

**REDUCING THE EXPENSE OF ELECTRONIC STRUCTURE
CALCULATIONS FOR LARGER MOLECULES: OPTIMIZED
AUXILIARY BASIS SETS, AND SYSTEM-SPECIFICALLY
REPARAMETRIZED SEMIEMPIRICAL METHODS**

Doctoral Thesis

submitted in partial fulfillment
of the requirements for the doctoral degree
“Dr. rer. nat.”

to the Faculty of Mathematics and Natural Science
of the Christian-Albrecht-University of Kiel

submitted by

Rungtiva Palangsuntikul

Institute of Physical Chemistry

Kiel

2005

RefereeProf. Dr. Bernd Hartke.....

Co-refereeProf. Dr. Friedrich Temps.....

Day Of The Exam

Authorised For Printing

Dean
.....

ABSTRACT

Optimization approaches using several global and local algorithms (genetic algorithms, direct search, simplex and implicit filtering) in the search for a global minimum are applied to optimize auxiliary basis sets for quantum chemistry *ab-initio* calculations and to reparametrize semiempirical methods.

We optimize auxiliary basis sets for RI-MP2 and RI-HF, by minimizing a suitable difference measure ΔI to the analogous calculations without the RI technique. It is shown that our methods of generating optimal auxiliary basis sets are more systematic and can be automatized more easily than the traditional approach. Hence, they can reasonably be expected to be faster and more reliable. At the same time, the quality of our basis sets is at least as good as that from the traditional approach. As an application, we present the first systematically optimized and complete set of mixed Poisson and density auxiliary basis sets for the atoms H, B, C, N, O and F, complementing the standard basis sets cc-pVXZ (X = D, T, Q and 5). As soon as efficient integral routines for this new basis function type become available, calculations with them will be much more efficient than with traditional basis sets.

Similarly, these global and local optimization methods are also employed to reparametrize semiempirical methods for a difficult double proton transfer system. System-specific reparametrization of the well-known AM1, PM3 and PM5 methods is done by minimizing the error of the semiempirical calculations compared to *ab-initio* reference data at the MP2/aug-cc-pVDZ level. This is done at a small set of selected geometries, leading to one- and two-dimensional potential energy surfaces that are quantitatively in agreement with the *ab-initio* data over a much broader range of geometries. With this system-specific adaption, these reparametrized methods lead to results far superior to those obtainable with standard parameters. Nevertheless, the full speed advantage of the semiempirical approach is retained, offering the possibility to do direct dynamics studies with the potential energy surface calculated on the fly at *ab-initio* quality but at a fraction of the *ab-initio* cost.

In both cases, our combination of genetic algorithm global search and Powell local search is the fastest and most robust choice for optimization, comparing with the other methods. Therefore, in these cases, a combination of global and local search is actually better than a purely local algorithm.

KURZZUSAMMENFASSUNG

Optimierungsmethoden, die verschiedene globale und lokale Optimierungsalgorithmen (genetische Algorithmen, direkte Suche, simplex und implizites Filtern) für die Suche nach dem globalen Minimum verwenden, werden hier dazu verwendet, um Basissätze für quantenchemische ab-initio Berechnungen zu optimieren und um semiempirische Methoden zu reparameterisieren.

Wir optimieren Basissätze für RI-MP2 und RI-HF durch Minimierung einer geeigneten Differenzmessung DI im Vergleich zu analogen Rechnungen ohne der RI-Methode. Es konnte gezeigt werden, dass unsere Methoden, nämlich das Generieren von optimalen Basissätzen, systematischer und einfacher zu automatisieren sind als bei der traditionellen Herangehensweise. Deshalb kann man begründeterweise erwarten, dass diese Methoden schneller und zuverlässiger sind. Gleichzeitig sind unsere Basissätze qualitativ mindestens genauso gut wie die der traditionellen Methoden. Als Anwendung präsentieren wir den ersten systematisch optimierten Satz von gemischten Poisson- und Dichte-Basissätzen für die Atome H, B, C, N, O und F, ergänzend zu den Standard-Basissätzen cc-pVXZ ($X = D, T, Q, 5$). Sobald effiziente Integrationsmethoden für diese neuen Typen von Basisfunktionen verfügbar sind, werden Rechnungen mit ihnen sehr viel effizienter als mit herkömmlichen Basissätzen.

Mit ähnlicher Vorgehensweise werden diese globalen und lokalen Optimierungsmethoden dafür verwendet, um semiempirische Methoden für ein schwieriges Protonen-Transfer-System zu reparameterisieren. Bei der system-spezifischen Reparameterisierung der gut bekannten Methoden AM1, PM3 und PM5 wird so vorgegangen, dass der Fehler bei den semiempirischen Rechnungen in Bezug auf ab-initio Referenzdaten auf MP2/(aug)-cc-pVDZ Niveau minimiert wird. Die Rechnungen wurden mit einem kleinen Satz ausgewählter Geometrien durchgeführt, was zu ein- und zwei-dimensionalen Potentialenergiehyperflächen führte, die in quantitativer Übereinstimmung mit den ab-initio Daten über eine sehr viel breitere Auswahl an Geometrien sind. Durch diese systemspezifische Adaption erhält man mit

diesen reparameterisierten Methoden sehr viel bessere Ergebnisse als mit den Standardparametern. Dennoch bleibt der Zeitvorteil des semiempirischen Ansatzes erhalten, was die Möglichkeit zu direkter Dynamik gibt, bei der die Potentialenergiehyperflächen mit ab-initio-vergleichbarem Niveau, aber mit einem Bruchteil der *ab-initio* Kosten, immer dann berechnet werden können, wenn sie benötigt werden.

In beiden Fällen ist unsere Kombination aus globaler Suche mittels genetischem Algorithmus und lokaler Suche mittels Powell die schnellere und robustere Wahl für die Optimierung verglichen mit den anderen Methoden. Deshalb ist in den oben aufgeführten Fällen die Kombination aus globaler und lokaler Suche tatsächlich besser als eine rein lokale Suche.

ACKNOWLEDGEMENTS

I would like to express my deep gratitude to my supervisor Prof. Dr. Bernd Hartke for giving me a precious chance to conduct my doctoral research in his group, as well as his excellent support, kind encouragement and valuable advices during my PhD. This thesis would never have become that extensive without the support from him.

I would like to give my gratitude to Prof. Dr. Friedrich Temps for correcting my thesis as a second referee and also to Prof. Dr. Karin Schwarz and Prof. Dr. Thisbe Kerstin Lindhorst for their acting as an assessor for my exam. Special thanks are due to my committees, Prof. Dr. Heinrich Mäder, Prof. Dr. Regine von Klitzing and Prof. Dr. Norbert Stock, for their extremely helpful comments and suggestions and for the time spent reading my thesis and for my exam.

I would like also to express my grateful acknowledges to PD Dr. Robert Polly who gave me a possibility to work on ‘RI auxiliary basis sets’ and for his valuable suggestions. I very much appreciated working with him.

I am also very grateful to Prof. Maurizio Persico (University of Pisa, Italy) for his kind guidance in reparametrized semiempirical methods, to PD Dr. Guntram Rauhut and Stefan Schweiger for their kind help and for giving us their structures of DPTRs, and to Dr. Giovanni Granucci who has always been answering my questions about MOPAC. It was a great pleasure to discuss with them, and I appreciate their generosity in every concern.

I would like to express my sincere thanks to Prof. Arnold Neumaier (Institute of Mathematics, University of Vienna, Austria) for many interesting discussions and valuable suggestions, and for his hospitality during a short stay of me in Vienna.

My special thank goes to Prof. Dr. Heinrich Mäder for his kindness and support. I appreciate it very much. A special thanks is also addressed to Dr. Franziska Schulz and her partner André Janz who provided help, support, friendship

and good times for me during my time in Germany until recently. In particular I wish to thank my colleagues, Dr. Adem Tekin, Bernhard Bandow, Frank von Horsten, Florian Koskowski for all their support, friendship and for the good times we had together. Special mention deserves to be made of my appreciation to, Ingo Berndt, Dr. Kai Berndel, Simmy Jacobsen, Stephan Wörmke that made me feel good in Kiel.

The financial support from Christian-Albrechts University of Kiel is gratefully acknowledged.

Here, my gratefulness is forwarded to my friends, Yingyot Poo-arporn, Chan Inntam, Pornpan Pungpo, Kanda Nivesanond, Somkiat Nokbin, Patchareenart Saparpakorn, Songwut Suramitr and Anwaraporn Nintarath for their friendship, support and always cheering me up as well as for those I did not mention by name.

Finally, I need to express my deepest gratitude and respect to my parents, Thada and Aai-ju, and my brothers, Sakda and Saksakul, whom their love, constant prayers, moral support, encouragement and continuous support tremendously helped me to finish my PhD work.

TABLE OF CONTENTS

	Page
TABLE OF CONTENTS	i
LIST OF TABLES	iii
LIST OF FIGURES	vii
1. INTRODUCTION	1
2. GLOBAL AND LOCAL OPTIMIZATIONS	
2.1 Genetic algorithm (GA)	5
2.2 Hybrid of GA and local search	8
2.3 Hybrid of Direct and local search	10
2.4 Implicit filtering for constrained optimization (IFFCO)	15
2.5 Downhill Simplex Method	18
2.6 Simulated annealing method	19
3. OPTIMIZED AUXILIARY BASIS SETS FOR RI-MP2 AND RI-HF CALCULATIONS	
3.1 INTRODUCTION	21
3.1.1 Hartree-Fock (HF) theory	24
3.1.2 Møller-Plesset Perturbation Theory	33
3.1.3 Resolution of the identity approximation	37
3.1.4 The Poisson method	38
3.1.5 Basis sets	41
3.2 METHOD OF CALCULATIONS	
▪ RI-MP2	49
▪ RI-HF	51
3.3 RESULTS AND DISCUSSION	
3.3.1 RI-MP2	54
3.3.2 RI-HF	83
3.4 CONCLUSIONS AND OUTLOOK	91

	Page
4. SYSTEM-SPECIFICALLY REPARAMETRIZED SEMIEMPIRICAL METHODS	
4.1 INTRODUCTION	93
4.1.1 Semi-empirical methods	97
4.1.2 Model system for re-optimizing semiempirical parameters	101
4.2 OPTIMIZATION METHODS APPLICATION	104
4.3 RESULTS AND DISCUSSION	107
4.3.1 Reparametrization geometries on the one-dimensional reaction path	108
4.3.2 Reparametrization on two-dimensional relaxed geometries	116
4.3.3 Reparametrization on two-dimensional rigid geometries	124
4.4 CONCLUSIONS AND OUTLOOK	133
5. BIBLIOGRAPHY	135
6. CURRICULUM VITAE	139
7. DECLARATION	141

LIST OF TABLES

Table		Page
1	Correlation consistent basis sets.	47
2	The composition and size of aug-cc-pVXZ, and cc-pCVXZ basis sets.	48
3	Specification of auxiliary basis sets, cc-pVXZ with X = D, T and Q for RI-MP2 calculations.	51
4	Molecules used to optimize the RI-HF auxiliary basis sets, cc-pVXZ series with X = D, T, Q and 5.	52
5	The sizes of auxiliary basis set cc-pVXZ with X = D, T, Q and 5, for RI-HF calculations.	52
6	Size n , centers α and ratios β of auxiliary basis set cc-pVTZ for the H ₂ O molecule, results obtained from the GA/Powell hybrid method.	54
7	The best objective function and CPU time (min.), comparing different optimization strategies of α and β parameters for H and O atoms. The calculations were performed using GA/Powell hybrid and Iffco methods for the H ₂ O molecule.	55
8	Summary of the best objective function values ($\Delta I/ E_{MP2} $) per atom, number of function evaluations and CPU time (min) for H ₂ , BH ₃ , CH ₄ , NH ₃ , H ₂ O and HF molecules, using the GA method.	56
9	Overall best values for the center parameters α of even-tempered auxiliary basis sets, complementing standard bases cc-pVXZ with X = D, T and Q for H ₂ , BH ₃ , CH ₄ , NH ₃ , H ₂ O and HF molecules, obtained from the genetic algorithm.	58
10	Summary of the best objective function values ($\Delta I/ E_{MP2} $) per atom, number of function evaluations and CPU time (min) for H ₂ , BH ₃ , CH ₄ , NH ₃ , H ₂ O and HF molecules, using the GA/Powell hybrid method.	59

Table		Page
11	Overall best values for the center parameters α of even-tempered auxiliary basis sets, complementing standard bases cc-pVXZ with X = D, T and Q for H ₂ , BH ₃ , CH ₄ , NH ₃ , H ₂ O and HF molecules, obtained from the GA/Powell hybrid method.	61
12	The error ($\Delta I/ E_{MP2} $) per atom, number of function evaluations and CPU time (min) of test runs for the H ₂ O molecule, using the Direct method alone.	62
13	Summary of the best objective function values ($\Delta I/ E_{MP2} $) per atom, number of function evaluations and CPU time (min) for H ₂ , BH ₃ , CH ₄ , NH ₃ , H ₂ O and HF molecules, using the Direct/Powell hybrid method.	63
14	Summary of the best objective function values ($\Delta I/ E_{MP2} $) per atom, number of function evaluations and CPU time (min) for H ₂ , BH ₃ , CH ₄ , NH ₃ , H ₂ O and HF molecules, using the Direct/Iffco hybrid method.	64
15	Overall best values for the center parameters α of even-tempered auxiliary basis sets, complementing standard bases cc-pVXZ with X = D, T and Q for H ₂ , BH ₃ , CH ₄ , NH ₃ , H ₂ O and HF molecules, obtained from the Direct/Powell hybrid method.	66
16	Overall best values for the center parameters α of even-tempered auxiliary basis sets, complementing standard bases cc-pVXZ with X = D, T and Q for H ₂ , BH ₃ , CH ₄ , NH ₃ , H ₂ O and HF molecules, obtained from the Direct/Iffco hybrid method.	67
17	Summary of the best objective function values ($\Delta I/ E_{MP2} $) per atom, number of function evaluations and CPU time (min) for H ₂ , BH ₃ , CH ₄ , NH ₃ , H ₂ O and HF molecules, using the Iffco method.	68

Table	Page
18 Overall best values for the center parameters α of even-tempered auxiliary basis sets, complementing standard bases cc-pVXZ with X = D, T and Q for H ₂ , BH ₃ , CH ₄ , NH ₃ , H ₂ O and HF molecules, obtained from the Iffco method.	70
19 The best objective function and CPU time (min). The calculations were performed by using GA (with maximum number of 1000 generations), GA/Powell hybrid and Iffco methods for H ₂ , BH ₃ , CH ₄ , NH ₃ , H ₂ O and HF molecules. All parameters in all atoms were optimized.	73
20 Overall best values for the optimized center parameters α and ratio parameters β of even-tempered auxiliary basis sets, complementing standard bases cc-pVXZ with X = D, T and Q for H ₂ , BH ₃ , CH ₄ , NH ₃ , H ₂ O and HF molecules, obtained from the GA/Powell hybrid method.	75
21 Summary of the best objective function values ($\Delta I/ E_{MP2} $) per atom, number of function evaluations and CPU time (min), optimizing parameters α and β for the B, C, N, O and F atoms (with α and β parameters for the H atom fixed).	77
22 Overall best values for the parameters α and parameters β , complementing standard bases cc-pVXZ with X = D, T and Q, obtained from optimization of α and β parameters for the B, C, N, O and F atoms only (with parameters for the H atom fixed).	79
23 Comparing the quality of the conventional auxiliary basis sets optimized by Weigend et al. and the mixed conventional/Poisson auxiliary basis sets optimized in this work. The errors ($\Delta I/ E_{MP2} $) are presented in ppb.	80
24 Test set of 40 sample molecules which contain only hydrogen and first row elements.	81

Table	Page	
25	Mean relative error $\bar{\Delta}E_{CORR} = (1/n) \sum (E_{Corr,RI-MP2}^i - E_{Corr,MP2}^i) / E_{Corr,MP2}^i$, the standard deviation and maximum error for a test set of 40 molecules.	82
26	Summary of the best objective function values (Δ RI) per atom for RI-HF calculations, for auxiliary basis sets complementing standard bases cc-pVXZ with X = D, T, Q and 5. The calculations were performed using the GA/Powell hybrid method.	84
27	Size n and overall best values for the center parameters α and ratio parameters β of even-tempered auxiliary basis sets for RI-HF calculations, complementing standard bases cc-pVXZ with X = D, T, Q and 5 for H, B, C, N, O and F atoms, obtained from the GA/Powell hybrid method using 20 sample molecules.	85
28	Summary of the best objective function values (Δ RI) per atom for RI-HF calculations, complementing standard bases cc-pVXZ with X = D, T, Q and 5. The calculations were performed using the Iffco method.	87
29	Size n and overall best values for the center parameters α and ratio parameters β of even-tempered auxiliary basis sets for RI-HF calculations, complementing standard bases cc-pVXZ with X = D, T, Q and 5 for H, B, C, N, O and F atoms, obtained from the Iffco method using 20 sample molecules.	89
30	Parameters used in semiempirical methods and their descriptions.	105

LIST OF FIGURES

Figure		Page
1	Simple reproduction allocates offspring strings using a roulette wheel with slots sized according to fitness.	7
2	Flowchart of the simple genetic algorithm.	8
3	Division of a hypercube. The numbers next to the points "o" are function values. Successive steps of the algorithm are shown in each row a, b, c from left to right. See text for further explanations.	12
4	Illustration of potentially optimal hyperrectangles (called boxes in this Figure). Note that $f^* = f_{\min} - \varepsilon f_{\min} $. Potentially optimal hyperrectangles are on the lower-right convex hull.	13
5	The 1s Slater-type orbital and the Gaussian-type orbital.	42
6	Comparison of 1s Slater-type orbital and Gaussian expansions with up to three terms.	43
7	Correlation functions added in shells.	46
8	Histogram of the energy differences (Δ RI) per atom between MP2 and RI-MP2 calculations, obtained from the genetic algorithm.	57
9	Histogram of the energy differences (Δ RI) per atom between MP2 and RI-MP2 calculations, obtained from the GA/Powell hybrid method.	60
10	Histogram of the energy differences (Δ RI) per atom between MP2 and RI MP2 calculations, obtained from the Direct/Powell hybrid method.	65
11	Histogram of the energy differences (Δ RI) per atom between MP2 and RI MP2 calculations, obtained from the Direct/Iffco hybrid method.	65

Figure		Page
12	Histogram of the energy differences (Δ RI) per atom between MP2 and RI MP2 calculations, obtained from the Iffco method.	69
13	Relative time spent for optimizing mixed Poisson/density auxiliary basis sets for the standard cc-pVTZ bases of H ₂ , BH ₃ , CH ₄ , NH ₃ , H ₂ O and HF molecules with different optimization methods, for clarity setting the time for the Direct with Powell method to 1.0 for each molecule. The actual absolute times for H ₂ , BH ₃ , CH ₄ , NH ₃ , H ₂ O and HF are 6, 61, 150, 54, 30 and 15 min., respectively.	72
14	Histogram of the energy differences (Δ RI) per atom between MP2 and RI-MP2 calculations, computed by using auxiliary basis sets which both optimized parameters α and β in all atoms.	74
15	Histogram of the energy differences between MP2 and RI-MP2 calculations per atom, optimizing both parameters α and β for B, C, N, O and F using the GA/Powell method while fixing α and β parameters of the H atom at their H ₂ optimized values.	78
16	Pyrazole-Guanidine model system.	101
17	Energy profile along the intrinsic reaction coordinate of the prototypical plateau reaction at the MP2 level.	102
18	Relaxed MP2/(aug)-cc-pVDZ potential energy surface of the double proton transfer reaction between pyrazole and guanidine.	103
19	Rigid MP2/(aug)-cc-pVDZ potential energy surface of the double proton transfer reaction between pyrazole and guanidine.	103
20	One-dimensional energy profiles of the DPTR, comparing two different sizes of reparametrized parameter set.	108

Figure		Page
21	One-dimensional energy profiles of the DPTR, obtained from the AM1, PM3 and PM5 methods in which a part of the parameter set was reparametrized on a subset of these geometries using the simplex algorithm.	109
22	One-dimensional energy profiles of the DPTR, obtained from the AM1, PM3 and PM5 methods in which a part of the parameter set was reparametrized on a subset of these geometries using the hybrid GA/Powell algorithm.	110
23	One-dimensional energy profiles of the DPTR, obtained from the AM1, PM3 and PM5 methods in which a part of the parameter set was reparametrized on a subset of these geometries using the Iffco algorithm.	111
24	One-dimensional energy profiles of the DPTR, obtained from the AM1, PM3 and PM5 methods in which all parameters were reparametrized on a subset of these geometries using the simplex algorithm.	113
25	One-dimensional energy profiles of the DPTR, obtained from the AM1, PM3 and PM5 methods in which all parameters were reparametrized on a subset of these geometries using the hybrid GA/Powell algorithm.	114
26	One-dimensional energy profiles of the DPTR, obtained from the AM1, PM3 and PM5 methods in which all parameters were reparametrized on a subset of these geometries using the Iffco algorithm.	115
27	Relaxed 2D energy profiles of the DPTR, obtained from the AM1, PM3 and PM5 methods for which a part of the parameter set was reparametrized on a subset of these geometries using the simplex algorithm.	117

Figure		Page
28	Relaxed 2D energy profiles of the DPTR, obtained from the AM1, PM3 and PM5 methods for which a part of the parameter set was reparametrized on a subset of these geometries using the GA/Powell algorithm.	118
29	Relaxed 2D energy profiles of the DPTR, obtained from the AM1, PM3 and PM5 methods for which a part of the parameter set was reparametrized on a subset of these geometries using the Iffco algorithm.	119
30	Relaxed 2D energy profiles of the DPTR, obtained from the AM1, PM3 and PM5 methods for which all parameters were reparametrized on a subset of these geometries using the simplex algorithm.	121
31	Relaxed 2D energy profiles of the DPTR, obtained from the AM1, PM3 and PM5 methods for which all parameters were reparametrized on a subset of these geometries using the GA/Powell algorithm.	122
32	Relaxed 2D energy profiles of the DPTR, obtained from the AM1, PM3 and PM5 methods for which all parameters were reparametrized on a subset of these geometries using the Iffco algorithm.	123
33	Rigid 2D energy profiles of the DPTR, obtained from the AM1, PM3 and PM5 methods for which a part of the parameter set was reparametrized on a subset of these geometries, using the simplex algorithm.	125
34	Rigid 2D energy profiles of the DPTR, obtained from the AM1, PM3 and PM5 methods for which a part of the parameter set was reparametrized on a subset of these geometries, using the GA/Powell algorithm.	126

Figure		Page
35	Rigid 2D energy profiles of the DPTR, obtained from the AM1, PM3 and PM5 methods for which a part of the parameter set was reparametrized on a subset of these geometries, using the Iffco algorithm.	127
36	Rigid 2D energy profiles of the DPTR, obtained from the AM1, PM3 and PM5 methods for which all parameters were reparametrized on a subset of these geometries, using the simplex algorithm.	128
37	Rigid 2D energy profiles of the DPTR, obtained from the AM1, PM3 and PM5 methods for which all parameters were reparametrized on a subset of these geometries, using the GA/Powell algorithm.	129
38	Rigid 2D energy profiles of the DPTR, obtained from the AM1, PM3 and PM5 methods for which all parameters were reparametrized on a subset of these geometries, using the Iffco algorithm.	130
39	2D energy profiles of the DPTR, obtained from the AM1 method for which all parameters were reparametrized on a subset of 1D geometries, using the GA/Powell algorithm.	131
40	1D and rigid 2D energy profiles of the DPTR, obtained from the AM1 method for which all parameters were reparametrized on a subset of 2D relaxed geometries, using the GA/Powell algorithm.	132

**REDUCING THE EXPENSE OF ELECTRONIC STRUCTURE
CALCULATIONS FOR LARGER MOLECULES: OPTIMIZED
AUXILIARY BASIS SETS, AND SYSTEM-SPECIFICALLY
REPARAMETRIZED SEMIEMPIRICAL METHODS**

1. INTRODUCTION

As a result of recent successes in understanding and predicting many properties of molecules and reactions at the atomic level, electronic structure calculations have become increasingly important in the fields of physics and chemistry over the past decade, especially with the advent of present-day, high performance computers. With electronic structure methods we are now capable of directly investigating many chemical phenomena of interest with an accuracy rivaling that of experiments. Moreover, they can now be used to reliably predict chemical phenomena in situations that are difficult or even impossible to observe directly.

Traditional implementation of electronic structure calculations involves an effort proportional to at least the cube of the number of electrons (N) in the system, effectively placing a limit on the scientific problems that can be tackled with these approaches. This undesirable expense/size scaling of the electronic calculations limits the size of the systems that can be studied to a few hundred atoms at most. In particular, explicit correlation methods, in which an enormous size of the correlation expansion is needed to achieve the desired high accuracy of the calculations, lead to extremely unfavorable scaling. Recent developments have shown that this unfavorable scaling behavior is artificial, arriving at several different linear scaling methods¹⁻² that ultimately rely on the locality of the problem. Linear scaling reduces the computational effort to only N^1 , with the theoretical ability to do calculations for much larger system. However, these methods are still in development, therefore the theoretical promise of linear scaling is not always realizable in practice and some technical problems still are unsolved (e.g. domain definitions in local correlation methods can induce discontinuities in the potential energy surface (PES); depending

on spatial extent and chemical connectivity, some systems are much better suited to linear scaling techniques than others; etc.). Furthermore, the linear scaling implementations generally reduce only the size scaling but not the prefactor; often they even increase the prefactor. The usefulness of any linear-scaling scheme is ultimately determined by its crossover point, namely the system size at which the method begins to be faster than conventional cubic-scaling approaches. This crossover depends largely on two factors: First the computational cost per iteration per atom, and second the number of iterations required to reach a given convergence threshold per atom. Even if a method is constructed in which the computational cost per iteration per atom is small and independent of system size, the number of iterations required may be so large that the minimization is prohibitively inefficient.

Therefore, it is very important to complement the linear scaling methods by other improvements, for example resolution of the identity (RI) methods (explained in more detail in section 3.1.3), that do not change the size scaling but reduce the prefactor considerably (by about one order of magnitude), with negligible loss of accuracy. Currently, they are being combined with linear scaling techniques in the latest implementations, which obviously combines the advantages of both approaches, i.e. both the prefactor and the size scaling are reduced.

The inherently high cost of electronic structure calculations arises from the huge number of two-electron integrals necessary for constructing the Fock matrix. Other approaches such as semiempirical methods, empirical potentials, force fields, etc. treat some or all of the electronic effects in a molecular system only implicitly. Therefore, they are relatively inexpensive computationally, which allows them to be used for very large systems containing many thousands of atoms. Although linear scaling methods can reduce the scaling of *ab-initio* methods, they still are orders of magnitude slower.

While semiempirical methods are orders of magnitude faster than *ab-initio* approaches, they typically are parametrized to work well for an application domain of chemical systems that is as broad and diverse as possible. Although the latest

parameterizations (e.g. PM5) achieve useful levels of accuracy in many cases, it is clear that this is always only a compromise, and that the accuracy for each system (or narrow class of systems) individually would be much better after a system-specific reparametrization, rivaling even high-level *ab-initio* calculations. Recently, this was demonstrated even for coupled electronically excited states by Persico et al.³

As on many other occasions in the applied sciences, it turns out that the realization of these techniques involves optimization tasks: RI methods require auxiliary basis sets, adapted (optimized) to the atoms and conventional basis sets present, by minimizing the error compared to the analogous calculation without the RI technique. System-specific reparametrization of semiempirical methods is a minimization of the error of the semiempirical calculation compared to reference data (e.g. *ab-initio* results and/or experimental data). In both cases, the optimization involves many parameters of different characteristics (linear/nonlinear, different sensitivity), leading to a difficult nonconvex multidimensional optimization problem. Traditionally, these optimization tasks were treated with simple standard local optimization methods, leading to lengthy multistart series, lack of reproducibility, lots of human effort, and still possibly suboptimal results. However, this problem class can also be treated with global optimization techniques (or at least with advanced local optimization methods that do not get stuck in small shallow sub-optima), which yield optimal solutions in a purely automatic way and in much shorter time. This makes these optimization tasks accessible to non-expert users, and, if necessary or desired, also enables them to redo these optimizations for each system under study.

The application of global (and advanced local) optimization methods to these two optimization problems is the topic of this work. It is shown that optimal auxiliary basis sets for RI-HF and RI-MP2 calculations can be generated (for hydrogen and first-row elements, as concrete examples) automatically and very quickly, but with a quality at least as good as that of the traditional, lengthy, local multistart approach. Similarly, these global optimization methods are also employed to reparametrize semiempirical methods for a specific, given system. This is done at a small set of selected geometries, leading to a potential energy surface that is

quantitatively valid over a much broader range of geometries but can be calculated at a fraction of the cost.

According to the two different tasks we have studied, this work is divided into two parts. In the first part, the optimization of mixed Poisson and density auxiliary basis sets, cc-pVXZ ($X = D, T, Q$ and 5) for RI-MP2 and RI-HF of H, B, C, N, O and F atoms are reported, testing the performance of several different optimization algorithms. In the second part, reparametrizations of semiempirical methods, AM1, PM3 and PM5, are performed and tested.

2. GLOBAL AND LOCAL OPTIMIZATIONS

A minimum of a function can either be global (the highest or lowest value over the whole region of interest) or local (the highest or lowest value over some small neighborhood). We are usually most interested in finding the global optimum but this can be very difficult. Often a problem will have many local optima which means that locating the single global optimum can be tricky.

The most suitable methods to locate minima depend upon the nature of the function we are dealing with. There are two broad classes of algorithms.

- Local minimizers that, given a point in a valley of the function, locate the lowest point in the valley.
- Global minimizers that range over a region of parameter space attempting to find the bottom of the deepest valley.

If a good estimate of the position of the global minimum exists we need only use a local minimizer to improve it and find the optimum choice of parameters. If no such estimate exists some global method must be used. The simplest would be to generate a set of possible starting points, locally optimize each and choose the best. However, this may not be the most efficient approach, in particular in search spaces of high dimensionality, since often the number of minima increases polynomially or even exponentially with the number of dimensions.

2.1 Genetic algorithm (GA)

Genetic algorithms⁴⁻⁶ are stochastic global search algorithms with heuristic ideas (operators) borrowed from the mechanisms of natural selection and natural genetics. These algorithms are computationally simple yet particularly powerful in the initial stages of their search. Furthermore, they are not fundamentally

limited by restrictive assumptions about the search space. The search is not exact meaning that there is no guarantee that the global minimum will be found. However, the result typically is at least a very low-valued local minimum.

The algorithm begins with a set of trial solutions (represented by individuals) called a population. The initial population is created from random numbers. A value for fitness is assigned to each individual depending on how close it actually is to solving the problem, or, more simply how much better it is than the other individuals. Individuals from one population are taken, modified, and used to form a new population according to their fitness. The more suitable they are the more chances they have to reproduce. A generation of the genetic algorithm begins with reproduction, which in turn involves selection and crossover (exchange of information between individuals). Reproduction proceeds in two steps:

- pairs of individuals are chosen as parents, where one parent is chosen according to fitness and the other at random. Choosing individuals according to their fitness means that individuals with a higher value have a higher probability of contributing one or more offspring in the next generation. The selection operator can be implemented by creating a biased roulette wheel where each current individual in the population has a roulette wheel slot sized in proportion of its fitness. A marble is thrown in the roulette wheel and the individual where it stops is selected. Clearly, the individuals with bigger fitness values will be selected more times. See an example in Figure 1.
- each parent pair generates two children, with a fixed probability for a one point crossover between the parent strings, with a randomly selected crossing site.

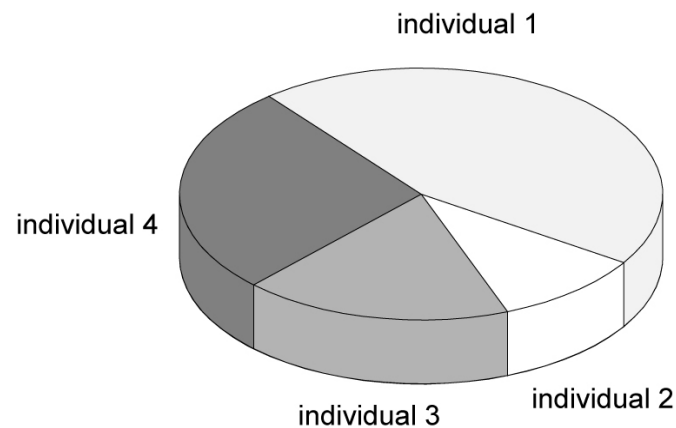


Figure 1 Simple reproduction allocates offspring strings using a roulette wheel with slots sized according to fitness.

One crossover point is selected, the genetic values are copied from the first parent till the crossover point, then the other parent is scanned.

The last operator, mutation, is performed with a fixed probability for mutation of each real number. In a mutation, the real number is modified by the addition of a random number evaluated from a normal Gaussian distribution. The new children constitute the following generation. Following reproduction, crossover and mutation, the new population is ready to be tested. This procedure is repeated for many generations until the global minimum or at least a very good local minimum has been found. The GA algorithm is illustrated by the flowchart shown in Figure 2.

We found that a population of 6 individuals per generation and a maximum number of 200 generations were sufficient for this study. Furthermore, crossover and mutation probabilities are 0.6 and 0.05, respectively, which correspond to standard values used in the literature.

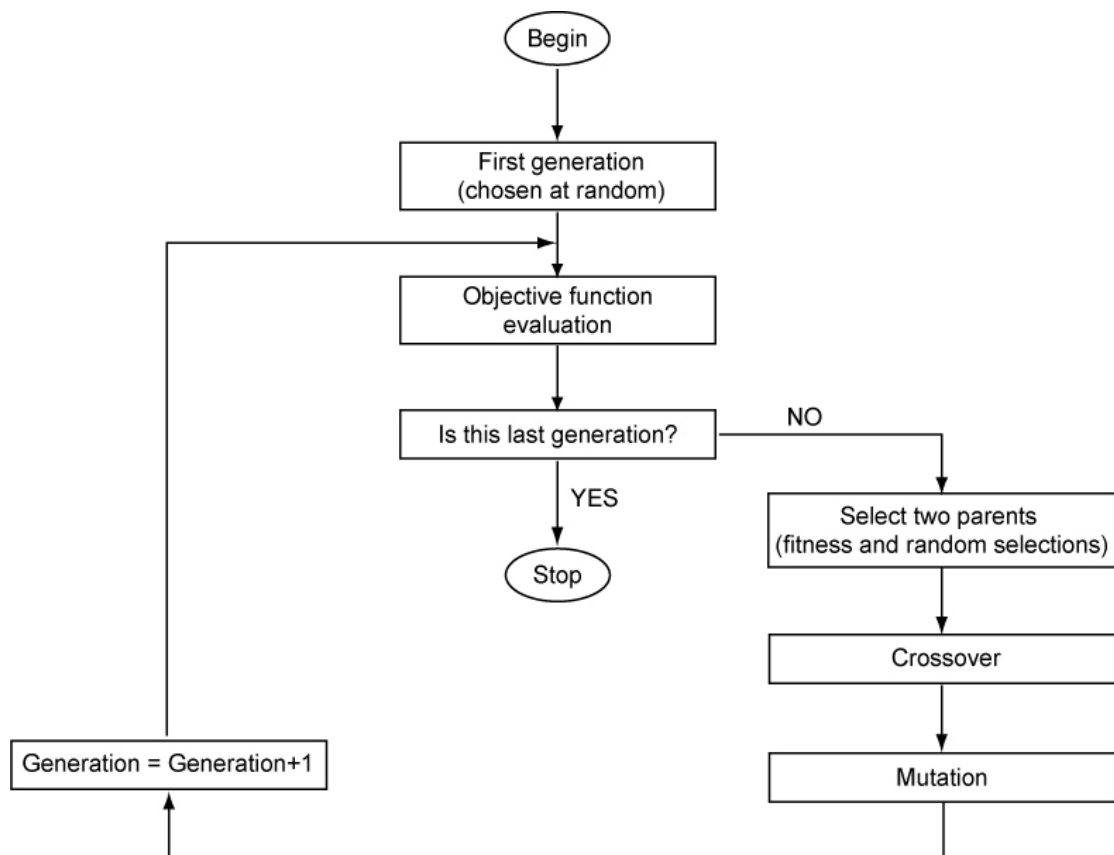


Figure 2 Flowchart of the simple genetic algorithm.

2.2 Hybrid of GA and local search

As has been observed many times, a simple standard GA is good at roughly locating promising regions of search space quickly but is then inefficient in narrowing down on the actual minima. Therefore, in many cases it turns out to be advantageous to combine the GA with a deterministic local search (see Gregurick *et al.*⁷ for an example). For smooth functions, efficient local search requires gradient or even second derivative information. Since this information may not be available analytically in all cases, we demonstrate here that the reduced efficiency resulting from this does not prevent such an approach from being used. We have therefore

decided to use Powell's method⁸⁻⁹, since it employs iterative steps along conjugate directions and hence can be expected to be fairly efficient. Nevertheless, the method is significantly more expensive than methods using gradient information, strictly ruling out excessive use of local search within the global search (as in Gregurick *et al.*⁷).

Therefore, our present hybrid of GA and Powell is rather crude and sparse: The best point from a short GA run, 30 generations, is used as starting point for a local Powell optimization, and the budget for Powell is limited by using a rather modest relative error value of 10^{-4} as convergence threshold.

The Powell implementation is described by Press *et al.*,⁸ given in algorithm 1. Powell first discovered a direction set method that does produce N mutually conjugate directions. The algorithm is initialized by setting the set of directions u_i to the basis vectors,

$$u_i = e_i \quad i = 1, \dots, N \quad (1)$$

Algorithm 1 Powell direction search

1. Save the starting position as P_0 .
 2. For $i = 1, \dots, N$, minimize the error function starting from P_{i-1} along the direction u_i and store the minimum as the next position P_i .
 3. Let u_i be the direction of the largest decrease. Now this direction u_i is replaced with the direction given by $(P_n - P_0)$. The assumption of the heuristic is that the substituted direction includes the replaced direction so that the resulting set of directions remains linearly independent.
 4. The iteration process continues with the new starting position $P_0 = P_n$, until the minimum is reached.
-

2.3 Hybrid of Direct and local search

Direct search methods¹⁰⁻¹¹ are bound constrained optimization techniques that do not explicitly use derivatives. The performance of a Direct implementation in real applications depends on the characteristics of the objective function, the problem dimension, and the desired solution accuracy. This algorithm is call Direct because it is an acronym for dividing rectangles, one of the primary operations in the procedure. Moreover, Direct search methods are reasonably straightforward to implement and can be applied almost immediately to many nonlinear optimization problems.

The general problem statement is

$$\min_{x \in D} f_0(x), \quad D = \{x \in D_0 \mid f_j(x) \leq 0, j = 1, \dots, J\} \quad (2)$$

where $D_0 = \{x \in E^n \mid l \leq x \leq u\}$ is a simple box constraint set. The objective function and constraints $f_j, j = 0, \dots, J$, must be Lipschitz-continuous on D_0 , satisfying

$$|f_j(x_1) - f_j(x_2)| \leq L_j \|x_1 - x_2\|, \quad \forall x_1, x_2 \in D_0 \quad (3)$$

This assumption means that the rates-of-change of the objective function f_0 and constraints f_1, \dots, f_J are bounded.

There are two main ingredients to this algorithm. The first is how to divide the domain and the second is how to decide which hyperrectangles we divide in the next iteration.

Dividing the domain

Division is based on N-dimensional trisection. The first one is dividing a hypercube. Let c be the center point of the hypercube. The algorithm evaluates the

function at the point $c \pm \delta e_i$ where δ equals 1/3 of the side length of the cube and e_i is the i -th Euclidean vector. Direct defines w_i by

$$w_i = \min\{f(c + \delta e_i), f(c - \delta e_i)\} \quad (4)$$

The algorithm then divides the hypercube in the order given by the w_i , starting with the lowest w_i . Direct divides the hypercube first perpendicular to the direction with the lowest w_i . Then it divides the remaining volume perpendicular to the direction of the second lowest w_i and so on until the hypercube is divided in all directions. This strategy puts c in the center of a hypercube with side length δ . Let $b = \arg \min\{f(c + \delta e_i), f(c - \delta e_i)\}$. b will be the center of a hyperrectangle with one side with a length of δ , the other $N-1$ sides will have a length of 3δ .

Figure 3a shows an example of the division of a hypercube. Here

$$\begin{aligned} w_1 &= \min\{5, 8\} = 5 \\ w_2 &= \min\{6, 2\} = 2 \end{aligned} \quad (5)$$

Therefore, we divide first perpendicular to the x_2 -axis, and in the second step the remaining rectangle is divided perpendicular to the x_1 -axis.

The second is dividing a hyperrectangle. In Direct, a hyperrectangle is only divided along its longest sides, which assures us that we get a decrease in the maximal side length of the hyperrectangle. Figure 3b represents the next step in the algorithm. Direct will divide the shaded area. The second box in Figure 3b shows where Direct samples the function, and the third box shows how the rectangle is only divided once. Figure 3c shows the third step in the algorithm for this example. In this step Direct will divide two rectangles, which are shaded. One of them is a square, therefore it is divided twice as described before. The larger area is again a rectangle and gets divided once.

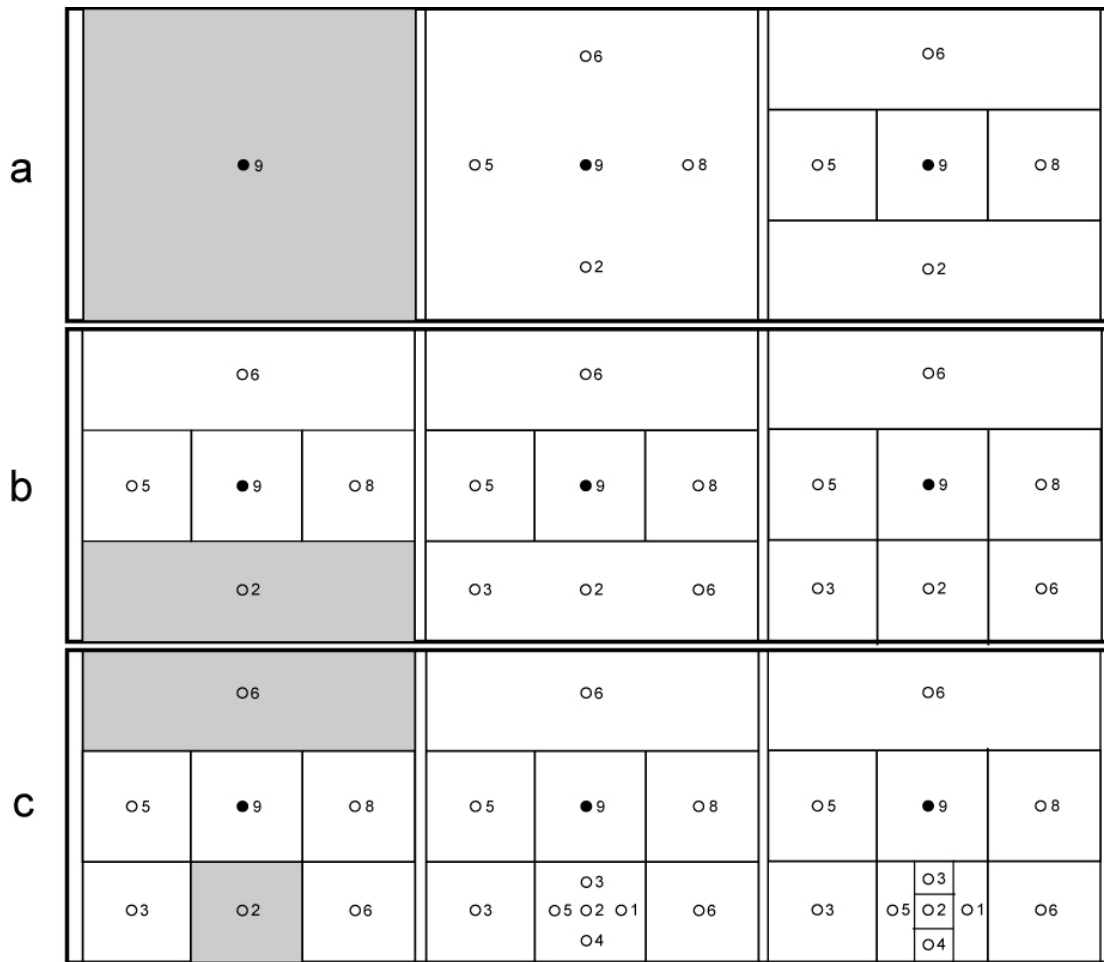


Figure 3 Division of a hypercube. The numbers next to the points "o" are function values. Successive steps of the algorithm are shown in each row a, b, c from left to right. See text for further explanations.

Potentially optimal hyperrectangles

Direct divides all potentially optimal hyperrectangles as defined in definition 1.

Definition 1 Let $\epsilon > 0$ be a positive constant and let f_{\min} be the currently best function value. A hyperrectangle j is said to be potentially optimal if there exists some $\tilde{K} > 0$ such that

$$f(c_j) - \tilde{K}d_j \leq f(c_i) - \tilde{K}d_i, \forall i, \quad \text{and} \quad (6)$$

$$f(c_j) - \tilde{K}d_j \leq f_{\min} - \varepsilon|f_{\min}| \quad (7)$$

In this definition c_j is the center of the hyperrectangle j , and d_j is the measure for this hyperrectangle. Jones *et.al.*¹¹ chose d_i to be the distance from the center of hyperrectangle i to its vertices. They divide all potentially optimal hyperrectangles in every iteration, even if two of them have the same measure and the same function value at the center. Figure 4 represents the set of hyperrectangles as points in a plane. The first inequality (6) screens out the boxes that are not on the lower right of the convex hull of the plotted points, as shown in Figure 4. Note that \tilde{K} plays the role of the (unknown) Lipschitz constant. The second inequality (7) prevents the search from becoming too local and ensures that a nontrivial improvement will (potentially) be found based on the current best solution. In Figure 4, f_{\min} is the current best solution, but its associated box is screened out of the potentially optimal box set due to the second inequality (7). This is illustrated by the dotted line in Figure 4.

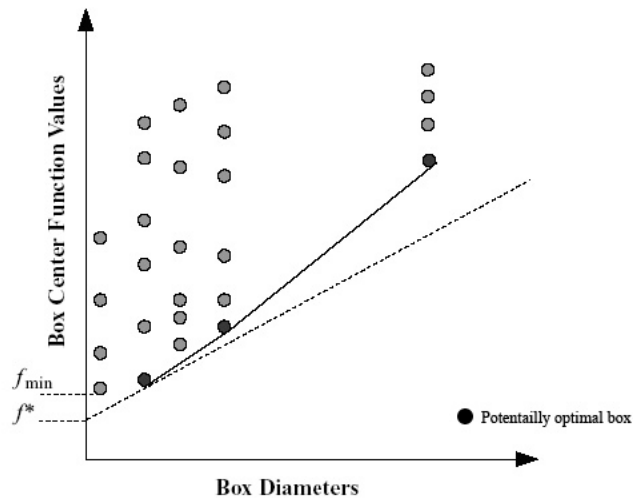


Figure 4 Illustration of potentially optimal hyperrectangles (called boxes in this Figure). Note that $f^* = f_{\min} - \varepsilon|f_{\min}|$. Potentially optimal hyperrectangles are on the lower-right convex hull.

The algorithm starts with mapping the hyperrectangle Q to the n -dimensional unit hypercube. Direct then samples the function at the center of this hypercube and at $2n$ more points, 2 in each coordinate direction. Using these function values, Direct then divides the domain into hyperrectangles, each having exactly one of the sampling points as its center. In each iteration, Direct chooses some of the existing hyperrectangles to be further divided. This algorithm is shown as algorithm 2.

Algorithm 2 Direct ($a, b, f, \varepsilon, numit, numfunc$)

1. Normalize the search space D to be the unit hypercube with center point c_1
 2. Evaluate $f(c_1)$, initialize $f_{\min} = f(c_1)$, evaluation counter $m = 1$, and iteration counter $t = 0$.
 3. while $t < numit$ and $m < numfunc$ do
 - 3.1. Identify the set S of the potentially optimal hyperrectangles
 - 3.2. while $S \neq \emptyset$ do
 - 3.2.1. Take $j \in S$
 - 3.2.2. Sample new points, evaluate f at the new points and divide the hyperrectangle j
 - i. Identify the set I of dimensions with the maximum side length. Let δ equal one-third of this maximum side length.
 - ii. Sample the function at the points $c \pm \delta e_i$ for all $i \in I$, where c is the center of the box and e_i is the i th unit vector.
 - iii. Divide the box j containing c into thirds along the dimensions in I , starting with the dimension with the lowest value of $w_i = \min\{f(c + \delta e_i), f(c - \delta e_i)\}$, and continuing to the dimension with the highest w_i .
 - 3.2.3. Update $f_{\min}, m = m + \Delta m$
 - 3.2.4. Set $S = S \setminus \{j\}$
 - 3.3. end while
 - 3.4. $t = t + 1$
 4. end while
-

Jones et al.¹¹ provide a good step-by-step example visualizing how DIRECT accomplishes the task of locating a global optimum. Step 3 forms a processing loop controlled by two stopping criteria: limits on iterations and function evaluations. Starting from the center of the initial hypercube, DIRECT makes exploratory moves across the design space by probing the potentially optimal subspaces. “Potentially optimal” is an important concept and has been explained above.

Direct is employed as described by Gablonsky.¹²⁻¹³ In test calculations performed with Direct alone, we found that for our present problem it is not competitive with the other methods presented here (explained in more detail in section 3.3.1). It typically needs about an order of magnitude more computing time or number of function evaluations, even compared to the simple GA. A possible reason could be that as the simple GA it does not use gradient information but its heuristics in choosing new points appear to be weaker and less global. Thus, for our present problem, local methods such as Powell and Iffco were used together with Direct. After the Direct optimization had found the best point for a limited budget, the local method used this point as a starting point for a local optimization. For this study, we have set the stopping criterion to 10^{-1} (it ensures sufficient decrease in function value when a new potentially optimal interval is chosen), the maximum number of iterations in Direct is 1000 and the termination tolerance for the local methods is set equal to 10^{-4} .

2.4 Implicit filtering for constrained optimization (IFFCO)

Iffco¹⁴⁻¹⁵ is a projected quasi-Newton method for bound constrained optimization problems. In contrast to simpler standard algorithms like conjugate gradients, the Iffco method uses a decreasing sequence of finite difference steps (scales) to approximate the gradient. The method is intended for noisy problems with many local minima. Specifically, Iffco is designed to solve problems where the function f has the form

$$f(x) = g(x) + o(x) \quad (8)$$

The function $g(x)$ is a smooth function with a simple form, such as a convex quadratic, within the hyper-box defined by the constraints. The function $o(x)$ is a low amplitude high frequency function. Iffco is especially effective on problems where the amplitude of $o(x)$ decays near minima of $g(x)$. However, Iffco has proven to be effective on more complex problems. Roughly, Iffco can be understood as a quasi-Newton method that employs a certain range of finite-difference approximations to the gradient, $d(x)$, (scales) and thus can be protected against getting stuck in small-scale local minima.

Because analytic second derivatives may not be available and because the cost of function evaluations may be too high to use the difference Hessian, SR1 (Symmetric Rank one) approximations and BFGS approximations to the Hessian can be used. The SR1 approximation is given by the formula:

$$S^c = S^- + \frac{rr^T}{r^T s} \quad (9)$$

where S^- is the previous SR1 approximation to the Hessian,

$$\begin{aligned} s &= x^c - x^- \\ r &= y - S^- s \\ y &= d(x^c) - d(x^-) \end{aligned} \quad (10)$$

and x^- is the previous iterate to x^c the current iterate. The BFGS approximation is given by the formula

$$B^c = \left(I - \frac{sy^T}{y^T s}\right) B^- \left(I - \frac{sy^T}{y^T s}\right) + \frac{ss^T}{y^T s} \quad (11)$$

where B^- is the previous BFGS approximation and I is a unit matrix. Unlike the SR1 approximation, with BFGS the approximate inverse Hessian is tracked. Here s and y are still defined as in the SR1 case. To accommodate the constraints a reduced approximation is used in both cases. The reduced approximation is obtained by setting the off-diagonal elements in the i th row and column to zero if the constraints on the i th variable are active. To keep the sequence of quasi-Newton steps within the hyper-box, each step is projected onto the hyper-box using the following formula

$$P(x)_i = \begin{cases} u_i, & \text{if } x_i > u_i \\ x_i, & \text{if } l_i \leq x_i \leq u_i \\ l_i, & \text{if } x_i < l_i \end{cases} \quad (12)$$

Iffco is a variation on the gradient projection method that uses a sequence of finite difference steps (scales) to approximate the gradient. A brief outline of the algorithm used in Iffco is given in Algorithm 3. A variation of Algorithm 3 is to start Algorithm 3 using the last point obtained in the iterative process as the next initial point until the point is obtained that satisfies the termination criteria at every scale. A point that satisfies the termination criteria at every scale is called a minimum at all scales.

For this study, Iffco is employed as described by Choi¹⁶. It was run with a termination tolerance of 10^{-6} , and optimization terminated once the maximum iterations on a given scale were exceeded or stencil failure occurred. The stencil is the set of points used to compute the difference gradient. A stencil failure occurs when all of the points in the (centered difference) stencil have a larger function value than the central point, which normally indicates that Iffco has succeeded in finding a (local) minimum and has no options to proceed further.

Algorithm 3 Iffco

1. Pick initial x and h ; find $f(x)$ and the difference gradient $\nabla_h f(x)$
 2. Initialize the model Hessian B to the identity
 3. while h and $\nabla_h f(x)$ satisfy conditions do
 - 3.1. use $\nabla_h f(x)$ and B to calculate a descent direction d , this step is a Quasi-Newton step.
 - 3.2. Perform a linesearch in the direction d , and signal success if some criteria are met.
 - 3.3. If linesearch was successful then
 - 3.3.1. Accept new point and project into the box Q .
 - 3.4. else
 - 3.4.1. $h \leftarrow h/2$
 - 3.5. end if
 - 3.6. Calculate the difference gradient $\nabla_h f(x)$
 - 3.7. Update B with either a rank-one SR1 update, or rank-two BFGS update
 4. end while
-

2.5 Downhill Simplex Method

A simplex⁸ is a geometrical figure consisting, in N dimensions, of $N + 1$ points (or vertices) and all their interconnecting line segments, polygonal faces, etc. In two dimensions, a simplex is a triangle. In three dimensions it is a tetrahedron, not necessarily the regular tetrahedron. In general simplexes are nondegenerate, i.e., they enclose a finite inner N -dimensional volume. If any point of a nondegenerate simplex is taken as the origin, then the N other points define vector directions that span the N -dimensional vector space. The algorithm starts with guessing an N -vector of independent variables as the first point to try. The algorithm is then supposed to make its own way downhill through the unimaginable complexity of an N -dimensional topography, until it encounters a (local, at least) minimum.

The downhill simplex method must be started not just with a single point, but with $N + 1$ points, defining an initial simplex. If one of these points (it does not matter which) is chosen as initial starting point P_0 , then the other N points are

$$P_i = P_0 + \lambda e_i \quad (13)$$

where the e_i 's are N unit vectors, and where λ is a constant which is a guess of the problem's characteristic length scale (or, it could be different λ_i 's for each vector direction.) The downhill simplex method consists of a series of steps, i.e., reflections and contractions. In a reflection step the algorithm moves the point of the simplex where the function is largest through the opposite face of the simplex to some lower point. If the algorithm reaches a "valley floor", the method contracts the simplex, i.e., the volume of the simplex decreases by moving one or several points, and moves along the valley.

2.6 Simulated annealing method

The method of simulated annealing⁸ is a global search technique that has attracted significant attention as being suitable for large optimization problems, especially ones where a desired global extremum is hidden among many local extrema. Given an essentially arbitrary way to generate steps of limited length in search space (for example, downhill and uphill simplex moves), the essential idea is to always accept downhill steps but to also accept uphill steps with a certain probability. This probability depends on the magnitude of the uphill move and on a control parameter T (Temperature). In the beginning, T should be so high that every point in search space is accessible. T is then lowered slowly, until the system is stuck in a single basin of attraction. One can prove that this final basin is the one of the global minimum, if the trajectory is of infinite length and the "cooling down" is slow enough. Obviously, this is hard to achieve in practice. Nevertheless, these basic methods as well as many sophisticated variations of it have been applied frequently with success.

3. OPTIMIZED AUXILIARY BASIS SETS FOR RI-MP2 AND RI-HF CALCULATIONS.

3.1 INTRODUCTION

Two factors limit the accuracy of traditional *ab initio* quantum chemical methods: One is the approximations inherent in the *ansatz* for the many-electron wavefunction itself, the other arises from its numerical implementation. The latter usually involves the use of finite sets of one-electron basis functions. These components have to be balanced so that no single component dominates either the computer time or the error. Unless general-purpose basis sets like plane waves are employed, these basis sets actually comprise only comparatively few functions, tailored to the expected form of the solution wavefunction. Hence, these basis functions have to be carefully selected and tuned. Surprisingly, this is still largely done by a combination of traditional empirical rules with educated trial and error. Such a procedure is not very attractive, in particular if basis sets are designed for other purposes, where the empirical rules may not apply. In such cases, crude optimization methods most probably yield only local minima of questionable quality. This is the situation in some recent variants of electronic structure methods, like the resolution-of-the-identity (RI) approximation where auxiliary basis sets are used to improve computational efficiency. With this work, we aim at introducing a more systematic and general approach to the task of generating such basis sets, applying different well established optimization algorithms capable of finding the global minima. As examples, we focus on RI-HF (sometimes also denoted density-fitting or Coulomb-fitting HF) and RI-MP2.

The amount of time spent computing the two-electron four-center integrals $(ai|bj)$ over basis functions a,b,i,j is a significant component of the total computer time of many *ab-initio* algorithms. For Hartree-Fock (HF) and second-order Møller-Plesset perturbation theory (MP2), the transformation of these integrals from atomic orbital (AO) to molecular orbital (MO) basis are N^4 and N^5 steps, respectively,

which constitutes the bottleneck of the whole method. Improvements in the computational efficiency of the basis integral evaluation algorithms have a significant effect on the overall speed of the calculation. An alternative, circumventing the problem completely, is the use of the resolution of the identity method (also referred to as density fitting; however, we here prefer to use the former name which leads to a less ambiguous abbreviation). It is a proven tool for obtaining an efficient and accurate approximation of the two-electron four-center integrals for HF¹⁷, MP2¹⁸⁻²¹ and also in density functional theory (DFT)²²⁻²³. The RI-MP2 and RI-HF algorithms use this approximation to reduce the overall computational cost by replacing the complete set of 4-index integrals with a sum of a smaller set of 3-index and 2-index integrals that are cheaper to compute. The evaluation in this way scales as N^2 for RI-HF. For RI-MP2, this scheme still scales with N^5 , but the most demanding N^5 step involves straightforward matrix multiplication and the prefactor of the N^5 step is very small. In contrast to conventional MP2, the N^5 step cannot be reduced by integral prescreening.

A new approach has recently been developed by Manby *et al.*²⁴⁻²⁵ and extended to MP2²⁶ and HF²⁷ (see also the earlier work in DFT²⁸). They use the Poisson equation connecting the density $\rho(r)$ and the resulting potential $v(r)$. The advantage of such a basis is that the Poisson equation can be used to reduce the 2-index and 3-index Coulomb integrals to simple overlap-like integrals, analogous in form to kinetic energy integrals. Therefore, this modified RI method carries the promise for even less expensive calculations. However, at the present time much less experience is available for the proper choice of Poisson auxiliary bases, making reliable and efficient optimization approach all the more valuable.

The nature of the RI approximation makes it especially useful in studies where large basis sets are used to obtain high-quality results.²⁹ Therefore, applications of RI methods require high-quality auxiliary basis sets to provide good approximations to the charge distributions and to introduce only small errors compared to the errors due to the finite standard basis set. In the last few years, the

series of the correlation consistent basis sets cc-pVXZ with $X = D, T, Q, \dots$ ³⁰⁻³¹ have become the most popular basis sets in correlated *ab-initio* methods. These are sets of Gaussian primitives and contractions that were optimized for electron correlation methods and are widely used in electronic structure calculations. The series cc-pVDZ, cc-pVTZ, cc-pVQZ, etc. leads to well-behaved, regular convergence of the energy towards the complete basis set limit. Auxiliary density basis sets complementing these standard cc-pVXZ basis sets for RI-MP2 were reported by Weigend et al.³² Since the object being expanded by an auxiliary basis is not the same as for standard basis sets, much less a priori information is available, making the task of generating an auxiliary basis even more difficult. Here we present a systematic generation of mixed density-Poisson auxiliary basis sets like those reported by Manby *et al.*²⁵, as an example for a more systematic approach to basis set generation.

Due to the high dimensionality of the search space, the mix of linear and nonlinear parameters and the non-trivial functional forms involved, the task of finding good (standard or auxiliary) basis sets^{17,21,32} with traditional local optimization methods was found to be expensive and unreliable. We expect that more refined local optimization methods and, at least for larger basis sets, global optimization methods³³ will perform better. The task is further complicated by the fact that no derivative information is available (definitely not if we plan to interface our optimization methods with arbitrary standard quantum chemistry packages). Also, some reasonable bounds on the ranges of the variables to be optimized can be guessed or found by easy experimentation, and the optimization method should be capable of obeying these bounds and utilizing their information.

Therefore, we have decided to test the performance of several different algorithms, belonging to different families of optimization strategies, for the task of generating auxiliary basis sets for RI-MP2 and RI-HF: Genetic Algorithms, Direct search and Implicit Filtering. The first two methods are global optimization algorithms. Direct search is deterministic and, with unlimited budget, exhaustive and hence globally convergent. From our experience with global optimization in search spaces of high dimensionality, we expect problems for such a deterministic method

for the optimization of larger basis sets. Therefore, we have also chosen Genetic Algorithms as a typical stochastic-heuristic global search for comparison. We expect advantages for such a method for larger problems, such as geometry optimization of clusters³⁴, crystal structure prediction³⁵ and protein folding³⁶. Moreover, a local direction set (Powell) method was used as a hybrid algorithm with GA and Direct. As explained below, in our work presented here, we are optimizing parameters of even-tempered sets of basis functions only. Therefore, the dimensionality and complexity of the search space are still moderate, and local methods may still be effective, provided they go beyond standard implementations of the simplest recipes. As an example of such a method, we decided to test the IFFCO program (Implicit Filtering For Constrained Optimization), a local optimization for noisy problems with constraints. Furthermore, this method has also been applied as a hybrid algorithm with Direct.

Proponents of linear scaling electron correlation methods speculate that ultimately it should be simpler to calculate electron correlation corrections (because electron correlation is a strictly local effect, falling off with r^{-6}) than to do the HF calculation. Actually, even with conventional implementations, the simplest levels of electron correlation, like MP2, can be less expensive than the HF step, for molecules of limited size. Therefore, it is important not only to improve the efficiency of the electron correlation methods but also that of the basic HF calculation. This is why we look at both RI-HF and RI-MP2 here.

3.1.1 Hartree-Fock (HF) theory

Hartree-Fock theory³⁷⁻³⁹ is fundamental to much of electronic structure theory because the equations are relatively simple and yet much profound chemistry can be described by them. The problem that quantum chemistry faces is the solution of the Schrödinger equation.

$$\hat{H}\Psi = E\Psi \tag{14}$$

\hat{H} is the Hamiltonian operator for the nuclei and electrons in a molecule

$$\hat{H} = \hat{T}_e + \hat{T}_n + \hat{V}_{ee} + \hat{V}_{en} + \hat{V}_{nn} \quad (15)$$

where \hat{T}_e , \hat{T}_n are kinetic energy of the electrons and nuclei, respectively. \hat{V}_{ee} , \hat{V}_{en} , \hat{V}_{nn} are electron-electron repulsion, electron-nuclear attraction and internuclear repulsion, respectively (written here in atomic units).

$$\hat{H} = \sum_{i=1}^N \left(-\frac{1}{2} \nabla_i^2 \right) + \sum_{a=1}^M \left(-\frac{1}{2M_a} \nabla_a^2 \right) + \sum_{i=1}^N \sum_{j>i}^N \frac{1}{|r_i - r_j|} - \sum_{i=1}^N \sum_{a=1}^M \frac{Z_a}{|r_i - R_a|} + \sum_{a=1}^M \sum_{b>a}^M \frac{Z_a Z_b}{|R_a - R_b|} \quad (16)$$

$$\hat{H}_e = \sum_{i=1}^N \hat{h}_i + \sum_{i=1}^N \sum_{j>i}^N \hat{g}_{ij} \quad (17)$$

The 1-electron terms;
$$\hat{h}_i = -\frac{1}{2} \nabla_i^2 - \sum_{a=1}^M \frac{Z_a}{|r_i - R_a|} \quad (18)$$

The 2-electron terms;
$$\hat{g}_{ij} = \frac{1}{|r_i - r_j|} \quad (19)$$

The molecular Schrödinger equation is commonly simplified using the Born-Oppenheimer separation, which derives from the observation that nuclei are much heavier than electrons, and consequently move much more slowly than do electrons. Therefore, the problem can be separated into two parts: The electronic problem, in which the nuclei are effectively stationary with respect to electron motion, and the nuclear problem, which can be treated after the electronic part is solved. The nuclear problem will not be considered further here. Thus, for the electronic problem, the molecular Hamiltonian reduces to \hat{H}_e , the electronic Hamiltonian, and is given by

$$\hat{H}_e = \hat{T}_e + \hat{V}_{ee} + \hat{V}_{en} \quad (20)$$

The solutions to \hat{H}_e are expressed using molecular spin orbitals φ_i . They are products of a spatial function ϕ_i and a spin function ω_i .

$$\varphi_i(x) = \phi_i(r)\omega_i(s) \quad (21)$$

For spin $-\frac{1}{2}$ electrons, only two choices of ω are possible, $\alpha(s)$ and $\beta(s)$, corresponding to spin up and down.

$$\omega(s) = \begin{cases} \alpha(s) \\ \beta(s) \end{cases} \quad (22)$$

These are functions of an unspecified spin variable s . From the operational point of view we need only specify that the two spin functions are complete and that they are orthonormal,

$$\langle \alpha | \alpha \rangle = \langle \beta | \beta \rangle = 1 \quad (23)$$

$$\langle \alpha | \beta \rangle = \langle \beta | \alpha \rangle = 0 \quad (24)$$

The spatial orbital ϕ_i is a function of the position r_i of electron i ($\vec{r}_i = (x_i, y_i, z_i)^T$; the vector notation is suppressed for clarity). ϕ_i describes the spatial distribution of electron i such that $|\phi_i(r)|^2 dr$ is the probability of finding the electron in the volume element dr . Spin orbitals of different spin functions are orthogonal, with $\langle \alpha | \beta \rangle = 0$

$$\varphi_i = \phi_i \alpha, \varphi_j = \phi_j \beta \quad \Rightarrow \quad \langle \varphi_i | \varphi_j \rangle = \langle \phi_i | \phi_j \rangle \langle \alpha | \beta \rangle = 0 \quad (25)$$

The spatial parts of an orthonormal set of spin orbitals with the same spin direction are orthonormal

$$\varphi_i = \phi_i \alpha, \varphi_j = \phi_j \alpha \Rightarrow \delta_{ij} = \langle \varphi_i | \varphi_j \rangle = \langle \phi_i | \phi_j \rangle \langle \alpha | \alpha \rangle = \langle \phi_i | \phi_j \rangle \quad (26)$$

The basic idea of Hartree-Fock theory assumes a system of N non-interacting electrons (i.e. without Coulomb repulsion among each other) that are characterized by a one-particle Hamiltonian \hat{h} ,

$$\hat{h}\varphi_j(x) = \varepsilon_j \varphi_j(x) \quad (27)$$

with

$$\hat{h} = -\frac{1}{2}\nabla^2 - \sum_{a=1}^M \frac{Z_a}{|r - R_a|}. \quad (28)$$

The corresponding many-particle Hamiltonian is $\hat{H} = \sum_{i=1}^N \hat{h}_i$. The resulting Schrödinger equation is solved exactly by a Hartree product of spin orbitals

$$\Psi_{HP}(x_1, x_2, \dots, x_N) = \varphi_1(x_1)\varphi_2(x_2)\dots\varphi_N(x_N) \quad (29)$$

The Hartree product fails to satisfy the anti-symmetry principle. The Pauli Exclusion Principle states that a wavefunction must change sign when any set of space spin coordinates of any two electrons are exchanged.

$$\Psi(\dots, x_i, \dots, x_k, \dots) = -\Psi(\dots, x_k, \dots, x_i, \dots) \quad (20)$$

It is a simple consequence of the fact that the elements are elementary particles and hence indistinguishable. The sign change is due to the fermionic nature of the electrons. A convenient method of making a simple product wavefunction anti-symmetric is to replace it by a Slater determinant of order N , which is an

antisymmetric linear combination of Hartree products that contain all possible electron permutations.

$$\Psi_{SD} = (N!)^{-1/2} \begin{vmatrix} \varphi_1(x_1) & \varphi_2(x_1) & \cdots & \varphi_N(x_1) \\ \varphi_1(x_2) & \varphi_2(x_2) & \cdots & \varphi_N(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_1(x_N) & \varphi_2(x_N) & \cdots & \varphi_N(x_N) \end{vmatrix} \quad (31)$$

The $(N!)^{-1/2}$ is the normalization factor. For convenience, two shorthand notations are often used for equation (31). The first is

$$\Psi_{SD}(1,2,\dots,N) = (N!)^{1/2} A[\varphi_1(x_1)\varphi_2(x_2)\dots\varphi_N(x_N)] \quad (32)$$

It uses the anti-symmetry operator A to represent the determinant and explicitly normalizes the wavefunction. The second uses Dirac bracket notation to represent both the Slater determinant and normalization constant.

$$\Psi_{SD}(1,2,\dots,N) = |\varphi_1(x_1)\varphi_2(x_2)\dots\varphi_N(x_N)\rangle \quad (33)$$

Given a closed-shell restricted wavefunction with N doubly-occupied orbitals, equation (33), by replacing φ_i from equation (21),

$$\Psi = |\phi_1(1)\alpha(1)\phi_1(2)\beta(2)\dots\phi_N(2N-1)\alpha(2N-1)\phi_N(2N)\beta(2N)\rangle \quad (34)$$

we can simplify the notation by writing

$$\Psi = |\phi_1\bar{\phi}_1 \dots \phi_N\bar{\phi}_N\rangle \quad (35)$$

where a bar over the orbital signifies spin down, and no bar signifies spin up. The Slater determinant is normalized when the set of constituting spin orbitals is

orthonormal.

$$\langle \varphi_i | \varphi_j \rangle = \delta_{ij} \quad \Rightarrow \quad \langle \Psi_{SD} | \Psi_{SD} \rangle = 1 \quad (36)$$

The Slater determinants still are exact eigenfunctions of the simplified Hamiltonian $\hat{H} = \sum_i \hat{h}_i$, i.e. neglecting electron-electron repulsion. Conversely, they can not be exact eigenfunctions in the presence of electron-electron repulsion. The energy of Ψ is given by

$$E = \frac{\langle \Psi | \hat{H} | \Psi \rangle}{\langle \Psi | \Psi \rangle} \quad (37)$$

The variational principle associated with this expression states that the energy is always an upper bound to the true energy. The best wavefunction of this functional form is the one which gives the lowest possible energy. If one inserts the simple Slater determinant form of Ψ into equation (37), one gets terms of the form $\langle \Psi | V_{ee} | \Psi \rangle$. The essence of the Hartree-Fock approximation thus is to replace the complicated many-electron problem by a one-electron problem in which electron-electron repulsion is treated in an average way, and to determine the orbitals φ_i not from the one-electron problem equation (27) but as variational items by minimizing E in the many-electron expression equation (37).

Inserting equation (35) into equation (37), we get:

$$E = \langle \phi_1 \bar{\phi}_1 \dots \phi_N \bar{\phi}_N | \sum_{i=1}^N \hat{h}_i + \sum_{i=1}^N \sum_{j>i}^N \hat{g}_{ij} | \phi_1 \bar{\phi}_1 \dots \phi_N \bar{\phi}_N \rangle \quad (38)$$

We start with the evaluation of the one-electron contribution and use the fact that the anti-symmetrization operator commutes with the 1- and 2-electron parts of the Hamiltonian. The complicated integrals over Slater determinants can be considerably

simplified to expressions involving only integrals over a few orbitals. This can be found as Slater-Condon rules in many textbooks and shall not be repeated here.

The 1-electron terms are given by

$$E = \langle \phi_1 \bar{\phi}_1 \dots \phi_N \bar{\phi}_N | \sum_{i=1}^N \hat{h}_i | \phi_1 \bar{\phi}_1 \dots \phi_N \bar{\phi}_N \rangle = \sum_{i=1}^N 2\hat{h}_{ii} \quad (39)$$

The 2-electron terms are given by

$$E = \langle \phi_1 \bar{\phi}_1 \dots \phi_N \bar{\phi}_N | \sum_{i=1}^N \sum_{j>i}^N \hat{g}_{ij} | \phi_1 \bar{\phi}_1 \dots \phi_N \bar{\phi}_N \rangle = \sum_{i,j=1}^N (2J_{ij} - K_{ij}) \quad (40)$$

The electron energy is given by

$$E = \sum_{i=1}^N 2\hat{h}_{ii} + \sum_{i,j=1}^N (2J_{ij} - K_{ij}) \quad (41)$$

The Coulomb operator,

$$J_i(1) = \int dr_2 \frac{\phi_i^*(2)\phi_i(2)}{r_{12}} \quad (42)$$

is an effective 1-electron operator, where the $\frac{1}{r_{12}}$ electron interaction is averaged over the probability distribution of electron 2. Its matrix element is the Coulomb integral:

$$J_{ij} = \int dr_1 \phi_j^*(1) J^i(1) \phi_j(1) = \int dr_1 \phi_i^*(1) J^j(1) \phi_i(1) \quad (43)$$

This 2-electron integral is commonly written $(ii|jj)$, where the first half of the symbol corresponds to electron 1 and the second part of the symbol corresponds to electron 2.

The Coulomb integral J_{ij} corresponds to the classical Coulomb interaction between the charge densities $\rho_i(r) = |\phi_i|^2$ and $\rho_j(r) = |\phi_j|^2$. Its value is independent of the spin parts of the spin orbitals ϕ_i and ϕ_j .

The K_{ij} terms are matrix elements of the exchange operator, which is purely a manifestation of the anti-symmetry of the wavefunction and has no macroscopic correspondence. The 1-particle exchange operator $K_i(1)$ is most easily defined in terms of its action on another orbital ϕ_j ;

$$K_i(1)\phi_j(1) = \left(\int dr_2 \frac{\phi_i^*(2)\phi_j(2)}{r_{12}} \right) \phi_i(1) \quad (44)$$

The K_{ij} matrix element is given by

$$K_{ij} = \left(\int dr_1 \int dr_2 \frac{\phi_i^*(1)\phi_j(1)\phi_j^*(2)\phi_i(2)}{r_{12}} \right) \quad (45)$$

This exchange integral is often written $(ij|ji)$. Its value is zero if the spin orbitals ϕ_i and ϕ_j have opposite spin directions.

The variation method to determine the optimal single-determinant wavefunction by minimizing the expectation value of the total energy is called Hartree-Fock method. The optimal orbitals ϕ_i are those that give the lowest energy with respect to changes in the orbital $(\phi_i + \delta\phi_i) = (\phi_i + \delta)$. This procedure is implemented by varying the spin orbital ϕ_i subject to the orthonormality constraints that ensure the normalization of the total wavefunction. The resulting Fock equation determines the canonical set of optimal spin orbitals

$$\hat{F}\phi_i(x) = \varepsilon_i\phi_i(x) \quad (46)$$

where \hat{F}_i is the Fock operator given by

$$\hat{F}_i = \hat{h} + \sum_{i=1}^N (\hat{J}_i - \hat{K}_i) \quad (47)$$

The Hartree-Fock equation (46) can not be solved directly since the operators \hat{J}_i , \hat{K}_i and also \hat{F} depend on the solutions ϕ_i . Therefore, the eigenvalue problem has to be solved iteratively: One guesses solutions $\phi_i^{(0)}$, constructs a Fock operator \hat{F} , solves the corresponding HF equation for new solutions $\phi_i^{(1)}$, and so on – until no significant changes occur. Then, one has reached self-consistency: The eigenvectors of the Fock operator generate the Fock operator. One has generated a self-consistent field (SCF).

Introducing a basis set transforms the Hartree-Fock equations into the Roothan equations. Denoting the atomic orbital basis function as χ_a , we have the expansion

$$\phi_i = \sum_{\alpha}^M C_{ai} \chi_{\alpha} \quad (48)$$

Now the Hartree-Fock-Roothan equations can be written in matrix form as

$$FC = SC\varepsilon \quad (49)$$

where ε is a diagonal matrix of the orbital energies ε_i . This is a generalized eigenvalue equation due to the overlap matrix S . To solve it, one performs a transformation of basis to go to an orthogonal basis to make S vanish. Then, it is an ordinary eigenvalue equation. However, since the Fock matrix depends on the solution matrix C , equation (49) can still only be solved iteratively. Therefore, the solution of the Hartree-Fock-Roothan equations is often called the SCF procedure.

3.1.2 Møller-Plesset Perturbation Theory

Hartree Fock theory assumes a single Slater determinant wavefunction appropriate for non-interacting particles, resulting in an averaged, effective single-particle treatment neglecting electron correlation. One way to account for electron correlation is to correct the HF results a posteriori, treating electron correlation as a perturbation. Perturbation theory³⁷⁻³⁹ is based upon dividing the Hamiltonian into two parts, a reference \hat{H}_0 , and a perturbation \hat{H}' .

$$\hat{H} = \hat{H}_0 + \lambda\hat{H}' \quad (50)$$

where \hat{H}_0 is an exactly solvable Hamiltonian, λ is a parameter formally controlling the size of the perturbation, and the \hat{H}' term is the perturbation. We assume that the energy eigenvalue problem for the Hamiltonian \hat{H}_0 has been solved exactly.

$$\hat{H}_0\phi_i = E_i\phi_i \quad , i = 0,1,2,\dots,\infty \quad (51)$$

The perturbed Schrödinger equation is

$$\hat{H}\Psi = E\Psi \quad (52)$$

As the perturbation is increased from zero to a finite value, the new energy and wavefunction must also change continuously; therefore, they can be written formally as series expression in λ :

$$(\hat{H}_0 + \lambda\hat{H}')\{\phi_0 + \lambda\phi_1 + \dots\} = (E_0 + \lambda E_1 + \lambda^2 E_2 + \dots)\{\phi_0 + \lambda\phi_1 + \dots\} \quad (53)$$

After expanding the products, we can equate the coefficients on each side of the equation for each power of λ , leading to a series of relations representing

successively higher orders of perturbation. E_1 is the first-order correction to the energy, E_2 is the second-order correction and so on.

$$\begin{aligned} E_0 &= \int \phi_0 \hat{H}_0 \phi_0 d\tau \\ E_1 &= \int \phi_0 \hat{H}' \phi_0 d\tau \\ E_2 &= \int \phi_0 \hat{H}' \phi_1 d\tau \end{aligned} \quad (54)$$

To determine the corrections to the energy it is therefore necessary to determine the wavefunctions to a given order. The choice of \hat{H}' is not unique. To achieve size-consistency, it is common to follow Møller and Plesset who proposed to take the unperturbed Hamiltonian \hat{H}_0 as the sum of the 1-electron Fock operators for the N electrons,

$$\hat{H}_0 = \sum_{i=1}^N \hat{F}_i = \sum_{i=1}^N \hat{h}_i + \sum_{i=1}^N \sum_{j=1}^N \langle \hat{g}_{ij} \rangle \quad (55)$$

The corresponding zeroth-order energy E_0 is equal to the sum of orbital energies for the occupied molecular orbitals

$$E_0 = \sum_{i=1}^N \varepsilon_i \quad (56)$$

The perturbed Hamiltonian \hat{H}' is given by

$$\hat{H}' = \hat{H} - \hat{H}_0 = \hat{H} - \sum_{i=1}^N \hat{F}_i = \sum_{i=1}^N \sum_{j>i}^N \hat{g}_{ij} - \sum_{i=1}^N \sum_{j=1}^N \langle \hat{g}_{ij} \rangle \quad (57)$$

which is the difference between the exact electron-electron repulsion and the HF averaged one.

The equation in λ^1 is the first-order perturbation equation;

$$E_1 = \langle \phi_0 | \hat{H}' | \phi_0 \rangle = \langle \phi_0 | \hat{H} | \phi_0 \rangle - \langle \phi_0 | \sum_{i=1}^N \hat{F}_i | \phi_0 \rangle \quad (58)$$

$$E_1 = E_{HF} - \sum_{i=1}^N \varepsilon_i = E_{HF} - E_0 \quad (59)$$

Therefore, $E_{HF} = E_0 - E_1$, and after having performed a HF calculation, the first correction due to electron correlation is E_2 . This level of theory is referred to as MP2 and involves the integral $\int \phi_0 \hat{H}' \phi_1 d\tau$. The higher-order wavefunction ϕ_1 is expressed as linear combinations of solutions to the zeroth-order Hamiltonian:

$$\phi_1 = \sum_n c_1^n \phi_0^n \quad (60)$$

By using the expansion in equation (53) and rearranging, an expression for the coefficients can be found:

$$c_1^n = - \frac{\langle \phi_0^0 | H' | \phi_0^n \rangle}{E_0^0 - E_0^n} \quad (61)$$

Then E_2 in equation (54) can be written:

$$E_2 = \sum_{n \neq 0} \frac{|\langle \phi_0^0 | \hat{H}' | \phi_0^n \rangle|^2}{E_0^0 - E_n^0}, \quad |\phi_0^0\rangle = \phi_{HF} \quad (62)$$

The equation in λ^2 , the second-order perturbation equation, involves matrix elements of the perturbation operator between the HF reference and all possible excited states. Since the perturbation is a 2-electron operator, all matrix elements involving triple-,

quadruple etc. excitations are zero. When canonical HF orbitals are used, matrix elements with singly excited states are also zero.

$$\begin{aligned}\langle \phi_0^0 | \hat{H}' | \phi_0^n \rangle &= \langle \phi_{HF} | \hat{H} - \sum_{j=1}^N \hat{F}_j | \phi_0^n \rangle \\ &= \langle \phi_{HF} | \hat{H} | \phi_0^n \rangle - \langle \phi_0^i | \hat{F} | \phi_0^n \rangle = 0\end{aligned}\quad (63)$$

The first term vanishes because of Brillouin's theorem; the second term vanishes since the two wavefunctions are eigenfunctions of the Fock operator but orthogonal to each other. The second-order correction to the energy, which is the first contribution to the correlation energy, therefore only involves a sum over doubly excited determinants. These can be generated from the HF determinant by promoting two electrons from occupied orbitals i and j to virtual orbitals a and b .

$$E_2 = \sum_{i < j}^{occ} \sum_{a < b}^{vir} \frac{\langle \phi_0 | \hat{H}' | \phi_{ij}^{ab} \rangle \langle \phi_{ij}^{ab} | \hat{H}' | \phi_0 \rangle}{E_0 - E_{ij}^{ab}} \quad (64)$$

The matrix elements between the HF and doubly excited states are given by 2-electron integrals over MOs. The difference in total energy between two Slater determinants becomes a difference in MO energies, and the explicit formula for the second-order Møller-Plesset correction is

$$E_{MP2} = \sum_{i < j}^{occ} \sum_{a < b}^{vir} \frac{[\langle \phi_i \phi_j | \phi_a \phi_b \rangle - \langle \phi_i \phi_j | \phi_b \phi_a \rangle]^2}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b} \quad (65)$$

Once the 2-electron integrals over MOs are available, the second-order energy correction can be calculated directly from equation (65) as a sum over such integrals. The advantage of this is that each order n of perturbation theory MP n is size consistent. However, the MP n formalism is not a variational method.

3.1.3 Resolution of the identity approximation

The 4-index integrals that appear in the Hartee-Fock (HF) and second-order Møller-Plesset perturbation theory (MP2) are problematic because their number increases as N^4 with basis set size N . This problem can be avoided by the resolution of the identity (RI) approximation. The basic approach of the RI method is to factor the 4-index integrals into two parts.¹⁷⁻²³

$$(ai|bj) = \int dr_1 \int dr_2 \frac{\phi_a(r_1)\phi_i(r_1)\phi_b(r_2)\phi_j(r_2)}{r_{12}} \quad (66)$$

$$(ai|bj) = \int dr_1 \int dr_2 \frac{\rho_{ia}(r_1)\rho_{jb}(r_2)}{r_{12}} \quad (67)$$

In the RI method, products of basis functions, $\rho_{ia}(r) = \phi_i(r)\phi_a(r)$, are approximated by a linear expansion of so-called auxiliary basis functions, $\Xi(r)$,

$$\tilde{\rho}_{ia}(r) = \sum_A d_A^{ia} \Xi_A(r) \quad (68)$$

Here and in the following "basis function" denotes the (usual) basis functions that are linearly combined to describe molecular orbitals; the fit functions in equation (68) are always termed "auxiliary basis functions". The expansion coefficients d_A^{ia} are obtained by minimizing the positive definite functional

$$\Delta = \int dr_1 \int dr_2 \frac{(\rho_{ia}(r_1) - \tilde{\rho}_{ia}(r_1))(\rho_{jb}(r_2) - \tilde{\rho}_{jb}(r_2))}{r_{12}} \quad (69)$$

$$\Delta = \int dr_1 \int dr_2 \frac{(\rho_{ia}(r_1) - \sum_A d_A^{ia} \Xi_A(r_1))(\rho_{jb}(r_2) - \sum_B d_B^{jb} \Xi_B(r_2))}{r_{12}} \quad (70)$$

This leads to

$$d_B^{ia} = \sum_A (ia|A) V_{AB}^{-1} \quad (71)$$

d_B^{ia} are fitting coefficients for densities $\rho_{ia}(r)$, V_{AB}^{-1} is an element of the inverse square root of the (positive definite) metric matrix. Then, the 4-integrals are approximated as

$$(ai|bj) \approx \sum_B d_B^{ia} (B|bj) \quad (72)$$

$$(ai|bj) \approx \sum_{AB} (ai|A) V_{AB}^{-1} (B|bj) \quad (73)$$

where,

$$V_{AB}^{-1} = (A|B) = \int dr_1 \int dr_2 \frac{\Xi_A(r_1) \Xi_B(r_2)}{r_{12}} \quad (74)$$

$$(ai|A) = \int dr_1 \int dr_2 \frac{\phi_a(r_1) \phi_i(r_1) \Xi_A(r_2)}{r_{12}} \quad (75)$$

The labels a, b refer to virtual MOs, while i, j refer to active occupied valence MOs. The right-hand side of equation (73) involves only 2-index, $(A|B)$, and 3-index integrals, $(ai|A)$, which replace the much more numerous 4-index integrals. Thus, substantial savings can be made.

3.1.4 The Poisson method

Recently it was observed that an intriguing simplification follows if the fitting basis functions, $\Xi(r)$, are chosen to have a form which relies on the Poisson equation^{24–28}

$$\hat{P}v(r) = \rho(r) \quad (76)$$

where $\hat{P} = -\frac{1}{4\pi}\nabla^2$, which relates the density $\rho(r)$ to the Coulomb potential $v(r)$.

The Coulomb potential is given by the three-dimensional integral

$$v(r_1) = \int dr_2 \frac{\rho(r_2)}{r_{12}} \quad (77)$$

which has to be performed at each grid point r_1 . The novel *ansatz* for the density is

$$\check{\rho}_{ia}(r) = \sum_A d_A^{ia} \hat{P}\Xi_A^P(r) \quad (78)$$

instead of equation (68). The Coulomb energy has the form

$$E = \frac{1}{2} \int dr_1 \int dr_2 \frac{\rho_{ia}(r_1)\rho_{jb}(r_2)}{r_{12}} \quad (79)$$

From equation (76) and (78) we can immediately write down the Coulomb energy, \check{E} , arising from the density $\check{\rho}(r)$,

$$\check{E} = \frac{1}{2} \sum_{AB} d_A^{ia}(r_1) [J_{AB}] d_B^{jb}(r_2) \quad (80)$$

with the Coulomb integral, J_{AB} ,

$$\check{J}_{AB} = \int dr_1 \int dr_2 \frac{(\hat{P}_1\Xi_A^P(r_1))(\hat{P}_2\Xi_B^P(r_2))}{r_{12}} \quad (81)$$

where \hat{P}_i acts on the function in the coordinates r_i . The integral identity of interest here can be obtained by inserting the Poisson equation (76) into the potential expression in equation (77), which holds exactly for any $f(r)$ that vanishes at large r faster than r^{-1}

$$\int dr_1 \frac{\hat{P}f(r_1)}{r_{12}} = f(r_2) \quad (82)$$

Inserting the integral identity and using the Hermiticity of \hat{P} reduces the Coulomb matrix elements (without approximation) in the auxiliary basis from 6- to 3-dimensional integrals.

$$\check{J}_{AB} = \int dr \Xi_A^P(r) \hat{P} \Xi_B^P(r) = (A | \hat{P} | B) \quad (83)$$

and the 3-index integrals become overlaps,

$$(A | ai) = \int dr_1 \int dr_2 \frac{\hat{P} \Xi_A^P(r_1) \phi_a(r_2) \phi_i(r_2)}{r_{12}} \quad (84)$$

$$(A | ai) = \int dr \Xi_A^P(r) \phi_a(r) \phi_i(r) \quad (85)$$

the Coulomb energy reduces to

$$\check{E} = \frac{1}{2} \sum_{AB} d_A^{ia}(r) d_B^{jb}(r) \int dr \Xi_A^P(r) \hat{P} \Xi_B^P(r) \quad (86)$$

This equation is interesting because it gives an exact expression for the Coulomb energy of the density $\check{\rho}(r)$ using only short range 3-dimensional integrals (see above), which differ from kinetic energy integrals only by factor of $(2\pi)^{-1}$.

3.1.5 Basis sets

The molecular orbitals are expanded as a linear combination of atomic orbitals as basis functions. The basis functions collectively are the basis set.

$$\phi_i = \sum_{\alpha}^M C_{\alpha i} \chi_{\alpha} \quad (87)$$

As $M \rightarrow \infty$, the complete basis set limit is reached, which is not directly possible in practical calculations. When M is finite, this representation is approximate. Therefore, basis sets should be chosen as large as one is able to afford, without risking an approximate linear dependency that leads to numerical instabilities. The more complex basis sets are more accurate but they use up a great deal of computing time. Two criteria for selecting basis function are: They should be physically meaningful, and computation of the integrals should be tractable.

The expression for a basis function given as a Slater Type Orbital (STO) is,

$$\chi_{\zeta,n,l,m}(r, \theta, \varphi) = NY_{l,m}(\theta, \varphi)r^{n-1}e^{-\zeta r} \quad (88)$$

where N is a normalization constant, $Y_{l,m}(\theta, \varphi)$ is a spherical harmonic and ζ is a parameter in the exponent of the radial function. STOs are constructed from a radial part describing the radial extent of the orbital and an angular part describing the shape of the orbital. STOs depend on quantum numbers n , l , m and exponent ζ . The expression (88) is similar to hydrogenic wavefunctions, but without radial nodes which can be introduced by making linear combinations of several STOs and ensures fairly rapid convergence with increasing number of functions. STOs are not very useful for electronic structure calculations because 2-electron integrals of polyatomic molecules are difficult to calculate when three and four center integrals in a general (nonlinear) configuration are involved.

Gaussian type orbitals can be written in terms of spherical coordinates

$$\chi_{\alpha,l,m}(r,\theta,\varphi) = NY_{l,m}(\theta,\varphi)r^l e^{-\alpha r^2} \quad (89)$$

or Cartesian coordinates

$$\chi_{\alpha,m,n,p}(r,\theta,\varphi) = Nx^m y^n z^p e^{-\alpha r^2} \quad (90)$$

The sum of exponents in Cartesian coordinates, $l = m+n+p$, is used analogously to the angular momentum quantum number of atoms, to mark functions as s-type ($l=0$), p-type ($l=1$), d-type ($l=2$), f-type ($l=3$) etc. Spherical GTOs can be formed by linear combination of Cartesian GTOs of the same angular momentum l . GTOs exhibit a wrong behavior near the nucleus, ($r \approx 0$). Unlike STOs, GTOs have no cusp at the origin and they diminish too rapidly with distance as shown in Figure 5.

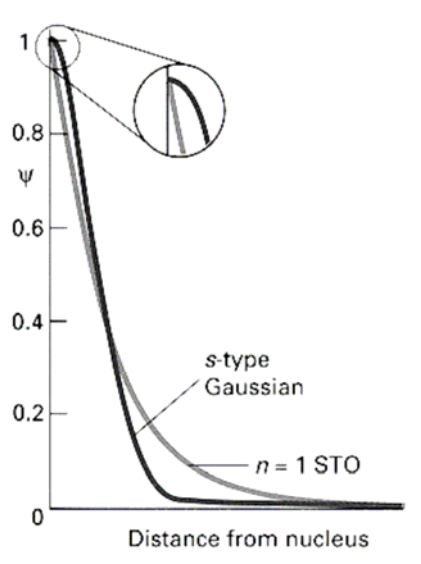


Figure 5 The 1s Slater-type orbital and the Gaussian-type orbital.

Therefore, many more GTOs are required to achieve the same accuracy as with STOs. Nevertheless, GTOs are by far the most common basis functions since

with these 4-center 2-electron integrals can be calculated almost analytically. One uses a linear combination of GTOs to overcome deficiencies for size and shape of the molecular charge distribution, as shown in Figure 6. It is clear that the fit improves as the number of Gaussian functions increases, but even so, the addition of many more GTOs can not properly describe the exponential tail in the true function and the cusp at the nucleus.

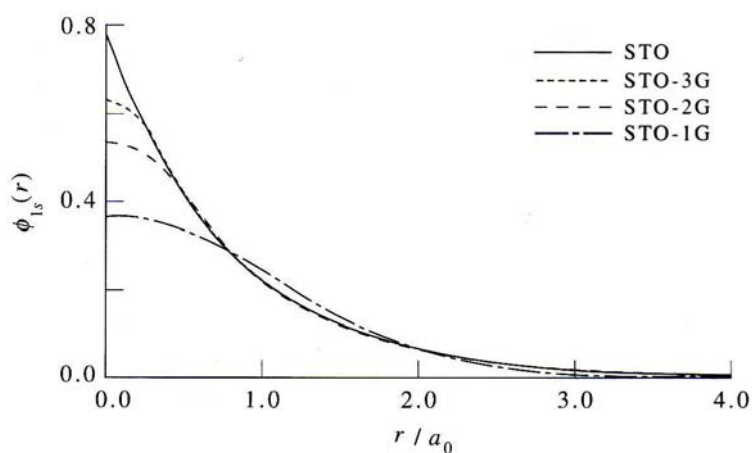


Figure 6 Comparison of 1s Slater-type orbital and Gaussian expansions with up to three terms.

For molecular calculations, we need to contract several Gaussian primitive functions together to form a single basis function. The term contraction means a linear combination of Gaussian primitives to be used as basis function. In the jargon, the functions that are contracted together are called primitive basis functions, and the resulting function is referred to as a contracted basis function. When quantum chemists refer to a basis function they typically mean a contracted basis function.

Types of basis sets

Minimal basis sets: Minimal basis sets contain only enough functions to accommodate the electrons of the neutral atoms, usually core plus valence orbitals, as in these examples:

H: 1s

C: 1s, 2s, 2p_x, 2p_y, 2p_z

Double Zeta (DZ) basis set: Double the number of all basis functions (two functions per shell or valence shell).

H: 1s, 1s'

C: 1s, 2s, 2s', 2p_x, 2p_y, 2p_z, 2p'_x, 2p'_y, 2p'_z

A DZ basis yields a better description of the charge distribution compared to a minimal basis.

Split valence basis set: Basis sets of this type are called "split valence" basis sets as the description of valence orbitals is split into two (or more) basis functions.

Example for carbon: 4-31 G

1s: 1 CGTO of 4 PGTOs,

2s/2p: 1 CGTO of 3 PGTOs for inner part plus 1 additional PGTO for outer part.

Doubling the number of functions provides a much better description of bonding in the valence region. Doubling the number of functions in the core region improves the description of energetically important but chemically uninteresting core electrons. Split valence basis sets improve the flexibility of the valence region and use a single (contracted) set of functions for the core.

Polarization functions: describe atomic orbital distortions due to bonding. Polarized basis sets add functions whose *l* quantum numbers are greater than the maximum *l* of the valence shell of the ground-state atom. Polarized basis sets add p functions for s-valence orbitals, d polarization functions for p-valence orbitals and f functions for d-valence orbitals.

Example carbon:

- 6-31G* \equiv 6-31G(d) : 6-31 basis plus 1d type polarization exponent
- 6-31G* * \equiv 6-31G(d,p) : 6-31 basis with 1d type polarization exponent for heavy elements and 1 p-type polarization exponent for hydrogen.

Diffuse functions: Diffuse basis functions are typically added as an additional set of uncontracted Gaussian functions of the same angular momentum as the valence electrons which allow orbitals to occupy a larger region of space. s-, p-, and d-functions with small exponents are usually added for systems where electrons are relatively far from the nucleus: molecules with lone pairs, anions and other systems with significant negative charge, systems in their excited states, systems with low ionization potentials, descriptions of absolute acidities, and so on.

6-31+G(d) : 6-31G(d) basis with diffuse functions added to heavy atoms.

6-31++G(d) : 6-31G(d) basis with diffuse functions added to the heavy atom and hydrogen atoms as well.

Diffuse functions on hydrogen atoms seldom make a significant difference in accuracy (for example, alkyl groups; but e.g. H-bonds are different)

High angular momentum basis sets: Even larger basis sets are now practical for many systems. Such basis sets add multiple polarization functions per atom to a double or triple zeta basis set. For example, the 6-31G(2d) basis set adds two d functions per heavy atom instead of just one, while the 6-311++G(3df,3pd) basis set contains three sets of valence region functions, diffuse functions on both heavy atoms and hydrogens, and multiple polarization functions: 3 d functions and 1 f function on heavy atoms and 3 p functions and 1 d function on hydrogen atoms. Such basis sets are useful for describing the interactions between electrons in electron correlation methods.

Correlation consistent basis sets: A series of basis sets for correlated calculations has also been developed by Dunning et al³⁰. Correlation consistent basis sets are built up by adding shells of functions to a core set of atomic Hartree-Fock functions. The name correlation consistent refers to the fact that the basis sets are designed such that each function in a shell contributes very similar amounts of correlation energy in an atomic calculation, independently of the function type. The 1st d-function provides a large energy lowering, but the contribution from a 2nd d-function is similar to that from the 1st f-function. The energy lowering from the 3rd d-function is similar to that from the 2nd f-function and the first g-function. Addition of polarization functions should therefore be done in the order: 1d, 2d1f, 3d2f1g. Several different sizes of cc basis sets are available in terms of final number of contracted functions. These are known by their acronyms: cc-pVDZ, cc-pVTZ, cc-pVQZ, cc-pV5Z and cc-pV6Z (correlation consistent-polarized valence double/ triple/ quadruple/ quintuple/ sextuple zeta).

The composition in terms of primitive and contracted functions is shown in Table 1. Moving one set up in terms of quality increases each type of basis function by one and adds a new type of higher order polarization function. For the 1st and 2nd row atoms, the cc-pVDZ basis set adds 1s, 1p, and 1d function. The cc-pVTZ set adds another s, p, d, and f function, etc., as show in the Figure 7.

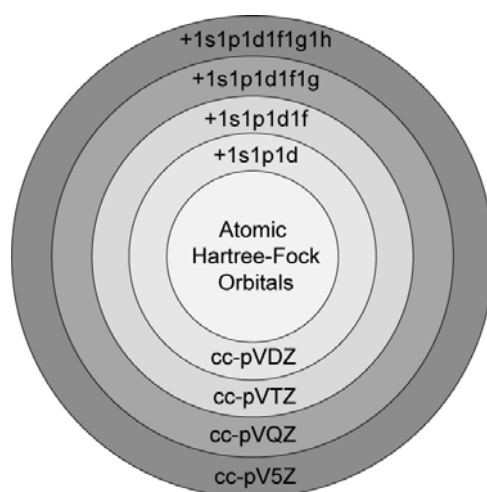


Figure 7 Correlation functions added in shells.

Table 1 Correlation consistent basis sets.

Basis	Primitive functions	Contracted functions
cc-pVDZ	9s,4p,1d	3s,2p,1d
cc-pVTZ	10s,5p,2d,1f	4s,3p,2d,1f
cc-pVQZ	12s,6p,3d,2f,1g	5s,4p,3d,2f,1g
cc-pV5Z	14s,9p,4d,3f,2g,1h	6s,5p,4d,3f,2g,1h
cc-pV6Z	16s,10p,5d,4f,3g,2h,1i	7s,6p,5d,4f,3g,2h,1i

Various augmentations to these cc-basis sets have also been developed. These include the addition of diffuse functions to better describe anions and weakly interacting molecules. A systematic extension is represented by the aug-cc-pVXZ series which derives from the cc-pVXZ series through addition of one set of diffuse functions with a smaller exponent for each angular momentum type already present in the respective cc-basis. The cc-pVDZ thus adds a diffuse, 1s, 1p, 1d, set to build the final, 4s, 3p, 2d, aug-cc-pVDZ set.

The cc-basis sets may also be augmented by additional tight functions (large exponent) if the interest is in recovering core-core and core-valence electron correlation, cc-pCVXZ and cc-pwCVXZ. There are two families of core-valence basis sets. The first one, adding core-core plus core-valence tight functions, yields the cc-pCVXZ family basis sets. The cc-pCVDZ has additionally 1s and 1p tight functions, the cc-pCVTZ has 2s, 2p and 1d tight functions and so on. The second one, in which just the tight functions of core-valence, plus only a few core-core functions, are added to the cc-pVXZ sets, yields other sets that are denoted as weighted core-valence basis sets, cc-pwCVXZ. The complete families of aug-cc-pVXZ, and cc-pCVXZ basis sets are presented in Table 2.

The main difficulty in using cc-basis sets is that each step up in quality almost doubles the number of basis functions, or equivalently, the number of functions increases as the third power of the highest angular momentum function. The main advantages of these basis set are that these sets recover a large fraction of the

correlation energy, provide systematic improvements that converge toward the complete basis set limit, and consistently reduce errors at both the HF and correlated levels with each step up in quality.

Table 2 The composition and size of aug-cc-pVXZ, and cc-pCVXZ basis sets.

Set	cc-pVXZ	aug-cc-pVXZ	cc-pCVXZ
DZ	3s,2p,1d	+ (1s,1p,1d)	+ (1s,1p)
TZ	4s,3p,2d,1f	+ (1s,1p,1d,1f)	+ (2s,2p,1d)
QZ	5s,4p,3d,2f,1g	+ (1s,1p,1d,1f,1g)	+ (3s,3p,2d,1f)
5Z	6s,5p,4d,3f,2g,1h	+ (1s,1p,1d,1f,1g,1h)	+ (4s,4p,3d,2f,1h)
6Z	7s,6p,5d,4f,3g,2h,1i	+ (1s,1p,1d,1f,1g,1h,1i)	–

Even- and Well-tempered Basis Set: The optimization of basis functions is a difficult problem. As the basis set becomes large, the basis functions become linearly dependent and the energy becomes a very flat function of the exponents. Furthermore, the multiple local minima problem is encountered. Empirical examination of results of such basis set optimizations has shown that in a set of such Gaussian functions, ordered by size of their exponents ζ , the ratio between two successive exponents often is constant. Therefore, orbital exponents ζ_k can be expressed as a function of two parameters α and β as follows:

$$\zeta_k = \alpha\beta^k, \quad k = 1, 2, 3, \dots, N \quad (91)$$

where α and β are fixed constants for a given type of function and nucleus charge. Taking this ratio to be constant (even-tempered), reduces the optimization problem to only two parameters for each type of basis function, independently of the size of the basis. In a well tempered basis set, the ratio is not taken evenly over the entire set of basis functions, but rather slanted so that the valence region is better covered.

3.2 METHOD OF CALCULATIONS

In the RI method the densities ρ_{ia} are approximated as a linear combination of auxiliary basis functions as shown in equation (68). Traditionally, the auxiliary basis set consists of primitive Gaussians. The size (number of basis functions per angular momentum and the highest angular momentum) of the auxiliary basis set depends on the underlying GTO basis set.²¹

In this work we do not use conventional ("density") auxiliary basis sets, but follow the *ansatz* proposed recently by Manby and Knowles.²⁴⁻²⁶ The auxiliary basis set consists of a few conventional auxiliary basis functions in order to be able to describe multipoles. This very small set of conventional functions is then augmented by a set of Poisson auxiliary functions, cf. equation (88). These Poisson functions help to provide an accurate description of the charge densities that are only very crudely described by the very few conventional basis functions.

RI-MP2

In this work, auxiliary basis sets for the atoms H, B, C, N, O, F in RI-MP2 calculations have been optimized for the standard basis set series cc-pVXZ with X = D, T and Q. The accuracy of the RI approximation is measured by the quantities ΔRI and ΔI .²¹ The energy difference (ΔRI),

$$\Delta RI = E_{MP2} - E_{MP2,RI} \quad (92)$$

is of prime importance for the accuracy of the RI approximation. Since E_{MP2} is neither an upper nor a lower bound to $E_{MP2,RI}$, minimizing ΔRI does not result in an optimal auxiliary basis set. One would rather like to consider $|\Delta RI|$, but this would not lead to a convenient procedure in the optimization of auxiliary basis sets. Instead of this we define

$$\Delta I = -\frac{1}{4} \sum_{i,a,j,b} \frac{|\langle ij||ab\rangle - \langle ij||ab\rangle_{RI}|^2}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b} \quad (93)$$

$$= 2\tilde{E} - E_{MP2} - E_{MP2,RI} \quad (94)$$

with

$$\tilde{E} = \sum_{\sigma_1, \sigma_2} \sum_{i,a \in \sigma_1} \sum_{j,b \in \sigma_2} (t_{ij}^{ab})(ia|jb)_{RI} \quad (95)$$

where σ_1 and σ_2 run over the spins and

$$t_{ij}^{ab} = \begin{cases} \frac{(ia|jb) - (ib|ja)}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b} & \text{if } \sigma_1 = \sigma_2 \\ \frac{(ia|jb)}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b} & \text{if } \sigma_1 \neq \sigma_2 \end{cases} \quad (96)$$

It is obvious from equation (94) that

$$\Delta I \geq 0, \text{ and } \Delta I = 0, \text{ if } E_{MP2} = E_{MP2,RI} \quad (97)$$

ΔI has all properties desired for an optimization procedure. It is always positive and, for a given set of MOs, ΔI is a function solely of the auxiliary basis functions.

All optimizations of the auxiliary basis sets for the atoms of the first row were carried out with calculations for the hydrides BH_3 , CH_4 , NH_3 , H_2O and HF . For H we choose the H_2 molecule.

Following Ref. 32, the sizes of the auxiliary basis sets corresponding to the standard cc-pVXZ series are as shown in Table 3. For RI-MP2, all these auxiliary s, p, d, f, g and h functions are Poisson functions with the exception of one p and one d

function for the H atom and two p and one d function for the other atoms, respectively, which are conventional Gaussian basis functions.

Table 3 Specification of auxiliary basis sets, cc-pVXZ with X = D, T and Q for RI-MP2 calculations.

atom	cc-pVDZ	cc-pVTZ	cc-pVQZ
H	3s, 2p, 1d	4s, 3p, 2d, 1f	5s, 4p, 3d, 2f, 1g
B-F	7s, 5p, 4d, 2f	8s, 6p, 5d, 3f, 1g	8s, 7p, 6d, 5f, 3g,

RI-HF

Additionally auxiliary basis sets for RI-HF calculations, complementing the standard bases cc-pVXZ with X = D, T, Q and 5, were optimized for H, B, C, N O and F atoms. The error of the exchange energy, ΔRI^{27} , was minimized in the optimization of RI-HF bases;

$$\Delta RI = \sum_{\mu\nu} (k_{\mu\nu} - k_{\mu\nu}^{RI}) \gamma_{\mu\nu} \geq 0 \quad (98)$$

where,

$$k_{\mu\nu}^{RI} = (\nu i | \mu j) \approx \sum_{AB} (\nu i | B) V_{AB}^{-1} (A | \mu j) \quad (99)$$

$k_{\mu\nu}$, is the corresponding exact exchange matrix. Here $\gamma_{\mu\nu} = 2 \sum_i c_{\mu i} c_{\nu i}$ is the one-electron closed-shell density matrix in the AO basis. In order to obtain balanced basis sets for different bonding situations, we selected for each atom a set of small molecules as presented in Table 4.

Table 4 Molecules used to optimize the RI-HF auxiliary basis sets, cc-pVXZ series with X = D, T, Q and 5.

atom	molecules
H	H ₂
B	BH ₃
C	CH ₄ , C ₂ H ₂ , C ₂ H ₄ , C ₂ H ₆
N	NH ₃ , N ₂ , CHN, NHO, CH ₃ N
O	H ₂ O, H ₂ O ₂ , CH ₂ O, CH ₃ OH
F	HF, F ₂ , HCF, HOF, CH ₃ F

The sizes of the auxiliary basis sets corresponding to the standard cc-pVXZ series with X = D, T, Q and 5, for RI-HF are shown in Table 5.

Table 5 The sizes of auxiliary basis set cc-pVXZ with X = D, T, Q and 5, for RI-HF calculations.

atom		cc-pVDZ	cc-pVTZ	cc-pVQZ	cc-pV5Z
H	GTO	1s,1p	1s,1p	1s,1p,1d	1s,1p,1d,1f
	Poisson	2s,1p,1d	4s,3p,2d,1f	4s,3p,3d,2f,1g	4s,2p,5d,3f,2g,1h
B-F	GTO	1s,1p,1d	1s,1p,1d	1s,1p,1d,1f	1s,1p,1d,1f,1g
	Poisson	6s,4p,3d,2f	10s,7p,5d,2f,1g	10s,7p,5d,3f,2g,1h	10s,7p,5d,4f,3g,2h,1i

All of these functions are left uncontracted and taken to be even tempered functions with centers α and ratios β , according to the *ansatz*:

$$\zeta_i = \alpha\beta^{(n+1)/2-i}, i = 1,2,3,\dots,N \quad (100)$$

Optimized values for all of these parameters are given in the results section below (where conventional Gaussian basis functions are called density functions and denoted as GTO functions).

Of course, ideally one should also optimize the number of auxiliary basis functions used, for RI-HF and RI-MP2. While this is indeed possible, we refrain from doing so in the present study, mainly to avoid the introduction of an even broader variety of optimization methods and of unavoidable arbitrariness in the definition of extended optimization targets that then would have to include computational efficiency measures.

The energies were calculated using MOLPRO⁴⁰ and ΔI was minimized globally by a (hybrid) GA or others methods, which were attached to MOLPRO as a user-defined subroutine. All calculations were performed serially on Intel Itanium 2 processors with 1.3 GHz. Adaption of the optimization algorithms discussed in chapter 2 to the present problem is obvious. For example, in a GA, the variables to be optimized, in our case the parameters α and β of the auxiliary basis sets under consideration, constitute the genes of each individual. Initially, the first generation is formed randomly. The error ΔI for each individual is calculated from the conventional MP2 and RI-MP2 energies. ΔI is interpreted as the fitness of the corresponding individual according to a fitness function. The other algorithms are adapted in an analogous fashion. The center parameters α and ratio parameters β were used as variables to be optimized by the algorithm and the error, ΔI in RI-MP2 or ΔRI in RI-HF, was a function value which was optimized until it satisfied the termination criteria in the algorithm.

3.3 RESULTS AND DISCUSSION

3.3.1 RI-MP2

Our primary goal is to reduce the error ΔI in RI-MP2 energy calculations. The accuracy of the optimized auxiliary basis sets were tested for some small molecules, H_2 , BH_3 , CH_4 , NH_3 , H_2O and HF . In our first investigations with the H_2O molecule, all parameters α and β of both H and O were optimized. We found that optimized ratio parameters β usually fall in the interval between 2.0 to 3.0. For example we show the parameters α and β of the auxiliary basis set cc-pVTZ in Table 6. Sizes n denote the size of the optimized even-tempered auxiliary basis sets, equation (100).

Table 6 Size n , centers α and ratios β of auxiliary basis set cc-pVTZ for the H_2O molecule, results obtained from the GA/Powell hybrid method.

atom		GTO function		n	Poisson function	
		α	β		α	β
H	s			4	1.143113	2.498312
	p	1.110210		2	1.270854	2.206321
	d	1.759650		1	1.322858	
	f			1	1.050087	
C	s			8	3.137808	2.273975
	p	0.590907	1.851995	4	3.330719	2.580808
	d	0.841943		4	1.955561	2.239383
	f			3	1.407368	2.287659
	g			1	1.692173	

Table 7 shows representative results for the performance of the GA/Powell hybrid and the Iffco method, for the H_2O molecule, comparing optimization of α and β parameters for H and O atoms, optimization of only α parameters (with β parameters fixed to 2.5) for H and O atoms, and optimization of α and β parameters for the O atom only (with α and β parameters for the H atom fixed at their best values obtained from calculations on the H_2 molecule).

Table 7 The best objective function and CPU time (min.), comparing different optimization strategies of α and β parameters for H and O atoms. The calculations were performed using GA/Powell hybrid and Iffco methods for the H₂O molecule.

basis set	GA with Powell hybrid method					
	optimized all		fixed β (opt. α for H and O)		fixed H (opt. α and β for O)	
H2O	$\Delta I/ E_{MP2} $	Time	$\Delta I/ E_{MP2} $	time	$\Delta I/ E_{MP2} $	time
cc-pVDZ	2.16764E-9	12	2.46097E-9	5	3.15680E-9	8
cc-pVTZ	4.65820E-9	61	5.85498E-9	21	5.36330E-9	20
cc-pVQZ	2.97926E-9	556	4.43396E-9	175	4.01914E-9	143
basis set	Iffco method					
	optimized all		fixed β (opt. α for H and O)		fixed H (opt. α and β for O)	
H2O	$\Delta I/ E_{MP2} $	Time	$\Delta I/ E_{MP2} $	time	$\Delta I/ E_{MP2} $	time
cc-pVDZ	2.08352E-9	19	2.44093E-9	7	2.99729E-9	14
cc-pVTZ	4.48755E-9	163	4.99874E-9	47	5.24212E-9	32
cc-pVQZ	2.02224E-9	1552	3.57325E-9	409	2.74926E-9	515

We found that the error, $\Delta I/|E_{MP2}|$, is reduced to a few ppb in all cases; the maximum error is 6 ppb. From these results we see that there is no significant loss of accuracy, neither by keeping the ratio parameter β fixed nor by keeping the H atom parameters fixed, but both restrictions speed up the calculations. Therefore, for further comparison of the performance of the different optimization algorithms used in this work, we decided to optimize only the α parameters for the different atoms while fixing the ratio parameters β to 2.5 and also fixing the α parameters of the H atoms at their H₂ optimized values.

We experimented with several algorithmic parameters and function evaluation budgets: Moreover, these calculations were repeated several times, since the GA algorithm is non-deterministic. One can not be sure if the obtained result is the global minimum or only a local minimum. Hence, we repeated several runs to arrive at a practical guarantee that we already found the global or at least a good local

minimum. Although the Iffco algorithm is deterministic, it starts with a random starting point thus we still need to repeat several runs to ensure the obtained result. The results from the GA method come from the best result of 30 repeated runs for 200 generations. Table 8 shows the best objective function values ($\Delta I/|E_{MP2}|$), number of function evaluations and time, obtained from the genetic algorithm.

Table 8 Summary of the best objective function values ($\Delta I/|E_{MP2}|$) per atom, number of function evaluations and CPU time (min) for H₂, BH₃, CH₄, NH₃, H₂O and HF molecules, using the GA method.

molecule	$\Delta I/ E_{MP2} $	f-eval	time(min.)
H ₂			
cc-pVDZ	5.40831E-09	1206	4
cc-pVTZ	5.65702E-09	1206	6
cc-pVQZ	1.32177E-08	1206	18
BH ₃			
cc-pVDZ	8.08008E-09	1206	10
cc-pVTZ	6.09501E-09	1206	44
cc-pVQZ	3.97226E-09	1206	280
CH ₄			
cc-pVDZ	4.99009E-09	1206	13
cc-pVTZ	5.48274E-09	1206	81
cc-pVQZ	4.02016E-09	1206	858
NH ₃			
cc-pVDZ	3.50725E-09	1206	10
cc-pVTZ	6.32096E-09	1206	46
cc-pVQZ	4.39202E-08	1206	433
H ₂ O			
cc-pVDZ	3.25629E-09	1206	13
cc-pVTZ	6.18952E-09	1206	29
cc-pVQZ	1.84922E-07	1206	188
HF			
cc-pVDZ	2.14798E-09	1206	9
cc-pVTZ	6.59593E-09	1206	14
cc-pVQZ	5.52736E-07	1206	74

As expected, the time used for optimization of auxiliary basis sets increase considerably with the size of auxiliary basis sets from cc-pVDZ, cc-pVTZ and cc-pVQZ, respectively. We found that the error $\Delta I/|E_{\text{MP2}}|$ in RI-MP2 calculations by using these auxiliary basis set is smaller than 0.55 ppm. In more than 77% of all cases the errors were smaller than 0.008 ppm.

For the energy difference per atom between MP2 and RI-MP2 (ΔRI), we observed that in all case the errors are smaller than $45 \mu E_h$, as shown in Figure 8.

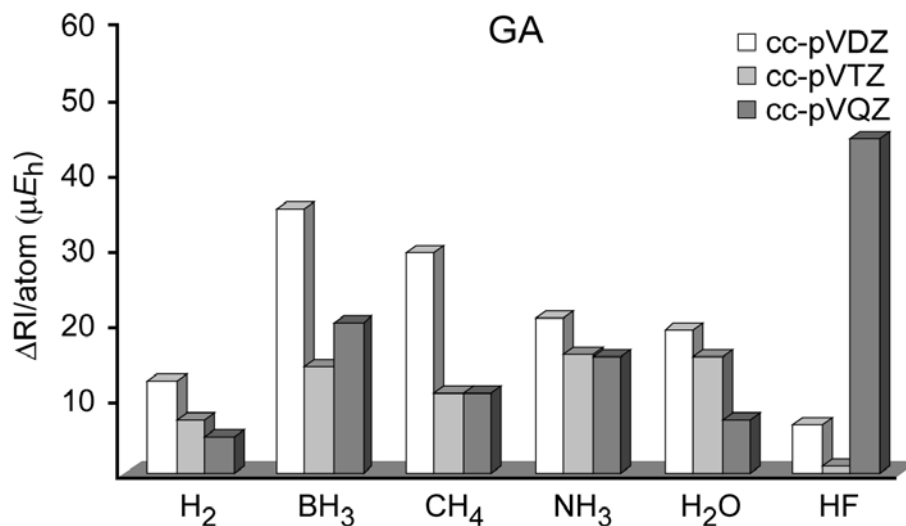


Figure 8 Histogram of the energy differences (ΔRI) per atom between MP2 and RI-MP2 calculations, obtained from the genetic algorithm.

The best values for the parameters α of B, C, N, O and F atoms, together with the fixed value of α for the H atom, are shown in Table 9. From the optimized parameters in Table 9, it can be seen that in many cases the center parameters α (Poisson function) of the even tempered series are systematically increasing from BH₃ to HF.

Table 9 Overall best values for the center parameters α of even-tempered auxiliary basis sets, complementing standard bases cc-pVXZ with X = D, T and Q for H₂, BH₃, CH₄, NH₃, H₂O and HF molecules, obtained from the genetic algorithm.

atom		GTO		Poisson			GTO		Poisson			GTO		Poisson																									
		α	n	α	n		α	n	α	n		α	n	α	n																								
H	s	0.9145667465 1.0404861603	3	1.1866126566	0.6062780374 0.9122881867	s	4	1.2344600891	0.5574584405 0.9551143795	s	5	2.0586415819	0.9048704659 1.3480301748 0.9266630370 1.5264999280	1	1.2955045884	p	2	1.5225972293	p	3	0.9048704659	d	2	1.3480301748	d	2	0.9266630370	f	1	1.2224902259	f	1	1.5264999280						
	p		1	1.2955045884		p	2	1.5225972293		p	3	0.9048704659																											
	d		1	1.2955045884		d	1	1.2955045884		d	1	1.2955045884																											
	f		1	1.2955045884		f	1	1.2955045884		f	1	1.2955045884																											
B	s	0.2398644970 0.8589971611	7	0.6754313467	0.5553534501 0.2858798955	s	8	2.3555592124	1.0737641838 0.6233882973	s	8	1.6732951322	8.1747910058 0.4347574215 0.6423873823	3	1.0437299545	p	4	1.0881836547	p	5	1.3510814698	d	5	0.6131902759	d	5	8.1747910058	f	3	0.4347574215	f	3	0.4347574215	g	1	0.6423873823	g	1	0.6423873823
	p		3	1.0437299545		p	4	1.0881836547		p	5	1.3510814698																											
	d		3	1.0437299545		d	4	1.0881836547		d	5	1.3510814698																											
	f		2	0.2754972420		f	3	0.3779651504		f	5	8.1747910058																											
C	s	0.5471556866 0.3483091633	7	0.8071344450	0.9681981022 0.3063460744	s	8	2.0830615014	0.8454595911 0.8031437055	s	8	2.0048936191	3.0586513558 0.7732051208 3.0230073194 1.3686921572 0.7167565563	3	1.6699256099	p	4	1.8510018827	p	5	3.0586513558	d	5	0.7732051208	d	5	3.0230073194	f	3	1.3686921572	f	3	1.3686921572	g	1	0.7167565563	g	1	0.7167565563
	p		3	1.6699256099		p	4	1.8510018827		p	5	3.0586513558																											
	d		3	1.6699256099		d	4	1.8510018827		d	5	3.0586513558																											
	f		2	0.6489262422		f	3	0.6577617957		f	5	3.0230073194																											
N	s	0.3944089712 0.4431423915	7	2.2111174204	0.5136887695 0.2147961563	s	8	8.2771953103	0.4616458261 0.9778503004	s	8	7.0117732351	1.7885504621 0.9731059536 2.0807652815 1.5524606566 1.2345433680	3	2.0782701875	p	4	2.7579157430	p	5	1.7885504621	d	5	0.9731059536	d	5	2.0807652815	f	3	1.5524606566	f	3	1.5524606566	g	1	1.2345433680	g	1	1.2345433680
	p		3	2.0782701875		p	4	2.7579157430		p	5	1.7885504621																											
	d		3	2.0782701875		d	4	2.7579157430		d	5	1.7885504621																											
	f		2	0.7894155057		f	3	1.1483626103		f	5	2.0807652815																											
O	s	0.9947861925 0.7146034923	7	2.7182435939	0.6181385759 0.5427359330	s	8	3.3793456373	1.3624639258 0.9968679263	s	8	2.6833017505	2.1428588331 2.4471650036 1.6642347101 1.5561695513 1.2087026432	3	2.8220221908	p	4	3.3777483336	p	5	2.1428588331	d	5	2.4471650036	d	5	1.6642347101	f	3	1.5561695513	f	3	1.5561695513	g	1	1.2087026432	g	1	1.2087026432
	p		3	2.8220221908		p	4	3.3777483336		p	5	2.1428588331																											
	d		3	2.8220221908		d	4	3.3777483336		d	5	2.1428588331																											
	f		2	1.1425276009		f	3	1.1554056066		f	5	1.6642347101																											
F	s	0.8953484271 0.7093342663	7	3.2597012702	0.6509654554 0.3631800036	s	8	3.4835422684	1.1086375272 3.4333491278	s	8	6.8120969753	4.0810149639 2.4896314681 1.8029397264 1.5882113358 1.2734770852	3	4.0231146805	p	4	4.2687857951	p	5	4.0810149639	d	5	2.4896314681	d	5	1.8029397264	f	3	1.5882113358	f	3	1.5882113358	g	1	1.2734770852	g	1	1.2734770852
	p		3	4.0231146805		p	4	4.2687857951		p	5	4.0810149639																											
	d		3	4.0231146805		d	4	4.2687857951		d	5	4.0810149639																											
	f		2	1.6810809914		f	3	1.5596795218		f	5	1.8029397264																											

Then, the GA was combined with Powell local optimization. The most efficient allocation of resources turned out to be 30 iterations of GA followed by just one iteration of Powell. These results also come from the best results of 30 repeated runs. The best objective functions ($\Delta I/|E_{MP2}|$) per atom, number of function evaluations and time, obtained from the GA/Powell hybrid method, are presented in Table 10.

Table 10 Summary of the best objective function values ($\Delta I/|E_{MP2}|$) per atom, number of function evaluations and CPU time (min) for H₂, BH₃, CH₄, NH₃, H₂O and HF molecules, using the GA/Powell hybrid method.

molecule	$\Delta I/ E_{MP2} $	f-eval	time(min.)
H ₂			
cc-pVDZ	5.38318E-09	475	1
cc-pVTZ	5.37480E-09	929	3
cc-pVQZ	8.78711E-09	960	14
BH ₃			
cc-pVDZ	7.92816E-09	1014	9
cc-pVTZ	5.87815E-09	984	36
cc-pVQZ	2.05384E-09	1503	486
CH ₄			
cc-pVDZ	4.70340E-09	1083	12
cc-pVTZ	5.39092E-09	1038	70
cc-pVQZ	2.85134E-09	1183	835
NH ₃			
cc-pVDZ	3.46337E-09	814	8
cc-pVTZ	6.18082E-09	693	26
cc-pVQZ	4.03259E-09	913	327
H ₂ O			
cc-pVDZ	3.21735E-09	907	6
cc-pVTZ	5.84535E-09	1092	23
cc-pVQZ	4.64457E-09	884	138
HF			
cc-pVDZ	2.10136E-09	564	5
cc-pVTZ	7.14264E-09	745	8
cc-pVQZ	5.65828E-09	940	53

These results show that the errors $\Delta I/|E_{\text{MP2}}|$ in RI-MP2 calculations are smaller than 0.009 ppm in all cases. Compared to Table 8, there are no significant differences in accuracy, and some small differences but no clear trend in computing times. The advantage of this combined method is a better chance of getting good results in a single run. In contrast, the standard GA alone may fail with premature convergence and hence needs several runs. The energy difference per atom between MP2 and RI-MP2 (ΔRI), are present in Figure 9. The histogram shows that the error, (ΔRI), is smaller than $35 \mu E_h$, slightly better than in Figure 8. It can be concluded that the combined method (Powell local optimization with GA global optimization) is the more effective method for all basis sets and molecules than GA alone.

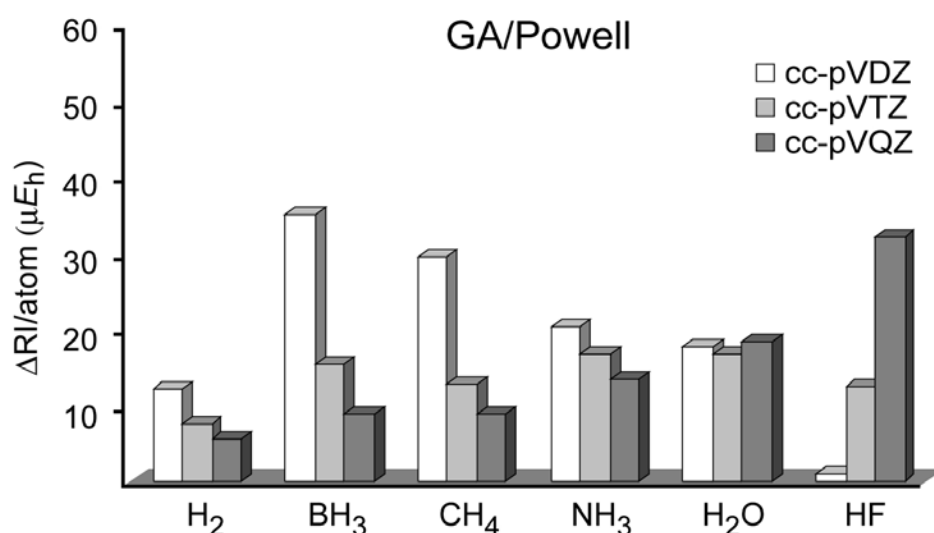


Figure 9 Histogram of the energy differences (ΔRI) per atom between MP2 and RI-MP2 calculations, obtained from the GA/Powell hybrid method.

The best values for the parameters α of B, C, N, O and F atoms, together with the fixed value of α for the H atom, obtained from the GA/Powell hybrid method, are shown in Table 11. Comparing the optimized parameters in Table 9 and 11, it can be seen that there are many similarities obtained from two different methods but also some differences. It follows that for our problem there are multiple similar minima.

Table 11 Overall best values for the center parameters α of even-tempered auxiliary basis sets, complementing standard bases cc-pVXZ with X = D, T and Q for H₂, BH₃, CH₄, NH₃, H₂O and HF molecules, obtained from the GA/Powell hybrid method.

atom		GTO		Poisson			GTO		Poisson			GTO		Poisson	
		α	n	α	n		α	n	α	n		α	n	α	n
H	s	0.9176170129 1.0325546012	3	1.1820003088	s p d f	0.6221983649 1.2740129165	4	1.1955496218	s p d f g	1.2633969703 0.8553503194	5	2.0408304652	0.8553503194	3	0.8883928575
	p		1	1.2951638519			2	1.4623821149			2	1.3221944024			
	d						1	1.5714344877			2	0.9372311764			
	f						1	1.1170729176			1	1.5142008296			
B	s	0.2041392723 0.2009565603	7	1.7232546393	s p d f g	0.4091517402 0.4971456014	8	5.6188081252	s p d f g h	0.3901039641 1.9219934490	8	1.4457775980	1.9219934490	5	1.5462196755
	p		3	1.0121092636			4	2.1718281233			5	0.6065968924			
	d		3	0.9915232893			4	0.5922982331			5	0.5078205979			
	f		2	0.2899091122			3	0.3632505730			3	0.7921199419			
C	s	0.5396092186 0.4830171316	7	2.0409709364	s p d f g	0.9636760822 0.3090122312	8	2.0932661145	s p d f g h	2.0204262377 1.2320817586	8	5.3225050079	1.2320817586	5	6.0398180574
	p		3	1.6560072686			4	1.8387115993			5	1.2188737740			
	d		3	1.0458224896			4	0.8670307616			5	0.7139542309			
	f		2	0.6459167163			3	0.6584644612			3	1.2671865288			
N	s	0.4058865516 0.4678833248	7	2.2154878782	s p d f g	0.4596903924 0.2312335289	8	2.5526778877	s p d f g h	0.5654751333 1.1882934880	8	3.3105176734	1.1882934880	5	2.2352380687
	p		3	2.0790638946			4	2.5007773495			5	1.1870419405			
	d		3	1.1613698123			4	1.0499207526			5	1.0356210597			
	f		2	0.7945176108			3	0.8251411485			3	1.8832091783			
O	s	0.9819809198 0.7174966756	7	2.7218779212	s p d f g	0.6191410296 0.5460104594	8	3.3036918921	s p d f g h	0.8852093744 1.0698810409	8	4.4812220786	1.0698810409	5	3.4999157998
	p		3	2.8445987189			4	3.3926278446			5	1.6811169767			
	d		3	1.7616894141			4	1.5777978766			5	1.4706706951			
	f		2	1.1543085397			3	1.1516114778			3	2.3222510573			
F	s	0.9051816291 0.6906592426	7	2.8906537331	s p d f g	0.3381904355 1.8952096347	8	8.5503345999	s p d f g h	1.0640732621 1.3089564617	8	5.4110739002	1.3089564617	5	4.2167739431
	p		3	4.1184898176			4	3.4719508257			5	5.2198267990			
	d		3	3.3259787494			4	2.0788869740			5	1.8913557253			
	f		2	1.6515689856			3	1.5805941112			3	2.9505834895			
							1	2.1750294869			1	2.6284320746			

The results from our first round of test runs for the H₂O molecule, where we optimized all even-tempered parameters for all atoms, using the Direct method are presented in Table 12. We found that in order to obtain reasonable results, the Direct method needs several long runs to optimize these auxiliary basis sets which constitutes a large amount of time spent for the optimization, even compared to the simple GA. A possible reason could be that as the simple GA it does not use gradient information but its heuristics in choosing new points appear to be weaker and less global.

Table 12 The error ($\Delta I/|E_{MP2}|$) per atom, number of function evaluations and CPU time (min) of test runs for the H₂O molecule, using the Direct method alone.

basis set	$\Delta I/ E_{MP2} $	f-eval	time(min.)
cc-pVDZ	8.57596E-09	18807	127
cc-pVTZ	9.96925E-09	20361	429
cc-pVQZ	1.97406E-08	20047	2780

As explained above, the performance of Direct alone is not competitive for our present problem. Therefore, hybrid methods of Direct and local search, such as Powell and Iffco, were investigated.

The best objective function values ($\Delta I/|E_{MP2}|$), number of function evaluations and time, obtained from the Direct/Powell method and the Direct/Iffco method, are presented in Table 13 and 14, respectively. The observed results from the Direct/Powell hybrid method show that the errors $\Delta I/|E_{MP2}|$ are smaller than 0.008 ppm, only the result from cc-pVQZ basis set for H₂ molecule is 0.01 ppm. The results from the Direct/Iffco hybrid method are also as good as those from the other hybrid methods. In all cases, the errors $\Delta I/|E_{MP2}|$ are found to be smaller than 0.009 ppm. These observations reveal that with the combination of Direct and local search (either Powell or Iffco) acceptable results can be obtained a lot faster than with the Direct method alone.

Table 13 Summary of the best objective function values ($\Delta I/|E_{MP2}|$) per atom, number of function evaluations and CPU time (min) for H₂, BH₃, CH₄, NH₃, H₂O and HF molecules, using the Direct/Powell hybrid method.

molecule	$\Delta I/ E_{MP2} $	f-eval	time(min.)
H ₂			
cc-pVDZ	5.80933E-09	1493	3
cc-pVTZ	5.52407E-09	1736	6
cc-pVQZ	1.00858E-08	1875	26
BH ₃			
cc-pVDZ	7.72574E-09	1449	11
cc-pVTZ	5.60127E-09	1697	61
cc-pVQZ	2.06880E-09	3418	1157
CH ₄			
cc-pVDZ	4.70348E-09	2038	21
cc-pVTZ	5.90174E-09	2217	150
cc-pVQZ	3.00895E-09	1999	1474
NH ₃			
cc-pVDZ	3.46331E-09	1331	10
cc-pVTZ	6.17533E-09	1417	54
cc-pVQZ	4.07009E-09	1723	643
H ₂ O			
cc-pVDZ	3.26673E-09	1394	8
cc-pVTZ	5.84251E-09	1450	30
cc-pVQZ	4.53176E-09	2031	321
HF			
cc-pVDZ	2.28780E-09	1806	11
cc-pVTZ	7.08149E-09	1424	15
cc-pVQZ	5.57010E-09	1983	110

Table 14 Summary of the best objective function values ($\Delta I/|E_{MP2}|$) per atom, number of function evaluations and CPU time (min) for H₂, BH₃, CH₄, NH₃, H₂O and HF molecules, using the Direct/Iffco hybrid method.

molecule	$\Delta I/ E_{MP2} $	f-eval	time(min.)
H ₂			
cc-pVDZ	5.81108E-09	1162	3
cc-pVTZ	5.52552E-09	1301	5
cc-pVQZ	8.77534E-09	1439	20
BH ₃			
cc-pVDZ	7.72594E-09	1361	10
cc-pVTZ	5.61757E-09	1410	51
cc-pVQZ	2.05395E-09	1837	598
CH ₄			
cc-pVDZ	4.70048E-09	1381	14
cc-pVTZ	5.40187E-09	1896	127
cc-pVQZ	3.00956E-09	1477	1049
NH ₃			
cc-pVDZ	3.47795E-09	1271	10
cc-pVTZ	6.17523E-09	1406	54
cc-pVQZ	3.99599E-09	1379	505
H ₂ O			
cc-pVDZ	3.26669E-09	1269	8
cc-pVTZ	5.84234E-09	1534	31
cc-pVQZ	4.58933E-09	1797	280
HF			
cc-pVDZ	2.28777E-09	1291	8
cc-pVTZ	7.08127E-09	1435	16
cc-pVQZ	5.48719E-09	1792	100

The energy differences per atom between MP2 and RI-MP2 (ΔRI), obtained from the Direct/Powell and Direct/Iffco methods, are presented in Figure 10 and 11, respectively. They show that the errors obtained from the two different hybrid methods are almost of the same quality, smaller than $35 \mu E_h$ and $36 \mu E_h$, respectively, and as good as in the previous cases, GA (Figure 8) and GA/Powell (Figure 9).

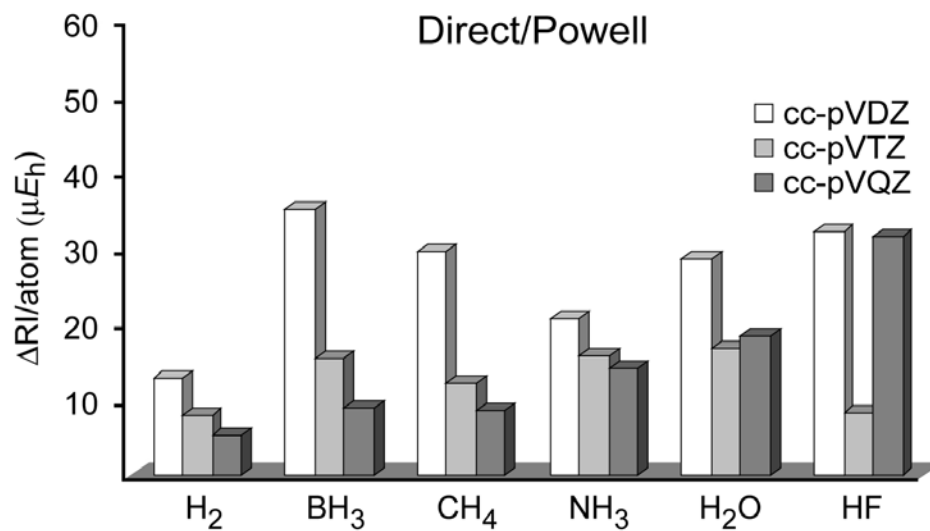


Figure 10 Histogram of the energy differences (ΔRI) per atom between MP2 and RI MP2 calculations, obtained from the Direct/Powell hybrid method.

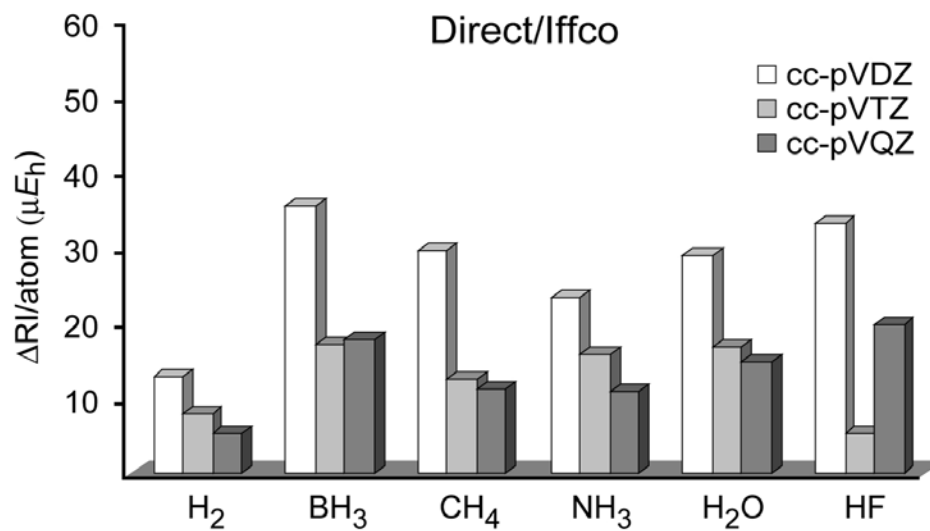


Figure 11 Histogram of the energy differences (ΔRI) per atom between MP2 and RI MP2 calculations, obtained from the Direct/Iffco hybrid method.

Table15 Overall best values for the center parameters α of even-tempered auxiliary basis sets, complementing standard bases cc-pVXZ with X = D, T and Q for H₂, BH₃, CH₄, NH₃, H₂O and HF molecules, obtained from the Direct/Powell hybrid method.

atom		GTO		Poisson			GTO		Poisson			GTO		Poisson	
		α	n	α	n		α	n	α	n		α	n	α	n
H	s	1.6404632575 1.0229313607	3	1.1790327318	s p d f	0.6260883922 1.7543994352	4	1.1967831329	s p d f g	0.4923396963 0.9829802587	5	1.8700579121			
	p		1	1.2727187874			2	1.4631673154			3	2.0086110186			
	d						1	1.5520686332			2	1.2649697749			
	f						1	1.1271358995			2	0.9301979995			
B	s	0.2044515056 0.2007450176	7	1.7241782657	s p d f g	0.5809711012 0.2707634993	8	0.9615136720	s p d f g h	0.3925656582 1.1783423676	8	1.4358505096			
	p		3	1.0103232122			4	1.1229458845			5	1.5516553572			
	d		3	0.9912979537			4	0.6088491140			5	0.5997033988			
	f		2	0.2898806120			3	0.3783522794			5	0.5071580376			
C	s	0.5383285674 0.4909432108	7	2.0355984440	s p d f g	0.8779357790 0.6194923255	8	2.0552140221	s p d f g h	0.6749523236 0.6708349331	8	5.3107102319			
	p		3	1.6518124771			4	1.6163263330			5	2.6020731431			
	d		3	1.0462476403			4	0.7145768167			5	0.7923339359			
	f		2	0.6470314081			3	0.6207753144			5	1.6757593871			
N	s	0.4058297916 0.4644295336	7	2.2167936793	s p d f g	0.5275867313 0.3141200103	8	1.6030028142	s p d f g h	0.5597117919 1.1573853829	8	7.6042029349			
	p		3	2.0791417506			4	2.8507772905			5	2.2124445309			
	d		3	1.1606217123			4	1.0819976599			5	1.1599216549			
	f		2	0.7945404539			3	1.1446304227			5	1.0434176848			
O	s	0.9757119449 0.7005032298	7	7.2152478934	s p d f g	0.6196551029 0.5644170973	8	3.2056018221	s p d f g h	0.8792789179 1.0608329515	8	4.4604582163			
	p		3	2.7966740059			4	3.3919595193			5	3.4766911256			
	d		3	1.7517919013			4	1.6005671246			5	1.6794292041			
	f		2	1.1523898528			3	1.1489516162			5	1.4696403544			
F	s	8.299480926 1.209443943	7	3.0999647558	s p d f g	9.8271799721 0.7412262858	8	8.2554172925	s p d f g h	3.5132037407 1.4574255065	8	5.7812446996			
	p		3	1.4981411547			4	2.8122825075			5	4.0523921400			
	d		3	2.4334593330			4	2.1021766967			5	2.0493573419			
	f		2	1.6591414560			3	1.5806339126			5	1.9112954695			
							1				3	1.6369161111			
											1	2.6381333745			

Table 16 Overall best values for the center parameters α of even-tempered auxiliary basis sets, complementing standard bases cc-pVXZ with X = D, T and Q for H₂, BH₃, CH₄, NH₃, H₂O and HF molecules, obtained from the Direct/Iffco hybrid method.

molecule		GTO		Poisson			GTO		Poisson			GTO		Poisson	
		α	n	α	n		α	n	α	n		α	n	α	n
H	s	1.6519526776 1.0234490477	3	1.1800985166	s p d f	0.6248467186 1.7483538244	4	1.1961152163	s p d f g	1.2628413066 0.9175723861	5	2.0397049206			
	p		1	1.2791937888			2	1.4630455657			3	0.8868924507			
	d						1	1.5471977402			2	1.3203171537			
	f						1	1.1239385290			2	0.9382567619			
B	s	0.2043261243 0.2015148494	7	1.7257704238	s p d f g	0.2504083267 0.2845453213	8	0.9650697042	s p d f g h	0.3948489703 1.8894877671	8	1.4279211926			
	p		3	1.0094960584			4	1.3158293990			5	1.5502332865			
	d		3	0.9935157818			4	0.6028750657			5	0.6008124900			
	f		2	0.2900471019			3	0.3805102292			5	0.5068006272			
C	s	0.5377938481 0.3442849141	7	2.0403545157	s p d f g	0.6458000738 0.1481535686	8	2.0927102444	s p d f g h	0.6409481874 0.8840131756	8	5.3111710239			
	p		3	1.6514671039			4	1.8977462494			5	2.6124224778			
	d		3	1.0441384006			4	0.8678148997			5	0.7879176294			
	f		2	0.6455540505			3	0.6576529272			5	1.6714336839			
N	s	0.7199609484 0.4673156177	7	2.2431745679	s p d f g	0.5270382152 0.2351886125	8	1.6034710482	s p d f g h	0.5825117617 1.1521749549	8	3.3328787314			
	p		3	2.0591852316			4	2.8504312152			5	2.2448310033			
	d		3	1.1677795879			4	1.0811732470			5	1.1825338830			
	f		2	0.7964604970			3	1.1443543328			5	1.0352493698			
O	s	0.9781026728 0.7000448438	7	7.222222189	s p d f g	0.6192236090 0.5695371006	8	3.1528713748	s p d f g h	0.8564822044 0.2156312718	8	4.3589918739			
	p		3	2.8031479574			4	3.3810576459			5	3.0971286977			
	d		3	1.7518135490			4	1.6078946516			5	1.7172907219			
	f		2	1.1526582439			3	1.1478238388			5	1.4674714878			
F	s	8.3221159534 1.2093797039	7	3.1090857138	s p d f g	9.8685856934 0.7404987943	8	8.2470687650	s p d f g h	0.9319622699 1.4355736272	8	5.7317907815			
	p		3	1.5019694202			4	2.8245996191			5	4.2802880073			
	d		3	2.4340930222			4	2.1013538550			5	2.0461630606			
	f		2	1.6593399594			3	1.5811709933			5	1.9109946506			
							1			3	1.6352662272				
										1	2.6381104173				

The best values for the parameters α of B, C, N, O and F atoms, together with the fixed value of α for the H atom, obtained from the Direct/Powell and Direct/Iffco methods, are shown in Table 15 and 16, respectively. From these Tables, we see that the optimized parameters obtained from Direct with the two local methods are almost the same in all cases. The parameters are also similar with optimized parameters obtained from the GA/Powell method in many cases.

Table 17 Summary of the best objective function values ($\Delta I/|E_{MP2}|$) per atom, number of function evaluations and CPU time (min) for H₂, BH₃, CH₄, NH₃, H₂O and HF molecules, using the Iffco method.

molecule	$\Delta I/ E_{MP2} $	f-eval	time(min.)
H ₂			
cc-pVDZ	5.38278E-09	256	1
c-pVTZ	5.36894E-09	509	2
cc-pVQZ	8.29373E-09	765	11
BH ₃			
cc-pVDZ	7.72594E-09	495	4
cc-pVTZ	5.59922E-09	718	26
cc-pVQZ	2.05385E-09	1317	451
CH ₄			
cc-pVDZ	4.70049E-09	524	6
cc-pVTZ	5.39194E-09	580	40
cc-pVQZ	2.58428E-09	913	647
NH ₃			
cc-pVDZ	3.46332E-09	509	5
cc-pVTZ	6.17533E-09	696	27
cc-pVQZ	3.99392E-09	961	363
H ₂ O			
cc-pVDZ	3.21713E-09	451	3
cc-pVTZ	5.84235E-09	748	18
cc-pVQZ	4.53119E-09	866	135
HF			
cc-pVDZ	2.10136E-09	486	3
cc-pVTZ	6.57268E-09	593	7
cc-pVQZ	5.32517E-09	628	37

Simple local optimization methods alone were not expected to be effective methods for our present problem, due to the presence of many local minima of varying quality. However, an advanced local method that does not get stuck in small unimportant minima may still work despite its lack of global convergence, because the dimensionality of the search space is not very large here and we assume that its structure is not complicated. Therefore, the Iffco method was also applied to optimize these auxiliary basis sets. These results also come from the best results of 30 repeated runs. Surprisingly, Iffco is fast and robust and as good as the other methods, even though Iffco formally is a local optimization algorithm. The results from Table 17 show that all errors $\Delta I/|E_{\text{MP2}}|$ per atom are found to be smaller than 0.009 ppm.

The energy difference per atom between MP2 and RI-MP2 (ΔRI), are presented in Figure 12. It shows that the error is smaller than $40 \mu E_h$ at most as good as for the previous methods, only the result from cc-pVTZ for the HF molecule is $82 \mu E_h$.

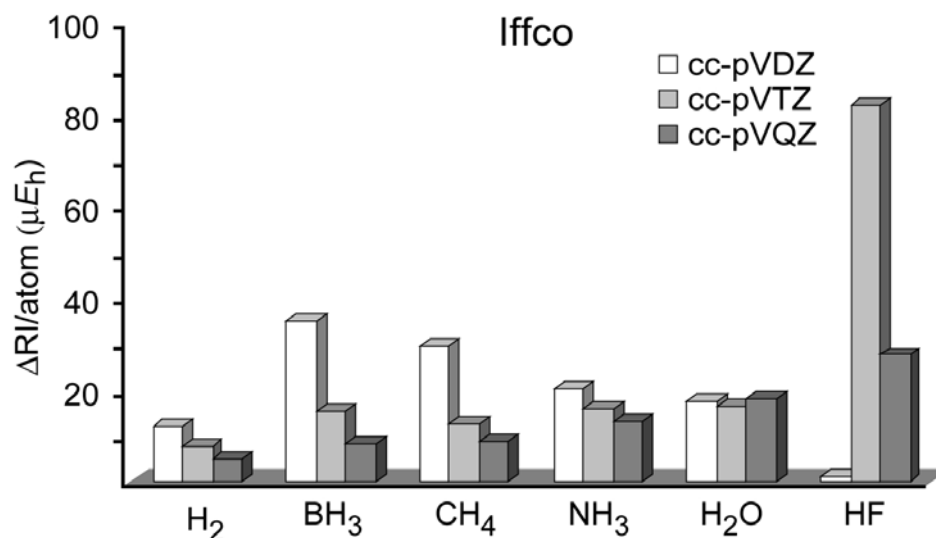


Figure 12 Histogram of the energy differences (ΔRI) per atom between MP2 and RI-MP2 calculations, obtained from the Iffco method.

Table 18 Overall best values for the center parameters α of even-tempered auxiliary basis sets, complementing standard bases cc-pVXZ with X = D, T and Q for H₂, BH₃, CH₄, NH₃, H₂O and HF molecules, obtained from the Iffco method.

molecule		GTO		Poisson			GTO		Poisson			GTO		Poisson	
		α	n	α	n		α	n	α	n		α	n	α	n
H	s	0.9183124008 1.0342218057	3	1.1824159507	s p d f	0.6243293977 1.3059533689	4	1.1959579465	s p d f g	0.5536679728 0.9841139919	5	2.0605277773	0.5536679728 0.9841139919	3	0.8968447143
	1		1.2966460989	2			1.4618406454	2			1.3375542833				
				1			1.6065071475	2			0.9283145793				
				1			1.1160345546	1			1.5089493897				
B	s	0.2041936557 0.2012444309	7	1.7244396034	s p d f g	0.6093348752 0.2687165914	8	2.3236520495	s p d f g h	0.3919868120 1.1682343195	8	1.4333300309	0.3919868120 1.1682343195	5	1.5505049474
	3		1.0102780229	4			1.1639274632	5			0.5988576172				
	3		0.9914821667	4			0.6073711101	5			0.5080356706				
	2		0.2899192672	3			0.3740983068	3			0.7900308003				
		1	0.5708313702	1	0.7638588656										
C	s	0.5400940491 0.3463775330	7	2.0379472091	s p d f g	0.9633645885 0.1467364587	8	2.0917293478	s p d f g h	1.7694320145 0.7005403688	8	2.0712708524	1.7694320145 0.7005403688	5	2.1215556667
	3		1.6548820788	4			1.8334091396	5			1.7506653549				
	3		1.0473169429	4			0.8598170015	5			0.7079103506				
	2		0.6474795403	3			0.6552076323	3			1.2759242161				
		1	0.9009471545	1	1.1932979194										
N	s	0.4048988811 0.4639062409	7	2.2162684813	s p d f g	0.5273726657 0.3129326314	8	1.6030683491	s p d f g h	0.4251257882 1.1865484036	8	3.2683630849	0.4251257882 1.1865484036	5	2.2116845220
	3		2.0788971872	4			2.8503485969	5			1.1857094278				
	3		1.1599805340	4			1.0817851750	5			1.0340972670				
	2		0.7945965209	3			1.1445770317	3			1.8800852406				
		1	1.3089286483	1	1.6105390745										
O	s	0.9804778691 0.7031070491	7	2.7213998569	s p d f g	0.6194366407 0.5687447980	8	3.1536018422	s p d f g h	0.8854458181 1.0710994212	8	4.4835127850	0.8854458181 1.0710994212	5	3.5016007507
	3		2.8432442263	4			3.3813015393	5			1.6815558738				
	3		1.7582748784	4			1.6073657359	5			1.4709914205				
	2		1.1540280373	3			1.1475149528	3			2.3218290989				
		1	1.6958256092	1	2.0642232519										
F	s	0.9045372537 0.6906826414	7	2.8912820018	s p d f g	0.6402660296 0.7430805629	8	3.5902858378	s p d f g h	0.7997414600 1.3818006856	8	5.5959539071	0.7997414600 1.3818006856	5	4.2907728963
	3		4.1152554296	4			4.2900928233	5			2.0468291630				
	3		3.3262591314	4			2.1122743399	5			1.9007856183				
	2		1.6515191908	3			1.5924268929	3			2.9521731600				
		1	2.1757801824	1	2.6314412498										

The best values for the parameters α of B, C, N, O and F atoms, together with the fixed value of α for the H atom, obtained from the Iffco method, are shown in Table 18. From the optimized parameters in Table 18, we still found the similarity of parameters in many cases comparing with those of the other methods. It can be concluded that the Iffco (local) method is able to find good minima although it lacks global convergence.

The best objective function values (errors ($\Delta I/|E_{mp2}|$) per atom), obtained from GA, GA/Powell hybrid, Direct/Powell hybrid, Direct/Iffco and Iffco, are summarized in Tables 8, 10, 13, 14 and 17, respectively. The errors using the density/Poisson auxiliary basis functions, are reduced to a few ppb at most and do not significantly depend on the method. Hence, the main criterion for method comparison is the time needed to get these results.

Most computational time was spent in evaluating energies by MOLPRO.⁴⁰ Therefore, the time needed is roughly proportional to the number of function evaluations. CPU time (min.) to optimize auxiliary basis sets for H₂, BH₃, CH₄, NH₃, H₂O and HF at the cc-pVTZ basis level is presented in Figure 13 (obtained with an Intel Itanium2/1.3 GHz).

As the data from this Figure and the errors obtained from GA and GA/Powell show, the GA/Powell hybrid can get results of the same quality as the GA alone but in less time. Thus, this hybrid is a good combination of global and local optimization methods. Although, the hybrids of Direct with Powell and Iffco also produced results as good as the other methods, the principal disadvantage of the hybrids of Direct and local search is that they are still slower than the other methods.

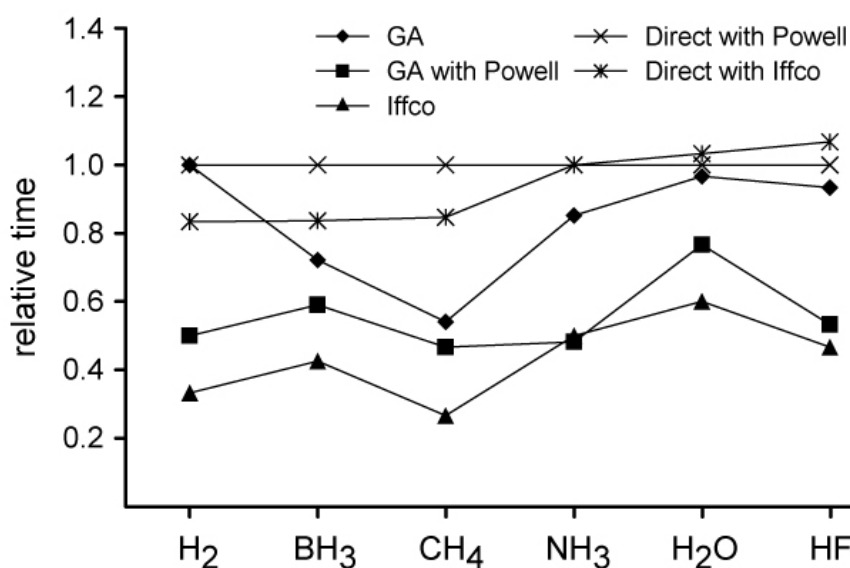


Figure 13 Relative time spent for optimizing mixed Poisson/density auxiliary basis sets for the standard cc-pVTZ bases of H₂, BH₃, CH₄, NH₃, H₂O and HF molecules with different optimization methods, for clarity setting the time for the Direct with Powell method to 1.0 for each molecule. The actual absolute times for H₂, BH₃, CH₄, NH₃, H₂O and HF are 6, 61, 150, 54, 30 and 15 min., respectively.

As we already remarked, it is surprising that a local optimization algorithm alone, Iffco, is as fast and robust and as good as the global optimization and hybrid methods. Not surprisingly, however, it is much faster for small search spaces than for large ones. This is demonstrated in Table 19, where we show results from our first round of test runs where we optimized all even-tempered parameters for all atoms. This leads to a maximum of 25 parameters to optimize for the molecules BH₃, CH₄, NH₃, H₂O and HF and consequently to very steeply increasing computation times for Iffco, more steeply than for the global methods.

Table 19 The best objective function and CPU time (min). The calculations were performed by using GA (with maximum number of 1000 generations), GA/Powell hybrid and Iffco methods for H₂, BH₃, CH₄, NH₃, H₂O and HF molecules. All parameters in all atoms were optimized.

	$\Delta I / E_{MP2} $					
	H ₂	BH ₃	CH ₄	NH ₃	H ₂ O	HF
GA						
cc-pVDZ	1.46854E-8	2.97798E-9	3.25660E-9	2.54273E-9	2.09606E-9	1.48109E-9
cc-pVTZ	3.48280E-8	2.33290E-9	4.58607E-9	3.60680E-9	4.66430E-9	6.13816E-9
cc-pVQZ	2.29836E-8	1.80359E-8	1.43607E-8	2.18190E-9	3.32640E-9	2.95210E-9
GA/Powell						
cc-pVDZ	1.46836E-8	2.51141E-9	3.21106E-9	2.52550E-9	2.08352E-9	1.49277E-9
cc-pVTZ	3.47187E-8	2.04734E-9	2.42343E-9	3.60269E-9	4.48755E-9	7.47548E-9
cc-pVQZ	2.21123E-8	1.96177E-9	2.73652E-9	1.80470E-9	2.02224E-9	2.74376E-9
Iffco						
cc-pVDZ	1.46836E-8	2.51141E-9	3.21106E-9	2.52550E-9	2.08352E-9	1.38642E-9
cc-pVTZ	3.47187E-8	2.04734E-9	2.42343E-9	3.60269E-9	4.48755E-9	5.81769E-9
cc-pVQZ	2.21123E-8	1.96177E-9	2.73652E-9	1.80470E-9	2.02224E-9	2.17334E-9
	Time (min.)					
	H ₂	BH ₃	CH ₄	NH ₃	H ₂ O	HF
GA						
cc-pVDZ	20	58	74	59	63	44
cc-pVTZ	28	219	406	230	138	90
cc-pVQZ	86	1986	4291	2190	969	380
GA/Powell						
cc-pVDZ	3	18	29	19	12	15
cc-pVTZ	5	114	221	99	61	23
cc-pVQZ	24	1185	2454	1090	556	179
Iffco						
cc-pVDZ	2	44	81	63	19	16
cc-pVTZ	9	219	450	216	163	81
cc-pVQZ	24	2052	8812	4560	1552	413

The errors $\Delta I / |E_{MP2}|$ obtained from different methods are smaller than 0.035 ppm in all cases. In more than 80% of all cases the errors were smaller than 0.008 ppm. Again, there are no significant accuracy differences between the methods.

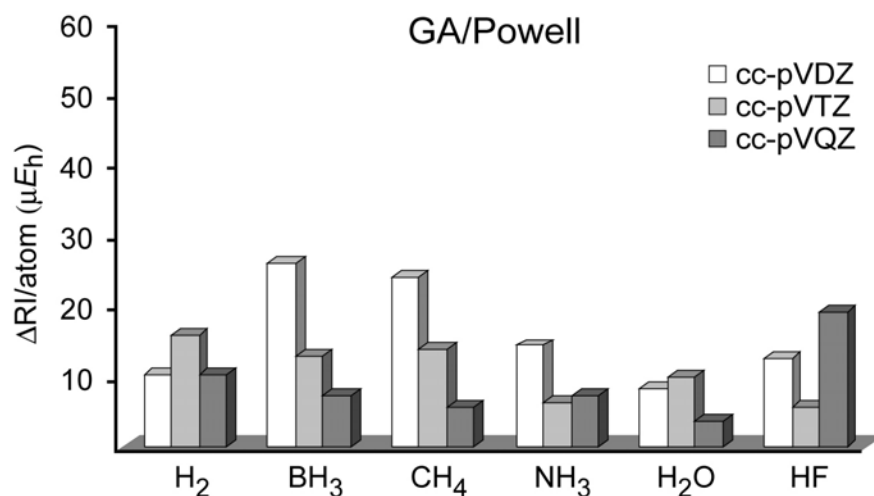


Figure 14 Histogram of the energy differences (Δ RI) per atom between MP2 and RI-MP2 calculations, computed by using auxiliary basis sets which both optimized parameters α and β in all atoms.

Figure 14 shows the energy differences between MP2 and RI-MP2 (Δ RI) calculations using the optimized parameters α and β of all atoms, obtained from the GA/Powell method only. It is found that the errors, in all cases, are smaller than $26 \mu E_h$. The optimized centers α and ratios β found in this way are presented in Table 20.

The optimized center parameters α of the Poisson functions systematically increase from B to F atoms, but not in all cases. As we mentioned above, the optimized ratio parameters β often fall between the values 2.0 to 3.0. Interestingly, the optimized center parameters α obtained from optimizing parameters α and β of all atoms are not similar with the results that we obtained from optimizing only parameters α , with fixed β and H atom parameters. Nevertheless, as shown above (Figure 14), this does not affect the quality of the results. It can be concluded that we found many local minima which gave a good objective function very close to the final minimum by different methods.

Table 20 Overall best values for the optimized center parameters α and ratio parameters β of even-tempered auxiliary basis sets, complementing standard bases cc-pVXZ with X = D, T and Q for H₂, BH₃, CH₄, NH₃, H₂O and HF molecules, obtained from the GA/Powell hybrid method

molecule	cc-pVDZ						cc-pVTZ						cc-pVOZ					
	GTO		n	Poisson			GTO		n	Poisson			GTO		n	Poisson		
	α	β		α	β		α	β		α	β		α	β				
H ₂	s p d	0.887147	3	1.491170	3.180388	s p d f	0.903054	4	1.292364	2.514385	s p d f g	0.860429	5	1.369173	2.366795			
			1	1.243454				2	1.143106	2.225097			3	1.259618	2.023611			
				0.575598				2	0.592045	3.311238			2	0.639799	3.137368			
								1	1.298437				2	0.960861	2.531998			
BH ₃	H	s p d	0.614275 0.944333	3	0.852984	3.117557	s p d f	0.705591 1.521546	4	1.500446	2.434974	s p d f g	1.296804 2.514366	5	1.732581	2.193398		
				1	0.985205				2	0.882145	2.650443			3	1.205523	1.928019		
									1	1.178896				2	0.898782	5.006531		
									1	0.920912				2	0.973487	2.388644		
	B	s p d f	3.182511 7.989573 0.254373	7	1.390198	2.095808	s p d f g	1.000958 3.719289 1.221589	8	1.847277	2.165843	p d f g h	0.509481 1.736804 0.715846	8	2.834136	2.401529		
				3	0.417868	2.077575			4	0.789115	3.886626			5	2.392312	2.686835		
				3	0.763991	2.195374			4	0.641712	2.055278			5	0.875928	1.861587		
				2	0.265767	3.402926			3	0.442047	2.240630			5	0.515536	1.997065		
									1	0.615952				3	0.558446	2.006949		
														1	0.788368			
CH ₄	H	s p d	0.541117 1.218158	3	0.696997	3.603043	s p d f	0.709910 0.965062	4	0.676131	3.000981	s p d f g	0.786009 0.951084	5	1.951875	2.116072		
				1	0.876742				2	0.815071	3.042045			3	1.138409	2.060456		
									1	1.309998				2	1.219176	2.674431		
									1	1.145048				2	1.153873	2.049141		
C	s p d f	1.134951 3.540316 1.537301	7	1.715752	2.369668	s p d f g	0.425313 4.376722 0.834583	8	3.249043	2.060022	p d f g h	8.100862 2.941786 2.090449	8	3.245700	2.696680			
			3	1.685792	3.968200			4	2.247295	2.418027			5	0.755098	2.177828			
			3	1.138214	2.153959			4	0.941438	2.100281			5	1.661054	1.804873			
			2	0.759515	2.176575			3	0.750970	2.192979			5	0.608064	1.941361			
								1	0.952001				3	0.981758	1.960013			
													1	1.212660				

Table 20 (continued)

molecule	cc-pVDZ						cc-pVTZ						cc-pVQZ					
	GTO		n	Poisson			GTO		n	Poisson			GTO		n	Poisson		
	α	β		α	β		α	β		α	β		α	β		α	β	
NH ₃																		
H	s		3	1.128374	2.849661	s		4	1.772967	2.311005	s		5	1.248716	2.221343			
	p	0.831991	1	1.150957		p	0.441937	2	0.916437	2.613555	p	1.050746	3	1.338172	1.869995			
	d	1.278749				d	1.741604	1	1.359591		d	2.529146	2	0.839168	5.685808			
								1	1.203565				2	1.101662	2.310156			
N	s		7	1.756778	2.258898	s		8	2.356672	2.245664	s		8	1.440817	2.122411			
	p	0.756340	3	1.756673	5.109962	p	0.641600	4	2.208575	3.138212	p	0.549108	5	4.664550	2.122411			
	d	1.964812	3	1.151113	2.624151	d	0.585366	4	1.321383	2.167491	d	2.539000	5	2.208765	2.526595			
	f		2	0.798102	2.519388	f		3	0.989160	2.204352	f	3.689084	5	1.427908	2.018782			
								1	1.321943				3	0.913845	1.992298			
													1	1.331633	1.953549			
													1	1.659822				
H ₂ O																		
H	s		3	1.198694	2.835421	s		4	1.143113	2.498312	s		5	1.987341	2.006880			
	p	0.797902	1	1.180395		p	1.110210	2	1.270854	2.206321	p	2.942041	3	1.157735	2.027932			
	d	1.323043				d	1.759650	1	1.322858		d	0.986433	2	1.559998	1.844751			
								1	1.050087				2	1.186611	2.264402			
O	s		7	2.342118	2.284600	s		8	3.137808	2.273975	s		8	1.382665	2.217501			
	p	0.666879	3	2.307590	5.338825	p	0.590907	4	3.330719	2.580808	p	11.901310	5	5.284883	2.217501			
	d	5.512445	3	2.303383	2.347358	d	0.841943	4	1.955561	2.239383	d	9.976264	5	1.484470	1.960116			
	f		2	1.098685	2.588346	f		3	1.407368	2.287659	f	4.438651	5	1.695248	2.033911			
								1	1.692173				5	1.141700	2.076417			
													3	1.745765	2.027028			
													1	2.118358				
HF																		
H	s		3	2.059053	2.131290	s		4	7.073615	3.564349	s		5	1.899505	2.079066			
	p	1.857025	1	1.396690		p	5.108017	2	2.247871	2.678533	p	3.774066	3	8.570042	8.405213			
	d	1.346959				d	0.978539	1	1.447889		d	2.906279	2	0.455569	2.931473			
								1	1.062350				2	1.210159	2.398416			
F	s		7	3.012295	2.270353	s		8	4.701826	2.200341	s		8	1.369359	2.130831			
	p	1.309047	3	3.104414	5.445520	p	1.331391	4	6.929997	4.672907	p	1.219918	5	7.916413	2.130831			
	d	2.419195	3	2.720683	2.290442	d	5.184381	4	2.654710	2.136179	d	1.408918	5	3.923869	2.609231			
	f		2	1.538955	2.615556	f		3	1.689149	2.420394	f	2.394304	5	3.659304	2.119237			
								1	2.173420				5	1.467933	2.105938			
													3	2.064969	2.153688			
													1	2.647511				

For reference purposes, we have also optimized both parameters α and β for B, C, N, O and F while fixing α and β parameters of the H atom at their H₂ optimized values, using the GA/Powell method only. We found that these results, as shown in Table 21, are almost of the same quality as the results where we optimized all parameters (Table 19).

Table 21 Summary of the best objective function values ($\Delta I/|E_{MP2}|$) per atom, number of function evaluations and CPU time (min), optimizing parameters α and β for the B, C, N, O and F atoms (with α and β parameters for the H atom fixed).

molecule	$\Delta I/ E_{MP2} $	f-eval	time(min.)
H ₂			
cc-pVDZ	3.21398E-09	844	3
cc-pVTZ	3.17014E-09	2871	11
cc-pVQZ	3.64545E-09	4695	66
BH ₃			
cc-pVDZ	4.54500E-09	2681	21
cc-pVTZ	4.47814E-09	3772	136
cc-pVQZ	1.72035E-09	2360	768
CH ₄			
cc-pVDZ	4.76659E-09	4862	51
cc-pVTZ	4.28002E-09	3382	227
cc-pVQZ	2.26327E-09	3485	2494
NH ₃			
cc-pVDZ	3.24644E-09	2970	28
cc-pVTZ	5.13607E-09	2578	109
cc-pVQZ	2.59434E-09	4037	1550
H ₂ O			
cc-pVDZ	2.87785E-09	2254	14
cc-pVTZ	5.22395E-09	4832	108
cc-pVQZ	2.70741E-09	3011	488
HF			
cc-pVDZ	1.87316E-09	3459	18
cc-pVTZ	6.75670E-09	2127	23
cc-pVQZ	2.72047E-09	3602	225

The results from Table 21 show that all errors $\Delta I/|E_{\text{MP2}}|$ are found to be smaller than 0.007 ppm. The energy difference between MP2 and RI-MP2 methods (ΔRI) are depicted graphically in Figure 15. We observed that in all cases the energy differences (ΔRI) between conventional MP2 and RI-MP2 calculations are smaller than $30 \mu E_h$.

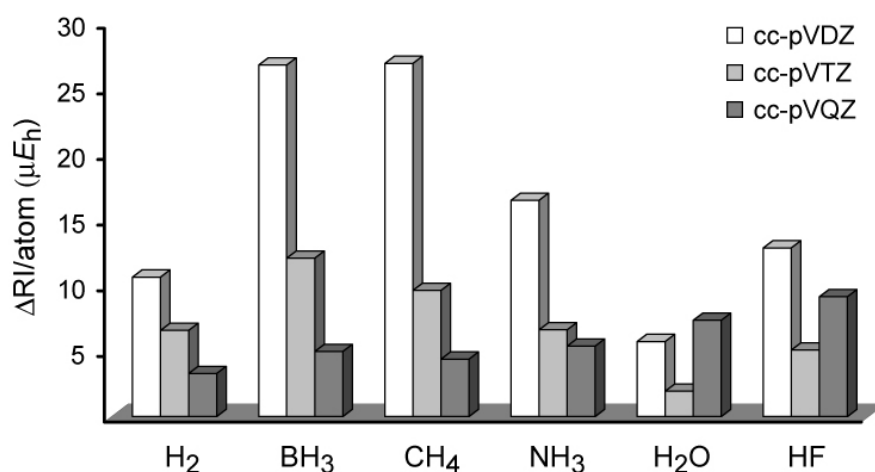


Figure 15 Histogram of the energy differences between MP2 and RI-MP2 calculations per atom, optimizing both parameters α and β for B, C, N, O and F using the GA/Powell method while fixing α and β parameters of the H atom at their H₂ optimized values.

For reference, the overall best values of this whole study for the parameters α and β are shown in Table 22. The optimized parameters α and β for the H atom are obtained from calculations on the H₂ molecule. The α and β parameters for the B, C, N, O and F atom are obtained from optimization of BH₃, CH₄, NH₃, H₂O and HF molecules using the fixed value of α and β for the H atom. These data can be used to reconstruct our best auxiliary basis sets.

Table 22 Overall best values for the parameters α and parameters β , complementing standard bases cc-pVXZ with X = D, T and Q, obtained from optimization of α and β parameters for the B, C, N, O and F atoms only (with parameters for the H atom fixed).

molecule	cc-pVDZ						cc-pVTZ						cc-pVQZ					
		GTO		n	Poisson			GTO		n	Poisson			GTO		n	Poisson	
		α	β		α	β		α	β		α	β		α	β			
H ₂	s			3	1.475591	3.157296	s			4	1.15024471	2.457073	s			5	1.469554	2.434055
	p	0.927004		1	1.309313		p	0.623428		2	1.31907721	2.158793	p	1.638598		3	1.333151	1.947071
	d	1.048366					d	1.910498		1	1.62886085		d	0.933403		2	1.721750	1.954350
	f						f			1	1.03631364		f			2	0.962757	2.439676
BH ₃	s			7	0.863907	2.158481	s			8	2.314043	2.034591	s			8	3.214398	2.010192
	p	0.297344	2.937930	3	1.107397	2.836344	p	0.493829	4.057568	4	1.416923	2.917074	p	0.709187	3.114376	5	5.019848	4.007095
	d	0.159235		3	0.620064	2.382984	d	0.399964		4	0.692161	1.937720	d	0.585814		5	0.806001	1.876939
	f			2	0.224741	4.227423	f			3	0.377221	2.311584	f			5	0.355522	1.992155
							g			1	0.581597		g			3	0.559375	2.026009
CH ₄	s			7	1.557120	2.103902	s			8	2.661624	2.062775	s			8	2.715941	2.233220
	p	0.488617	2.005066	3	1.520215	3.608923	p	0.485676	4.014175	4	1.743234	3.219092	p	0.697189	6.229326	5	2.416480	2.015606
	d	0.363170		3	0.998089	2.285215	d	0.685291		4	0.959973	2.033676	d	5.834248		5	0.768619	1.570624
	f			2	0.636956	2.211102	f			3	0.477768	2.750334	f			5	0.532591	2.010747
							g			1	0.920917		g			3	0.957231	2.010119
NH ₃	s			7	1.806426	2.208373	s			8	3.961446	2.042071	s			8	4.264942	2.222851
	p	0.588211	3.027054	3	1.807148	4.473839	p	0.630491	3.518095	4	2.296725	3.096615	p	0.679637	1.639650	5	2.187809	2.537142
	d	0.487539		3	1.154937	2.564671	d	0.851197		4	1.435752	2.063024	d	0.281665		5	1.544660	1.972353
	f			2	0.797408	2.486125	f			3	0.982256	2.226152	f			5	0.827150	2.066045
							g			1	1.330653		g			3	1.262024	1.985455
H ₂ O	s			7	2.179294	2.262251	s			8	5.194436	2.064022	s			8	5.781340	2.243238
	p	0.869947	2.590126	3	2.314203	4.833912	p	1.147036	2.233712	4	2.978112	3.159964	p	0.892709	1.851888	5	3.092883	2.544185
	d	0.537849		3	1.456908	2.772817	d	0.736632		4	1.756002	2.302378	d	0.465167		5	1.752955	2.192444
	f			2	1.147301	2.443979	f			3	1.216391	2.387487	f			5	1.069945	2.109672
							g			1	1.718261		g			3	1.777062	2.005021
HF	s			7	4.834444	1.979656	s			8	3.740462	2.330844	s			8	7.901731	2.139694
	p	1.048992	3.790130	3	3.166857	4.950369	p	0.628267	2.402664	4	4.298211	2.573587	p	1.186611	1.676028	5	3.973864	2.583769
	d	0.784467		3	2.151777	2.658653	d	4.260591		4	2.083167	2.297166	d	1.445148		5	3.293375	1.980698
	f			2	1.659248	2.520184	f			3	1.592359	2.500087	f			5	1.487316	2.117671
							g			1	2.180601		g			3	2.202148	2.097104
												h			1	2.636022		

In further calculations, the accuracy of the optimized auxiliary basis sets, complementing the standard sets cc-pVDZ, cc-pVTZ and cc-pVQZ, was systematically tested for the various optimization methods and the test molecules H₂, BH₃, CH₄, NH₃, H₂O and HF. In order to assess the quality of the mixed conventional/Poisson auxiliary basis sets determined in this work we want to compare with the purely conventional auxiliary basis sets optimized by Weigend et al.³² The errors ($\Delta I/|E_{MP2}|$) obtained from these two different auxiliary basis sets are shown in Table 23.

Table 23 Comparing the quality of the conventional auxiliary basis sets optimized by Weigend et al. and the mixed conventional/Poisson auxiliary basis sets optimized in this work. The errors ($\Delta I/|E_{MP2}|$) are presented in ppb.

atom	$\Delta I/ E_{MP2} $	
	mixed auxiliary basis sets	conventional auxiliary basis sets
H		
cc-pVDZ	3.21398	130.90500
cc-pVTZ	3.17014	34.48600
cc-pVQZ	3.64545	118.96800
B		
cc-pVDZ	4.54500	8.57600
cc-pVTZ	4.47814	20.56200
cc-pVQZ	1.72035	14.79900
C		
cc-pVDZ	4.76659	1.39500
cc-pVTZ	4.28002	15.19200
cc-pVQZ	2.26327	18.36500
N		
cc-pVDZ	3.24644	0.66000
cc-pVTZ	5.13607	11.00800
cc-pVQZ	2.59434	17.69100
O		
cc-pVDZ	2.87785	1.18700
cc-pVTZ	5.22395	108.68300
cc-pVQZ	2.70741	46.52900
F		
cc-pVDZ	1.87316	1.19600
cc-pVTZ	6.75670	151.77300
cc-pVQZ	2.72047	43.25600

Although we used only even tempered basis sets, in contrast to Weigend et al. who optimized all exponents, we found that the performance of the basis sets optimized in this work is as good (for cc-pVDZ) or even a lot better (for cc-pVTZ and cc-pVQZ) than that of the auxiliary basis sets optimized by Weigend et al.³² Moreover, we followed the procedure suggested by them and calculated several correlation energy error measures for a test set of 40 sample molecules, which is listed in Table 24.

Table 24 Test set of 40 sample molecules which contain only hydrogen and first row elements.

Sample molecules
B ₂ H ₆ , B ₄ H ₄ , BF ₃ , BH ₃
BH ₃ CO, BH ₃ NH ₃ , B ₂ N ₃ H ₆
C ₂ H ₂ , C ₂ H ₃ N, C ₂ H ₄ , C ₂ H ₆
C ₄ H ₄ , C ₆ H ₆ , CF ₄ , CH ₂ O
CH ₂ O ₂ , CH ₃ N, CH ₃ OH, CH ₄
CO, CO ₂ , F ₂ , H ₂
H ₂ CO ₃ , H ₂ O ₂ , HCN
HF, HCN, HNO, HNO ₂
HNO ₃ , N ₂ , N ₂ H ₂
N ₂ H ₄ , N ₄ , NF ₃ , NH ₃
NH ₄ F, OF ₂

These correlation energy errors are shown in Table 25, in direct comparison to the corresponding values obtained by Weigend et al.³²

Table 25 Mean relative error $\overline{\Delta E}_{CORR} = (1/n) \sum (E_{Corr,RI-MP2}^i - E_{Corr,MP2}^i) / E_{Corr,MP2}^i$, the standard deviation and maximum error for a test set of 40 molecules.

basis set	cc-pVDZ	cc-pVTZ	cc-pVQZ
Conventional RI-MP2 ^a			
$\overline{\Delta E}_{Corr}$	-0.008126	-0.005897	-0.003157
$\overline{\Delta E}_{Corr}^{Std}$	0.005613	0.003708	0.001976
$\overline{\Delta E}_{Corr}^{Max}$	-0.020375	-0.01185	-0.006668
RI-MP2 with mixed auxiliary basis set			
$\overline{\Delta E}_{Corr}$	-0.008207	-0.003115	-0.002492
$\overline{\Delta E}_{Corr}^{Std}$	0.005722	0.005242	0.002340
$\overline{\Delta E}_{Corr}^{Max}$	-0.020829	-0.011850	-0.006507

^a Ref. 28

From these results, the performance of the basis sets optimized in this work is as good (for cc-pVDZ) or even slightly better (for cc-pVTZ and cc-pVQZ) than that of the auxiliary basis sets optimized by Weigend et al. Accordingly, applying systematic optimization methods is an advantage over traditional optimization “by hand”. Another result which should be mentioned is that in the course of the optimization we found many local minima which gave a value of the objective function very close to the final minimum. Hence many of these local minima would yield very accurate auxiliary basis sets of high quality, too. Of course, the number of low-quality local minima is still much larger, such that a superior performance of global or at least more sophisticated local optimization methods is to be expected. Comparing with the traditional approach to optimize auxiliary basis sets for RI-MP2 calculations, our algorithms are more systematic and can be automatized more easily. Hence, they can reasonably be expected to be faster and more reliable. In the publication of Weigend et al. there is no indication on how much time their basis set optimizations took. From informal, indirect information we know that their calculations ran for many months, with sporadic irregular occurrence of small improvements. In contrast, with our methods, good results can be obtained reliably in minutes or hours.

3.3.2 RI-HF

The resolution of the identity approximation is also applied to generate the Fock matrix in Hartree-Fock calculations,¹⁷ and the Poisson method is also used to replace almost all Coulomb integrals with simple overlaps²⁷. Here we decided to generate auxiliary basis sets for RI-HF by using the GA/Powell and Iffco methods. To access the error in the HF exchange energies, equation (98), with the mixed conventional/Poisson basis sets, we have set up a test set of 20 sample molecules. This test set of small molecules, which is listed in Table 4, was chosen for each element H, B, C, N, O and F, in order to obtain balanced basis sets for different bonding situations.

In our investigations, all parameters α and β of H were optimized on the H₂ molecule, and for the B, C, N, O and F atoms, all parameters α and β were optimized on the sample molecules in Table 4, while fixing the α and β parameters of the H atoms at their H₂ optimized values. These mixed conventional/Poisson basis sets are used to compute the exchange part of the Fock matrix. The accuracy of the optimized auxiliary basis sets, complementing the standard sets cc-pVDZ, cc-pVTZ, cc-pVQZ and cc-pV5Z, is tested.

As we remarked in the previous section, the GA/Powell hybrid method is an efficient algorithm for finding good minima. The performance of the GA/Powell hybrid method is demonstrated in Table 26.

Table 26 Summary of the best objective function values (Δ RI) per atom for RI-HF calculations, for auxiliary basis sets complementing standard bases cc-pVXZ with X = D, T, Q and 5. The calculations were performed using the GA/Powell hybrid method.

molecule	cc-pVDZ	cc-pVTZ	cc-pVQZ	cc-pV5Z
H ₂	4.23502E-06	2.03032E-07	1.61342E-07	2.71996E-08
BH ₃	1.41593E-05	1.39626E-06	2.89464E-07	1.50704E-07
CH ₄	2.27640E-05	2.74824E-06	5.97006E-07	3.17088E-07
C ₂ H ₂	9.18700E-05	1.23558E-05	5.95586E-06	2.18234E-06
C ₂ H ₄	6.06550E-05	6.06504E-06	3.20614E-06	1.60125E-06
C ₂ H ₆	3.44463E-05	4.14515E-06	6.90253E-07	3.45291E-07
H ₂ O	1.26847E-04	8.38378E-06	4.11164E-06	3.77682E-06
H ₂ O ₂	2.27530E-04	2.74725E-05	6.61660E-06	5.09195E-06
CH ₂ O	1.51833E-04	1.62403E-05	5.01759E-06	4.37734E-06
CH ₃ OH	8.93800E-05	8.66829E-06	2.24981E-06	1.81375E-06
NH ₃	5.92275E-05	5.98586E-06	3.99826E-06	1.40783E-06
N ₂	3.60155E-04	4.60397E-05	2.52680E-05	7.16901E-06
CHN	1.75020E-04	2.21351E-05	5.79198E-06	2.60910E-06
NHO	2.43040E-04	3.26108E-05	1.01403E-05	5.52275E-06
CH ₃ N	9.09240E-05	9.95513E-06	3.64135E-06	1.38599E-06
HF	2.02765E-04	2.80137E-05	8.40512E-06	5.46092E-06
F ₂	4.81905E-04	5.36950E-05	1.44079E-05	1.86816E-05
HCF	2.37197E-04	2.79173E-05	1.65865E-05	1.30904E-05
HOF	3.03450E-04	3.43367E-05	9.39439E-06	7.03209E-06
CH ₃ F	2.10342E-04	8.14800E-05	1.72867E-05	3.14805E-06

From these results, it can be seen that the error computed with these mixed auxiliary basis sets is small and somewhat sensitive to the basis set sizes. We observe a decrease of the error for the cc-pVXZ basis sets with increasing cardinal number X, the only slightly different result is found for the F₂ molecule. Moreover, we see that the error is reduced to a few microhartree at most, for cc-pVQZ and cc-pV5Z basis sets. The optimized values of α and β parameters for H, B, C, N, O and F atoms, obtained from the GA/Powell hybrid method using 20 sample molecules, are shown in Table 27.

Table 27 Size n and overall best values for the center parameters α and ratio parameters β of even-tempered auxiliary basis sets for RI-HF calculations, complementing standard bases cc-pVXZ with $X = D, T, Q$ and 5 for H, B, C, N, O and F atoms, obtained from the GA/Powell hybrid method using 20 sample molecules.

atom	cc-pVDZ					cc-pVTZ					
		GTO		Poisson			GTO		Poisson		GTO
		α	n	α	β		α	n	α	β	
H	s	1.005752	2	1.261755	5.318587	s	0.737542	4	2.341904	3.143875	
	p	1.725496	1	1.226951		p	0.406173	3	1.003177	2.468621	
	d		1	1.269532		d		2	0.759498	5.368544	
						f		1	0.779707		
B	s	0.447898	6	9.194377	2.781455	s	1.795998	10	17.524311	2.128044	
	p	0.824156	4	2.818668	2.814522	p	0.774770	7	2.779521	2.217297	
	d	0.562341	3	1.114788	3.144812	d	0.812340	5	1.212003	1.994157	
	f		2	0.684652	3.073939	f		2	0.583101	2.479241	
						g		1	0.950981		
C	s	1.361895	6	11.691273	2.790423	s	0.438631	10	12.080716	2.284049	
	p	0.870642	4	2.957805	3.015756	p	0.537671	7	3.883550	2.245892	
	d	0.345502	3	1.506009	2.521351	d	0.995338	5	1.558007	2.085788	
	f		2	0.442970	3.275615	f		2	0.651163	2.724703	
						g		1	1.014407		
N	s	1.103251	6	10.199401	3.021271	s	15.160095	10	15.500703	2.230528	
	p	1.243492	4	4.142005	3.006849	p	1.513228	7	4.987278	2.162258	
	d	0.567403	3	2.600942	2.635302	d	0.449413	5	2.921522	2.118873	
	f		2	0.698815	2.895883	f		2	0.918269	2.864964	
						g		1	1.419844		
O	s	2.235033	6	19.244294	2.823827	s	19.684415	10	20.193645	2.226266	
	p	1.325685	4	4.569546	3.209070	p	1.931420	7	6.487155	2.180970	
	d	1.311409	3	2.522249	2.873828	d	0.472485	5	3.483835	2.210440	
	f		2	0.709076	3.323529	f		2	1.251700	2.753998	
						g		1	1.262651		
F	s	1.727505	6	15.924445	3.052968	s	9.628639	10	21.480488	2.219365	
	p	1.939272	4	6.472626	3.119759	p	2.451822	7	8.323853	2.198496	
	d	0.922916	3	4.552007	2.705142	d	0.839203	5	3.108354	2.338414	
	f		2	0.944605	2.829576	f		2	0.859840	2.756664	
						g		1	1.232155		

Table 27 (continued)

atom	cc-pVQZ					cc-pV5Z					
		GTO		Poisson			GTO		Poisson		GTO
		α	n	α	β		α	n	α	β	
H	s	2.725658	4	2.807335	2.230375	s	1.412399	4	1.809116		3.963564
	p	0.219693	3	1.503848	2.894843	p	5.394865	2	1.465322		7.438865
	d	1.218135	3	0.821734	1.860239	d	0.213523	5	0.668475		2.360170
	f		2	0.395885	2.950831	f	0.430319	3	1.495753		2.053091
	g		1	5.632509		g		2	1.349480		2.451896
						h		1	2.615684		
B	s	13.863795	10	14.201503	2.106662	s	0.600591	10	14.498795		2.193083
	p	0.668059	7	4.925288	2.147424	p	0.422416	7	3.503128		2.345170
	d	0.464589	5	1.558556	2.150786	d	0.428659	5	1.482025		2.184947
	f	0.966247	3	1.554229	4.250284	f	0.186167	4	1.105106		2.099512
	g		2	0.690622	2.043551	g	0.518577	3	0.842952		2.027043
	h		1	0.923090		h		2	1.466985		2.514636
					i		1	1.128161			
C	s	7.364298	10	16.108998	2.160494	s	0.620995	10	28.377458		2.127211
	p	0.765701	7	5.744972	2.180794	p	4.663487	7	6.843844		1.943244
	d	0.397938	5	2.587434	2.010520	d	0.624811	5	2.087694		2.250298
	f	0.476346	3	1.585068	2.007090	f	0.290683	4	1.032415		1.884760
	g		2	0.792092	2.057494	g	1.024548	3	2.036622		2.436954
	h		1	1.162299		h		2	1.333163		2.063181
					i		1	1.160100			
N	s	5.508034	10	12.842150	2.266411	s	2.033658	10	11.796289		2.338065
	p	0.941170	7	3.708244	2.405103	p	0.813938	7	6.763712		2.274171
	d	3.121065	5	2.366591	1.924562	d	1.181734	5	2.050014		2.444951
	f	1.602421	3	1.187336	2.476880	f	0.507738	4	1.176001		2.132646
	g		2	0.477662	10.333203	g	0.282345	3	0.877004		2.143812
	h		1	2.625467		h		2	0.896627		4.251305
					i		1	1.345357			
O	s	27.693869	10	28.371129	2.142224	s	21.150427	10	21.660815		2.196694
	p	1.329971	7	5.374222	2.414091	p	11.837554	7	8.882952		1.940745
	d	0.885435	5	3.305534	2.274288	d	0.897996	5	5.894670		2.038108
	f	1.752272	3	1.358057	3.306055	f	0.313570	4	1.110847		2.229831
	g		2	1.036412	2.434700	g	0.334388	3	0.859991		2.571642
	h		1	1.608335		h		2	0.562672		3.637969
					i		1	1.312050			
F	s	21.346019	10	21.658999	2.231176	s	31.985950	10	32.589653		2.138367
	p	2.555269	7	8.688610	2.198120	p	2.147685	7	8.292033		2.354466
	d	0.689257	5	5.021554	2.184871	d	0.648006	5	4.990389		2.210528
	f	1.291627	3	2.072321	2.178462	f	0.626290	4	1.557693		2.030042
	g		2	1.049051	2.354056	g	0.733903	3	1.192668		3.529522
	h		1	1.373545		h		2	0.732416		3.033723
					i		1	1.412287			

The performance of the Iffco method in optimization of these auxiliary basis sets, cc-pVDZ, cc-pVTZ, cc-pVQZ and cc-pV5Z, for the H, B, C, N, O and F atoms, is shown in Table 28.

Table 28 Summary of the best objective function values (Δ RI) per atom for RI-HF calculations, complementing standard bases cc-pVXZ with X = D, T, Q and 5. The calculations were performed using the Iffco method.

molecule	cc-pVDZ	cc-pVTZ	cc-pVQZ	cc-pV5Z
H ₂	4.20191E-06	1.78484E-07	6.76034E-08	6.92227E-08
BH ₃	1.42204E-05	1.87757E-06	9.83072E-07	5.45547E-07
CH ₄	2.19780E-05	2.88941E-06	6.64751E-07	4.76057E-07
C ₂ H ₂	1.17430E-04	9.51724E-06	4.28173E-06	1.10830E-05
C ₂ H ₄	6.66217E-05	5.80762E-06	2.06272E-06	3.95825E-06
C ₂ H ₆	3.33350E-05	4.59126E-06	8.73909E-07	4.45510E-07
H ₂ O	1.49043E-04	9.78437E-06	4.14062E-06	1.84128E-06
H ₂ O ₂	2.71595E-04	2.98725E-05	6.29350E-06	3.20831E-06
CH ₂ O	1.73048E-04	1.64281E-05	4.86020E-06	3.16092E-06
CH ₃ OH	1.06295E-04	9.13730E-06	2.45238E-06	1.20229E-06
NH ₃	7.19150E-05	6.35168E-06	2.98258E-06	1.88677E-06
N ₂	4.35510E-04	5.46750E-05	1.80560E-05	2.63106E-05
CHN	1.91140E-04	2.05025E-05	6.76751E-06	1.09664E-05
NHO	3.07180E-04	3.52700E-05	8.55960E-06	6.52046E-06
CH ₃ N	1.04712E-04	1.05795E-05	3.44700E-06	2.08744E-06
HF	2.03895E-04	2.31986E-05	8.45887E-06	6.14442E-06
F ₂	4.88350E-04	6.29650E-05	1.82763E-05	1.46432E-05
HCF	2.39937E-04	2.75928E-05	1.04202E-05	2.01487E-05
HOF	3.41857E-04	3.95033E-05	9.58095E-06	7.07712E-06
CH ₃ F	2.24866E-04	5.67420E-05	1.71591E-05	6.54994E-06

In most cases the errors obtained from the Iffco method are of similar magnitude as the errors from the GA/Powell hybrid method. We again found a decrease of the error for the cc-pVXZ basis sets with increasing basis set size. However, there are some differences for the H₂, C₂H₂, C₂H₄, N₂, CHN and HCF molecules, which do not correspond to the results from the GA/Powell method. This effect is probably due to

Iffco being only a local optimization algorithm. As in the case of RI-MP2, a problem in the optimization is the occurrence of multiple local minima. Even though Iffco can find a good minimum, it is not good enough compared with the minimum found by the GA/Powell method. Therefore, in this case the GA/Powell hybrid method appears to be slightly better than the Iffco method. The optimized values of α and β parameters for H, B, C, N, O and F atoms, obtained from the Iffco method using 20 sample molecules, are shown in Table 29.

The optimized center parameters α and ratio parameters β of auxiliary basis set RI-HF series showed the similarity of parameters obtained from two different methods but not in all cases. The ratio parameters β often fall into the interval 2.0 to 3.0 as we already observed in the previous section (RI-MP2). In this case we also found multiple similar minima that give good objective function values close to the global minimum.

Table 29 Size n and overall best values for the center parameters α and ratio parameters β of even-tempered auxiliary basis sets for RI-HF calculations, complementing standard bases cc-pVXZ with $X = D, T, Q$ and 5 for H, B, C, N, O and F atoms, obtained from the Iffco method using 20 sample molecules.

atom	cc-pVDZ					cc-pVTZ				
	GTO		Poisson			GTO		Poisson		
	α	n	α	β	α	n	α	β		
H	s	0.999860	2	1.250883	5.311154	s	1.245973	4	1.535926	4.843246
	p	0.948655	1	1.826418		p	0.685924	3	1.476952	3.933093
	d		1	1.225185		d		2	0.904518	3.918480
						f		1	0.494869	
B	s	9.414832	6	9.734672	2.838115	s	13.761984	10	13.996552	2.119248
	p	0.482052	4	3.393803	2.647697	p	0.666081	7	2.398246	2.263293
	d	0.557571	3	1.086714	3.072803	d	0.762018	5	1.143284	2.003114
	f		2	0.640745	2.800037	f		2	0.549020	2.323626
						g		1	0.787736	
C	s	1.515851	6	12.502552	2.763487	s	14.587810	10	14.797219	2.158565
	p	0.884017	4	3.072766	3.013831	p	1.239585	7	3.985846	2.094597
	d	0.396536	3	1.752990	2.601726	d	0.363994	5	1.276056	2.130796
	f		2	0.428473	3.309639	f		2	0.637462	2.757695
						g		1	1.093562	
N	s	15.618802	6	16.066775	2.759103	s	5.848601	10	13.251158	2.246478
	p	1.024173	4	3.525014	3.196138	p	1.317417	7	4.615892	2.224925
	d	1.050477	3	1.885336	2.666376	d	0.653805	5	2.188086	2.163728
	f		2	0.614006	2.927273	f		2	0.842834	2.754488
						g		1	1.251371	
O	s	4.070172	6	12.496961	2.921618	s	7.742664	10	17.279021	2.224644
	p	0.816888	4	6.476044	2.810142	p	1.801252	7	6.217856	2.207128
	d	1.246778	3	2.360814	2.861595	d	0.487620	5	3.447321	2.174571
	f		2	0.761623	2.972477	f		2	1.267082	2.696000
						g		1	1.240888	
F	s	1.662621	6	15.548668	3.067167	s	8.195508	10	18.702822	2.255796
	p	1.873718	4	6.317637	3.148071	p	2.022683	7	7.340051	2.277656
	d	0.883430	3	4.356640	2.710675	d	2.522412	5	3.940849	2.091191
	f		2	0.937183	2.969605	f		2	0.986039	2.876474
						g		1	1.413646	

Table 29 (continued)

atom	cc-pVQZ					cc-pV5Z				
	GTO		Poisson			GTO		Poisson		GTO
	α	n	α	β	α	n	α	β	β	
H	s	1.827747	4	2.451554	3.762117	s	2.367381	4	2.683450	2.498038
	p	2.191277	3	0.805459	1.992165	p	5.912622	2	0.688984	3.553879
	d	11.810830	3	1.600377	3.540091	d	0.366902	5	1.518297	1.971150
	f		2	1.461072	8.447121	f	0.282137	3	2.668474	3.167651
	g		1	0.582619		g		2	2.330735	8.177349
						h		1	0.744798	
B	s	9.576391	10	20.130214	2.050894	s	28.912236	10	14.214744	2.060064
	p	0.319169	7	2.599948	2.314528	p	0.528589	7	2.324893	2.455386
	d	1.409894	5	1.077327	1.835815	d	14.410539	5	1.219621	1.909280
	f	0.744280	3	1.443083	9.250572	f	1.076840	4	1.325908	2.382148
	g		2	5.693061	34.110882	g	0.263703	3	1.317312	2.722813
	h		1	0.377612		h		2	1.264850	47.600102
C	s	23.261446	10	23.743320	2.058785	s	4.782860	10	22.185418	2.104918
	p	0.711879	7	5.527563	2.209562	p	0.725617	7	6.227530	2.273528
	d	0.483039	5	1.855997	2.349234	d	24.389341	5	1.502319	2.156084
	f	0.884755	3	0.475864	4.073114	f	1.357675	4	0.952420	2.079213
	g		2	0.705374	2.401218	g	0.938739	3	0.514913	2.821510
	h		1	1.229552		h		2	0.488467	5.748482
N	s	37.842290	10	18.018285	2.133356	s	7.996495	10	17.971955	2.192818
	p	1.058529	7	3.988306	2.328198	p	8.281786	7	6.196282	1.952806
	d	1.189319	5	1.925169	2.190323	d	0.525911	5	3.607653	2.098622
	f	1.311751	3	1.013487	2.242655	f	0.275032	4	1.391097	2.191534
	g		2	10.529832	65.648853	g	0.287325	3	1.146243	2.413814
	h		1	1.690946		h		2	9.818647	44.270768
O	s	26.811677	10	27.434632	2.140270	s	2.178170	10	25.207661	2.234169
	p	1.547082	7	5.776002	2.307491	p	1.691253	7	6.822173	2.380213
	d	0.526649	5	3.802868	2.170779	d	0.576092	5	4.310122	2.202946
	f	2.026692	3	1.555693	4.747740	f	1.326376	4	8.551304	7.487635
	g		2	1.039594	2.246091	g	0.550595	3	1.007445	2.318424
	h		1	1.337613		h		2	5.686375	21.169699
F	s	46.704403	10	21.530196	2.212382	s	22.070601	10	22.435074	2.231502
	p	0.725814	7	10.478223	2.150446	p	2.243796	7	8.138636	2.284126
	d	0.719232	5	5.328678	2.180587	d	0.738801	5	5.478919	2.181349
	f	1.169714	3	2.552224	6.904273	f	2.896960	4	4.979194	5.007822
	g		2	0.915825	2.500858	g	1.298063	3	0.838437	3.906227
	h		1	1.103236		h		2	0.604396	5.410107
					g		1	0.399913		

3.4 CONCLUSIONS AND OUTLOOK

In this work, global and local optimization approaches were presented to optimize mixed Poisson/density auxiliary basis sets for H, B, C, N, O and F atoms, complementing standard basis sets cc-pVXZ with X = D, T and Q in RI-MP2 and X = D, T, Q and 5 in RI-HF calculations. For the RI-MP2 calculations, we observed that in more than 94% of all cases the errors ($|\Delta I|/E_{\text{MP2}}$) were smaller than 0.009 ppm. Moreover, the energy differences, $E_{\text{MP2}} - E_{\text{MP2-RI}}$, in more than 93% of all cases were smaller than $35 \mu E_h$. These results are of a quality comparable with corresponding values reported by Weigend et al.³² for pure density auxiliary bases. Comparing with the traditional approach to optimize auxiliary basis sets our algorithms are more systematic and can be automatized more easily. Hence, they can reasonably be expected to be faster and more reliable. Considering the CPU times to obtain results of this quality with different optimization methods, we found that the Iffco and GA/Powell hybrid methods are the fastest and most robust choices. This indicates that the size of the search spaces and the complexity of the optimization landscapes encountered here are at the borderline between requiring local and global optimization methods. For the RI-HF calculations, we found that the error computed with these mixed auxiliary basis sets is sensitive to the basis set sizes. We observed a decrease of the error for the cc-pVXZ basis sets with increasing the basis set sizes from cc-pVDZ, cc-pVTZ, cc-pVQZ to cc-pV5Z, respectively. Moreover, we can see that the error is reduced to a few microhartree in most cases, for cc-pVQZ and cc-pV5Z basis sets. Comparing the performance of the two different methods, it is found that the results from the GA/Powell hybrid method are better than the results from the Iffco method. Therefore, in this case, a combination of global and local search is actually better than a purely local algorithm.

At the time of this work, implementation of integral evaluation of Poisson functions is not yet very efficient in Molpro, therefore actual running times in calculations are not yet as good as they could be. However, comparing to the successful implementations of RI methods with purely conventional auxiliary bases in Turbomole (where these methods gain a factor of about 10 compared to traditional

calculations), it is clear that there is a big advantage in the method by construction. This will be realized as soon as the Molpro development team implements more efficient integral evaluation techniques. Then, the auxiliary bases developed in this work can be used with great advantage. Another important issue is that the numbers of conventional and Poisson auxiliary bases function were fixed at pre-defined values in the present work. Obviously, these numbers should also be optimized. While it is not very difficult to include these numbers into the optimization for certain methods (particularly for GAs), it does become much more difficult and also controversial to construct a good objective function in this case, since the target has to be an ill-defined combination of minimizing the error (which drives the basis set size to infinity) and minimizing the computational effort (which drives the basis set size to zero and strongly depends on personal resources and the actual case at hand). With such a method in place, however, one could completely delegate the tedious human work of searching for an optimal basis set for each new computational chemistry project to the computer.

4. SYSTEM-SPECIFICALLY REPARAMETRIZED SEMIEMPIRICAL METHODS

4.1 INTRODUCTION

Since *ab-initio* calculations lead to huge computational expenses for medium or large molecules, many semiempirical methods have been developed to treat this problem. Semiempirical methods are characterized by their use of parameters derived from experimental data or from the results of *ab-initio* calculations, in order to simplify the calculations. One advantage of methods parametrized using experimental data is their implicit inclusion of electron correlation effects. As such, they are relatively inexpensive and can be practically applied to very large molecules. There are a variety of semi-empirical methods and parametrizations such as MNDO⁴¹, MINDO/3⁴², AM1⁴³, PM3⁴⁴ and PM5⁴⁵. However, dependence on experimental data means that semiempirical methods can not be expected to perform well on unusual types of molecules for which no data are available from which to construct parameters.

Typically, these semiempirical parameters are supposed to be determined specifically for each atom and then to be applicable universally to all other systems. Many successful realisations of this approach show that it can work quite well. Simultaneously, however, it is obvious that the accuracy and reliability of this method is confined, partly because specific atomic parameters necessarily are a compromise between various systems and even between different bonding arrangements (one wants to eliminate the proliferation of parameters which is typical for force fields) and partly because one does not want the normal user to recalibrate the parameter each time.

System-specific parametrization of semiempirical methods has been done before: Rossi and Thular⁴⁶ have suggested to undergo a partial re-optimization of AM1-Parameters with the aid of several reference points on the *ab-initio* reaction

path. The resulting reparametrized NDDO-method then allows for a fast but nevertheless sufficiently accurate computation of the potential energy surface in the surroundings of the reaction path. This method of specific reaction parameters (SRP) has been used a couple of times since then.⁴⁷⁻⁴⁹ Bartlett and Thiel⁵⁰ did further improvements on this method through the application of a coupled-cluster-like Hamiltonian.

This approach has already been used to improve the weaknesses of semiempirical methods. Voityuk et al.⁵¹ reparametrized MNDO in such a way that hydrogen bonds are better reproduced. Other authors have shown that similar effects could be obtained through improved core-core interactions,^{47,52} amongst other things through pair-specific dispersion terms. Jorgensen et al.⁵³ analyzed this situation further and improved the so-called PDDG-variants of semiempirical methods.

In general and also for other situations this approach seems to be far less common. In the field of clusters, Hartke⁵⁴⁻⁵⁶ introduced cluster size specific reparametrization of empirical potential functions. This has been recently extended by Ge and Head to the reparametrization of semiempirical methods⁵⁷⁻⁵⁸.

More sporadic attempts of this kind shall not be accounted for here. In general, in these cases, standard optimization procedures are employed and the electronic ground state is examined, but these are exceptions: Persico et al.³ recently demonstrated that semiempirical reparametrization can lead to a considerable improvement of modeling excited states. Although this is of enormous importance for future work of our group, the specific difficulties in excited state studies will be to select a suitable set of reference data from *ab-initio* calculations and to perform these *ab-initio* calculations at a sufficiently high level. The challenges of the parameter optimization themselves can be expected to be similar for ground and excited states. Therefore, we again focus on a ground-state case here, and, in line with the first part of this work, check if advanced methods of local and global optimization can again transform the reparametrization from the tedious trial-and-error procedure to a simple black-box task, even for a difficult molecular system.

Of course it is obvious that system-specific optimization in the very strict sense cannot be reasonable. In order to do the reparametrization there have to be comparative calculations but then of course one does not need the reparametrized calculations for exactly the same conditions any more. Hence a certain minimum amount of transferability has to be given. Instead of making the field of applicability of a semiempirical parameter set as large as possible, the idea therefore is to keep it as narrow as it is reasonable, which maximizes its accuracy. As in the SRP methods, the reparametrization will be done for a reaction system at a few key geometries and these parameters can then be used for the same system at many other geometries. In this work, such a reparametrization of semiempirical parameters is done for a reaction system that is very challenging for semiempirical methods, for various reasons that will be explained below. Nevertheless, the aim is to obtain quantitative agreement with high-quality *ab-initio* data, such that the reparametrized method can then be used in direct dynamics calculations.

Proton transfers are fundamental processes in nature. Investigations of proton transfer reaction mechanisms are important for understanding many basic biological, physical, and chemical processes such as solvation, photosynthesis, acid-base neutralization, and enzymatic reactions.⁵⁹⁻⁶² Many experimental and computational studies on molecular cluster systems found that the most prominent examples in that respect are nucleic acid base pairs in which two monomers are linked by two hydrogen bonds.⁶³⁻⁶⁶ This directly leads to double proton transfer reactions (DPTRs) which therefore are of importance for biochemical processes and thus life science in general. Hence, both models of base pairs and actual base pairs have been studied by many people both experimentally and theoretically.⁶⁷⁻⁷⁰ Unfortunately, both experimental and theoretical works mostly focus on structures and properties of DNA/RNA, not on the proton/H transfer. Nevertheless, DPTRs have extensively been studied by Limbach et al.,⁶⁷ presenting kinetic isotope effects by using diffraction techniques and proton-transfer dynamics by high-resolution NMR spectroscopy for solid pyrazoles. Moreover, solid pyrazoles are also explored by De Paz et al.⁶⁸ They presented combined computational and experimental works about proton transfer on the pyrazole cyclic dimer, trimer and tetramers as well as on linear oligomers. In other

computational studies, Douhal et al.⁶⁹ investigated the nature of the double proton-transfer process of the 7-azaindole base-pair and Peeters et al.⁷⁰ computed the proton affinity of guanidine and relate system at the MP2/6-31G** level.

Rauhut⁷¹⁻⁷⁴ discovered that certain DPTRs exhibit curious energy profiles, lacking a well-defined saddle point and showing a broad, flat plateau instead. Rauhut's plateau energy profiles happened to be found for DPTRs, but they could also occur in other reaction systems involving at least two steps. Since the energy profile of these reactions is very different from the standard textbook (Eckart) profiles, the expectation is that the dynamical characteristics of these reactions are quite different from that of reactions with Eckart profiles. This is just beginning to be studied. In order to simulate the dynamics theoretically, one has to be able to calculate very many points on the potential energy surface very quickly, which is impossible to do at the MP2/(aug)-cc-pVDZ level (or any other adequate level). On the other hand, one has to achieve a very high degree of quantitative agreement with the *ab-initio* data points, otherwise the plateau may disappear and deform into a non flat shape with different characteristics. For this difficult task, system-specifically reparametrized semiempirical methods may be a possibility. As we know, semiempirical methods have a number of well-known limitations. Types of problems on which they do not perform well include hydrogen bonding, transition structures, molecules containing atoms for which they are poorly parametrized, and so on. The AM1 method does better than MNDO for hydrogen bonding, but still has serious failings⁷⁵. For species with CHO hydrogen bonds, PM3 performed rather poorly but AM1 did fairly well, but for species with HOH hydrogen bond both method had serious problems.⁷⁶ Obviously, the systems of Rauhut constitute a very difficult case for semiempirical methods. Therefore, it is an interesting challenge to check if reparametrization can overcome these difficulties.

Our aim of this work is to check if the reparametrized semiempirical energy profiles are as accurate as the best available *ab-initio* ones (MP2/(aug)-cc-pVDZ level). By their semiempirical nature, they greatly speed up the computing time in any case. Therefore, in this work several system-specifically reparametrized

semiempirical methods were tested to generate one- and two-dimensional potential energy surfaces (PES) for a double proton transfer reaction in a base pair model system consisting of one pyrazole entity and a guanidine molecule, and these were compared with the energy profiles from Rauhut's work.^{71,74}

4.1.1 Semi-empirical methods

Semiempirical^{37,77} approaches are normally formulated within the same conceptual framework as *ab-initio* methods, but they neglect many smaller integrals to speed up the calculations. In *ab-initio* methods, the two-electron multi-center integrals J_{ij} (equation 43) and K_{ij} (equation 45) are solved explicitly. These can take a considerable amount of time to calculate on a computer. In semi-empirical methods these integrals are neglected or parameterized, and only valence shell electrons are considered. In *ab-initio* methods the following equation is solved:

$$Fc = Sc\varepsilon \quad (101)$$

c : eigenvector; F : Fock matrix; ε : eigenvalue; S : overlap matrix. The total electronic energy of the system is given by

$$E = \frac{1}{2} P(H + F) \quad (102)$$

in which P : density matrix; H : one-electron matrix. In semiempirical methods, all overlap integrals arising from the overlap of two different atomic orbitals are neglected. This reduces the overlap matrix to a unit matrix. The secular equation thus reduces to

$$Fc = c\varepsilon \quad (103)$$

These methods employ Slater-type orbitals (STOs), equation (88), as basis set functions and make the following simplifying approximation. (NDDO: neglect of diatomic differential overlap)

$$\int dr_1 \int dr_2 \frac{\phi_\mu^*(1)\phi_\nu(1)\phi_\lambda^*(2)\phi_\sigma(2)}{r_{12}} = \delta_{\mu\nu}\delta_{\lambda\sigma}(\mu\nu|\lambda\sigma) \quad (104)$$

Here $\delta_{\mu\nu} = 1$ if $\mu = \nu$ or if $\mu \neq \nu$ and the functions ϕ_μ and ϕ_ν are on the same atom. In all other cases $\delta_{\mu\nu} = 0$. Likewise, $\delta_{\lambda\sigma} = 1$ if $\lambda = \sigma$ or if $\lambda \neq \sigma$ and the functions ϕ_λ and ϕ_σ are on the same atom, and $\delta_{\lambda\sigma} = 0$ in all other cases. The notation $(\mu\nu|\lambda\sigma)$ refers to the two-electron interaction integral.

We write equation (47) for a Fock matrix element in AO basis

$$F_{\mu\nu}^\alpha = H_{\mu\nu} + \sum_\lambda \sum_\sigma \left[P_{\lambda\sigma}^{\alpha+\beta} \langle \mu\nu|\lambda\sigma \rangle - P_{\lambda\sigma}^\alpha \langle \mu\lambda|\nu\sigma \rangle \right] \quad (105)$$

The methods all use a minimum basis set consisting of a maximum of one atomic orbital for each angular quantum number. The normal basis set for any atom consists of one s and three p orbitals (p_x , p_y and p_z). An important exception to this rule is the calculation of the one-electron two-center integral $H_{\mu\nu}$, which is approximated by:

$$H_{\mu\nu} = S_{\mu\nu} \frac{1}{2} (U_{\mu\mu} + U_{\nu\nu}) \quad (106)$$

where $S_{\mu\nu}$ is the overlap integral between atomic orbital ϕ_μ on an atom and ϕ_ν on another atom, and the U values are atomic orbital constants, supplied as data. The density matrix elements $P_{\lambda\sigma}$, are defined for closed-shell configurations as

$$P_{\lambda\sigma} = 2 \sum_i^{OCC} c_{\lambda i} c_{\sigma i} \quad (107)$$

Continuing with the neglect of differential overlap, all two-electron integrals involving charge clouds arising from the overlap of two atomic orbitals on different centers are ignored. In the Fock matrix, if ϕ_μ and ϕ_ν are on different centers the NDDO matrix element $F_{\mu\nu}^\alpha$ reduces to

$$F_{\mu\nu}^\alpha = H_{\mu\nu} - \sum_{\lambda}^A \sum_{\sigma}^B P_{\lambda\sigma}^\alpha \langle \mu\lambda | \nu\sigma \rangle \quad (108)$$

Thus no Coulombic terms are present in the two-center Fock matrix elements.

If ϕ_μ and ϕ_ν are different but on the same center, then, since a minimal basis set is being used, all integrals of the type $\langle \mu\nu | \lambda\sigma \rangle$ are zero by the orthogonality of the atomic orbitals unless $\mu = \nu$ and $\lambda = \sigma$, or $\mu = \lambda$ and $\mu = \sigma$. The off-diagonal one-center NDDO Fock matrix elements become:

$$F_{\mu\nu}^\alpha = H_{\mu\nu} + 2P_{\mu\nu}^{\alpha+\beta} \langle \mu\nu | \mu\nu \rangle - P_{\mu\nu}^\alpha (\langle \mu\nu | \mu\nu \rangle + \langle \mu\mu | \nu\nu \rangle) \quad (109)$$

If ϕ_μ is the same as ϕ_ν , then, because of the symmetry of the two-electron integrals, the diagonal NDDO Fock matrix elements reduce to:

$$F_{\mu\mu}^\alpha = H_{\mu\mu} + \sum_{\nu}^A (P_{\nu\nu}^{\alpha+\beta} \langle \mu\mu | \nu\nu \rangle - P_{\nu\nu}^\alpha (\langle \mu\nu | \mu\nu \rangle)) + \sum_{\lambda}^B \sum_{\sigma}^B P_{\lambda\sigma}^{\alpha+\beta} \langle \mu\mu | \lambda\sigma \rangle \quad (110)$$

All one-center, two-electron integrals $\langle \mu\mu | \nu\nu \rangle$ and $\langle \mu\nu | \mu\nu \rangle$ are evaluated by a procedure that involves the fitting of the theoretical energies of the atoms to spectroscopic data. The values of these one-center, two-electron interaction integrals and the internuclear distances are used to compute the two-center, two-electron interaction integrals $\langle \mu\nu | \lambda\sigma \rangle$ which are approximated by classical multipole-multipole charge interactions between atoms A and B. The multipole charge

separations within an atom are treated as adjustable parameters, i.e. they are optimised to fit the experimentally derived one-center integrals.

The total energy of the molecule, E_{total} , is the sum of the total valence electronic energy, E_{el} , and the energy of repulsion between the cores on atoms A and B.

$$E_{\text{total}} = E_{\text{el}} + \sum_{B>A} \sum_A (Z_A Z_B \langle AA|BB \rangle + E_N(A, B)) \quad (111)$$

In MNDO⁴¹ (Modified Neglect of Diatomic Overlap)

$$E_N(A, B) = Z_A Z_B \langle AA|BB \rangle (e^{-\alpha_A R_{AB}} + e^{-\alpha_B R_{AB}}) \quad (112)$$

where a_A and a_B are parameters and R_{AB} is the internuclear distance.

In AM1⁴³ (Austin Model 1), PM3⁴⁴ (Modified Neglect of Diatomic Overlap, Parametric Method Number 3) and MNDO- d (MNDO with d orbitals modifications to the core-core term), these modifications are the same as that for MNDO with the addition of an extra term to reduce the excessive core-core repulsions just outside bonding distances. The additional term may be considered as a van der Waal's attraction term. The AM1 and PM3 core-core terms are:

$$E_N(A, B) = E_N^{\text{MNDO}}(A, B) + \frac{Z_A Z_B}{R_{AB}} \left(\sum_k a_{kA} e^{-b_{kA}(R_{AB}-c_{kA})^2} + \sum_k a_{kB} e^{-b_{kB}(R_{AB}-c_{kB})^2} \right) \quad (113)$$

The extra terms define spherical Gaussian functions, the a , b , and c are adjustable parameters. PM3 has two Gaussians per atom, while AM1 has between two and four.

The MNDO, AM1, and MNDO- d one-center two-electron integrals are derived from experimental data on isolated atoms. Most were taken from Oleari's⁷⁸ work, but a few were obtained by optimization to fit molecular properties. The values of PM3 one-center two-electron integrals were optimized to reproduce experimental molecular

properties. The recently generated PM5⁴⁵, a reparametrization of PM3, appears to considerably reduce the error of many calculated properties. The PM5 method proves to be up to four times more accurate for heats of formation than current methods, such as AM1, PM3, and MNDO. For MINDO/3 and PM5 atomic and diatomic parameters exist, while MNDO, AM1, PM3 and MNDO-d use only single-atom parameters.

4.1.2 Model system for re-optimizing semiempirical parameters

In Rauhut's work,⁷¹⁻⁷⁴ double proton transfer reactions of model base pair systems pyrazole-guanidine, as shown in Figure 16, have been studied by computational methods at the MP2/(aug)-cc-pVDZ level, with various sets of substituents ("aug" in parentheses indicates that augmented functions were used only on some atoms.). They showed an unusual plateau of almost constant energy in the region of the (formal) transition state. The plateau occurs because the system is at the transition from a concerted to a stepwise mechanism. Additionally, the plateau is not limited to just one dimension but rather extends at least in two dimensions.

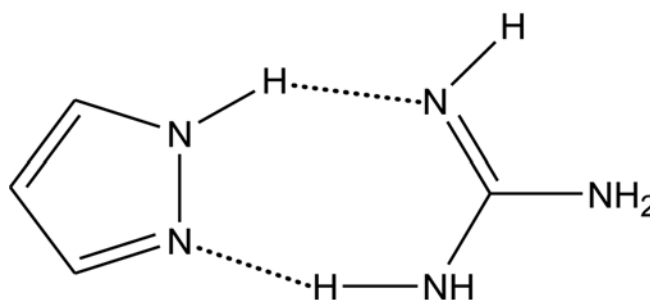


Figure 16 Pyrazole-Guanidine model system.

The reaction profile of the prototype system in one dimension is given in Figure 17. The intrinsic reaction coordinate (IRC) has been traced by the Gonzalez-Schlegel algorithm⁷⁹⁻⁸⁰ with a step width of 0.05 amu^{1/2}·Bohr. If not otherwise noted, the reaction coordinates used in this study refer to mass-weighted coordinates, and the *ab-*

initio data are taken from Rauhut and Schweiger. The one-dimensional energy profiles have been obtained with relaxation of all other degrees of freedom.

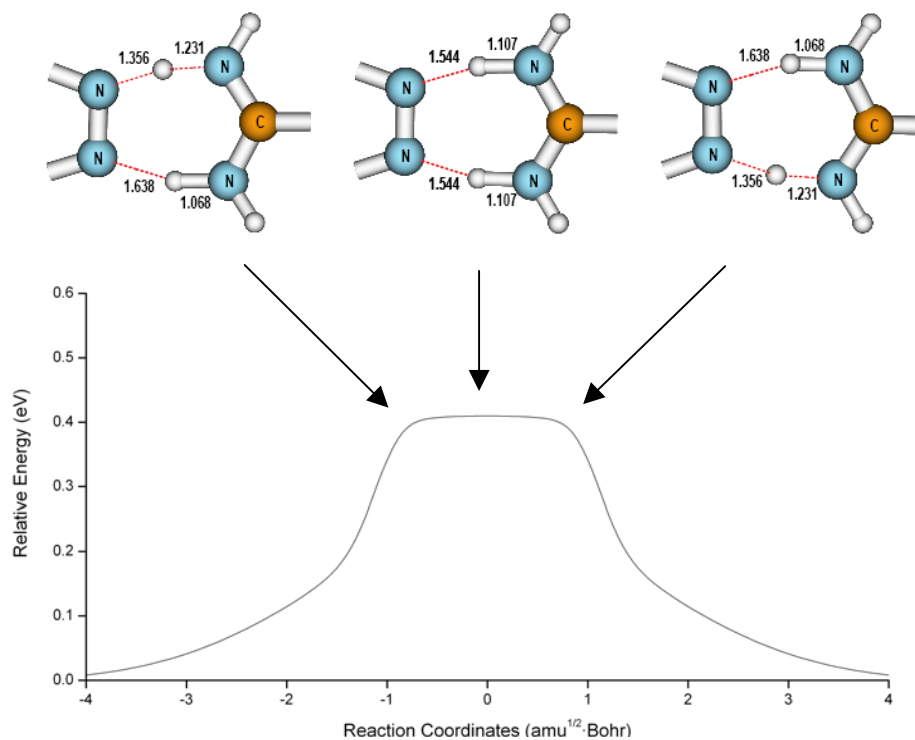


Figure 17 Energy profile along the intrinsic reaction coordinate of the prototypical plateau reaction at the MP2 level.

Two-dimensional potential energy surfaces spanned by the two N-H distances between the pyrazole unit and the two protons are shown in Figure 18 and 19, with the other degrees of freedom relaxed and kept rigid, respectively. Contour lines are separated by about 0.08 and 0.12 eV for relaxed and rigid potential energy surfaces, respectively. Here, in all contour potential energy surfaces, the energies are presented in eV and distances are presented in Å unit.

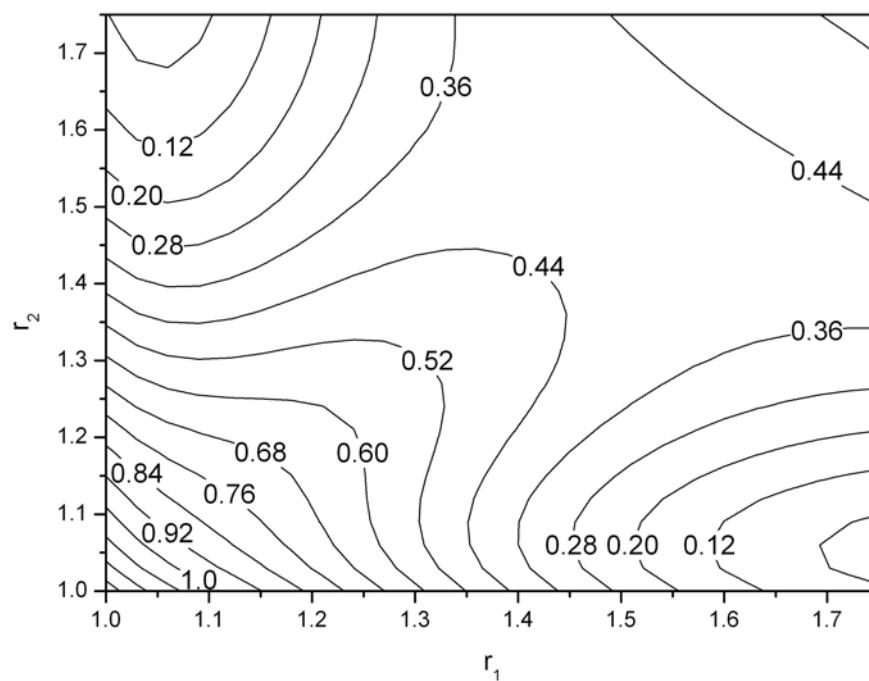


Figure 18 Relaxed MP2/(aug)-cc-pVDZ potential energy surface of the double proton transfer reaction between pyrazole and guanidine.

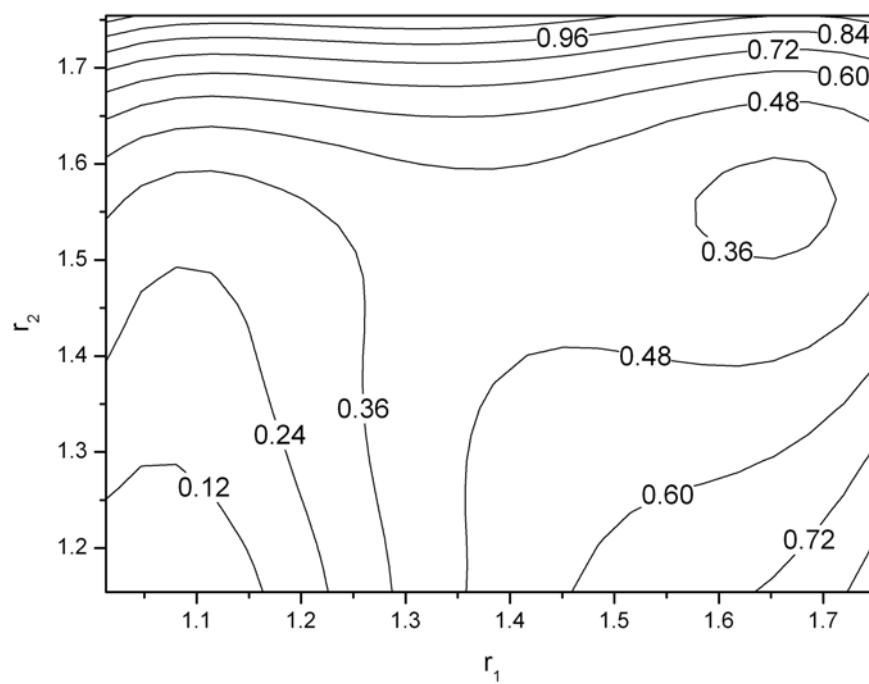


Figure 19 Rigid MP2/(aug)-cc-pVDZ potential energy surface of the double proton transfer reaction between pyrazole and guanidine.

4.2 OPTIMIZATION METHODS APPLICATION

The optimization of semiempirical parameters took as a starting point the standard semiempirical AM1, PM3 and PM5 methods as implemented in MOPAC.⁸¹ The semiempirical parameters are optimized using MOPAC and the error (ΔE) is minimized by searching for minima of the weighted sum of square error³

$$\Delta E(P) = \left[\sum_i \left(\frac{V_{s,i}(P) - V_{t,i}}{V_{t,i}} \right)^2 W_i \right] \left[\sum_i W_i \right]^{-1} \quad (115)$$

where $V_{t,i}$ are the *ab-initio* energy data to be reproduced (target values), $V_{s,i}$ are their semiempirically computed counterparts, which depend on the parameter set P , and W_i are weights.

All geometries and all target values were taken from Rauhut et al.^{71,74} In this work, only some small sets of geometries were chosen for reparametrization and these parameters can then be used for the whole set of geometries. There are 160 reference geometries for the one-dimensional IRC spanned by N-H distances with a step width of $0.05 \text{ amu}^{1/2} \cdot \text{Bohr}$. Changing the N-H distances between the pyrazole unit and the two protons from 1.00 to 1.75 Å with a step length of 0.03 Å for the relaxed two-dimensional case leads to 676 different relaxed reference geometries. For the rigid case spanned by two N-H distances with a step length of 0.02 Å from 1.01 to 1.75 Å, 1169 geometries are available.

In some of the calculations, only a part of the parameter set was reparametrized, with the exclusion of the Gaussian functions, in order to reduce the size of the search space. However, the whole parameter set was also reparametrized to compare the results from two different sizes of parameter sets. The list of parameters used in the semiempirical methods AM1, PM3 and PM5 and their descriptions are given in Table 30.

Table 30 Parameters used in semiempirical methods and their descriptions.

Parameter	Description
U_{ss}	<i>s</i> atomic orbital one-electron one-center integrals
U_{pp}	<i>p</i> atomic orbital one-electron one-center integrals
β_s	<i>s</i> atomic orbital one-electron two-center resonance integral terms
β_p	<i>p</i> atomic orbital one-electron two-center resonance integral terms
ξ_s	<i>s</i> -type Slater atomic orbital exponent
ξ_p	<i>p</i> -type Slater atomic orbital exponent
α_p	Atom P core-core repulsion term
GSS	<i>s-s</i> atomic orbital one center two electron repulsion integral
GPP	<i>p-p</i> atomic orbital one center two electron repulsion integral
GSP	<i>s-p</i> atomic orbital one center two electron repulsion integral
GP2	<i>p-p'</i> atomic orbital one center two electron repulsion integral
HSP	<i>s-p</i> atomic orbital one-center two-electron exchange integral
FN1n, K_{nA} or a_{nA} ; m=1 to 4	A Gaussian multiplier for <i>n</i> th Gaussian of atom A
FN2n, L_{nA} or b_{nA} ; m=1 to 4	A Gaussian exponent multiplier Gaussian of atom A
FN3n, M_{nA} or c_{nA} ; m=1 to 4	A Radius of center of <i>n</i> th Gaussian of atom A

The minimization was performed by means of the simplex method, combined with simulated annealing (SA),⁸ which is a general optimizing program combined with the MOPAC package by Persico et al.,³ as well as with the hybrid GA/Powell and the Iffco methods which were already used with success in the first part of this work. The hybrid simplex/SA method is employed as described by Press et al. (for more details, see Ref.8). It is possible to terminate when the vector distance

moved in a step is fractionally smaller in magnitude than some tolerance tol ; here a value of 10^{-4} was used. Alternatively, we require that the decrease in the function value in the terminating step be fractionally smaller than some tolerance; for this a value of 10^{-6} was used.

The optimized GA/Powell hybrid and Iffco methods are attached to MOPAC as user-defined subroutine. For the hybrid GA/Powell method, the best point from a GA run of 500 generations is used as starting point for a local Powell optimization, and the budget for Powell is limited by using a rather modest relative error value of 10^{-4} as convergence threshold. The Iffco method was run with a termination tolerance of 10^{-6} .

4.3 RESULTS AND DISCUSSION

In this work, the system-specifically reparametrized semiempirical methods AM1, PM3 and PM5 were tested for their ability to correctly reproduce one- and two-dimensional potential energy surfaces (PES) for a double proton transfer reaction in a base pair model system consisting of one pyrazole entity and a guanidine molecule, comparing the energy profiles with *ab-initio* data at the MP2/(aug)-cc-pVDZ level from Rauhut's work.^{71,74} In order to make the usual semiempirical energy output (standard enthalpies of formation) comparable to the *ab-initio* output (absolute electronic energies), the parameter sets of the AM1, PM3 and PM5 methods were reparametrized setting relative energies between the educt/product asymptote and the origin of the reaction coordinate as the a target values.

First we consider the DPTR along the one-dimensional reaction path. We started with comparing two different sizes of the parameter set for H, N and C atoms, reparametrizing all parameters of H, N and C atoms and reparametrizing all parameters of N and C with fixed H parameters. The parameter sets of the AM1 method were tested, reparametrizing by the simplex algorithm. These results are shown in Figure 20.

We found that the energy profiles obtained from these two different sizes of the parameter set do not lead to significant loss in qualitative agreement with the energy profile found at the MP2/(aug)-cc-pVDZ level, even with the H parameters fixed. Therefore, for our first round of tests, we kept the parameters of the H atom fixed.

In this work, in some of the calculations, a part of the parameter set of the N and C atoms (24 parameters), with exclusion of the Gaussian function, were reparametrized, comparing the performance to other cases in which all parameters of the C and N atoms (45, 36 and 30 parameters of AM1, PM3 and PM5, respectively) were reparametrized.

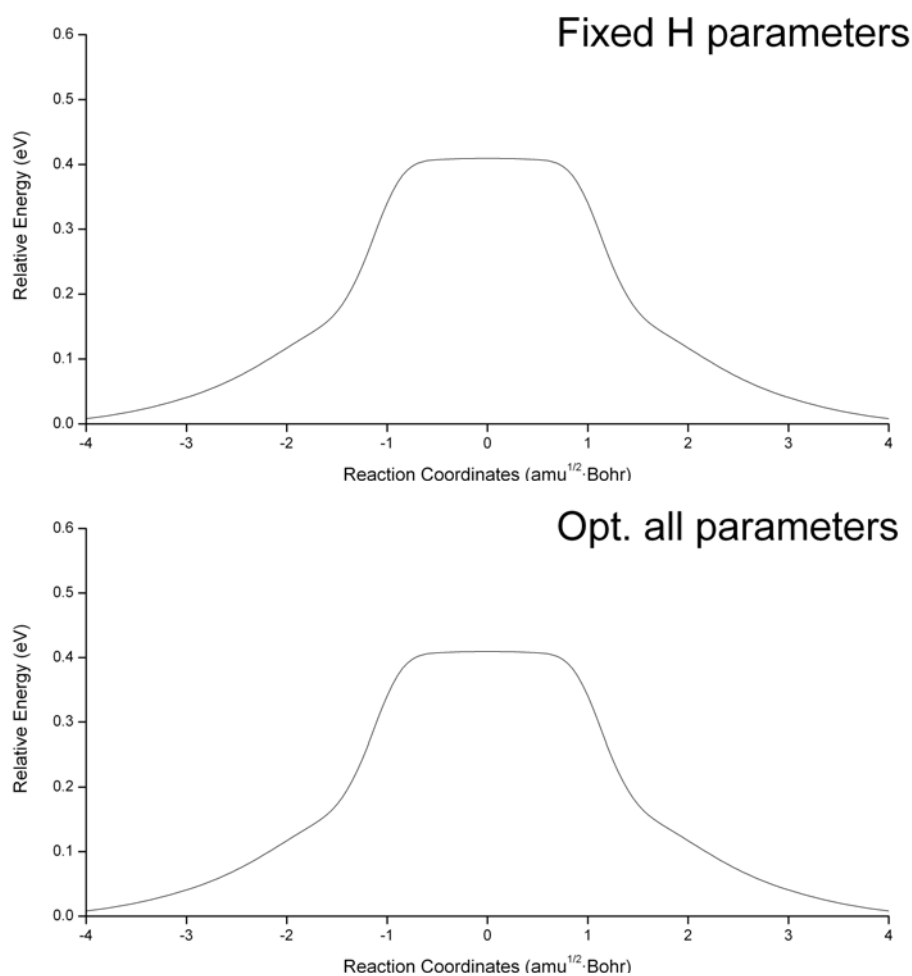


Figure 20 One-dimensional energy profiles of the DPTR, comparing two different sizes of reparametrized parameter set.

4.3.1 Reparametrization geometries on the one-dimensional reaction path

The reparametrization was done on 21 one-dimensional geometries and the resulting parameters then were used to recompute energies for all 160 one-dimensional geometries. These geometries were selected as such that they are evenly distributed on the reaction path. We started with reparametrization on a part of the parameter set of the AM1, PM3 and PM5 methods, using simplex, hybrid GA/Powell and Iffco algorithms. This leads to the energy profiles depicted in Figures 21, 22 and 23, respectively.

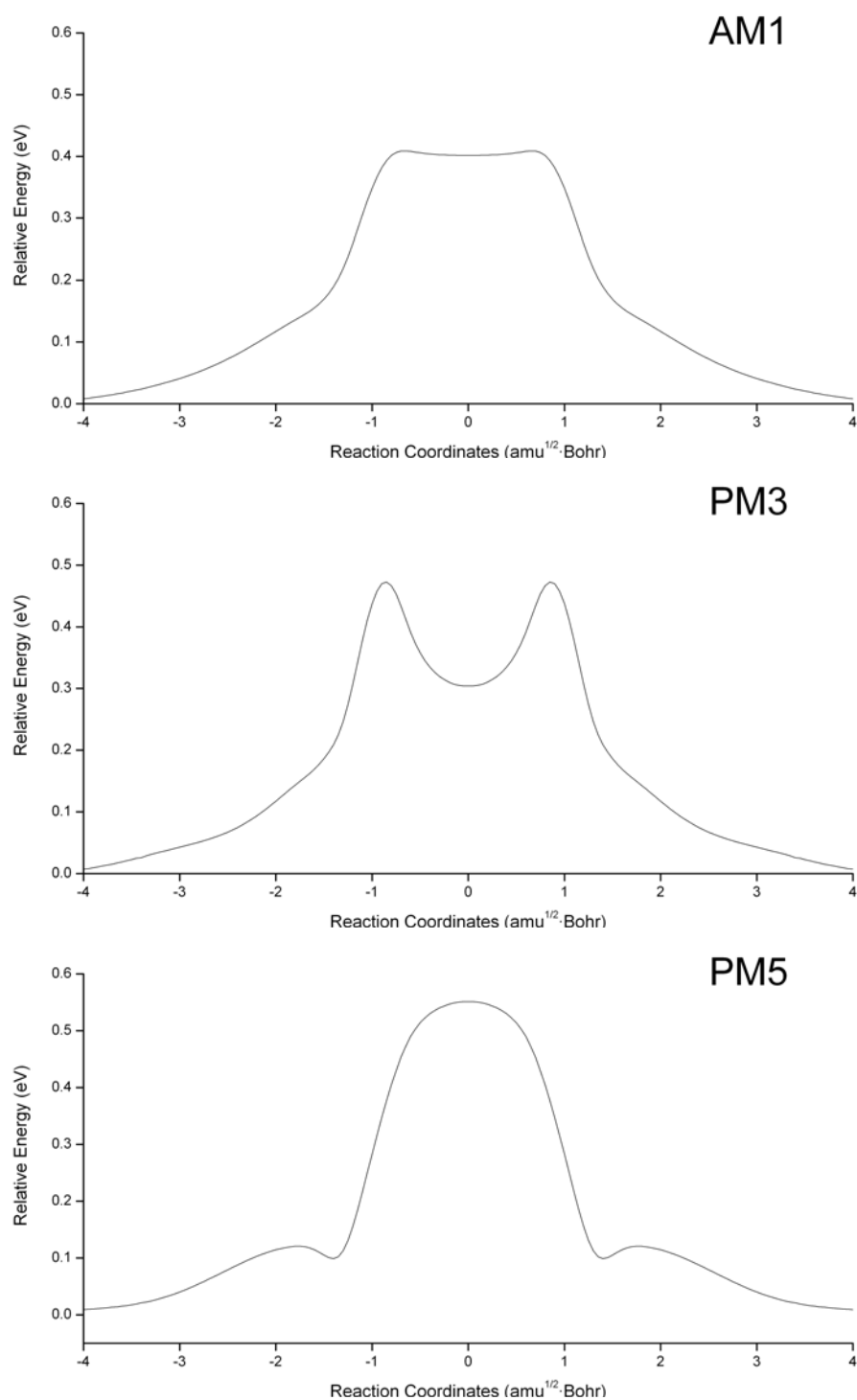


Figure 21 One-dimensional energy profiles of the DPTR, obtained from the AM1, PM3 and PM5 methods in which a part of the parameter set was reparametrized on a subset of these geometries using the simplex algorithm.

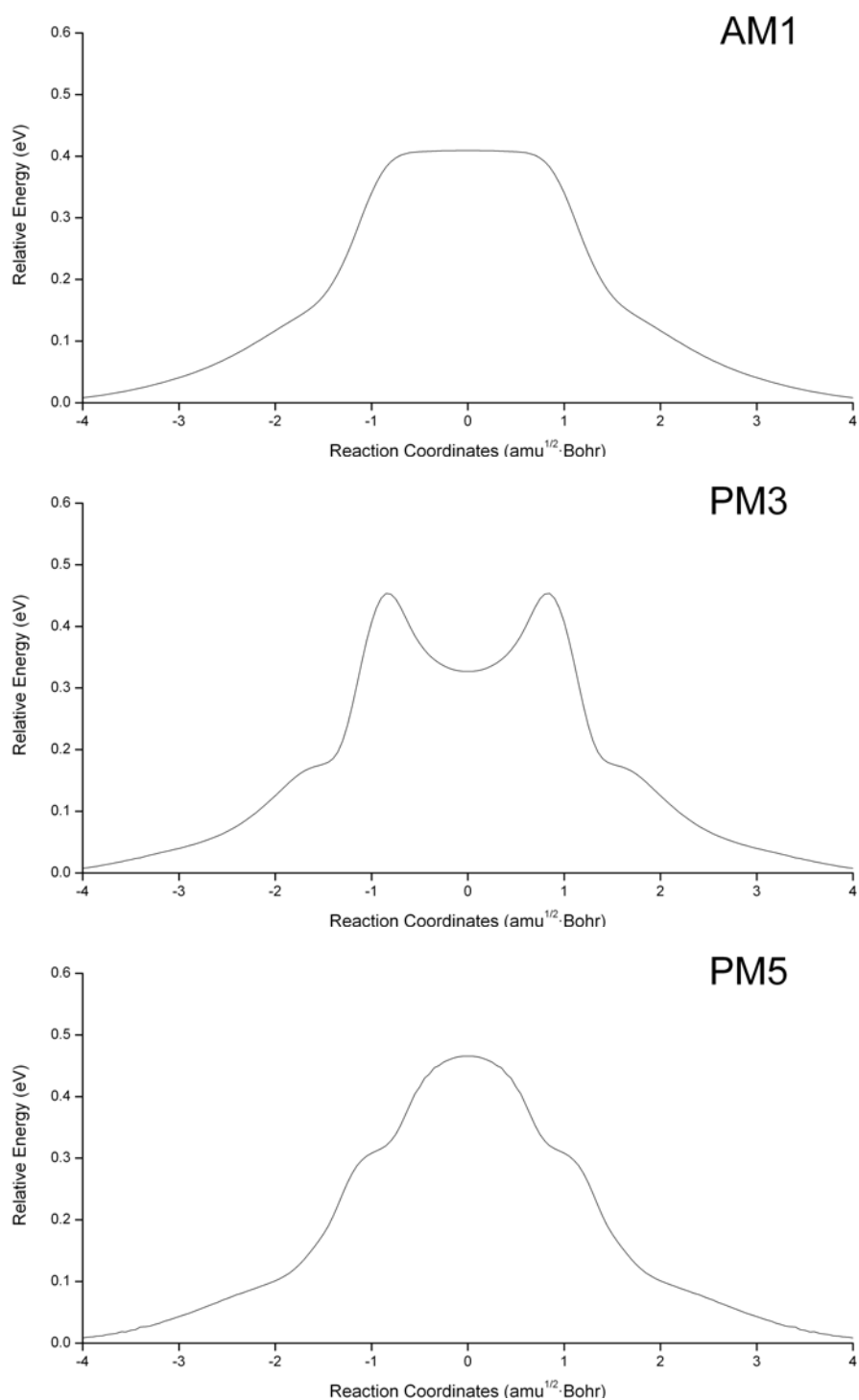


Figure 22 One-dimensional energy profiles of the DPTR, obtained from the AM1, PM3 and PM5 methods in which a part of the parameter set was reparametrized on a subset of these geometries using the hybrid GA/Powell algorithm.

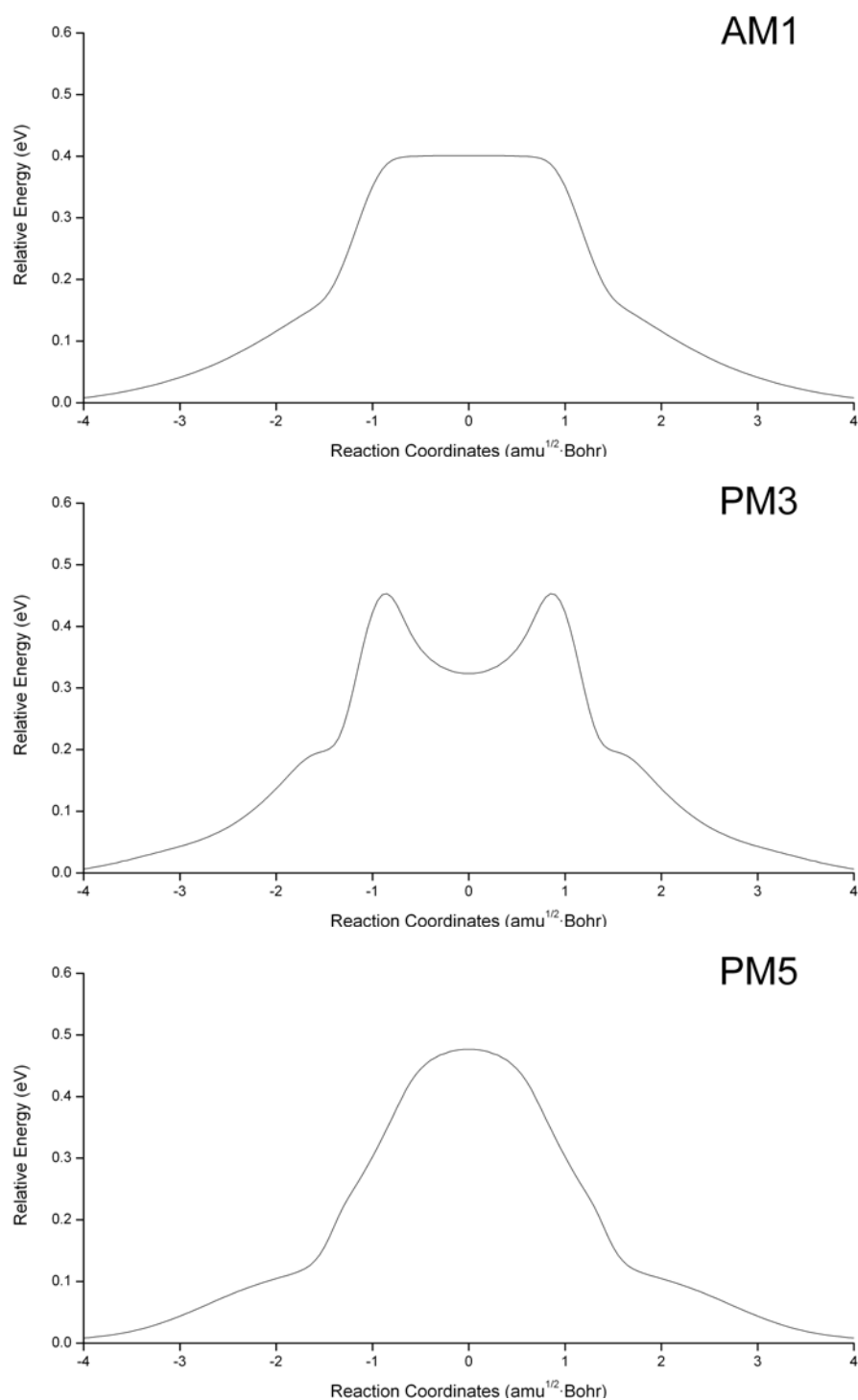


Figure 23 One-dimensional energy profiles of the DPTR, obtained from the AM1, PM3 and PM5 methods in which a part of the parameter set was reparametrized on a subset of these geometries using the Iffco algorithm.

Comparison of these data for the same semiempirical method but for different optimization algorithms reveals a high degree of similarity. Comparison within any of the optimization algorithms and across semiempirical methods, however, leads to a totally different picture: We found that only the energy profiles predicted from the reparametrized AM1 method yield a plateau characteristic. In fact they even yield a quantitatively accurate match to the energy profile found at the MP2/(aug)-cc-pVDZ level (Figure 17). On the other hand, the energy profiles found from the reparametrized PM3 and PM5 methods exhibit a qualitatively wrong behavior: The plateau disappears and deforms into a non flat shape with very different characteristics at the transition state.

For comparison, also the whole parameter sets of the AM1, PM3 and PM5 methods were reparametrized on the same 21 one-dimensional geometries, using again the simplex, hybrid GA/Powell and Iffco methods. These PESs are presented in the Figures 24, 25 and 26.

These Figures show that in all cases the energy profiles for which all parameters were reparametrized are closer to the target than the energy profiles found from the cases where only part of them was reparametrized. For the AM1 case, the improvements are small but quantitatively important, compare e.g. the slight “dip” on the plateau in Figure 21 with the truly flat shape in Figure 24. We now also found acceptable PESs from the PM3 method reparametrized by the simplex and GA/Powell algorithms. Reparametrization of the PM5 method with the simplex and Iffco algorithms still leads to wrong characteristics of the energy profile. But even in this apparently difficult case, our GA/Powell algorithm can find good parameter sets that produce at least qualitatively acceptable energy profiles.

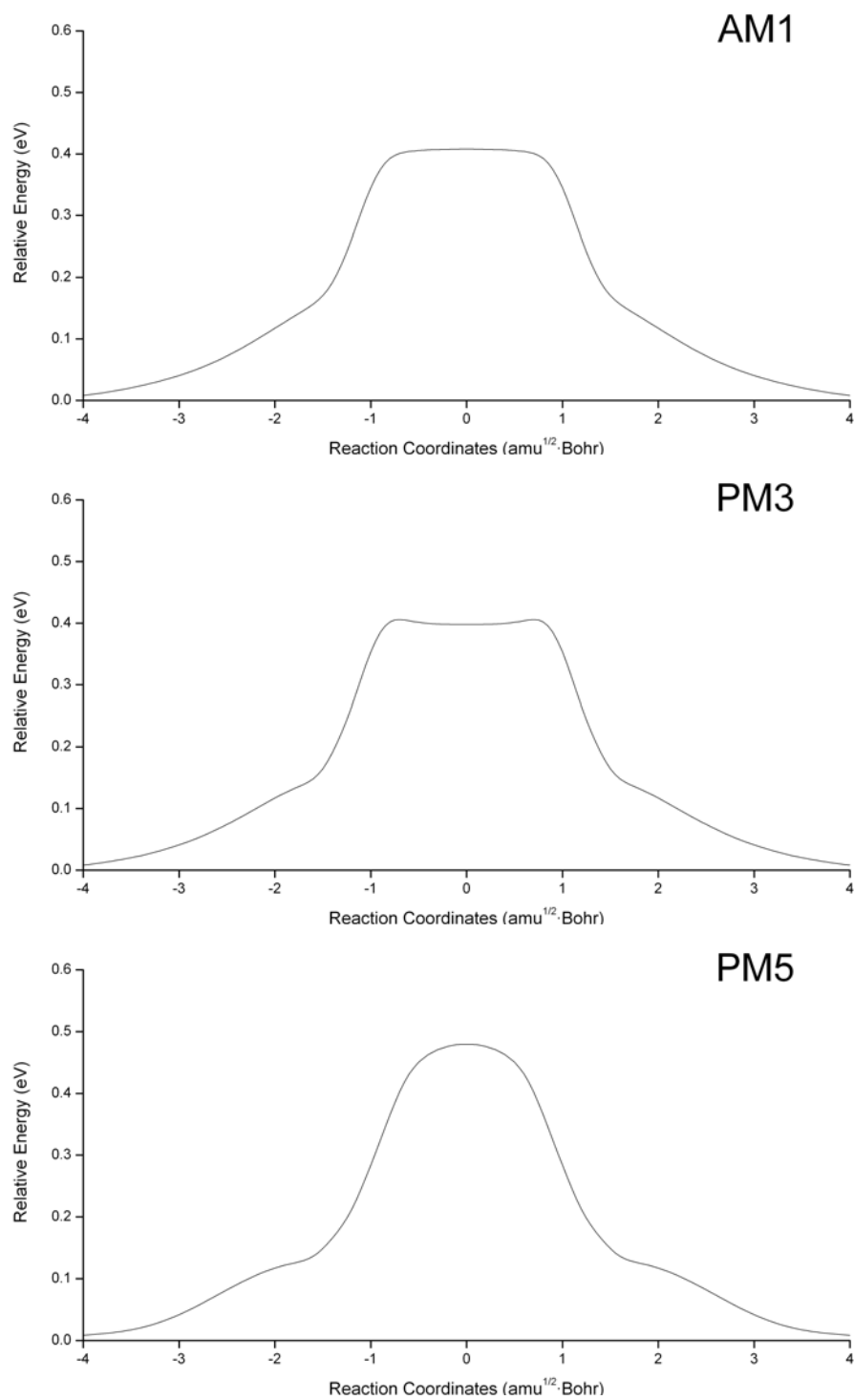


Figure 24 One-dimensional energy profiles of the DPTR, obtained from the AM1, PM3 and PM5 methods in which all parameters were reparametrized on a subset of these geometries using the simplex algorithm.

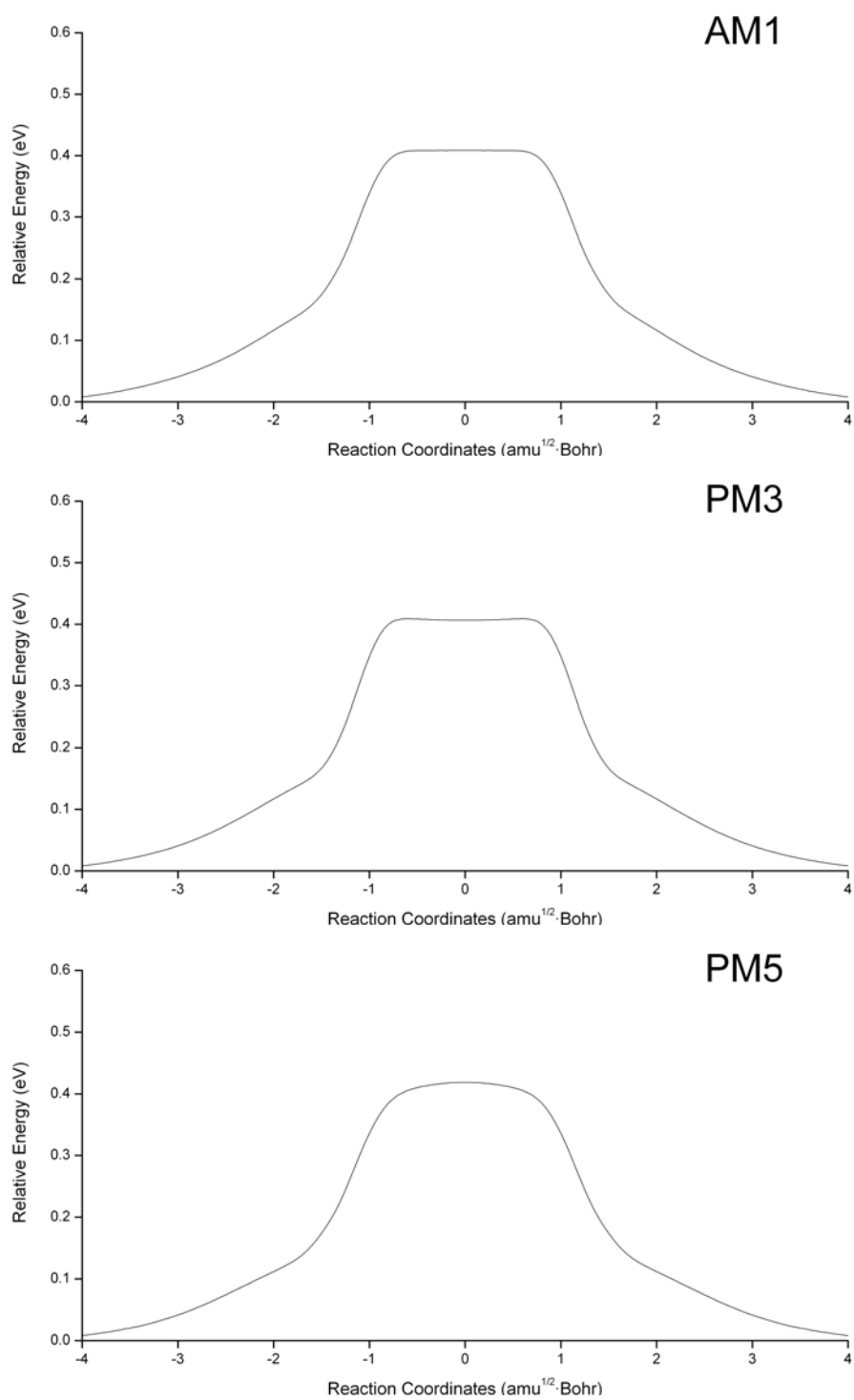


Figure 25 One-dimensional energy profiles of the DPTR, obtained from the AM1, PM3 and PM5 methods in which all parameters were reparametrized on a subset of these geometries using the hybrid GA/Powell algorithm.

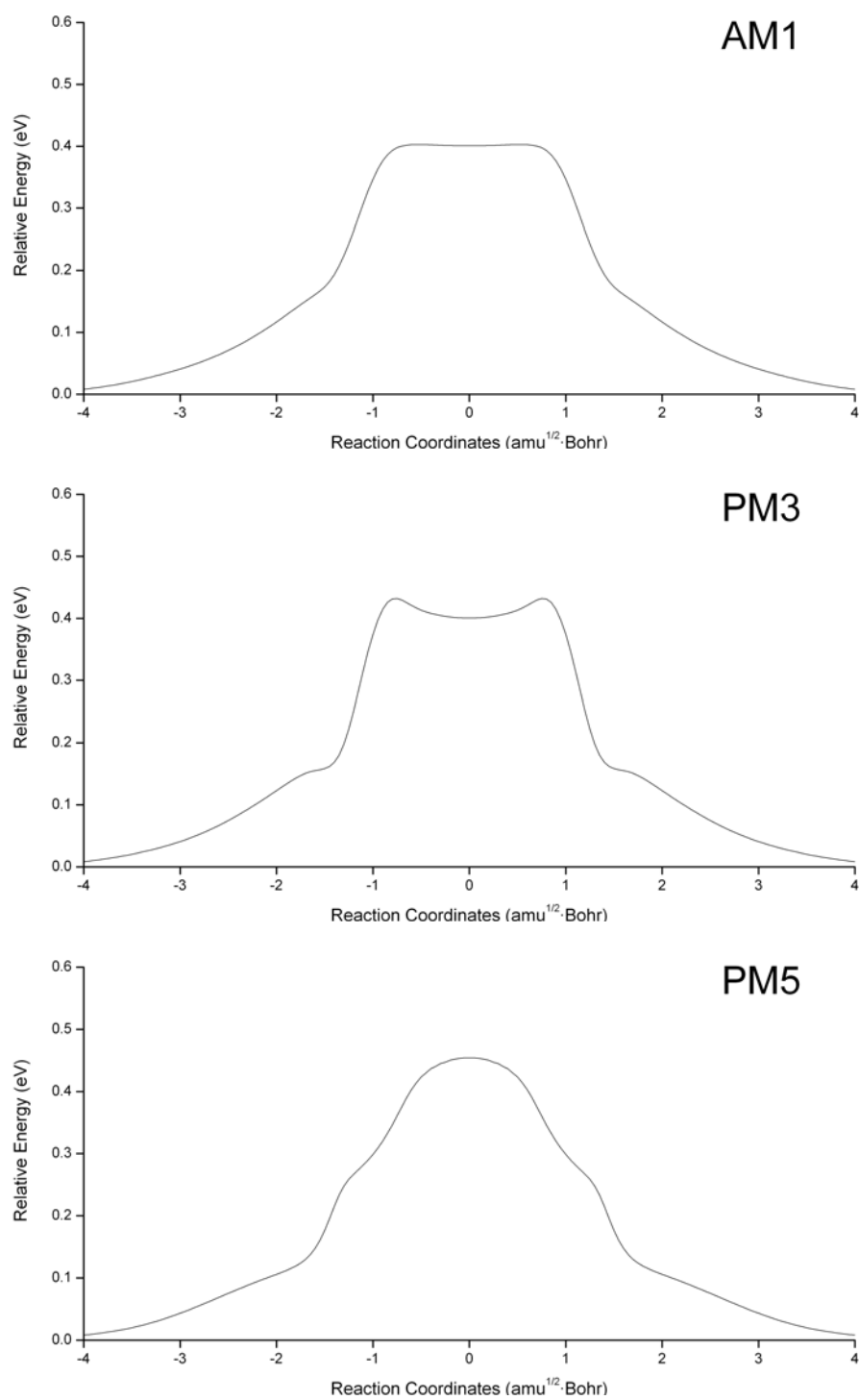


Figure 26 One-dimensional energy profiles of the DPTR, obtained from the AM1, PM3 and PM5 methods in which all parameters were reparametrized on a subset of these geometries using the Iffco algorithm.

4.3.2 Reparametrization on two-dimensional relaxed geometries

The reparametrization was done on 144 two-dimensional relaxed geometries and these parameters then were used to recalculate energies for all 676 geometries. These geometries were selected as such that they are evenly distributed on the reaction path. As in the one-dimensional case, at first only part of the parameter sets of the AM1, PM3 and PM5 methods was reparametrized using the simplex, hybrid GA/Powell and Iffco algorithms. These were then used to predict the energy profiles shown in Figures 27, 28 and 29, respectively.

Comparing to the energy profile found at the MP2/(aug)-cc-pVDZ level for the relaxed geometries (Figure 18), we observe that only the reparametrized AM1 method can produce quantitatively accurate energy profiles, independent of the choice of optimization algorithm. The reparametrized PM3 and PM5 methods show not only slightly incorrect relative energies but also wrong qualitative features of the energy profiles. Nevertheless, the reparametrized PM5 method obtained from the hybrid GA/Powell algorithm shows only slight deviations.

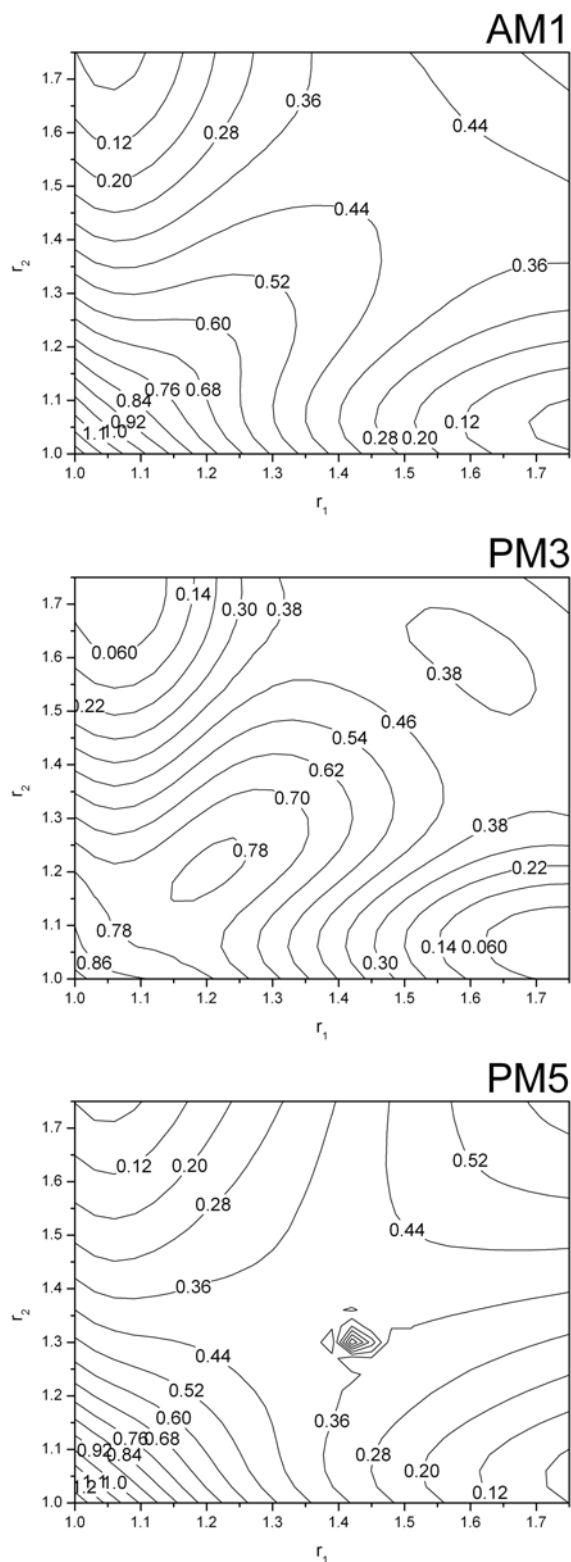


Figure 27 Relaxed 2D energy profiles of the DPTR, obtained from the AM1, PM3 and PM5 methods for which a part of the parameter set was reparametrized on a subset of these geometries using the simplex algorithm.

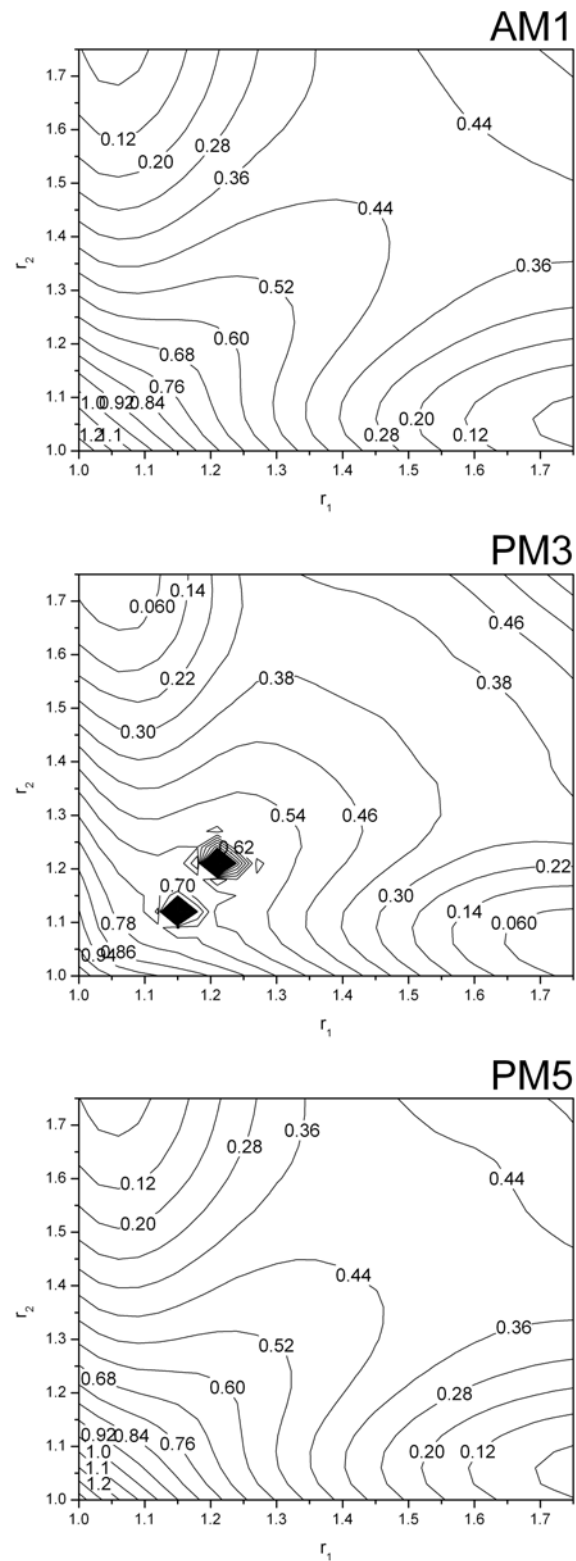


Figure 28 Relaxed 2D energy profiles of the DPTR, obtained from the AM1, PM3 and PM5 methods for which a part of the parameter set was reparametrized on a subset of these geometries using the GA/Powell algorithm.

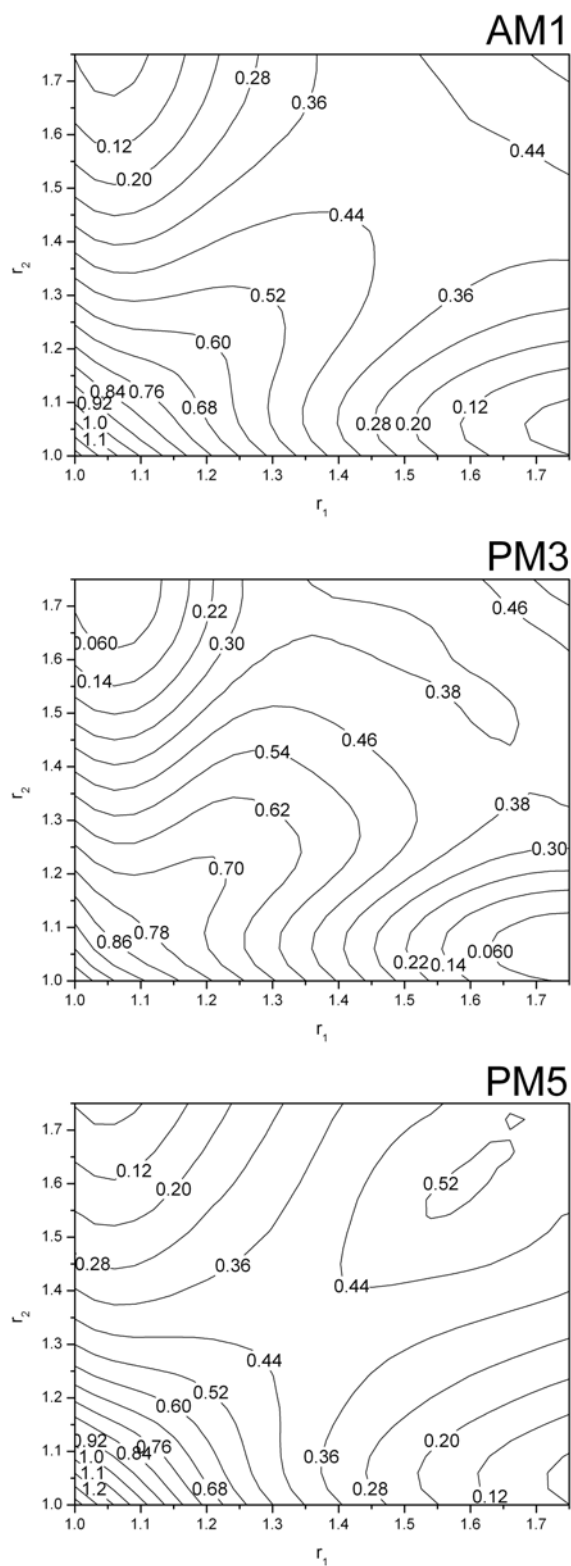


Figure 29 Relaxed 2D energy profiles of the DPTR, obtained from the AM1, PM3 and PM5 methods for which a part of the parameter set was reparametrized on a subset of these geometries using the Iffco algorithm.

Figures 30, 31 and 32 show energy profiles produced from the AM1, PM3 and PM5 methods for which all parameters were reparametrized using the simplex, hybrid GA/Powell and Iffco algorithms, respectively.

As observed in the one-dimensional case, the energy profiles predicted from optimization of all parameters show correct relative energies and better features for the PM3 method, in addition to the quantitatively accurate energy profiles predicted from the reparametrized AM1 method. Unfortunately, the energy profiles found from the reparametrized PM5 method, obtained from the simplex and Iffco algorithms, still exhibit wrong features.

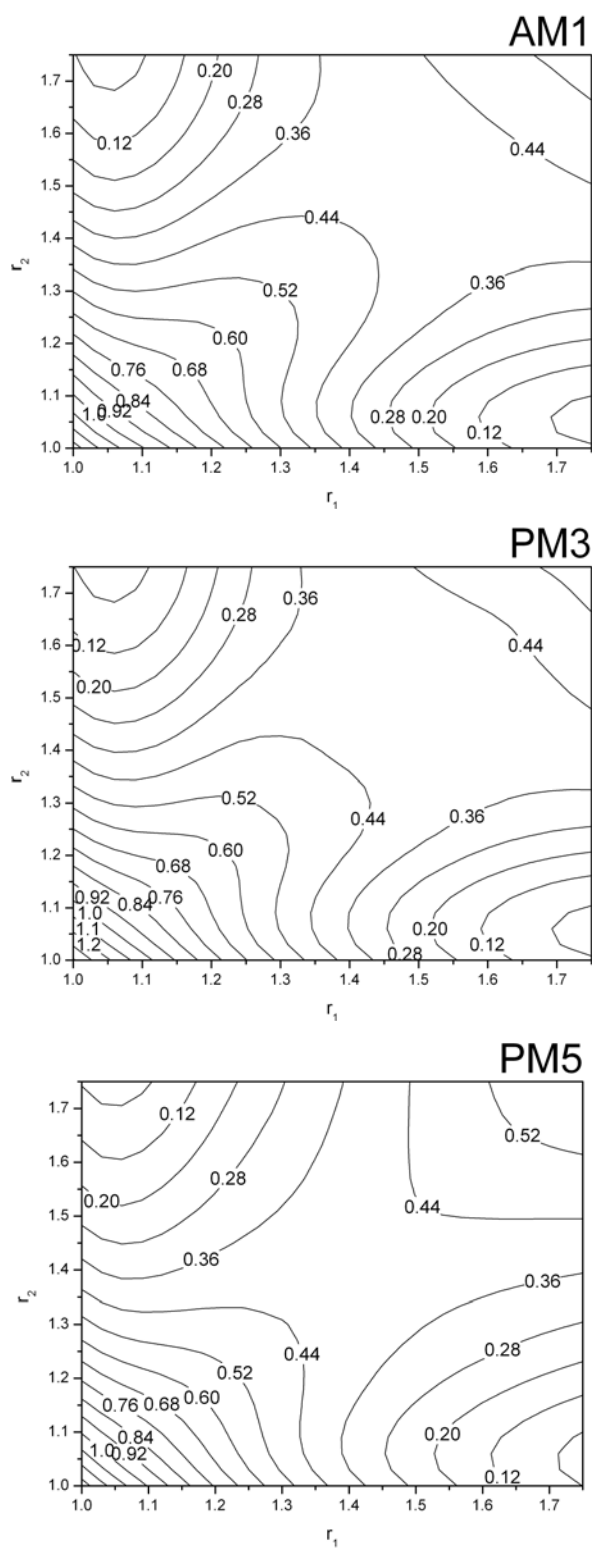


Figure 30 Relaxed 2D energy profiles of the DPTR, obtained from the AM1, PM3 and PM5 methods for which all parameters were reparametrized on a subset of these geometries using the simplex algorithm.

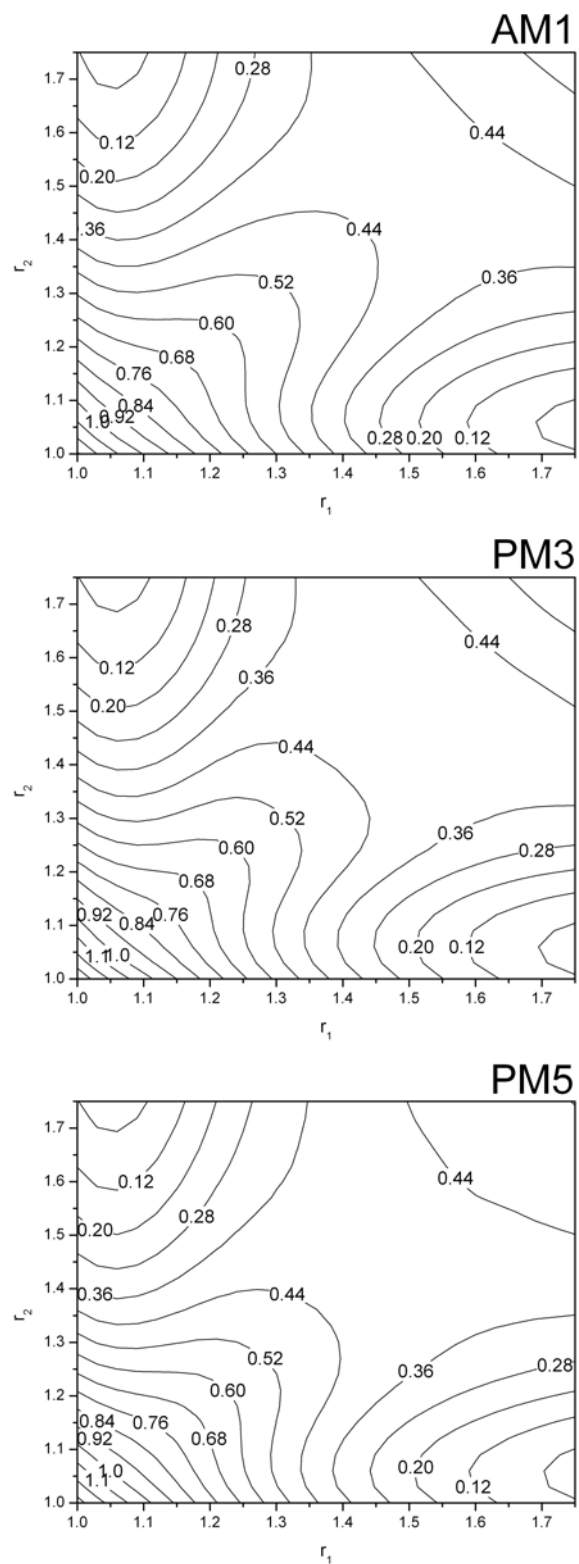


Figure 31 Relaxed 2D energy profiles of the DPTR, obtained from the AM1, PM3 and PM5 methods for which all parameters were reparametrized on a subset of these geometries using the GA/Powell algorithm.

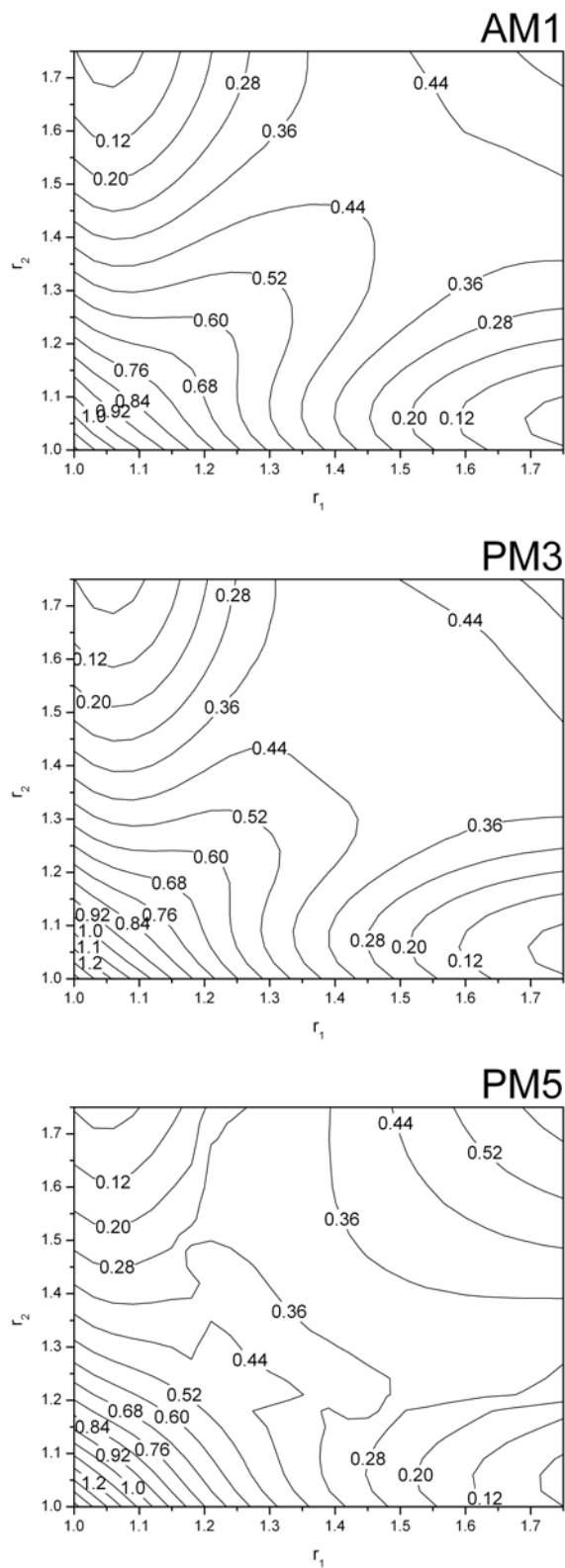


Figure 32 Relaxed 2D energy profiles of the DPTR, obtained from the AM1, PM3 and PM5 methods for which all parameters were reparametrized on a subset of these geometries using the Iffco algorithm.

4.3.3 Reparametrization on two-dimensional rigid geometries

The reparametrization was done on 136 two-dimensional rigid geometries and these parameters then were used to recalculate energies for all 1169 geometries. These geometries were selected as such that they are evenly distributed on the reaction path. Mirroring our investigations of relaxed two-dimensional geometries, we again checked the performance of AM1, PM3 and PM5 after reparametrization of part of the parameter sets (Figures 33-35) and of the whole parameter sets (Figures 36-38). This time, the deviations from the *ab-initio* target are generally smaller, but similar trends prevail: Reparametrized AM1 leads to quantitative agreement independent of the optimization method. With PM3 and PM5 we have more problem, but performance improves both with inclusion of all parameters into the optimization and with using our global GA/Powell method.

Therefore, we can conclude that the parametrized AM1 method, using three different algorithms, can generate one- and two-dimensional potential energy surfaces as good as the energy profiles found at the MP2/(aug)-cc-pVDZ level from Rauhut's work,^{71,74} even if the Gaussian functions were excluded from the reparametrization. For PM3, all parameters need to be reparametrized to produce acceptable energy profiles. However, the reparametrization of all parameters in the PM5 method using the simplex and Iffco methods still generates slightly wrong features of the energy profiles, while the reparametrized PM5 method obtained from our hybrid GA/Powell algorithm can produce acceptable energy profiles in all cases. Thus, it can be concluded that the GA/Powell algorithm is the best algorithm for finding good parameter sets for all semiempirical methods, comparing with the others methods tested here.

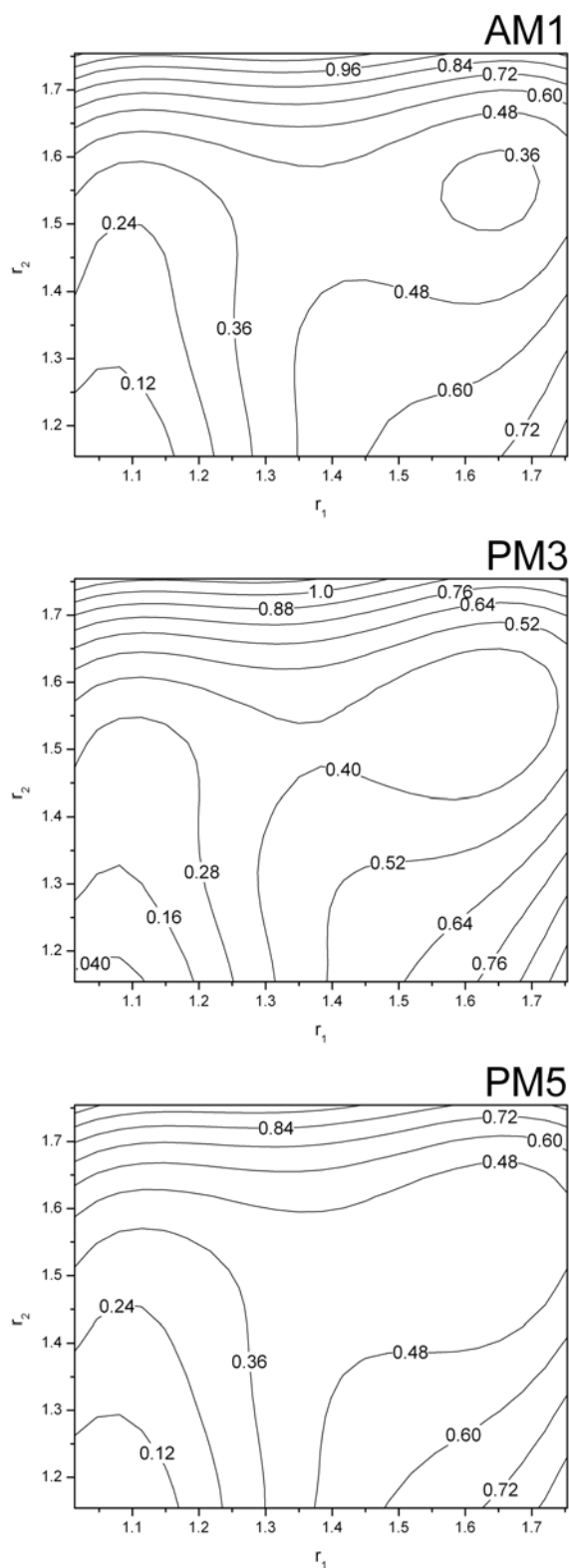


Figure 33 Rigid 2D energy profiles of the DPTR, obtained from the AM1, PM3 and PM5 methods for which a part of the parameter set was reparametrized on a subset of these geometries, using the simplex algorithm.

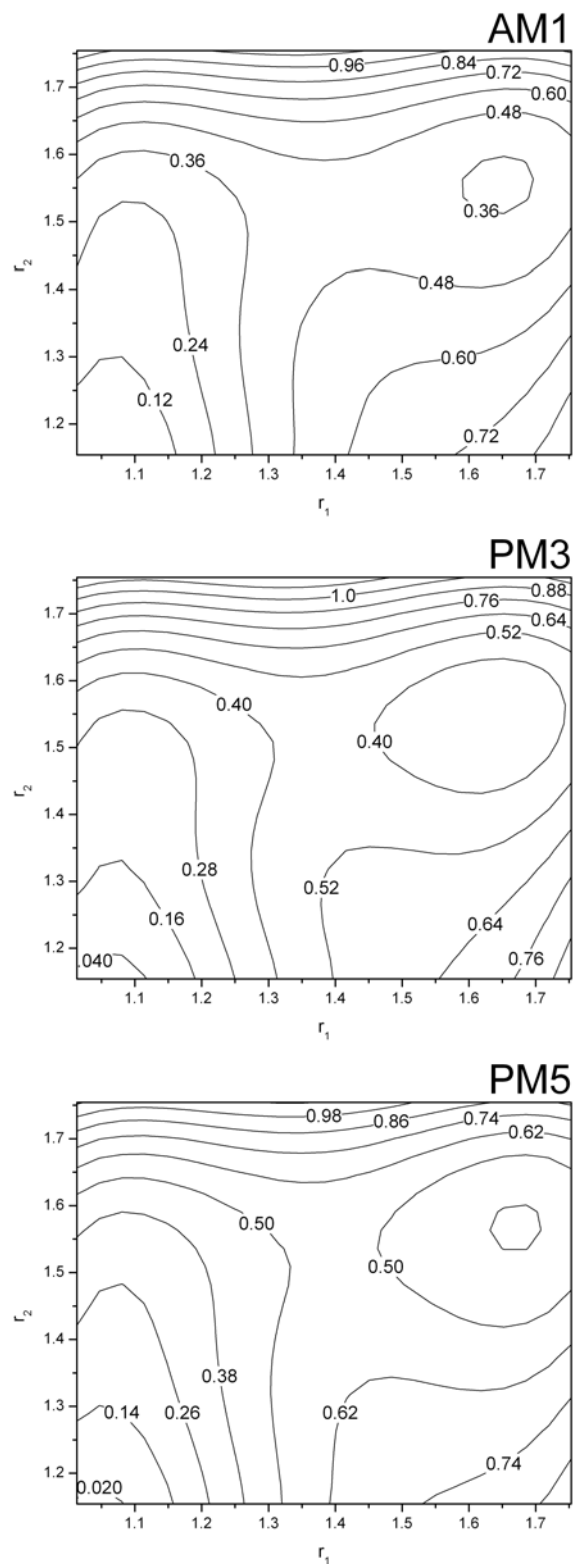


Figure 34 Rigid 2D energy profiles of the DPTR, obtained from the AM1, PM3 and PM5 methods for which a part of the parameter set was reparametrized on a subset of these geometries, using the GA/Powell algorithm.

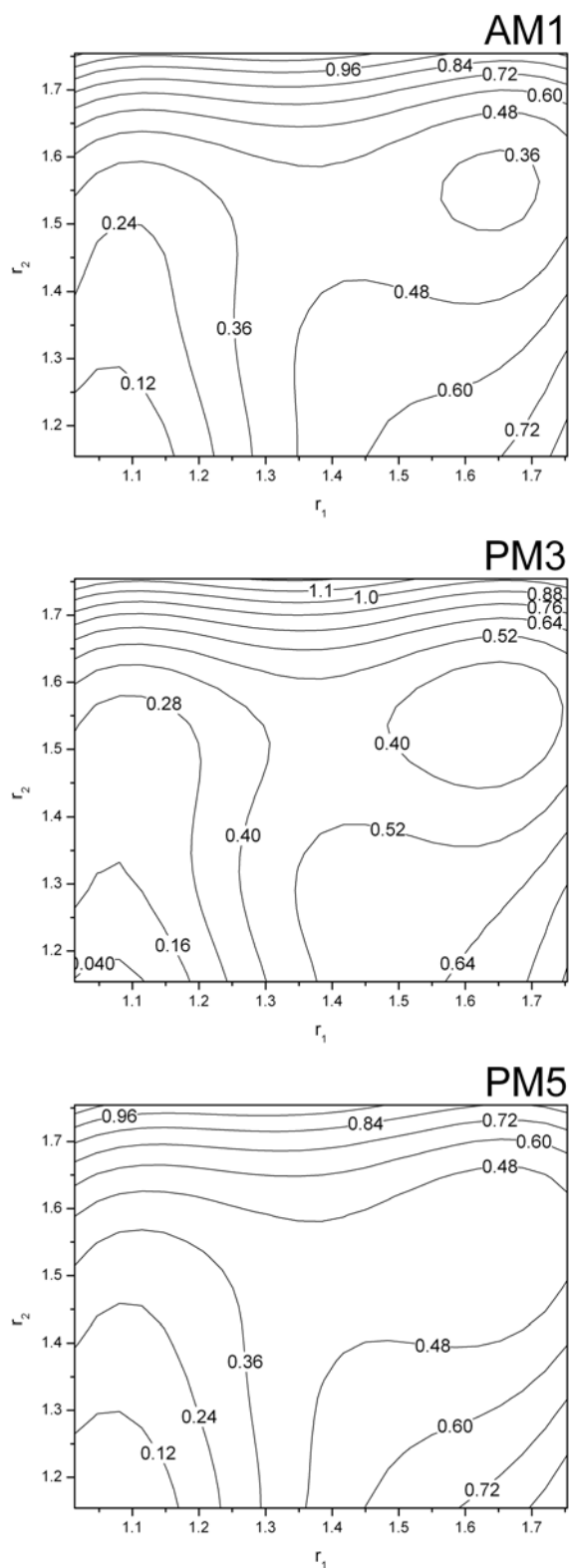


Figure 35 Rigid 2D energy profiles of the DPTR, obtained from the AM1, PM3 and PM5 methods for which a part of the parameter set was reparametrized on a subset of these geometries, using the Iffco algorithm.

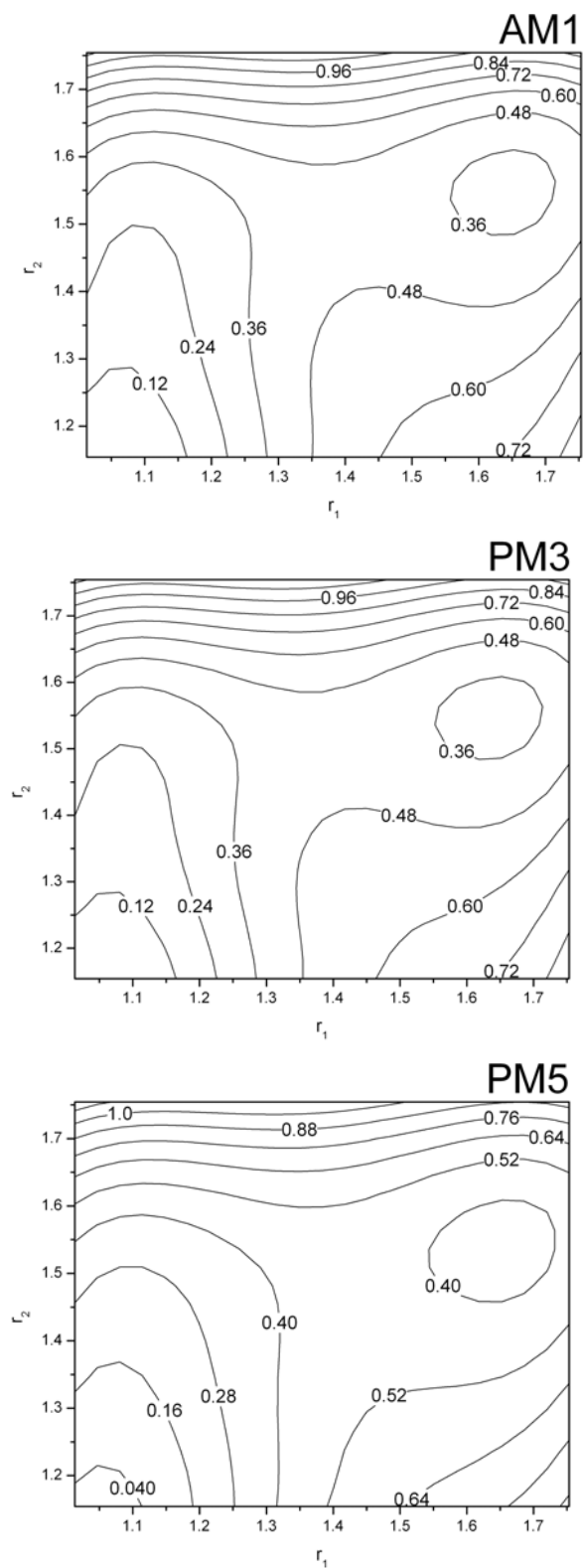


Figure 36 Rigid 2D energy profiles of the DPTR, obtained from the AM1, PM3 and PM5 methods for which all parameters were reparametrized on a subset of these geometries, using the simplex algorithm.

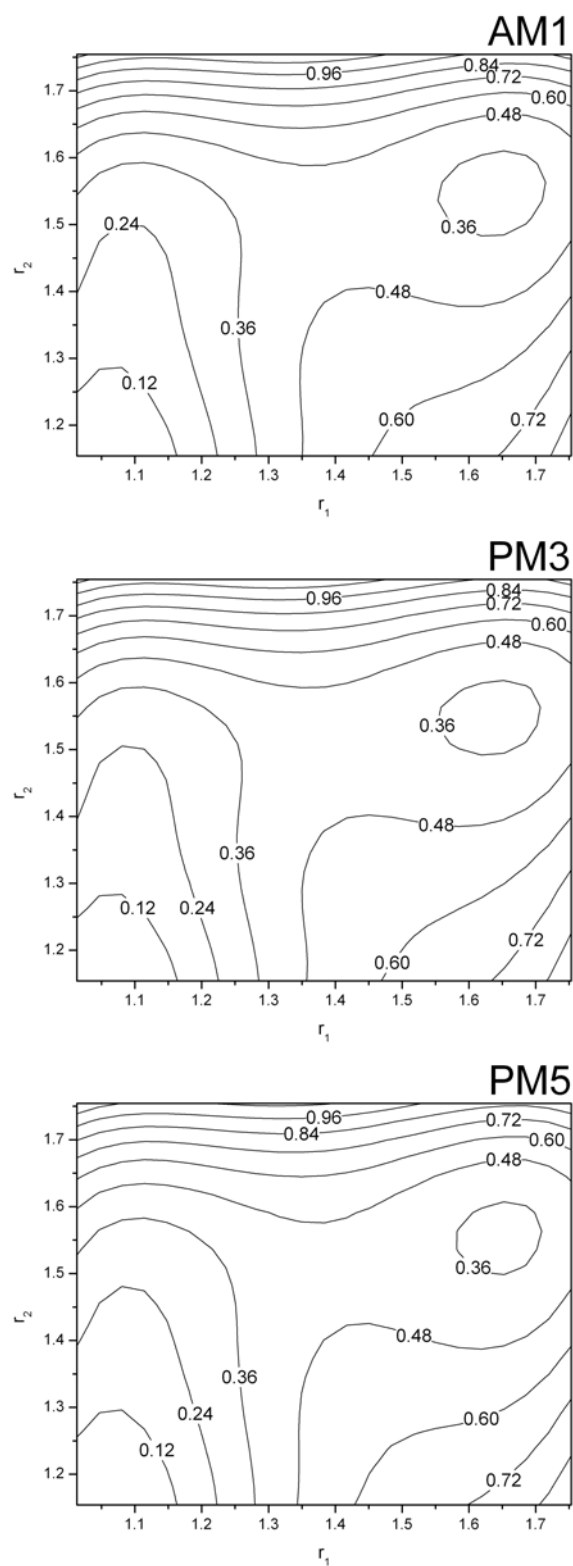


Figure 37 Rigid 2D energy profiles of the DPTR, obtained from the AM1, PM3 and PM5 methods for which all parameters were reparametrized on a subset of these geometries, using the GA/Powell algorithm.

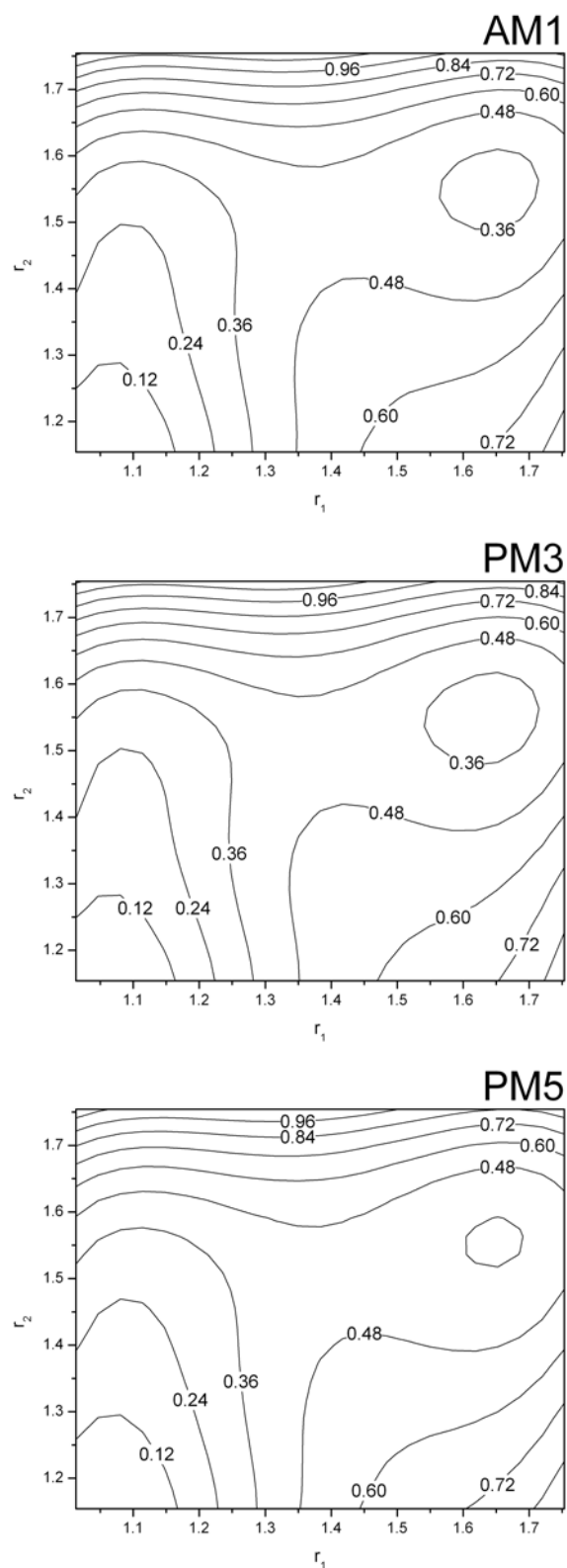


Figure 38 Rigid 2D energy profiles of the DPTR, obtained from the AM1, PM3 and PM5 methods for which all parameters were reparametrized on a subset of these geometries, using the Iffco algorithm.

An important test for the realistic applicability of such a reparametrization scheme is to extend the area of comparison between semiempirical and ab-initio data far beyond the region of the fitted reference points, even into other degrees of freedom. Due to limited time, we could only perform a few first checks of this type. The AM1 method, reparametrized with our GA/Powell algorithm for a small set of points along the 1D reaction path, was used to recalculate energies for all 2D relaxed and rigid geometries. The results in Figure 39 show largely correct relative energies but a slightly incorrect qualitative shape of the surfaces. Of course, a degradation of the fit quality is to be expected. In our opinion, it is encouraging that the degradation is not larger than this.

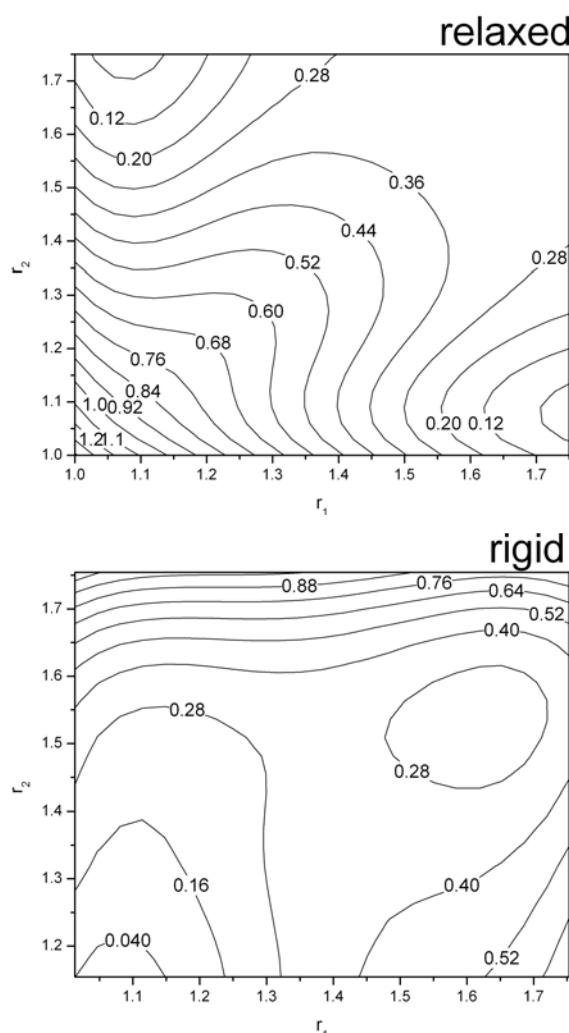


Figure 39 2D energy profiles of the DPTR, obtained from the AM1 method for which all parameters were reparametrized on a subset of 1D geometries, using the GA/Powell algorithm.

Similarly, the AM1 method, reparametrized with our GA/Powell algorithm for a small set of 2D relaxed geometries, was used to recalculate energies for the 1D reaction path and for the 2D rigid geometries. Figure 40 shows that in this case the results are quantitatively accurate. This is to be expected, since this test is less stringent than the one before.

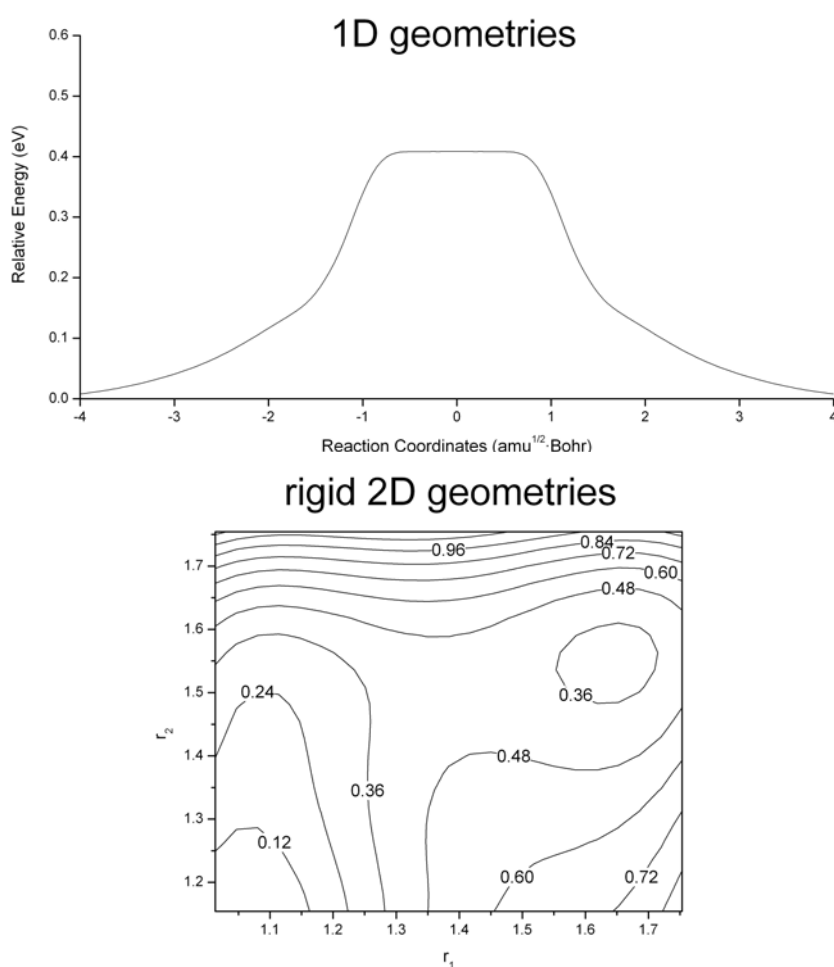


Figure 40 1D and rigid 2D energy profiles of the DPTR, obtained from the AM1 method for which all parameters were reparametrized on a subset of 2D relaxed geometries, using the GA/Powell algorithm.

4.4 CONCLUSIONS AND OUTLOOK

Semiempirical calculations are relatively inexpensive and provide reasonable qualitative descriptions of molecular systems and accurate quantitative predictions of energies and structures for systems where good parameter sets exist. Therefore, system-specifically reparametrized semiempirical methods are the promising tool to calculate high-accuracy potential energy surface quickly. Clearly, the practical applicability of this strategy hinges upon the ability to do the parameter reoptimization quickly, reliably and successfully. In this work, several algorithms, simplex, hybrid GA/Powell and Iffco, were used to reoptimize parameter sets of semiempirical method for a specific system. We found that for the optimization of only a part of the parameter sets of the AM1, PM3 and PM5 methods, only AM1 leads to quantitatively accurate potential energy surfaces in all cases, compared to the MP2/(aug)-cc-pVDZ results from Rauhut's work.^{71,74} If all parameters were optimized, we found acceptable results for all the details of the PES for AM1 and PM3. For PM5, only our GA/Powell algorithms could generate acceptable energy profiles. Comparing to the first part of this work, optimization of auxiliary basis sets, this suggests that here the optimization problem has even less local character, and hence it can only be solved reliably with a method that incorporates truly global convergence.

The specific system studies have was a double-proton transfer reaction of a model base pair pyrazole-guanidine. This clearly is a challenging system for any standard semiempirical method. Therefore, it is gratifying to see that after reparametrized with our global approach, reparametrized semiempirical calculation can actually compete in accuracy with the MP2/(aug)-cc-pVDZ results. This suggests that standard prejudices on the limitations of semiempirical methods can be obliterated by system-specific global reparametrization. Clearly, however, the present findings are only the first steps of work in progress, and many areas for further studies are left open. One of these issues is the choice of reference data, another one is how far away from the reference the fit carries. The fundamental idea in reparametrization is that through the use of a limited set of reference points. It is possible to create a set

of parameters such that all points on the potential energy surface can be accurately represented using those parameters. In the present work, the evaluation of the reparametrized semiempirical methods was done largely within the area spanned by the reference points, that is, the semiempirical method was used mainly in an interpolatory manner. Clearly, for large systems with many degrees of freedom, one can not afford to put reference points into all these degrees of freedom. Therefore, the next step is to check how a reparametrized semiempirical method fares in an extrapolatory usage, first within a given set of degree of freedom but then also across degrees of freedom into “uncharted territory”. First tests in this direction indicate successes but also failures: After reparametrization with reference points only along one-dimensional reaction path, the two-dimensional cuts shown in this work can be reproduced at least qualitatively for the case of AM1 method for which all parameters were reparametrized using the GA/Powell algorithm. On the other hand, for the system under study here, there is also a large amount of *ab-initio* frequency data available. After fitting of the parameters only to the energy data (as done here), these frequencies are not reproduce with sufficient accuracy by the semiempirical methods. Therefore, we are currently working towards including also gradient and frequency data into the reparametrization. Another important issue is to check for systematic breakdowns of semiempirical approaches, and to test in how far these can be avoided by adding extra empirical terms, the parameters of which can easily be included into the system-specific parameter optimization.

Nevertheless, the degree of quantitative agreement in one-dimensional and two-dimensional energy data achieved in this work already is much better than what many people expect from semiempirical methods. In addition, we have shown that such results can be achieved reliable, easily and quickly with a proper global optimization algorithm. Since the reparametrized semiempirical methods loose nothing of their great speed advantage, this opens up the possibility to use this approach e.g. in direct classical and quantum dynamics studies, with the potential energy surface calculated on the fly, at the cost of semiempirical methods but with an accuracy as good as *ab-initio* calculations. Work along these lines is in progress in our lab.

5. BIBLIOGRAPHY

1. G. Galli, *Curr. Opin. Solid State Mater. Sci.* 1 (1996) 864.
2. S. Goedecker, *Rev. Mod. Phys.* 71 (1999) 1085.
3. C. Ciminelli, G. Granucci and M. Persico, *Chem. Eur. J.* 10 (2004) 2327.
4. J. H. Holland, *Adaptation in natural and artificial system*, Ann Arbor, The University of Michigan Press, 1975.
5. D. E. Goldberg, *Genetic Algorithms in search, optimization, and machine learning*, Addison Wesley, 1989.
6. D. E. Goldberg, *The Design of Innovation – Lessons from and for Competent Genetic Algorithms*, Kluwer, Dordrecht, 2002.
7. S. K. Gregurick, M. H. Alexander and B. Hartke, *J. Chem. Phys.* 104 (1996) 2684.
8. W. H. Press, B. P. Flannery, S. A. Teukolsky and W. T. Vetterling, *Numerical Recipes in Fortran*, Cambridge University Press, Cambridge, U.K., 1989.
9. M. J. D. Powell, *Comput. J.* 7 (1964) 155.
10. R. Hooke and T. A. Jeeves, *J. Assoc. Comput. Mach.* 8 (1961) 212.
11. D. R. Jones, C. D. Perttunen and B. E. Stuckman, *J. Optimz. Theory App.* 79 (1993) 157.
12. J. M. Gablonsky, *An implementation of the DIRECT algorithm*, Technical Report CRSC-TR98-29, Center for Research in Scientific Computation, North Carolina State University, August 1998.
13. J. M. Gablonsky, *Direct version 2.0 userguide*, Technical Report, CRSC-TR01-08, Center for Research in Scientific Computation, North Carolina State University, April 2001.
14. R. G. Carter, J. M. Gablonsky, H. A. Patrick, C. T. Kelley and O. J. Eslinger, *Optimization and Engineering.* 2 (2002) 157.
15. P. A. Gilmore and C. T. Kelley, *SIAM J. Opt.* 4 (1995) 269.
16. T. D. Choi, O. J. Eslinger, P. A. Gilmore, C. T. Kelley and H. A. Patrick, *User's Guide to IFFCO*. Technical Report CRSC-TR01-17, Center for Research in Scientific Computation, North Carolina State University, July 2001.
17. F. Weigend, *Phys. Chem. Chem. Phys.* 4 (2002) 4285.

18. O. Vratras, J. E. Almloef and M. W. Feyereisen, *Chem. Phys. Lett.* 213 (1993) 514.
19. M. W. Feyereisen, G. Fitzgerald and A. Komornicki, *Chem. Phys. Lett.* 208 (1993) 359.
20. F. Weigend and M. Häser, *Theor. Chem. Acc.* 97 (1997) 331.
21. F. Weigend, M. Häser, H. Patzelt and R. Ahlrichs, *Chem. Phys. Lett.* 294 (1998) 143.
22. E. J. Baerends, D. E. Ellis and P. Ros, *Chem. Phys.* 2 (1973) 41.
23. B. I. Dunlap, *Phys. Chem. Chem. Phys.* 2 (2000) 2113.
24. F. R. Manby and P. J. Knowles, *Phys. Rev. Lett.* 87 (2001) 163001.
25. F. R. Manby, P. J. Knowles and A. W. Lloyd, *J. Chem. Phys.* 115 (2001) 9144.
26. R. Polly, H. -J. Werner, F. Manby and P. J. Knowles, to be published.
27. R. Polly, H. -J. Werner, F. Manby and P. J. Knowles, *Mol. Phys.* 102 (2004) 2311.
28. J. W. Mintmire and B. I. Dunlap, *Phys. Rev. A* 25 (1982) 88.
29. H. -J. Werner, F. R. Manby and P. J. Knowles, *J. Chem. Phys.* 118 (2003) 8149.
30. T. H. Dunning Jr, *J. Chem. Phys.* 90 (1989) 1007.
31. A. K. Wilson, T. van Mourik and T. H. Dunning, *J. Mol. Struct. (THEOCHEM)* 388 (1996) 339.
32. F. Weigend, A. Köhn and C. Hättig, *J. Chem. Phys.* 116 (2002) 3175.
33. R. Horst, P. M. Pardalos and N. V. Thoai, *Introduction to Global Optimization*, Kluwer, Dordrecht, 2000.
34. B. Hartke, *Angew. Chem. Int. Ed.* 41 (2002) 1468.
35. V. E. Bazterra, M. B. Ferraro and J. C. Facelli, *J. Chem. Phys.* 116 (2002) 5984.
36. F. Koskowski and B. Hartke, *J. Comput. Chem.*, in press (2005).
37. F. Jensen, *Introduction to Computational Chemistry*, Wiley, Chichester, 1999.
38. I. N. Levine, *Quantum Chemistry*, 5th ed., Prentice Hall, Upper saddle River, 2000.
39. P. W. Atkins and R. S. Friedman, *Molecular Quantum Mechanics*, 3rd ed., Oxford University Press, New York, 1996.
40. H. -J. Werner, P. J. Knowles, M. Schütz, R. Lindh, P. Celani, T. Korona, G. Rauhut, R. D. Amos, A. Bernhardsson, A. Berning, D. L. Cooper, M. J. O.

- Deegan, A. J. Dobbyn, F. Eckert, C. Hampel, G. Hetzer, A. W. Lloyd, S. J. McNicholas, F. R. Manby, W. Meyer, M. E. Mura, A. Nicklaß, P. Palmieri, R. Pitzer, U. Schumann, H. Stoll, A. J. Stone, R. Tarroni and T. Thorsteinsson, Molpro, a package of *ab initio* programs designed by H. -J. Werner and P. J. Knowles, version 2002.4.
41. M. J. S. Dewar and W. Thiel, *J. Am. Chem. Soc.* 99 (1977) 4899.
 42. R. C. Bingham, M. J. S. Dewar and D. H. Lo, *J. Am. Chem. Soc.* 97 (1975) 1285.
 43. M. J. S. Dewar, E. G. Zoebisch, E. F. Healy and J. J. P. Stewart, *J. Am. Chem. Soc.* 107 (1985) 3902.
 44. J. J. P. Stewart, *J. Comp. Chem.* 10 (1989) 209.
 45. MOPAC2002: <http://www.cachesoftware.com/mopac/index.shtml>.
 46. I. Rossi and D. G. Truhlar, *Chem. Phys. Lett.* 233 (1995) 231.
 47. G. Kovacevic, T. Hrenar and N. Doslic, *Chem. Phys.* 293 (2003) 41.
 48. P. L. Fast, N. E. Schultz and D. G. Truhlar, *J. Phys. Chem. A* 105 (2001) 4143.
 49. S. Sekusak, M. G. Cory, R. J. Barlett and A. Sabljic, *J. Phys. Chem. A* 103 (1999) 11394.
 50. C. E. Taylor, M. G. Cory, R. J. Bartlett and W. Thiel, *Comput. Mater. Sci.* 27 (1993) 204.
 51. A. A. Voityuk and A. A. Blizniuk, *Theor. Chim. Acta.* 72 (1987) 223.
 52. M. I. Bernal-Uruchurtu and M. F. Ruiz-López, *Chem. Phys. Lett.* 330 (2000) 118.
 53. M. P. Repasky, J. Chandrasekhar and W. L. Jorgensen, *J. Comput. Chem.* 23 (2002) 1601.
 54. B. Hartke, *Chem. Phys. Lett.* 258 (1996) 144.
 55. B. Hartke, *Theor. Chem. Acc.* 99 (1998) 241.
 56. A. Tekin and B. Hartke, *Phys. Chem. Chem. Phys.* 6 (2004) 503.
 57. Y. Ge and J. D. Head, *Inter. J. Quant. Chem.* 95 (2003) 617.
 58. Y. Ge and J. D. Head, *J. Phys. Chem. B* 108 (2004) 6025.
 59. E. F. Caldin and V. Gold, *Proton Transfer Reactions*, Chapman and Hall, London, 1975.
 60. E. F. Caldin, *Chem. Rev.* 69 (1969) 135.
 61. M. Rini, B. -Z. Magnes, E. Pines, and E. T. J. Nibbering, *Science* 301 (2003) 349.

62. B. G. Malmstrom, *Chem. Rev.* 90 (1990) 1247.
63. M. Kratochvil, J. Sponer and P. Hobza, *J. Am. Chem. Soc.* 122 (2000) 3495.
64. J. Smets, L. Adamowicz and G. Maes, *G. J. Phys. Chem.* 100 (1996) 6434.
65. A. K. Chandra, M. T. Nguyen and T. Zeegers-Huyskens, *J. Mol. Struct. (THEOCHEM)* 519 (2000) 1.
66. C. Alhambra, F. J. Luque, F. Gago and M. Orozco, *J. Phys. Chem. B* 101 (1997) 10075.
67. F. Aguilar-Parrilla, G. Scherer, H. H Limbach, M. C. Foces-Foces, F. H Cano, J. A. S. Smith, C. Toiron and J. Elguero, *J. Am. Chem. Soc.* 114 (1992) 9657.
68. J. L. G. de Paz, J. Elguero, M. C. Foces-Foces, A. L. Llamas-Saiz, F. Aguilar-Parrilla, O. Klein and H. H. J Limbach, *Chem. Soc., Perkin Trans. 2* (1997) 101.
69. A. Douhal, V. Guallar, M. Moreno and J. M. Lluch, *Chem. Phys. Lett.* 256 (1996) 370.
70. D. Peeters, G. Leroy and C. Wilante, *J. Mol. Struct. (THEOCHEM)* 416 (1997) 21.
71. S. Schweiger and G. Rauhut, *J. Phys. Chem. A* 107 (2003) 9668.
72. S. Schweiger, B. Hartke and G. Rauhut, *Phys. Chem. Chem. Phys.* 6 (2004) 3341.
73. S. Schweiger, B. Hartke and G. Rauhut, *Phys. Chem. Chem. Phys.* 7 (2005) 49.
74. G. Rauhut and S. Schweiger; in *High Performance Computing in Science and Engineering '04*, E. Krause, W. Jäger, M. Resch (Eds.), Springer, Heidelberg, (2005) 323.
75. G. Buemi, F. Zuccarello and A. Raudino, *J. Mol. Struct. (THEOCHEM)* 164 (1988) 379.
76. J. J. Dannenberg, *J. Mol. Struct. (THEOCHEM)* 401 (1997) 279.
77. A. R. Leach, *Molecular Modelling Principle and Applications*, Longman Press, Harlow, U.K., 1996.
78. L. Oleari, L. DiSipio and G. DeMichells, *Molec. Phys.* 10 (1966) 97.
79. C. Gonzalez and H. B. Schlegel, *J. Chem. Phys.* 90 (1989) 2154.
80. C. Gonzalez and H. B. Schlegel, *J. Phys. Chem.* 94 (1990) 5523.
81. J. J. P. Stewart, *MOPAC 2002*, Fujitsu Limited, Tokyo (Japan), 2001.

CURRICULUM VITAE

PERSONAL INFORMATION

Name	Miss Rungtiva Palangsuntikul
Nationality	Thai
Religion	Buddhism
Birth Date	Dec 27, 1976
Age	28 Years
Status	Single
Address	Olshausenstrasse 66/114 Kiel, 24118 Germany Tel. (0431) 880-1759
E-mail	palang@phc.uni-kiel.de

GRADUATE STUDIES

2002-Present	Ph.D candidate Christian-Albrecht University of Kiel, Germany
1999-2002	Master of Sciences in Chemistry Kasetsart University, Bangkok, Thailand Completed 14 courses in Physical Chemistry. GPA is 3.64
1995-1999	Bachelor of Sciences in Chemistry Kasetsart University, Bangkok, Thailand Completed 145 credits in chemistry and related subjects

CONFERENCES-ORAL-POSTER PRESENTATIONS- WORKSHOPS

19 - 23 Sep, 2004	<i>Poster presentation</i> in the 40th Symposium for Theoretical Chemistry in Suhl, Germany.
29 - 31 May, 2003	<i>Poster presentation</i> in the Bunsentagung 2003, Kiel, Germany.
6 - 8 Feb, 2002	<i>Poster presentation</i> in the proceeding of The International Conference on Bioinformatics 2002 : North - South Networking, Le Royal Meridien, Bangkok, Thailand.
16 - 18 Oct, 2001	<i>Poster presentation</i> in the proceeding of The 27 th Congress on Science and Technology of Thailand, Lee Gardens Plaza Hotel, Had Yai, Songkla, Thailand.

- 19 - 20 Jun, 2001 *Oral presentation* in the proceeding of The Fifth Annual National Symposium, on Computational Science and Engineering (ANSCSE 2001), Bangkok Convention Center(BCC), Central Plaza, Bangkok, Thailand.
- 18 - 20 Oct, 2000 26th Congress on Science and Technology of Thailand, Queen Sirikit National Convention Center, Bangkok, Thailand.
- 18 - 22 Sep, 2000 The First Workshop on Biological Physics 2000 (BP2K), Chulalongkorn University, Bangkok, Thailand.
- 27 - 29 Mar, 2000 The Fourth Annual National Symposium, On Computational Science and Engineering (ANSCSE 2000), Kasetsart University, Bangkok, Thailand.
- 4 Mar, 2000 *Poster presentation* in Post-4th Eurasia Symposium, Chulalongkorn University, Bangkok, Thailand.
- 20 - 22 Oct, 1999 25th Congress on Science and Technology of Thailand, Amarin Lagoon Hotel, Pitsanuloke, Thailand.

PUBLICATIONS

1. R. Palangsuntikul and S. Hannongbua, The Interaction Energy of the Water Molecule with Amino Acids in the Active Site of HIV-1 Reverse Transcriptase, Based on Quantum Chemical Calculations., in the Proceeding of The Fifth Annual National Symposium, on Computational Science and Engineering (ANSCSE 2001), Bangkok Convention Center(BCC), Central Plaza, Bangkok, Thailand., 19-20 Jun, 2001, 324.
 2. R. Palangsuntikul and S. Hannongbua, Movement, Orientation and Interaction Energy Study of the Water Molecule in the HIV-1 Reverse Transcriptase Active Site by Quantum Chemical Calculations., in the Proceeding of 27th Congress on Science and Technology of Thailand, Lee Gardens Plaza Hotel, Had Yai, Songkla, Thailand., 16-18 Oct, 2001, 267.
 3. R. Palangsuntikul and S. Hannongbua, Investigation on Movement, Orientation, Interaction Energy and IR spectroscopy study of the Water Molecule in the HIV-1 Reverse Transcriptase Active Site by Quantum Chemical Calculations., in the Proceeding of The International Conference on Bioinformatics 2002 : North-South Networking (InCoB2002), Le Royal Meridien, Bangkok, Thailand, 6-8 Feb, 2002, 110.
 4. M. Kuno, R. Palangsuntikul and S. Hannongbua, Investigation on an Orientation and Interaction Energy of the Water Molecule in the HIV-1 Reverse Transcriptase Active Site by Quantum Chemical Calculations, Journal of Chemical Information and Computer Sciences. 43 (2003) 1584.
 5. R. Palangsuntikul, R. Polly and B. Hartke, Global and local optimization of auxiliary basis sets for RI-MP2 calculations, Phys.Chem.Chem.Phys. 6 (2004) 5456.
-

7. DECLARATION

Hereby I declare that this dissertation was written independently by me and represents my own work. This work has not been submitted for any other degree or qualification. Information derived from the literature or unpublished work of others has been acknowledged in the text and by a list of references provided.

Kiel, 10.05.2005

Rungtiva Palangsuntikul