

A new method for locating stationary points in enzymatic catalysis

Xavier PRAT-RESINA¹, Gérald MONARD², Mireia GARCIA-VILOCA¹, Josep Maria BOFILL³, Àngels GONZÁLEZ-LAFONT¹ and José Maria LLUCH¹

¹Departament de Química, Universitat Autònoma de Barcelona
08193 Bellaterra (Barcelona) - SPAIN

²Université Henri Poincaré - Nancy I, UMR 7565 - Groupe Chimie théorique
Faculté des Sciences - B.P. 239, F54506 Vandoeuvre-lès-Nancy - FRANCE

³Departament de Química Orgànica and Centre Especial de Recerca en Química Teòrica
Universitat de Barcelona, Martí i Franquès 1, 08028 Barcelona - SPAIN

Abstract

QM/MM methods have been widely used during these last years for the study of very large reactive systems. The elucidation of enzymatic reactivity is one of their main successful applications, and this study can not be carried out without an accurate location of minima and transition states.

The methods for stationary point search in reactive systems of few atoms have been deeply studied and validated. However those methods for big size systems like enzymes have not yet a general strategy or a standard protocol of application. The main problem lies on the high computational efforts required when effective methods such as Newton-Raphson are used. The manipulation of a Hessian, *i.e.* its storage and diagonalization, implies too much memory and CPU time consuming when a system of thousands of atoms are moved. Then a suitable method should be effective enough to avoid too many steps during optimization on the QM/MM surface but without any manipulation of a very big Hessian matrix.

In this poster we present the development of a new useful method for the search of minima and mainly of transition states in a high dimensioned QM/MM surface. The original feature is, from one part, to avoid the storage of the full Hessian taking advantage of its sparsity. And from the other part, during the optimization process we prevent the full diagonalization process thanks to the combination of the Rational Function Optimization method with an iterative Lanczos-type diagonalization of the corresponding Augmented Hessian. Some examples for real systems are presented and discussed.

Standard second order methods

•Newton Raphson

$$\Delta q_k = -B_k^{-1} g_k$$

B : Hessian
 Δq : displacement
 g : gradient

•Rational Function Optimization

Augmented Hessian (AH): $\begin{pmatrix} 0 & \mathbf{g}_k^T \\ \mathbf{g}_k & \mathbf{B}_k \end{pmatrix}$ \mathbf{v} : eigenvector of AH

$$\Delta q_k = \frac{1}{\mathbf{v}_{1,v}^{(k)}} \mathbf{v}_v^{(k)} \begin{cases} v=1 \text{ for minima} \\ v=2 \text{ for transition states} \end{cases}$$

$$\mathbf{v}_v^{(k)} = (\mathbf{v}_{2,v}^{(k)}, \dots, \mathbf{v}_{n+1,v}^{(k)})$$

Optimization in reactive enzymatic systems:

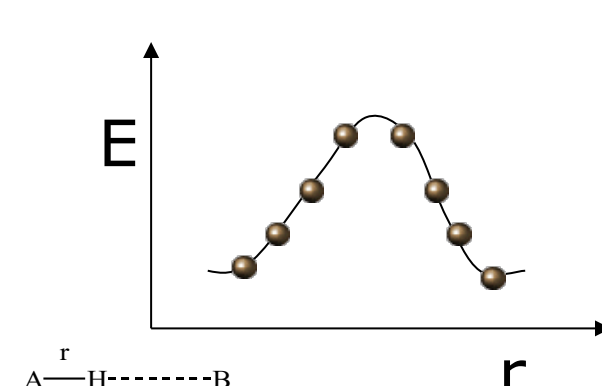
- Thousands of atoms have to be moved
- Process takes a lot of steps to converge
- When energy evaluation is not cheap (e.g. QM/MM) the algorithm must be effective: we need 2nd derivative methods \Rightarrow the manipulation of a Hessian is needed

Drawbacks when size increases

- Calculation of initial Hessian matrix $O(N^2)$
Hessian has to be good enough, this is crucial when we look for TS
- Diagonalization scales $O(N^3)$
A fast process in low-dimension systems can be a problem when dealing with thousands of degrees of freedom
- Storage $O(N^2)$
The storage of the Hessian, the eigenvectors and the update requires a lot of computer memory

Solutions proposed up to now

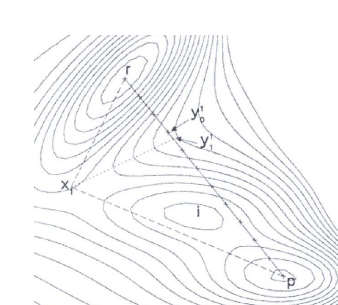
methods



Coordinate scan

Easy to perform

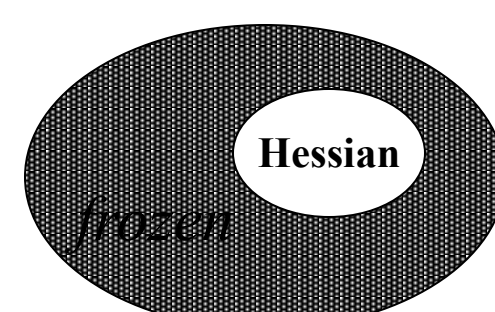
Not always so intuitive
TS may not exist



Conjugate Peak refinement^a

Direct location without
Hessian manipulation

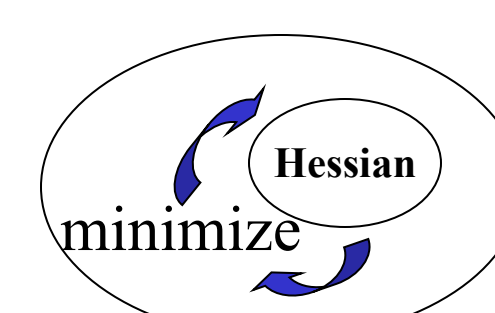
There is no transition
vector



Hessian for a core
(the environment frozen)

2nd order direct location

The environment is
never relaxed



Iterative
core/environment^b

2nd order direct location.
Permits the environment
to relax

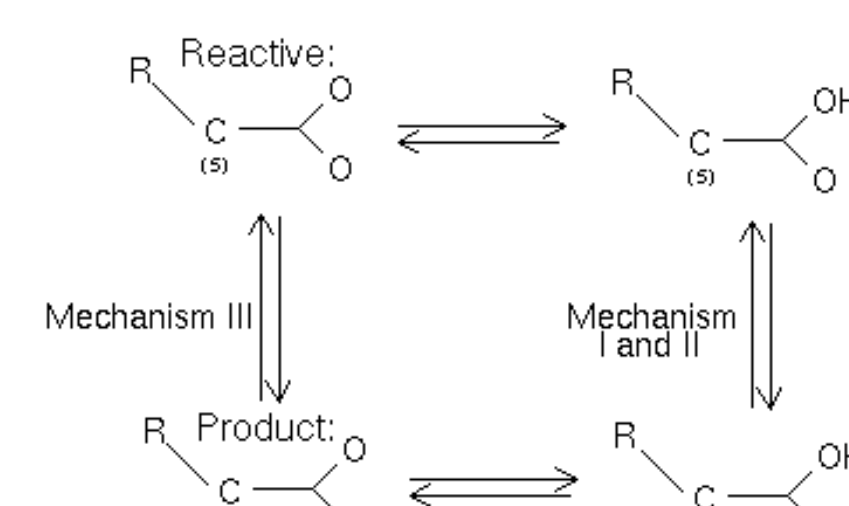
If core is not big there
will be coupling
between the two zones

a) S. Fischer, M. Karplus, Chem. Phys. Lett., 194, 3, 252 (1992)

b) A. J. Turner, V. Moliner, I. H. Williams; Phys Chem Chem Phys, 1, 1323 (1999); S.R. Billeter, A.J. Turner, W. Thiel, Phys.Chem.Chem.Phys., 2,2117 (2000); Y. Zhang, H. Liu, W. Yang, J. Chem. Phys. 112, 3483 (2000)

Application to enzyme reactivity

We have used the iterative core/environment algorithm to elucidate the reaction mechanism of mandelate racemase (MR)^a. Our QM/MM model is constituted of 3963 atoms, 81 of them are treated quantum mechanically with PM3, and the rest with AMBER forcefield. Three substrates have been studied^b: Mandelate (R:phenyl), Propargyl glycolate (R:ethynyl) and Vinyl glycolate (R:ethenyl)

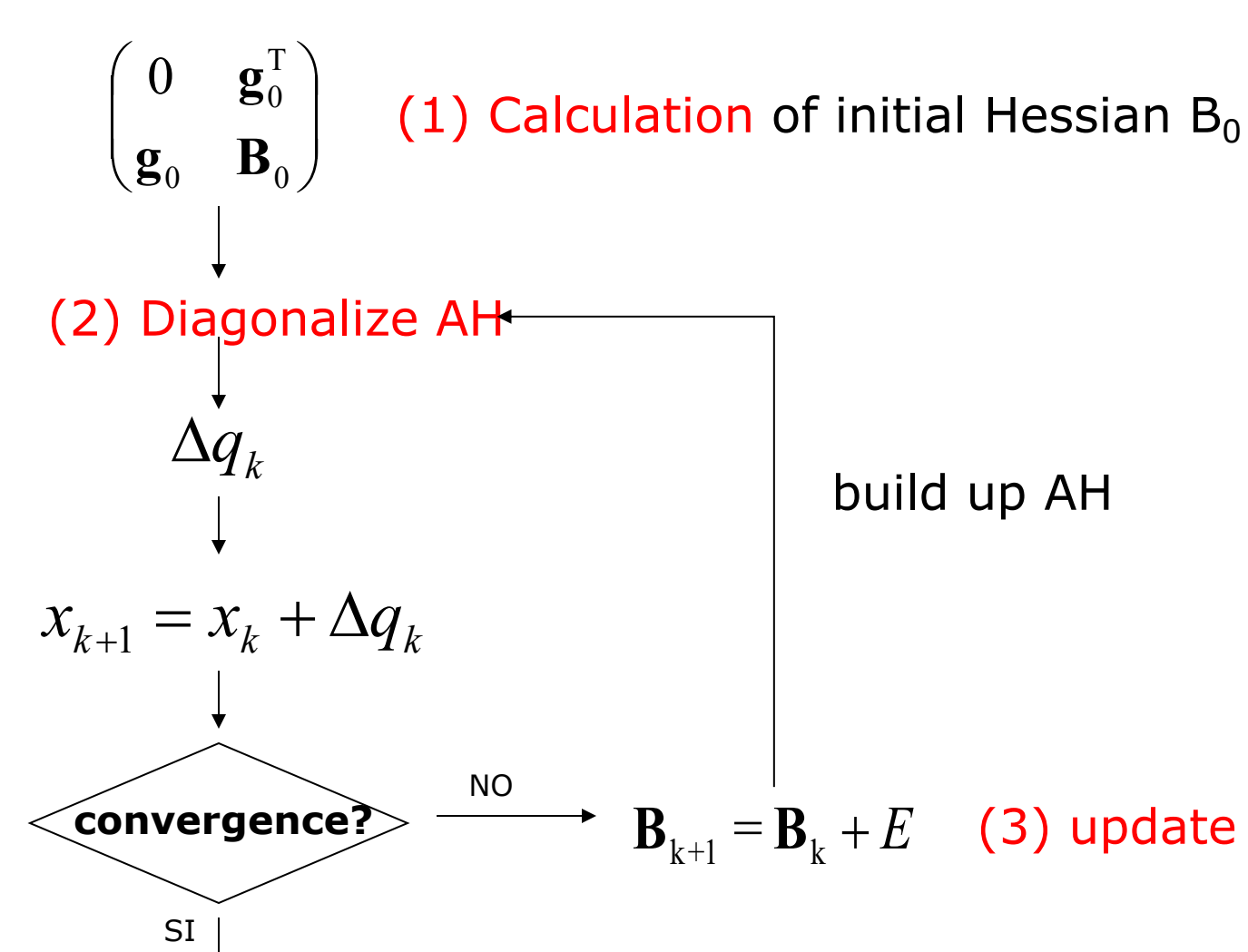


When applying the core/environment algorithm we have found coupling between the two zones when the convergence criteria for optimization is tight. This is solved by increasing the core size but then the problems mentioned above arise.

a) to be published

b) M. Garcia-Viloca, A. González-Lafont, J.M. Lluch, J.Am.Chem.Soc.,123,709 (2001); X. Prat-Resina, M. Garcia-Viloca, A. González-Lafont, J.M. Lluch, Phys.Chem.Chem.Phys, submitted

Possible solutions to the problems



1. Approximated initial Hessian calculation

Second derivatives calculation can be accelerated in QM/MM systems:

- Only 1 SCF cycle for environment atoms
- Pure Molecular Mechanics potential

2. A fast diagonalization

RFO permits the usage of iterative diagonalization process (Lanczos) since we need to extract only few eigenpairs.

3. Avoid Hessian storage combined with a fast diagonalization process^a.

Starting from an approximated initial Hessian, the full Hessian is never build up. Only gradient and displacement are stored for update, and a diagonalization algorithm^b that only needs matrix vector product (used in CI hamiltonians) is used to diagonalize AH.(work in progress)

a) X. Prat-Resina, M. Garcia-Viloca, G. Monard, A. González-Lafont, J.M. Lluch, J.M. Bofill, J.M. Anglada, Theor.Chem.Acc., 107, 147 (2002)
b)J.M. Anglada, J.M. Bofill, E. Besalú, Theor. Chem. Acc. 104, 414 (2000)

Full diagonalization vs Partial diagonalization

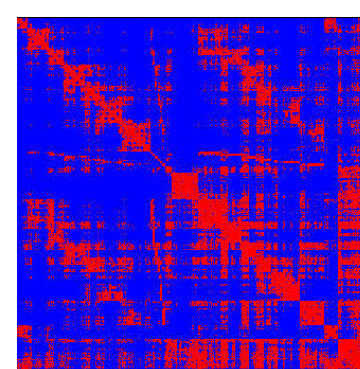
Lapack^a subroutines DSPEV and DSPEVX are used to perform complete diagonalization and partial diagonalization (extracting only the first 5 eigenpairs) respectively of the Augmented Hessian. The computational effort increases rapidly when a full diagonalization is carried out. Partial diagonalization is adequate since both procedures gives the same result, *i.e.* it takes the same number of steps to converge.

Dimension of the system (coordinates)	Full diagonalization	Partial diagonalization
99	0.03 (0.24%) ^b	0.01 (0.08%)
285	1.9 (13.4%)	0.16 (1.3%)
420	6.2 (32.3%)	1.1 (8.0%)
978	92.8 (87.96%)	17.3 (57.7%)
3864	15360.2 (99.9%)	1195.1 (98.3%)

a) E. Anderson et. al. LAPACK User's Guide, Third Edition, SIAM Press 1999
b) percentage of time occupied for diagonalization in a generic step

Influence of partial storage of the Hessian in optimization

Hessian matrix of 444 cartesian coordinates of MR active center. In red the matrix elements bigger than 0.1 kcal/mol/Å².(25.32%)



Test in a decapeptide TS optimization (270 coordinates)

Minimum value stored	10 (6.2%)	1 (10.5%)	1.E-1 (22.1%)	1.E-2 (41.2%)	1.E-3 (65.3%)	1.E-4 (85.8%)	1.E-5 (96.4%)	1.E-6 (99.5%)
steps required to converge	300	25	24	9	9	9	9	9

\Rightarrow We have the same efficiency storing only elements bigger than 1.E-2

Approximated initial Hessian

The calculation of the initial Hessian is the most expensive process in the optimization. The environment can have its 2nd derivatives approximated.

number of coordinates (exact/approx)	CPU time required (s)			time required for a step (s)
	QM/MM (exact)	MM	QM/MM (1 SCF)	
DHAP (15/21)	4.5	2.3	4.4	0.08
PHTAL (27/21)	6.9	4.3	6.7	0.08
MR (240/3654)	92196	10438		20.7

If the environment has not an important role in the reaction the number of steps required to converge is not very different. In this case the pure MM approximation is the most adequate.