SOFTWARE METAPAPER

ConfBuster: Open-Source Tools for Macrocycle Conformational Search and Analysis

Xavier Barbeau[1], Antony T. Vincent[2] and Patrick Lagüe[3]

[1] The Quebec Network for Research on Protein Function, Engineering, and Applications (PROTEO), Institute for Integrative Systems Biology (IBIS) and Department of Chemistry, Faculty of Sciences and Engineering, Laval University, Quebec, CA

[2] Institute for Integrative Systems Biology (IBIS) and Department of Biochemistry, Microbiology and Bioinformatics, Faculty of Sciences and Engineering, Laval University, Quebec, CA

[3] The Quebec Network for Research on Protein Function, Engineering, and Applications (PROTEO), Institute for Integrative Systems Biology (IBIS) and Department of Biochemistry, Microbiology and Bioinformatics, Faculty of Sciences and Engineering, Laval University, Quebec, CA

Corresponding author: Patrick Lagüe (Patrick.Lague@bcm.ulaval.ca)

Macrocycles are cyclic macromolecules that have gained an increased interest in drug development. To our knowledge, the current bioinformatics tools that are available to investigate and predict macrocycles 3D conformations are limited in their availability. In this paper, we introduce ConfBuster, a suite of tools written in Python with the goal of sampling the lower energy conformations of macrocycles. The suite also includes tools for the analysis and visualisation of the conformational search results. Coordinate sets of single molecules in MOL2 or PDB format are required as input, and a set of lower energy conformation coordinates is returned as output, as well as PyMOL script and graphics for results analysis. In addition to Python and the optional R programming languages with freely available packages, the tools require Open Babel and PyMOL to work properly. For several examples, ConfBuster found macrocycle conformations that are within few tenths of Å of the experimental structures in minutes. To our knowledge, this is the only open-source tools for macrocycle conformational search available to the scientific community.

Keywords: Macrocycles; conformational search; conformational sampling; molecular modelling; open-source tools; click chemistry tools; cheminformatics; rational drug design; drug development

(1) Overview

Introduction

Macrocycles are cyclic macromolecules containing rings of 8–12 or more atoms [1, 2]. These compounds are found in natural products such as macrolides, that consist of a large macrocyclic lactone ring to which one or more deoxy sugars may be attached, which have antibiotic or antifungal activities and are used as pharmaceutical drugs [3]. Due to their unique properties [4, 5], such as of their ability in achieving protein-protein inhibition [6, 7] and cell permeability [8], this class of compounds has gained an increased interest in drug development. As a consequence, many macrocycles are currently in clinical development [4], supporting the development of macrocycle-based pharmaceutical companies (Cyclenium pharma, Circle pharma).

The knowledge of the 3D structure is fundamental to the rational design of biologically active molecules. Ideally, these structures would be determined experimentally, either by X-ray crystallography or nuclear magnetic resonance spectroscopy, but this could sometimes be a laborious, time-consuming costly avenue (if possible at all). Fortunately, molecular modelling techniques were developed over the past decades to assess the conformational sampling of chemical structures with a reliable accuracy [9]. To our knowledge, the current bioinformatics tools that are available to investigate and predict macrocycles 3D conformations are limited in their availability, precluding the development of macrocycles, as they are commercially distributed (MOE [10], MacroModel [11], Tork [12], Corina [13] or Schrödinger [14]) or not available to the public (SOS [15] and ForceGen 3D [16]).

In this Software Metapaper, we present an open-source tool suite called ConfBuster, providing tools for macrocycles conformational search, analysis and visualisation of results. The tool suite is based on Python and the NetworkX [19] python package, and on Open Babel [17] chemical toolbox and the PyMOL [18] molecular viewer. For the optional analysis of the results, ConfBuster also relies on the R programming language and the ComplexHeatmap [20] and circlize [21] packages. For several examples, ConfBuster found macrocycle conformations that are within few tenths of Å of the experimental structures in minutes. To our knowledge, this is the only open-source

tools for macrocycle conformational search available to the scientific community.

Implementation and architecture

ConfBuster is composed of 4 Python scripts with the goal of finding the lowest energy conformation of single macrocycles, given a set of molecular coordinates provided as MOL2 or PDB format. The relation between the ConfBuster scripts is illustrated in **Figure 1**, along with their respective third-party program and package dependencies. The primary script (`ConfBuster-Macrocycle-Linear-Sampling.py`) performs a conformational search by cleaving the macrocycle at different positions and, for each of the linear molecule created from the cleavage of a bond, calls the secondary scripts to perform a rotational search (`ConfBuster-Rotamer-Search.py`) and energy minimisations (`ConfBuster-Single-Molecule-Minimization.py`). Further, the primary script provides PyMOL input files for visualisation of the search progress and results. As an optional step, the script `ConfBuster-Analysis.py` generates the plot of the root mean square deviation (RMSD)-based hierarchical clustering of the resulting conformations and of the conformational energies.

Dependencies. ConfBuster depends on Open Babel [17], either for file format conversion as well as several functions for conformational sampling and energy minimisation. ConfBuster also depends on PyMOL [18] for RMSD calculations, for a dihedral sampling protocol and for visualisation. The Python package NetworkX [19] is used to identify atoms belonging to cycles and macrocycles. In addition to standard R packages, the optional ConfBuster analysis relies on the ComplexHeatmap package [20] available through the Bioconductor software project (bioconductor.org) and the circlize package [21]. The respective dependencies of ConfBuster scripts are illustrated in **Figure 1**.

ConfBuster-Single-Molecule-Minimization.py. This script uses the command-line program `obminimize` from the Open Babel package to optimize the geometry and minimise the energy of the molecule. Then, the `obenergy` command-line program is used to calculate the final energy, stored in the title of the final MOL2 file. The script command line arguments are the mandatory input filename (-i *name*) and the prefix of the output files (-o [default: replace input file]).

ConfBuster-Rotamer-Search.py. This script uses the `obabel` command-line program with the `--conformer` option to generate a number of conformations. These conformations are then all evaluated with the `obenergy` program. Corresponding energy is stored in the title of the resulting molecule coordinate file. The script command line arguments are the mandatory input filename (-i *name*), the number of generations for the genetic algorithm search (-g [default: 100]), the energy cutoff used to discriminate conformations in units of kcal/mol (-e [default: 50]), the output directory name (-d [default: use the prefix of the input filename] and the format of the outputted molecules (-f [xyz or default: mol2]).

ConfBuster-Macrocycle-Linear-Sampling.py. This is the main script of the tool suite that performs the macrocycle conformational search. The search is achieved by cleaving the macrocycle in linear molecules (**Figure 2A**). For each cleavable bond (any single bond between two atoms that are not chiral centres), a conformational search is performed. First, the bond is removed and hydrogens are adjusted on the two terminal atoms using PyMOL. This linear molecule is then sampled *n* times to identify *n* low
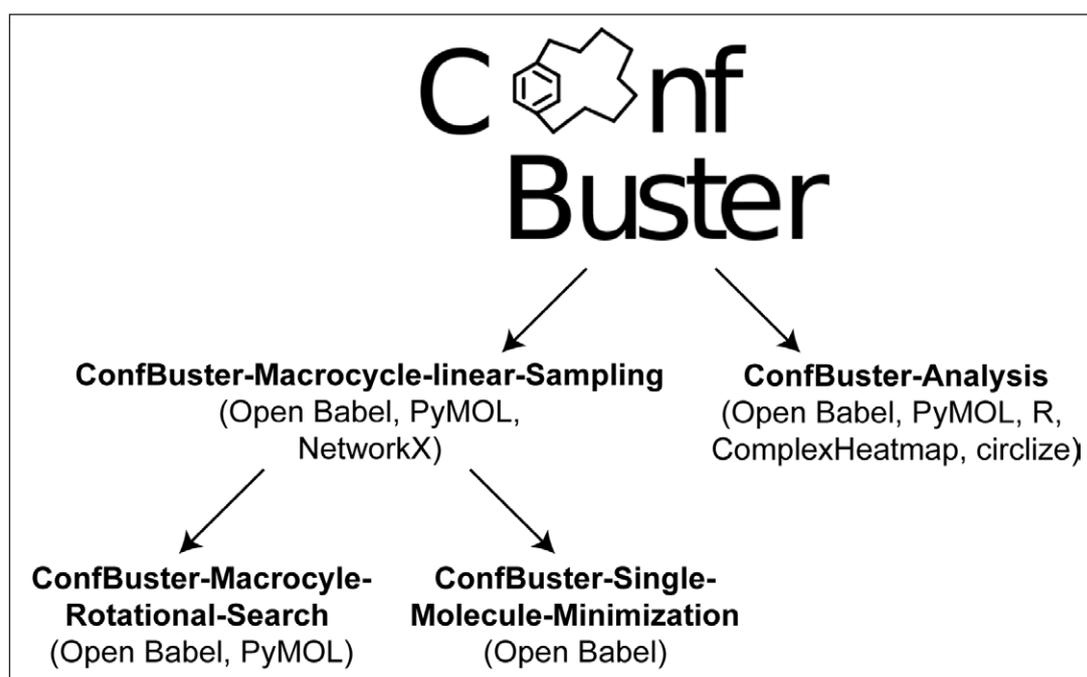
Figure 1: Hierarcherical relation between the ConfBuster scripts. The third-party program and package dependencies of each script are indicated between parenthesis.

energy conformations. For each of these samplings, new conformations are generated from systematic rotations of all the dihedrals angles and all possible pairs of dihedral angles. *N* clash-free conformations are selected, from the shortest to the longest distance between the cleaved atoms, to be cyclized back and minimized (**Figure 2B**), resulting in a total of *n*N* cyclized conformations per cleavable bond. The script command line arguments are the mandatory input filename (-i *name*), the root mean square deviation cutoff in units of Å (-r [default: 0.5]), the number of rotamer searches for each cleavable bond (-n [default: 5]), the number of conformations kept from each rotamer search (-N [default: 5]) and the mandatory output directory name (-o *name* [default: prefix of the input filename]).

ConfBuster-Analysis.py. In this optional step, the RMSD values between all the conformations are calculated. The RMSD- and energy-based classifications are listed in text files and the hierarchical clustering of the conformations based on the Euclidean distances between the RMSD values is plotted in a tree and a 2D matrix, and the energy of each conformation is plotted using a colour scale (see **Figure 2D**). This publication-quality figure, stored in a PDF file, allows the quick identification of the lowest energy conformation as well as an RMSD clustering of the best-energy conformations from the search. The script command line arguments are the mandatory directory name of the search results (-i *name*), the root mean square deviation cutoff in units of Å (-r [default: 0.5]), the number of conformations to include in the analysis (-n [default: use all the conformations present in the directory]) and the mid-point value of the energy color scale in units of kcal/mol (-e [default: 0]).

Results. **Figure 2B** and **D** presents the search results for the macrocycle sopharen A from PDB 1W96 [22]. **Figure 2B** exposes the 73 conformations identified from the search, **Figure 2C** presents the lowest energy conformation aligned to the reference structure from the PDB file (the RMSD between the two conformations is 0.405 Å), while **Figure 2D** displays the RMSD hierarchical clustering and the energy-based classification of the conformations identified from the search. This last figure allows the identification of the lowest energy conformation as well as to evaluate the relation between the conformations. Several examples of a molecular conformational search are provided with the ConfBuster distribution. In these examples, the RMSD values between the best search results and their respective reference conformations range between 0.010 Å (PDB 3R92 [23]) and 2.728 Å (PDB 3MT6 [24]).
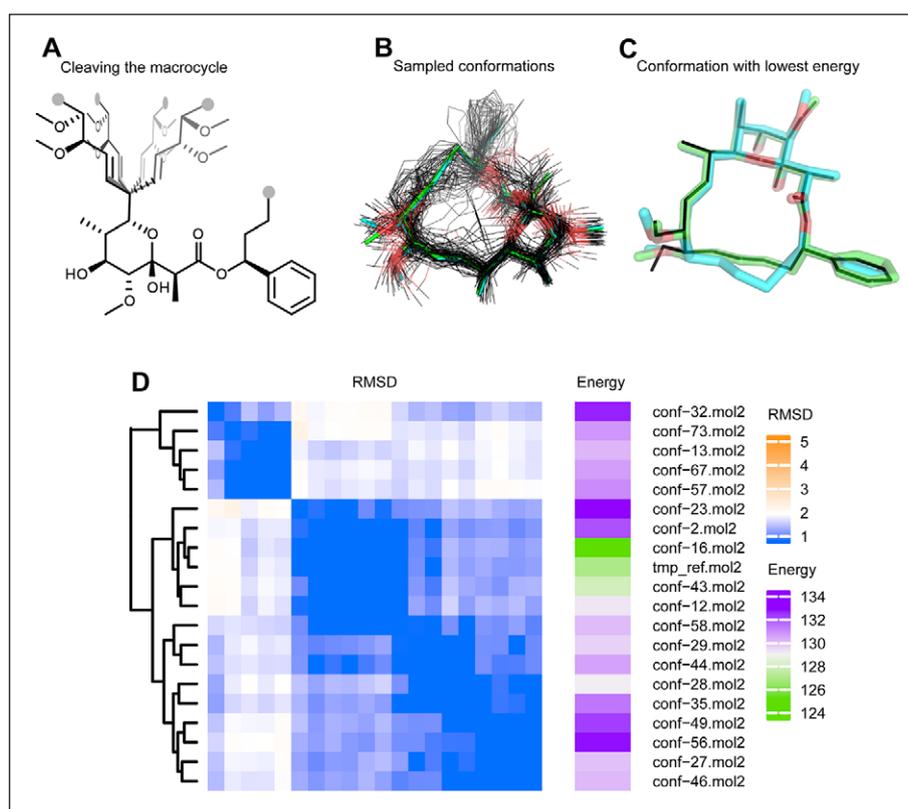
Figure 2: Overview of the ConfBuster conformational search process, illustrated for the macrocycle sopharen A (PDB 1W96). **A)** The macrocycle is cleaved for the conformational search (only one cleavage is presented for the sake of simplicity, in practice all the single bonds between two atoms that are not chiral centers are cleaved during the search), **B)** The results include several different conformations, which allows to evaluate the conformational space of the macrocycle, **C)** The lowest energy conformation identified from the search superposed to the experimental structure, and **D)** Classification of the conformations identified from the search based on the RMSD clustering and using an energy color scale.

Quality control

ConfBuster installation, dependencies and running instructions are provided in details with the distribution on the Github repository. Additionally, several examples including all the command lines and required files to run macrocycle conformational searches are also included with the distribution, in the *examples* folder and in the *examples/Instructions.pdf* file. As for the sopharen A example above, the macrocycle molecule was extracted from the PDB 1W96, and the molecule was validated for the correct bond orders and the hydrogens were added and saved in the file examples/1w96/macro-1w96.pdb. Then, a minimisation was performed, using the following command in a terminal window:

```
$ ConfBuster-Single-Molecule-Minimization.py -i
macro-1w96.pdb
```

This was followed by a macrocyclic conformational search:

```
$ ConfBuster-Macrocycle-Linear-Sampling.py -i
macro-1w96.mol2 -n 5 -N 5 -r 0.5
```

The progress of the search may be monitored using the `run` command in PyMOL:

```
(in PyMOL) run Follow-macro-1w96.py
```

Finally, the analysis of the search results was performed as follows:

```
$ ConfBuster-Analysis.py -i macro-1w96 -R
macro-1w96.mol2 -n 20
```

which built the *Heatmap_20.pdf* file with the content presented in **Figure 2D**.

Therefore, the users can assess that their installation is working properly by cross-checking their results against those presents in the distribution. Further, the users can also validate the results obtained using other molecules against experimental results whenever available. However, as the searches involve random parameters and different molecular complexity, the results might be slightly different than those included in the code distribution or from the experimental values. Running multiple conformational searches may increase the liability of the results. In any case, support is available and provided via GitHub Issues.

(2) Availability

Operating system

ConfBuster is able to function on any operating system that supports standard Python and R installations and the dependent packages, which includes Linux, Windows and macOS (tested on Linux Ubuntu 14.04 LTS).

Programming languages

Python == 2.7 and (optional) R ≥ 3.0.0 (tested with Python 2.7.6 and R 3.4.1).

Additional system requirements

There is no special system requirement. However, hardware requirements in terms of processor power, memory capacity, etc. depend primarily on the complexity of the molecules that are processed.

Dependencies

The following software is a required dependency for all the ConfBuster scripts:

Open Babel == 2.4.1 (will not work with older or more recent versions, tested with Open Babel version 2.4.1).

The following Python package is a required dependency for ConfBuster-Macrocycle-Linear-Sampling.py:
NetworkX (tested with version 1.11).

The following software is a required dependency for ConfBuster-Macrocycle-Linear-Sampling.py and ConfBuster-Analysis.py, and optional for the other ConfBuster scripts:
PyMOL ≥ 1.8 (tested with version 1.8.5.0).

The following R package is a required dependency for ConfBuster-Analysis.py:
ComplexHeatmap (tested with ComplexHeatmap version 1.14.0, from Bioconductor release 3.5 (bioconductor.org)).
Circlize (tested with version 0.3.10).

List of contributors

1. Xavier Barbeau: Development.
2. Antony T. Vincent: Contributions, documentation and testing and code review.
3. Patrick Lagüe: Documentation and testing.

Software location

Archive

Name: GitHub
Persistent identifier: https://github.com/patricklague/ConfBuster/releases/tag/v1.0
Licence: GPL-3.0
Publisher: Antony T. Vincent
Version published: 1.0
Date published: 22/08/2017

Code repository

Name: GitHub
Persistent identifier: https://github.com/patricklague/ConfBuster
Licence: GPL-3.0
Date published: 22/08/2017

Language

English

(3) Reuse potential

The knowledge of the low-energy 3D conformation is fundamental to the rational design of biologically active molecules. The availability of ConfBuster to the drug developer community will provide an access to a free powerfull tool that will guide macrocycle drug design using any low-cost computer or supercomputers from national research facilities. It is important to note that, to our knowledge, ConfBuster is actually the only free conformational search tool for macrocycles. To help users that are less familiar with scientific computing, the package distribution includes the step by step installation instructions and a number of examples to demonstrate its use and the analysis of the results. The

`ConfBuster-Analysis.py` script provide tools to help the validation and publication of the conformational search results.

The ConfBuster code has been implemented with the focus on readability and modularity, simplifying the reuse of the modules independently in different projects. Further, the modularity of the package facilitates its inclusion in a potential conformational search plugin that can be implemented in molecular viewers, such as PyMOL. Finally, the package can be used as an engine of a molecular conformational search server with a graphical user interface, which would eliminate software installation and the use of command lines, that would be useful to chemists with reduced computer skills or knowledge.

Competing Interests
The authors have no competing interests to declare.

References

1. **Dunitz, J D** 1967 Perspectives in Structural Chemistry. Dunitz, J D and Ibers, J A (eds), 2. J. Wiley, New York: NY, USA.

2. **Still, W C** and **Galynker, I** 1981 Chemical consequences of conformation in macrocyclic compounds. *Tetrahedron*, 37: 3981–3996. DOI: https://doi.org/10.1016/S0040-4020(01)93273-9

3. **Ganesan, A** 2008 The impact of natural products upon modern drug discovery. *Current Opinion in Chemical Biology*, 12: 306–317. DOI: https://doi.org/10.1016/j.cbpa.2008.03.016

4. **Giordanetto, F** and **Kihlberg, J** 2014 Macrocyclic Drugs and Clinical Candidates: What Can Medicinal Chemists Learn from Their Properties? *Journal of Medicinal Chemistry*, 57: 278–295. DOI: https://doi.org/10.1021/jm400887j

5. **Marsault, E** and **Peterson, M L** 2011 Macrocycles Are Great Cycles: Applications, Opportunities, and Challenges of Synthetic Macrocycles in Drug Discovery. *Journal of Medicinal Chemistry*, 54: 1961–2004. DOI: https://doi.org/10.1021/jm1012374

6. **Heinis, C** 2014 Drug discovery: Tools and rules for macrocycles. *Nature Chemical Biology*, 10: 696–698. DOI: https://doi.org/10.1038/nchembio.1605

7. **Yudin, A K** 2015 Macrocycles: lessons from the distant past, recent developments, and future directions. *Chem. Sci.*, 6: 30–49. DOI: https://doi.org/10.1039/C4SC03089C

8. **Over, B, Matsson, P, Tyrchan, C, Artursson, P, Doak, B C, Foley, M A, Hilgendorf, C, Johnston, S E, Lee, M D, Lewis, R J, McCarren, P, Muncipinto, G, Norinder, U, Perry, M W D, Duvall, J R** and **Kihlberg, J** 2016 Structural and conformational determinants of macrocycle cell permeability. *Nature Chemical Biology*, 12: 1065–1074. DOI: https://doi.org/10.1038/nchembio.2203

9. **Leach, A** 2009 Molecular modelling: principles and applications. Pearson Prentice Hall.

10. **Chemical Computing Group ULC.** 2017.

11. **Watts, K S, Dalal, P, Tebben, A J, Cheney, D L** and **Shelley, J C** 2014 Macrocycle Conformational Sampling with MacroModel. *Journal of Chemical Information and Modeling*, 54: 2680–2696. DOI: https://doi.org/10.1021/ci5001696

12. **Chang, C-E** and **Gilson, M K** 2003 Tork: Conformational analysis method for molecules and complexes. *Journal of Computational Chemistry*, 24: 1987–1998. DOI: https://doi.org/10.1002/jcc.10325

13. **Sadowski, P** and **Baldi, P** 2013 Small-Molecule 3D Structure Prediction Using Open Crystallography Data. *Journal of Chemical Information and Modeling*, 53: 3127–3130. DOI: https://doi.org/10.1021/ci4005282

14. **Sindhikara, D, Spronk, S A, Day, T, Borrelli, K, Cheney, D L** and **Posy, S L** 2017 Improving Accuracy, Diversity, and Speed with Prime Macrocycle Conformational Sampling. *Journal of Chemical Information and Modeling*. DOI: https://doi.org/10.1021/acs.jcim.7b00052

15. **Bonnet, P, Agrafiotis, D K, Zhu, F** and **Martin, E** 2009 Conformational Analysis of Macrocycles: Finding What Common Search Methods Miss. *Journal of Chemical Information and Modeling*, 49: 2242–2259. DOI: https://doi.org/10.1021/ci900238a

16. **Cleves, A E** and **Jain, A N** 2017 ForceGen 3D structure and conformer generation: from small lead-like molecules to macrocyclic drugs. *Journal of Computer-Aided Molecular Design*, 31: 419–439. DOI: https://doi.org/10.1007/s10822-017-0015-8

17. **O'Boyle, N M, Banck, M, James, C A, Morley, C, Vandermeersch, T** and **Hutchison, G R** 2011 Open Babel: An open chemical toolbox. *Journal of Cheminformatics*, 3: 33. DOI: https://doi.org/10.1186/1758-2946-3-33

18. The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.

19. **Hagberg, A A, Schult, D A** and **Swart, P J** 2008 Exploring network structure, dynamics, and function using NetworkX. In: *Proceedings of the 7th Python in Science Conference (SciPy2008)*, 11–15. Pasadena, CA USA.

20. **Gu, Z, Eils, R** and **Schlesner, M** 2016 Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*, 32: 2847–2849. DOI: https://doi.org/10.1093/bioinformatics/btw313

21. **Gu, Z, Gu, L, Eils, R, Schlesner, M** and **Brors, B** 2014 circlize implements and enhances circular visualization in R. *Bioinformatics*, 30: 2811–2812. DOI: https://doi.org/10.1093/bioinformatics/btu393

22. **Shen, Y, Volrath, S L, Weatherly, S C, Elich, T D** and **Tong, L** 2004 A Mechanism for the Potent Inhibition of Eukaryotic Acetyl-Coenzyme A Carboxylase by Soraphen A, a Macrocyclic Polyketide Natural Product. *Molecular Cell*, 16: 881–891. DOI: https://doi.org/10.1016/j.molcel.2004.11.034

23. **Zapf, C W, Bloom, J D, McBean, J L, Dushin, R G, Golas, J M, Liu, H, Lucas, J, Boschelli, F, Vogan, E** and **Levin, J I** 2011 Discovery of a macrocyclic o-aminobenzamide Hsp90 inhibitor with heterocyclic

tether that shows extended biomarker activity and in vivo efficacy in a mouse xenograft model. *Bioorganic & Medicinal Chemistry Letters*, 21: 3627–3631. DOI: https://doi.org/10.1016/j.bmcl.2011.04.102

24. **Li, D H S, Chung, Y S, Gloyd, M, Joseph, E, Ghirlando, R, Wright, G D, Cheng, Y-Q, Maurizi, M R, Guarné, A** and **Ortega, J** 2010 Acyldepsipeptide Antibiotics Induce the Formation of a Structured Axial Channel in ClpP: A Model for the ClpX/ClpA-Bound State of ClpP. *Chemistry & Biology*, 17: 959–969. DOI: https://doi.org/10.1016/j.chembiol.2010.07.008