COMPUTATIONAL CHEMISTRY CAPACITY BUILDING IN AN UNDERPRIVILEGED CONTEXT: CHALLENGES, OUTCOMES AND PERSPECTIVES

L. Mammino
Department of Chemistry, University of Venda, Thohoyandou 0950, South Africa
e-mails: liliana@univen.ac.za, sasdestria@yahoo.com

ABSTRACT
Computational chemistry is a fast developing branch of modern chemistry, focusing on the study of molecules to enable better understanding of the properties of substances. Its applications comprise a variety of fields, from drug design to the design of compounds with desired properties (e.g., catalysts with specific actions) and to ample overlaps with nanotechnology. However, despite its relevance, the presence of computational chemistry research in Sub-Saharan Africa is still scarce, practically absent in many institutions. Bridging the gap with the other continents requires the identification of capacity building options that may enable a rapid growth. The recent experience at the University of Venda (a historically disadvantaged and under-resourced university in South Africa), where computational chemistry research capacity has been built “from scratch” up to currently promising levels, testifies the feasibility of such capacity building also in conditions of limited resources and may constitute a reference for other institutions. The paper highlights the main stages of the capacity building process through an overview of the research themes and the corresponding results. Specific attention is given to the major features that have characterised the capacity building process – the key role of human resources, the feasibility within limited infrastructure and financial resources, and the integration of research and training since the very first steps and throughout the process.

INTRODUCTION
Computational chemistry is a fast developing branch of modern chemistry. Since the properties of substances depend on the properties of their molecules, computational chemistry studies the properties of molecules to understand the properties of substances. The more we know and understand about the properties of molecules, the more we can understand the properties of existing substances and be able to predict the properties of substances that have not yet been synthesised. Thus, computational chemistry has continuously expanding applications, from drug design to the design of other compounds with desired properties (e.g., catalysts with specific actions, or substances that are more environmentally-benign than corresponding ones in current use). It extensively interfaces/overlaps with nanotechnology, as the latter implies managing molecules at the nanoscale, which implies knowing about molecules and their quantum behaviour.

Despite its relevance, the presence of computational chemistry research in Sub-Saharan Africa is still scarce. It is not yet present in a number of countries (in their universities or other research institutions), which implies a large gap with the other continents, including developing ones (Asia, Latin America, and also part of Northern Africa), where computational chemistry research is extensively present.

Developing computational chemistry in Sub-Saharan Africa may bring important contributions to the overall chemistry research capacity:

• By providing interpretation tools and pathways for the results obtained from
research in other areas of chemistry;
• By offering the possibility of effectively including computational chemistry as part of chemistry students’ training, in line with what is already common practice in other continents;
• Because of the potential interfaces with many research areas, including areas more directly related to industry like drug design, design of catalysts or nanotechnology.

In particular, it can be expected to contribute to retaining in the continent important steps of the development of drugs from biologically active molecules identified from materials utilized in traditional medicine (Mammino 2005), as the presence of adequate computational chemistry expertise is fundamental for the development of viable drugs from identified lead compounds. This would have specific relevance for research on endemic diseases in general and for valuing African biologically active molecules. The importance of lead compounds from natural materials is expected to increase sharply in the next years, as the lower-than-exception performance of combinatorial chemistry and high-throughput screening prompts new interest in the potentialities of compounds of natural origin, in view of their expected higher organism- and target-compatibility stemming from the fact that they are already parts of living organisms. Therefore, the next years can offer important roles to computational chemistry in African institutions.

All these aspects stress the importance of identifying capacity building options that may enable rapid growth of computational chemistry research and that are realistic under existing conditions. The characteristics of computational chemistry research enable reasonable development even in conditions of limited resources. Computational chemistry is an area with comparatively low financial demands (considerably lower than in other areas of chemistry) and in which the dominant factor are human resources. This implies that the major pre-requisite for its development is the training of specialists.

A realistic feasibility assessment for the initialization and development of computational chemistry research under available conditions can be based on concrete cases. This paper considers the example of the development of computational chemistry at the University of Venda (UNIVEN) – a chronically under-resourced Historically Black University located in a rural area of South Africa – where computational chemistry research has developed steadily in recent years.

FEASIBILITY ASSESSMENT THROUGH A CONCRETE EXAMPLE
Contextual situation and associated challenges
The initialization and development of computational chemistry research at UNIVEN can be viewed as representative (or particularly apt) for feasibility assessment because:
• The institution is a non-privileged one, experiencing realities and constraints frequent in several other non-privileged contexts.
• Initialization and development had to build “from scratch”, as there had been no prior research activity in this area in the institution.
• The development has been realised in a strictly “economical” manner, i.e., with very little funding.
Therefore, the experience at UNIVEN may have the role of possible model for other institutions in similar situations.

The main factors that have been essential for the development have been the presence of a specialist, the presence of a dedicated postgraduate student and the existence of linkages with a long-established group,
which enabled interactions and exchanges of views and also contributed technical support. The main difficulty was the minimal size of the research group (one professor and one student); however, such difficulty has been largely overcome through frequent contacts with other experts on the basis of the just-mentioned linkage and of frequent conference attendance.

The pattern selected for capacity building tried to respond in the most apt way to the major challenges, i.e., having to build “from scratch” and having to train new specialists. It was assumed that “building from scratch” should ideally comprise:

- Selecting a research topic whose investigation is realistic under the given circumstances and is tuned to some existing needs, either interfacing with ongoing research or responding to major interests and concerns.
- Fully integrating the training of new specialists into the capacity building process.
- Producing results in such a way that they constitute proof of existence of expertise and of the participation of the trainees in the capacity building process.

The steps of the development will be described rather in detail, together with the mention of the corresponding outputs, as such review can offer concrete testimonial of the actual feasibility of the capacity building process.

**Building research capacity through a project’s articulation**

The research area selected was the computational study of biologically active molecules. This can be viewed as an ideal area because it responds to the widely-spread interest for the development of drugs for endemic diseases and because it may easily interface with information from Indigenous Knowledge Systems (IKS) studies, e.g., by studying compounds identified from materials utilised in traditional medicine.

The patterns for capacity building starting “from scratch” can be easily matched with the standard patterns for the investigation of a biologically active compound or a given class of active compounds. The steps that have already been largely realised are the following:

- The study of a selected molecule in vacuo (that may involve also the selection and study of convenient model structures).
- The study of the same molecule in water solution and/or in other solvents that may be of interest.
- The study of the parent compound.
- The study of a sufficiently representative number of compounds of the same class.
- The study of derivatives of the same parent compound that may be interesting for comparisons, e.g., to obtain independent confirmation of identified trends.

The practical implementation of each of these steps is outlined in the next paragraphs. All the calculations were performed using GAUSSIAN 03, Revision D 01 (Frisch et al. 2003).

The first compound investigated was caespitate (Fig. 1), a compound isolated from a plant utilised in traditional medicine in South Africa and exhibiting antituberculosis, antibacterial and antifungal activities (Mathekga et al. 2000, Mathekga 2001, Meyer et al. 2002). Its study expanded to a rather detailed investigation of intramolecular hydrogen bonding, which, in turn, prompted the study of some model structures for the investigation of the intramolecular hydrogen bond between the carbonyl O of the acyl chain (COCH(CH₃)₂) and a neighbouring phenolic OH. Preliminary results were initially presented at conferences (Mammino and Kabanda...
2005a, Mammino and Kabanda 2005b); then the results in vacuo were published (Mammino and Kabanda 2007a). Major aspects of the approach were further presented individually at conferences (Mammino and Kabanda 2007b, Mammino and Kabanda 2007c, Mammino and Kabanda 2007d) to have the opportunity of proposing them for specific discussion, also in view of the expansion of the study to other molecules. The molecule was then investigated in solution, utilising the Polarisable Continuum Model (PCM, Miertus et al. 1981, Miertus and Tomasi 1985, Pascual-Ahuir et al. 1987, Alagona et al. 1990, Pascual-Ahuir and Silla 1990, Bonaccorsi et al. 1991, Silla et al. 1991, Pascual-Ahuir et al. 1994, Cossi et al. 1996, Barone and Cossi 1997, Cancès et al. 1997, Barone et al. 1998, Amovilli et al. 1999, Tomasi et al. 1999, Cossi et al. 2002, Tomasi et al. 2005) and selecting three solvents with different polarities and hydrogen bonding abilities – chloroform, acetonitrile and water – to mimic the range of possible media in which the molecule may act in a living organism. For the case of water, adducts with explicit water molecules (Fig. 1) were also considered, to complement the information obtained from PCM calculations (Mammino and Kabanda 2007e, Mammino and Kabanda 2008a).

![Figure 1](image1.png)

**Figure 1:** Two low-energy conformers of caespitate (a and b) and an adduct of conformer (a) with explicit water molecules (c).

![Figure 2](image2.png)

**Figure 2:** The two conformers of phloroglucinol (a and b) and a representation (c) of the best adduct of the lower energy conformer (a) with explicit water molecules.
If the ultimate objective of an investigation is that of utilising computational information for the study of biological activities, it is important to have information on an adequate number of compounds of the same class. In the case of caespitate, the class is that of acylphloroglucinols (Fig. 3), derivatives of phloroglucinol (1,3,5-trihydroxybenzene) characterised by the presence of at least one COR group (acyl group).

The preliminary step was the study of the parent compound (Fig. 2), in vacuo and in solution (Mammino and Kabanda 2006, Mammino and Kabanda 2008b), to identify features that might be relevant for the conformational preferences of its derivatives. Phloroglucinol was also compared to the other polyphenols, through a systematic study of polyhydroxybenzenes (Mammino and Kabanda 2011a) which extended also to their dimers (Mammino and Kabanda 2012a).

The core task was the study of a sufficiently representative number of acylphloroglucinols to be able to identify patterns in their conformational preferences and the factors influencing them. One hundred eighteen molecular structures were studied, comprising structures reported in the major review on naturally occurring phloroglucinols (Singh and Bharate 2006) and model structures meant to represent the most common features that may have significant roles in determining their conformational preferences. Thus, the investigated structures included: different R; the possibility, for each R, that R′ = H or that R′ ≠ H (modelled by R′ = CH₃); cases in which one or more phenolic OH are replaced by OCH₃ groups; cases in which an ortho phenolic OH is replaced by a keto O; and cases with R′ ≠ H more complex than CH₃.

Different levels of theory were utilised (HF, DFT/B3LYP, MP2) to assess the performance of the weaker ones against the results of the more sophisticated ones. The results highlight patterns that remain consistent across structures and across levels of theory and, therefore, they constitute an ensemble of information that can be utilised to predict conformational preferences of other acylphloroglucinols.

![Figure 3: General structure of acylphloroglucinols.](image)

![Figure 4: The two possibilities for formation of the intramolecular hydrogen bond between the sp² O of the COR group and an ortho OH in acylphloroglucinols (Mammino and Kabanda 2009).](image)
Given the huge amount of information made available by a complete conformational study of such a large number of structures, each of them in four media (vacuum, chloroform, acetonitrile and water), and given the fact that several sets of results prompted theoretical reflections at interpretation level, the results for different major features were presented separately, often maintaining the practice of initially presenting them at conferences (to have the opportunity of preliminary discussions and feedback) and then through a publication.

The main results can be categorised as follows:

• Results concerning the intramolecular hydrogen bond between the sp$^2$ O of the COR group and an ortho phenolic OH (first IHB, Fig. 4). This IHB has a fundamental role in determining conformational preferences (Mammino and Kabanda 2009a, Mammino and Kabanda 2009b). It is present in all populated conformers and remains present in solution, including water solution (Mammino and Kabanda 2009c). Its characteristics show consistent patterns (Mammino and Kabanda 2009b, Mammino and Kabanda 2009c).

• Results concerning the role of C−H⋯O intramolecular hydrogen bonds (Fig. 5) in influencing conformational preferences of different conformers in which the first IHB is present (Mammino and Kabanda 2012b).

• Results concerning the roles of features characterizing relevant subsets of acylphloroglucinols, like the possibility of O−H⋯π interactions (Mammino and Kabanda: in press 1).

**Figure 5:** Illustration of selected aspects of the study of acylphloroglucinols. (a): One of the patterns for the formation of C−H⋯O intramolecular hydrogen bonds. b, c: Adducts with explicit water molecules for a compound with R = CH$_3$ and R′ = H (b) and for a compound with R = CH$_3$ and R′ = CH$_3$ (c).
Results concerning the major trends in conformational preferences, in vacuo and in solution, considering all the factors influencing them (Mammino and Kabanda 2008c, Mammino and Kabanda 2011b, Mammino and Kabanda 2012c). The factors include features like preferences in the orientation of the phenolic OH – e.g., the preference (similarly to the parent compound) for uniform mutual orientation of the three phenolic OH (as in both geometries shown in Fig. 4). In solution, the factors include aspects like the effect of the solvent polarity on the conformers’ energy gaps.

- Results on the energetics of the solution process, in the three solvents considered (Mammino and Kabanda 2012c).
- Results highlighted by the study of adducts with explicit water molecules, Fig. 5, (Mammino and Kabanda 2009d, Mammino and Kabanda 2010). These include the preference for specific arrangements of water molecules around the hydrophobic region of the first IHB or around free phenolic OH and the influence of the orientation of free phenolic OH on the arrangements of water molecules.

The use of models, like modelling R’ through a methyl group, requires validation of the reliability of the modelling option. This was carried out through extensive comparisons of model structures with R’ = CH₃ and structures with longer and more complex R’ and the same R. Some pairs are shown in Fig. 6. Even the study of a rather complex structure like nodifloridin B (Mammino and Kabanda 2009e, Mammino and Kabanda 2009f) confirmed the predictability of the patterns concerning the region of the phloroglucinol moiety and the first IHB on the basis of the results obtained for simpler acylphloroglucinols.

![Figure 6: Examples of pairs of simplified/model and more complex acylphloroglucinol structures](image)
A derivative of the same parent compound that appeared interesting for comparisons, and for obtaining separate confirmation of some of the identified trends, was the carboxylic acid of phloroglucinol (Fig. 7), as the carboxylic functional group contains an sp\(^2\) O and can, therefore, form an IHB with an ortho OH having interesting analogies with the first IHB of acylphloroglucinols. The study expanded beyond the comparison with acylphloroglucinols to include the features that are typical of the study of carboxylic acid, from the prediction of the dissociation constant to the consideration of possible dimers (Mammino and Kabanda 2008d, Mammino and Kabanda 2010).

**Integration of research and education**

Each new step in the development of the study outlined in the previous section required new concepts and new tools: learning to choose model molecules apt for the study of specific features; learning how to study molecules in solution; learning to compare results in different media, results for different structures, and results obtained with different calculation methods; and learning to identify the theoretical issues that might be of interest (e.g., the study of intramolecular hydrogen-bonding, specific aspects of the effect of solvents, etc.). They corresponded to new steps (often milestones) in a student’s training and acquisition of expertise.

Integrating research and education has several advantages. A student is involved in active research since the very beginning (what is particularly important for postgraduate students). The development of expertise is hands-on, i.e., the student learns theory and concepts when the need for applying them arises, and, therefore, theories and concepts acquire a concrete dimension that has an important role in students’ perceptions. The active participation in the capacity building process heightens the student’s motivations, as he/she perceives the relevance of the research and shares in the emotions of “building something”. The production of research outputs constitutes an important recognition of the student’s progress – each output practically becoming a sort of milestones marking a set of achievements.

**Additional considerations on feasibility and motivations**

Some concerns and some hesitations may deserve *ad hoc* consideration, so as to complete the picture of the relevance and feasibility of the development of computational chemistry. An idea that appears frequently among expressed
concerns is that this type of research is not suitable for developing institutions, but only for “first world” countries. However, computational chemistry research is well active in all other developing contexts (Asia, Latin America, Northern Africa), producing a high number of interesting results. Thus, there is no reason for a continuation of its “scarce-skills” or “nearly absence” status in Sub-Saharan Africa.

Experience shows that postgraduate students going overseas to train in this field rarely return home after completion, not only because of economic reasons, but also because they are not sure that they can continue doing research in this area after returning to their home countries. A possibility for tackling this problem may involve the following set of options:

- Training students in the continent, possibly in conditions not too different from those in which they will work on coming back home, so that they are emotionally and technically ready to carry out research within those conditions;
- Training them to be initiators, so that they feel ready for the challenges of being the first to initialize research in this area on coming back home;
- Building linkages and networks, so that young specialists undertaking the initialization of capacity building in their institutions can have opportunities for extensive exchanges of views and information, sharing intellectual and investigation challenges and relying on the support by more experienced researchers.

The current major difficulty is the scarcity of experts continent-wide. The number of experts in computational chemistry is seriously inadequate. This affects all levels of activities, dramatically reducing the possibility of offering exposure to chemistry students, of preparing new specialists, and of initializing and developing research. The design of feasible options for somehow “sharing” the currently available experts (Mammino and Khanra 1995) can be the best short-term solution: it would enable the dissemination of information about the roles of theoretical and computational chemistry, the realisation of some computational chemistry exposure for chemistry undergraduates, and the possibility of training selected students. The preparation of new specialists is the major key to ensure development. The integration of research and education can greatly help prepare new specialists that will also be able to be initiators (i.e., capable of initialising and developing computational chemistry research in their institutions) and trainers (i.e., capable of, in turn, training other new specialists). Learning through an approach that gives adequate attention to education aspects is optimal to ensure that a student becomes an initiator and a trainer.

CONCLUSIONS
Computational chemistry research is still inadequately present in many institutions in Sub-Saharan Africa. The key to a rapid growth under current circumstances resides in the training of an adequate number of new specialists capable of being initiators and trainers. The integration of research and education can be particularly extensive in the early stages of a capacity building process. The outcomes of the capacity building process at UNIVEN show that all this is feasible even in institutions with limited resources.

In computational chemistry, human resources are the key factor. Ensuring the preparation of human resources is the key to a development that can simultaneously be fast and realistic. Under the current circumstances, “preparation of human resources” means training new specialists that can be ready to undertake the initiatives for capacity building.
It is possible to produce interesting results even in the early stages of the capacity building process, and even while working with limited resources. For instance, although it is widely acknowledged, in principle, that the study of biologically active molecules requires the availability of information on many related compounds, it appears that the computational study of acylphloroglucinols carried out at UNIVEN is the first systematic study comprising a high number of different compounds of the same class and analysing both the common features and the features related to the individuality of each compound. This makes the development of computational chemistry research an endeavour with appealing potentialities for research outputs since its very beginning.

REFERENCES


Mammino L and Kabanda MM 2009a A study of the intramolecular hydrogen bond characterising acylphloroglucinols. 13th ICQC, International Congress of
of *Helichrysum* species and the isolation of a new phloroglucinol from *Helichrysum Caespititium*. Doctoral thesis at the University of Pretoria.

Meyer JJ, Lall N and Mathegka ADM 2002

Miertus S, Scrocco E and Tomasi J 1981


