

Journal of

**Applied
Crystallography**

ISSN 0021-8898

Editor: **Anke R. Kayser-Pyzalla**

CLUSTERGEN: a program for molecular cluster generation from crystallographic data

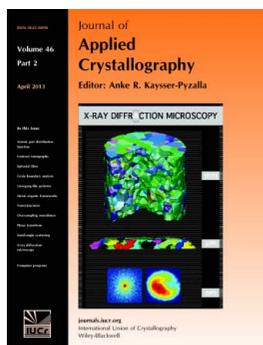
Radosław Kamiński, Katarzyna N. Jarzemska and Sławomir Domagała

J. Appl. Cryst. (2013). **46**, 540–543

Copyright © International Union of Crystallography

Author(s) of this paper may load this reprint on their own web site or institutional repository provided that this cover page is retained. Republication of this article or its storage in electronic databases other than as specified above is not permitted without prior permission in writing from the IUCr.

For further information see <http://journals.iucr.org/services/authorrights.html>



Journal of Applied Crystallography covers a wide range of crystallographic topics from the viewpoints of both techniques and theory. The journal presents papers on the application of crystallographic techniques and on the related apparatus and computer software. For many years, the *Journal of Applied Crystallography* has been the main vehicle for the publication of small-angle scattering papers and powder diffraction techniques. The journal is the primary place where crystallographic computer program information is published.

Crystallography Journals **Online** is available from journals.iucr.org

CLUSTERGEN: a program for molecular cluster generation from crystallographic data

Radosław Kamiński,* Katarzyna N. Jarzemska and Sławomir Domagała

Department of Chemistry, University of Warsaw, Pasteura 1, Warszawa, 02-093, Poland. Correspondence e-mail: rkaminski@chem.uw.edu.pl

A new program, *CLUSTERGEN*, for molecular cluster generation is introduced. *CLUSTERGEN* provides the quantum mechanics/molecular mechanics (QM/MM) input files for program packages such as *ADF* [Baerends *et al.* (2012). Vrije Universiteit, Amsterdam, The Netherlands] and *GAUSSIAN* [Frisch *et al.* (2009). Gaussian Inc., Pittsburgh, Pennsylvania, USA]. Additionally, it prints out a standard *CRYSTAL* [Dovesi *et al.* (2009). University of Turin, Italy] input and, in general, facilitates file-format manipulation. The *CLUSTERGEN* program is supported by an extensive manual and a user-friendly graphical interface. The code is freely available and carefully commented, which makes it easily modifiable. Exemplary applications of *CLUSTERGEN* concerning QM/MM calculations and derivation of nucleus-independent chemical shift indices are demonstrated.

© 2013 International Union of Crystallography
Printed in Singapore – all rights reserved

1. Introduction

Recent advances in solid-state experimental techniques have allowed the investigation of light-induced species in molecular crystals (Coppens, 2011). Nevertheless, there are few theoretical methods and tools providing reliable reference results. The current state is a consequence of severe difficulties concerning proper modelling of partially excited crystals, *i.e.* extended systems composed of both ground and excited state molecules. One of the most common approaches, which makes use of the assumption of low excitation percentage, employs the so-called embedded model. The idea is to treat an excited state species as to some extent independent from its ground state surrounding. Such an approach can be realized using various quantum mechanics/molecular mechanics (QM/MM) schemes (Bjornsson & Bühl, 2012). It has been recently shown that constrained QM/MM geometry optimizations of excited state species provide almost quantitative agreement with the experimental results (Benedict *et al.*, 2011; Makal *et al.*, 2012, 2011; Terentjevs *et al.*, 2012). Moreover, this method is computationally cheap and robust. Additionally, the QM/MM approach has been applied in order to estimate atomic displacement parameters in molecular crystals, so the gained information may serve as a supplementary input in charge-density or high-pressure X-ray diffraction studies (Dittrich *et al.*, 2012).

Despite the increased interest in QM/MM studies of molecular crystals, there are few computer programs that write the appropriate input files (Dittrich *et al.*, 2012). The existing ones are often not user friendly and cannot be easily modified, as the source code is rarely available. Therefore, here we present a new program, *CLUSTERGEN*, designed especially for the purpose of creating input for computational studies. The main feature of the program is the easy generation of various molecular clusters, using the available crystallographic data, which can serve for QM/MM calculations. *CLUSTERGEN* produces input designed for the *GAUSSIAN* (Frisch *et al.*, 2009), *ADF* (Baerends *et al.*, 2012; Velde *et al.*, 2001) and *CRYSTAL* (Dovesi *et al.*, 2009, 2005) packages.

2. Program description

CLUSTERGEN uses linear algebra concerning crystal symmetry and metrics. The main features of the *CLUSTERGEN* program are listed below:

- (1) Molecular cluster shell building around a selected molecule, a fragment or a given point in space.
- (2) Molecular fragment completion based on space-group symmetry.
- (3) Automatic and manual *X*–H bond-length standardization (where *X* is a non-H atom).
- (4) Automatic fragment search.
- (5) Generation of *SHELX*-type (Sheldrick, 2008) RES and XYZ output files.
- (6) Generation of QM/MM input files suitable for *GAUSSIAN* or *ADF* programs.
- (7) Other options, such as the assignment of point positions for computation of nucleus-independent chemical shift (NICS) for a specified aromatic ring *etc.*

2.1. Fragment search

Molecular fragments can be defined manually or detected by the automatic fragment search procedure. The program gradually checks which atoms are bonded together, using a covalent radius library and a given tolerance, and reconstructs a molecule atom by atom. This procedure is thus a convenient way of dividing a given structure into well defined molecular fragments on the basis of the connectivity table. Additionally, it is possible to specify atomic pairs not to be treated as bonded during the search procedure.

2.2. Molecule completion

The *CLUSTERGEN* program uses the crystal space-group symmetry and, if needed, can complete a molecular fragment. Generally, symmetry-equivalent atoms that constitute the same molecular fragment are located within adjacent unit cells. It is

checked whether an analysed atom is close enough to any atom belonging to the current fragment, *i.e.* within the sum of the corresponding atomic covalent radii with a defined tolerance. When such an atom is found, it is included in the fragment atom list and the search is repeated until no new atoms are detected. If the number of new atoms is greater than the number of atoms in the whole unit cell (*e.g.* a structure is a coordination polymer) the program stops and prints out an appropriate warning. In such a case the user may consider manual assignment of individual fragments or may exclude some of the bonds during the automatic fragmentation procedure.

2.3. X–H bond standardization

For the purpose of many QM/MM calculations, X–H bonds should be adjusted to the standard neutron distances. Therefore an automatic standardization routine, using the X–H bond lengths recently published by Allen & Bruno (2010), was implemented in *CLUSTERGEN*. The algorithm consists of the following stages: (1) identification of the heavy atom, X, bound directly to the analysed H atom; (2) analysis of atoms connected to the X atom in order to correctly assign the bond type (in the case of aromatic C atoms and N atoms, further analysis of the X-atom coordination geometry is considered); and (3) standardization of the X–H bond to the tabulated value by shifting the H atom along the bond direction. When no proper bond type is specified in the code, selected H atoms can be manually shifted to a chosen distance. According to our recent publications, we emphasize the importance of the X–H adjustment in the solid-state QM calculations (Jarzemska *et al.*, 2012; Durka *et al.*, 2012).

2.4. Cluster building

The main purpose of the program is to build a cluster around a selected point in space, atom or molecular fragment, within a specified radius (d_{\max}). The first step of the procedure is the analysis of the number of unit cells that need to be taken into account. The second step is the generation of molecules belonging to the selected unit cells, and verification of whether they are within a chosen distance from the central point, atom or molecular fragment. This distance is defined as the sum of the specified maximal distance (d_{\max}) and molecular radius (d_{mol}). Three options are available: (1) $d_{\text{mol}} = 0$, the selected area constitutes a sphere based only on the d_{\max} value (Fig. 1a); (2) d_{mol} is calculated as the distance between the geometrical centre and the most distant atom of the central fragment (Fig. 1b); and (3) d_{mol} is calculated as the distance between the geometrical centre and the most distant atom of the central molecular fragment in a given direction (this model takes into account the molecular shape) (Fig. 1c).

2.5. Output options

Currently, the program prints out several types of output files:

(1) XYZ file – a common format containing Cartesian coordinates for a given cluster.

(2) RES file – a crystallographic format containing fractional coordinates of a studied system, similar to the output file produced by the *SHELX* programs (Sheldrick, 2008).

(3) *ADF* QM/MM input file – the format contains fractional coordinates of both the central fragment and shell molecules. In this case *CLUSTERGEN* writes the full connectivity information for a molecular mechanics calculation together with the corresponding atomic types. For the purpose of QM/MM computation such atomic types can be assigned by specific software (*e.g.* *AMBERTOOLS*;

Case *et al.*, 2012) or specified by the user for a chosen force field. If requested, the program can also assign and print out atomic charges.

(4) *GAUSSIAN* QM/MM input file – a format similar to the *ADF* input file.

(5) D12 file – the format read by the *CRYSTAL* program, used for periodic calculations and evaluation of properties in the solid state.

2.6. Program code details

CLUSTERGEN is written in the Fortran90 programming language. It consists of a main program body and a number of minor subroutines. This enables its easy modification in order to implement new features or change the existing ones. The program is available on request from the authors in the form of a source code or can be downloaded directly from the web site <http://www.chem.uw.edu.pl/people/RKaminski/software.html>. We provide also an extensive manual where all the functions and capabilities of the program are described. The graphical user interface (GUI) (Fig. 2) for the *CLUSTERGEN* program is also freely available. The GUI is designed in Java 1.6 using lightweight Swing components and JGoodies forms (<http://www.jgoodies.com>). The Fortran executable is invoked *via* the underlying operating system's native commands. The output from the Fortran executable is redirected to the GUI. The current version is capable of importing CIFs and XYZ files, as well as saving and loading the native *CLUSTERGEN* input files. All the commands used in the *CLUSTERGEN* Fortran program can be controlled through the graphical interface. The GUI is a multi-plat-

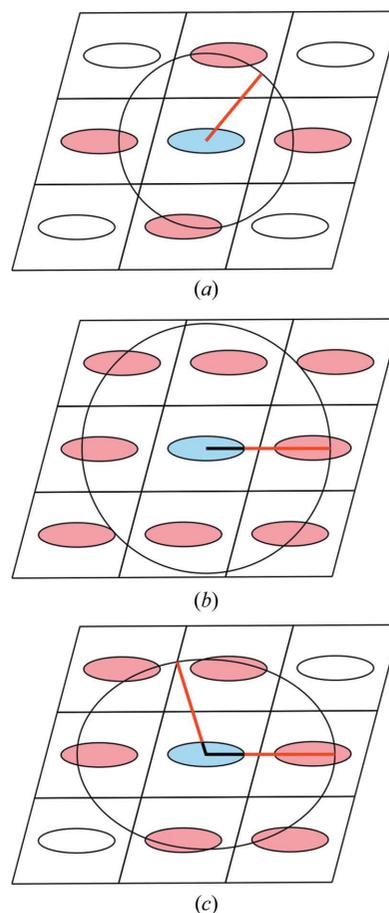


Figure 1
Schematic representation of three cluster building models: (a) d_{\max} (isotropic model); (b) $d_{\max} + d_{\text{mol}}^{\text{iso}}$ (isotropic model); (c) $d_{\max} + d_{\text{mol}}^{\text{aniso}}$ (anisotropic model).

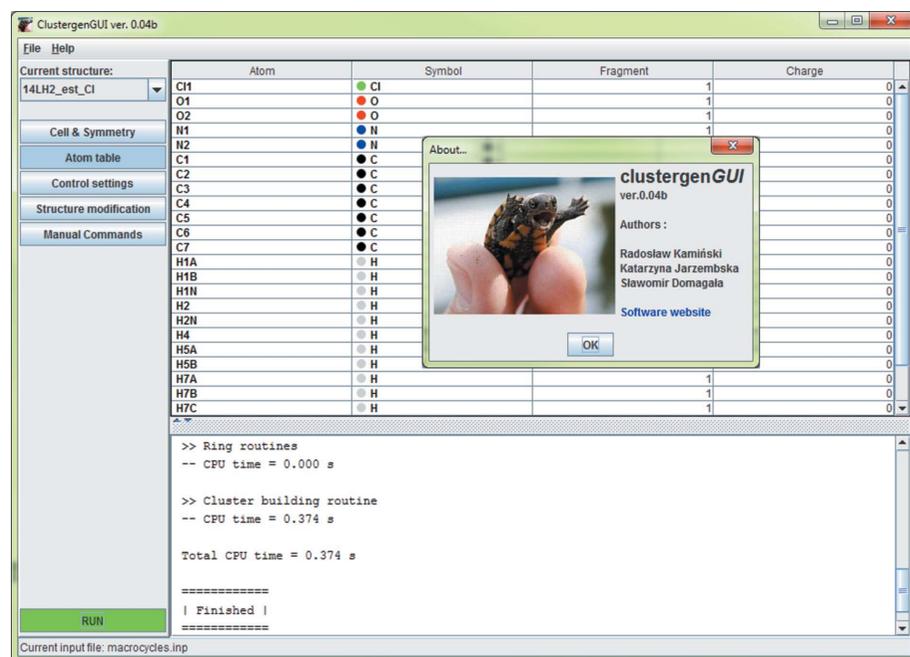


Figure 2
Graphical user interface for the *CLUSTERGEN* program.

form program and requires only the Java virtual machine to be installed (version 1.6 or later).

3. Computational examples

The main functionality of the program is presented in a simple example of a QM/MM calculation. For this demonstrative study we chose the ionic 1,2-azaborine complex (Fig. 3*a*), first synthesized by Marwitz *et al.* (2010), which exhibits interesting fluorescence properties in the solid state. Following the strategy presented in the literature (Kamiński *et al.*, 2010) we computed the excited state geometry of the 1,2-azaborine-type cation. The *CLUSTERGEN* program was used to build up a cluster around the cation moiety. The central part was surrounded by one shell of ions, and all C–H bond lengths were adjusted to standardized neutron distances. QM/MM calculations were performed with the *ADF* package using the BP86 functional (Becke, 1988; Perdew, 1986) and triple- ζ basis set for the high-level region (*i.e.* central moiety). The low-level region (molecular mechanics) was modelled by atomic point charges and force-field parameters transferred from the Generalized Amber Force Field (GAFF) library (Wang *et al.*, 2006, 2004) attributed to the fixed molecular geometry. The difference between ground and excited state geometries is presented in Fig. 3(*b*). It is clearly visible that in the solid state the excited molecule is characterized by a significant twist of the middle aromatic ring with respect to the ground state geometry. The remaining part of the molecule is rather inert to the excitation. The largest C-atom displacement is around 0.4 Å, which suggests that such an effect could be observed experimentally using pump-probe photocrystallographic experiments. It should also be emphasized that the substantial structural changes make such systems potentially attractive in materials science, especially in electronics applications.

Another application of our program that is worth mentioning is the calculation of points in space where the so-called NICS (Chen *et al.*,

2005; Schleyer *et al.*, 1996) values can be computed. This is a handy feature of *CLUSTERGEN* owing to the lack of other tools for automatic generation of NICS-point positions. At first, the molecular ring should be manually specified so as to activate the NICS-point calculation option. The procedure utilizes a least-squares fit of a plane to the selected atomic positions. Then, two vectors, perpendicular to the plane, are calculated. The number of points to be generated along these two vectors up to a considered distance from the ring centroid can be specified. As an example, we chose two compounds: *p*-phenylenediboronic acid and its tetrafluoro derivative (Durka *et al.*, 2012). Using *CLUSTERGEN* we generated ten points up to 2.0 Å distance from the aromatic ring centroid (Fig. 4*a*). Automatically generated input was then submitted to *GAUSSIAN* (B3LYP/aug-cc-pVDZ level of theory). The computational results are presented in Fig. 4(*b*). As clearly seen, the two compounds behave somewhat differently, especially in the vicinity of the aromatic ring centres. The unsubstituted acid exhibits a well defined minimum at

about 0.8 Å distance from the aromatic ring centre, whereas no such feature is observed in the case of the second compound. Interpretation of such results can be found in the vast literature treating aromaticity indices and their applications (Cyrański, 2005; Krygowski & Cyrański, 2001; Schleyer *et al.*, 2001). Generally, this example demonstrates the usefulness of our program in studies regarding magnetic properties of aromatic molecules.

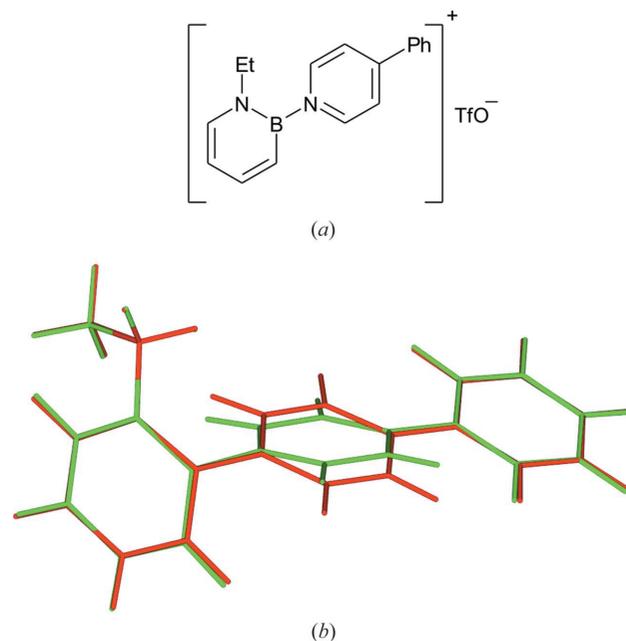


Figure 3
(*a*) Schematic representation of the studied system. (*b*) Difference between the ground (light grey; green in the electronic version of the journal) and excited (dark grey or red) state geometries computed using the QM/MM approach.

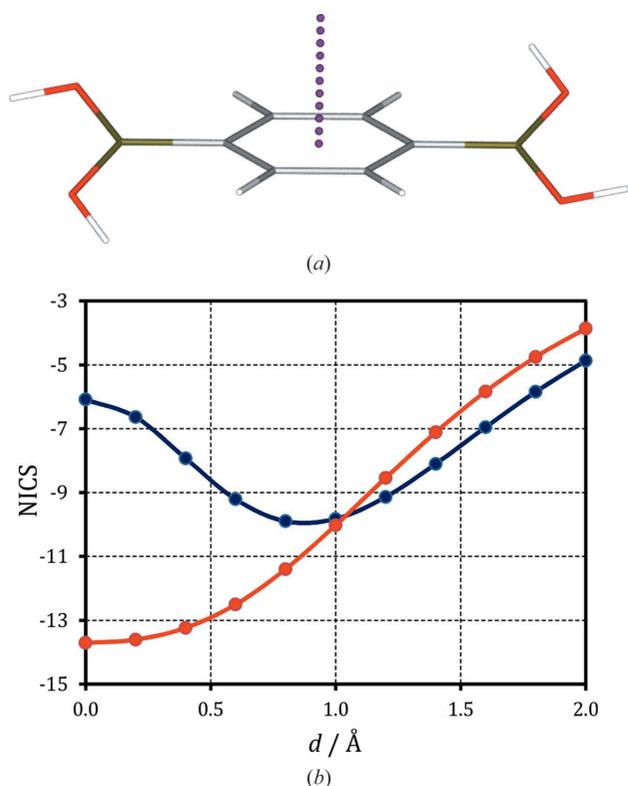


Figure 4

(a) NICS-point positions generated by the program for the unsubstituted acid molecule (only NICS points at one side of the ring are shown). (b) NICS values scanned along the distance measured from the ring centre [dark grey points (blue in the electronic version of the journal): unsubstituted acid; light grey (red) points: tetrafluoro derivative].

4. Conclusions

CLUSTERGEN is a robust program that provides the QM/MM input files for program packages such as *ADF*, *GAUSSIAN* and *CRYSTAL*. *CLUSTERGEN* also facilitates file-format manipulation. The code is freely available and carefully commented, which makes it easily modifiable. The *CLUSTERGEN* program is supported by an extensive manual and a user-friendly GUI, which may serve as a basis for future development of more advanced support tools for quantum chemistry and crystallography.

The authors would like to thank Philip Coppens and Jesse Sokolow (Buffalo, USA) for inspiration and fruitful discussions, and Krzysztof Woźniak (Warszawa, Poland) for encouragement and support. RK would like to thank the Foundation for Polish Science for financial support within the ‘International PhD Projects’ and START

programmes. RK and KNJ also acknowledge the Ministry of Science and Higher Education in Poland for financial support within the Inventus Plus grant (120000-501/59 PM-121).

References

- Allen, F. H. & Bruno, I. J. (2010). *Acta Cryst.* **B66**, 380–386.
- Baerends, E. J. *et al.* (2012). *ADF2012*. Vrije Universiteit, Amsterdam, The Netherlands.
- Becke, A. D. (1988). *Phys. Rev. A*, **38**, 3098–3100.
- Benedict, J. B., Makal, A., Sokolow, J. D., Trzop, E., Scheins, S., Henning, R., Graber, T. & Coppens, P. (2011). *Chem. Commun.* **47**, 1704–1706.
- Bjornsson, R. & Bühl, M. (2012). *J. Chem. Theory Comput.* **8**, 498–508.
- Case, D. A. *et al.* (2012). *AMBER12*. University of California, San Francisco, California, USA.
- Chen, Z., Wannere, C. S., Corminboeuf, C., Puchta, R. & Schleyer, P. V. (2005). *Chem. Rev.* **105**, 3842–3888.
- Coppens, P. (2011). *J. Phys. Chem. Lett.* **2**, 616–621.
- Cyrański, M. K. (2005). *Chem. Rev.* **105**, 3773–3811.
- Dittrich, B., Pfitzenreuter, S. & Hübschle, C. B. (2012). *Acta Cryst.* **A68**, 110–116.
- Dovesi, R., Orlando, R., Civalieri, B., Roetti, C., Saunders, V. R. & Zicovich-Wilson, C. M. (2005). *Z. Kristallogr.* **220**, 571–573.
- Dovesi, R., Saunders, V. R., Roetti, C., Orlando, R., Zicovich-Wilson, C. M., Pascale, F., Civalieri, B., Doll, K., Harrison, N. M., Bush, I. J., D’Arco, P. & Llunell, M. (2009). *CRYSTAL09*. University of Turin, Italy.
- Durka, K., Jarzemska, K. N., Kamiński, R., Luliński, S., Serwatowski, J. & Woźniak, K. (2012). *Cryst. Growth Des.* **12**, 3720–3734.
- Frisch, M. J. *et al.* (2009). *GAUSSIAN09*. Gaussian Inc., Pittsburgh, Pennsylvania, USA.
- Jarzemska, K. N., Kubsik, M., Kamiński, R., Woźniak, K. & Dominiak, P. M. (2012). *Cryst. Growth Des.* **12**, 2508–2524.
- Kamiński, R., Schmökel, M. S. & Coppens, P. (2010). *J. Phys. Chem. Lett.* **1**, 2349–2353.
- Krygowski, T. M. & Cyrański, M. K. (2001). *Chem. Rev.* **101**, 1385–1419.
- Makal, A., Benedict, J., Trzop, E., Sokolow, J., Fournier, B., Chen, Y., Kalinowski, J. A., Graber, T., Henning, R. & Coppens, P. (2012). *J. Phys. Chem. A*, **116**, 3359–3365.
- Makal, A., Trzop, E., Sokolow, J., Kalinowski, J., Benedict, J. & Coppens, P. (2011). *Acta Cryst.* **A67**, 319–326.
- Marwitz, A. J. V., Jenkins, J. T., Zakharov, L. N. & Liu, S. (2010). *Angew. Chem. Int. Ed.* **49**, 7444–7447.
- Perdew, J. (1986). *Phys. Rev. B*, **33**, 8822–8824.
- Schleyer, P. von R., Maerker, C., Dransfeld, A., Jiao, H. & Hommes, N. J. R. van E. (1996). *J. Am. Chem. Soc.* **118**, 6317–6318.
- Schleyer, P. von R., Manoharan, M., Wang, Z.-X., Kiran, B., Jiao, H., Puchta, R. & Hommes, N. J. R. van E. (2001). *Org. Lett.* **3**, 2465–2468.
- Sheldrick, G. M. (2008). *Acta Cryst.* **A64**, 112–122.
- Terentjevs, A., Fabiano, E. & Della Sala, F. (2012). *Theor. Chem. Acc.* **131**, 1154–1161.
- Velde, G. te, Bickelhaupt, F. M., van Gisbergen, S. J. A., Guerra, C. F., Baerends, E. J., Snijders, J. G. & Ziegler, T. (2001). *J. Comput. Chem.* **22**, 931–967.
- Wang, J., Wang, W., Kollman, P. A. & Case, D. A. (2006). *J. Mol. Graph.* **25**, 247–260.
- Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. (2004). *J. Comput. Chem.* **25**, 1157–1174.