.170

Table **3** – Polarizability Ellipsoids x 10^{41} / $C^2m^2J^{-1}$

		parallel	perpendicular
carbon sp^3	hydrogen	6.356	6.356
carbon sp^3	fluorine	6.661	6.661
carbon sp^3	chlorine	40.834	23.143
carbon sp^3	bromine	51.738	34.270
carbon sp^3	iodine	75.772	51.850

In the calculations in **Table 1** an electric field is placed along the C-X bond in an energetically unfavourable direction with the bond dipole. The first column is the 1st order interaction with the dipole, then the 2nd order with the polarizability tensor elements with the correct symmetry *i.e.* the parallel diagonal element. The total is compared with the bond energy. The bond dipoles and polarizabilities are taken from references [19] and [20] and tabulated in **Tables 2** and **3**. For the smaller field 0.04 the energy enhancement only exceeds the bond energy for the reactive C-I bond. However for the larger field 0.08 once we are passed the unreactive C-H and C-F bonds all the bonds can be readily broken. In this limited case the simple model predicts chemical commonsense.

It has been hypothesised that when the polarizability of a system increases due to geometrical distortion more of the polarizability increase goes with the anion than with the rest of the molecule. This seems like common sense and the following *ab initio* calculation of distributed polarizabilities [21] as a function of bond length **(Table 4)** shows this is indeed the case for the alkyl halides at least.

Table 4 – Distributed Polarizabilities and Gradients for Alkyl Halides

	Pol / $C^2m^2J^{-1}$			Gradients in $C^2m^2J^{-1}$ per Ångstrom		
	C	Hal	H	C	Hal	H
CH_3F	11.1185	6.8111	3.4750	4.8672	6.6776	-0.9843
	(¶) 39.21	24.02	12.26	(§) 14.51	7.09	4.95
CH_3Cl	10.4130	27.1259	3.6153	5.1063	12.5835	0.1467
	21.52	56.06	7.47	10.55	17.31	2.31
CH_3Br	8.9095	14.5604	2.7945	5.9372	7.8251	0.7007
	27.97	45.71	8.77	4.78	8.90	0.16
CH_3I	9.2996	22.6666	3.0880	6.6231	8.2934	0.9893
	22.56	54.98	7.49	5.11	0.90	0.00

(¶) - The following 3 numbers are the percentage of the total molecular polarizability.

(§) - The following 3 numbers are the 2^{nd} gradients of the polarizability.

The calculations are inconsistent in that it is not possible to have a uniform quality of basis set whilst going down a column of the periodic table from F to I. CH_3F and CH_3Cl are calculated using Sadlej's medium polarized basis set[22]. CH_3Br and CH_3I use only 3-21G[23]. However we see the polarizability gradient is always largest for the halogen and is more or less constant for the carbon in CH_3F and CH_3Cl. If we had a better Hartree-Fock basis for CH_3Br and CH_3I all gradients would presumably be larger and the carbon begins to *own* more of the polarizability because it has some of the iodine's polarization in its partitioning polyhedra because the iodine is so much larger and more polarizable. The integration over the Voronoi polyhedra is not an exact partitioning whatever *exact* means in this context. These simple calculations confirm what might be expected. As the bond which creates the leaving group is stretched, the leaving group becomes more polarizable relative to the rest of the molecule. At a transition state the whole complex has become more polarizable, (only one gradient is negative, that corresponding to the unreactive C-H bond).

It can be noticed that chlorine has both a large 1st and 2nd gradient of the polarizability. This is analagous to the known behaviour where in some reactions chlorine behaves as though it were almost as electronegative as fluorine dues to its much smaller value of the hardness parameter *b*. This is particularly apparant if one uses the more recent values of *a* and *b* from Politzer *et al.* (Table 5) rather than the earlier values which are in many textbooks.

Table 5 – Values of Electronegativity Parameters *a* and *b*

Units of Volts / electron		
	a	*b*
fluorine	12.18	17.36
chlorine	9.38	11.30
bromine	8.40	9.40
iodine	8.10	9.15
Data from Huheey [19]		
fluorine	10.41	14.03
chlorine	8.29	9.35
bromine	7.59	8.48
iodine	6.76	7.41
neon	10.78	21.55
argon	7.88	15.75
krypton	7.00	14.01
xenon	6.07	12.14
Data from Politzer *et al.* [24]		

4 CONCLUSIONS

We have suggested a model which may be useful in visualising catalysis but is also numerical and it does not treat every molecular system as a unique case where any properties would have to be calculated from first principles. The ubiquity of additive and transportable concepts in chemistry such as bond energies and infrared vibration tables seems to indicate that properties of functional groups and so some extent bonds and atoms, are translatable from molecule to molecule. In a sense, a molecule behaves like a composite structure. This is precisely the success of philosophy behind the force-field approach to molecular mechanics, which acknowledges the fact that quantum mechanical systems, at some basic level, behave like classical ones. The cases where this is not applicable are what we would refer to as *interesting* chemistry. We hope we have added another idea to the tools available for rationalizing reactions.

Acknowledgment

We would like to thank the following people for their help and support during the project which lead to this communication: Prof. David E. Fenton, Dr. Nicholas H. Williams and Prof. Barry T. Pickup. We would also like to thank the developers of the CADPAC program for making the distributed polarizability computations possible. We wish to acknowledge the use of the EPSRC's Columbus Service at the Rutherford Laboratory.

5 REFERENCES

[1] A. L. Tchougréeff, Quantum mechanical models for organometallic reactivity*, Intl. J. Quantum Chem.*, **1996** *58*, 67-84.
[2] D. E. Fenton, Biocoordination Chemistry, **1995** Oxford University Press.
[3] R. J. P. Williams, Energized (entatic) states of groups and of secondary structures in proteins and metalloproteins *, Eur. J. Biochem,* **1995** *234* 363-381.
[4] T. M. Trnka and R. H. Grubbs, The development of L2X2Ru = CHR olefin metathesis catalysts: An organometallic success story, *Acc. Chem. Res.*, **2001** *34*, 18-29. http://www.ilpi.com/organomet/romp.html.
[5] N. H. Williams, and P. Wyman, Phosphate diester hydrolysis within a highly reactive dinuclear cobalt(III) complex. Ligand effect on reactivity, transition state and dissociation*, J. Chem. Soc., Perkin Trans*. 2, **2001** 2068-2073.
[6] R. Wolfenden, and M. J. Snider, The depth of chemical time and the power of enzymes as catalysts*, Acc. Chem. Res.*, **2001** *34*, 938-945.
[7] B. L. Vallee, and R.J.P. Williams, *Proc. Natl. Acad. Sci. USA*, **1968** *59* 498.
[8] J. B. S. Haldane, Enzymes, **1930** London. Green & Co.
[9] L. Pauling, *Chem. Eng. News*, **1946** *24*, 1375-1377; L. Pauling, *Nature*, **1948** *161*, 707-714.

[10] J. Gao, P. Amara, C. Alhambra, and M. J. Field, A generalized hybrid orbital (GHO) method for the treatment of boundary atoms in combined QM/MM calculations**,** *J. Phys. Chem. A*, **1998**, *102* , 4714 -4721.
[11] D. G. Truhlar, J. Gao, C. Alhambra, M. Garcia-Viloca, J. Corchado, M. Luz Sanchez, M. and J. Villa, The incorporation of quantum effects in enzyme kinetics modeling, *Acc. Chem. Res.*, **2002** *35*, 341-349.
[12] P. Fariselli, A. Bottoni, F. Bernadi and R. Casadio, Quantum mechanical analysis of oxygenated and deoxygenated states of hemocyanin: Theoretical clues for a plausible allosteric model of oxygen binding, *Protein Sci.* **1999** *8* 1546-1550.
[13] M. Grayson, *The Estimation of Molecular Polarizability and Magnetizability*, ed. S.G. Pandalai, *Recent Research Developments in Physical Chemistry*, Transworld Research Network, Trivandrum, India, **2002**, *6*, 437-455.
[14] W. J. Horsley and H. Sternlicht H. *J. Amer. Chem. Soc.*, **1968** *90*, 3738.
[15] J. Augspurger, J. G. Pearson, E. Oldfield, C. E. Dykstra, K. Deok Park and D. Schwartz, Chemical shift ranges in proteins, *J. Magn. Reson.*, **1992**, *100*, 342-357.
[16] R. Z. Khaliullin, A. T. Bell, V. B. Kazansky, An experimental and density functional theory study of the interactions of CH_4 with H-ZSM-5 *J. Phys. Chem. A*, **2001** *105* 10454-10461.
[17] A. Warshel, Electrostatic basis of structure-function correlation in proteins *, Acc. Chem. Res.*, **1981** *14* 284-290.
[18] M.J. Sykes, B.T. Pickup, J.A. Grant, C.A. Kitchen, and A. Nicholls, Approximate Solvation models, Presented at the 4th Openeye Scientific Software users meeting in Sante Fe, New Mexico, Feb 2003: www.eyesopen.com.
[19] J. E. Huheey, *Inorganic Chemistry*, 3rd edition, **1983** Harper International.
[20] M. Grayson, Calculating and Estimating Polarizabilities, in press: *Int. J. Mol. Sci.*, **2003**.
[21] A. J. Stone 1996, *The Theory of Intermolecular Forces*, **1996** Clarendon, Oxford.
[22] A. J. Sadlej., Medium-size Polarized Basis Sets for High-level Correlated Calculations of Molecular Electric Properties, *Collect. Czech. Chem. Commun.*, **1998** *53* 1995-2016.
[23] J. B. Foresman and A. Frisch, *Exploring Chemistry with Electronic Structure Methods, Second edition, Gaussian, Inc., Pittsburgh, PA, 15106 USA,* **1993.**
[24] P. Politzer, J. S. Murray and M. E. Grice*, Structure and Bonding, 80, Chemical Hardness,* **1993** 101.

Biographies

Martin Grayson is Senior Programmer in the Chemistry Department at the University of Sheffield. He has published 30 papers, mostly in computational quantum mechanics.

Steafan J. Janucz researched a 4th year undergraduate project on catalysis with Dr. Grayson and is now in the 1st year of a PhD on nano-technology at Sheffield.