

MolTK



User's Manual

Release 0.3.4

December 20, 2007

Website: rotatingpenguin.com/moltk

Copyright and Permission Notice

Copyright (c) 2011 Christopher M. Bruns
Contributors: Cami K. Bruns

Permission is hereby granted, free of charge, to any person obtaining a copy of this document (the "Document"), to deal in the Document without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Document, and to permit persons to whom the Document is furnished to do so, subject to the following conditions:

This copyright and permission notice shall be included in all copies or substantial portions of the Document.

THE DOCUMENT IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS, CONTRIBUTORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE DOCUMENT OR THE USE OR OTHER DEALINGS IN THE DOCUMENT.

Acknowledgments

Special thanks to Joy Ku who created the beautiful template for this manual. Thanks also to Christopher D. Putnam and Robert Edgar for helpful discussions about sequence alignment.

Table of Contents

1	INTRODUCTION TO MOLTK	1
1.1	Overview	1
1.1.1	<i>What is MolTK?</i>	1
1.1.2	<i>Design Goals of MolTK</i>	1
1.2	Installing MolTK	2
1.2.1	<i>Installing the MolTK Application</i>	2
1.2.2	<i>Installing the MolTK Python Module</i>	2
1.2.3	<i>Installing the MolTK C++ SDK</i>	2
1.2.4	<i>Building MolTK from Source Code</i>	2
1.3	Software License	2
1.4	Resources.....	3
2	THE MOLTK APPLICATION	5
3	PROGRAMMING MOLTK	7
3.1	Python.....	7
3.1.1	<i>Tutorial</i>	7
3.1.2	<i>MolTK for SEQUOIA Users</i>	7
3.2	C++	7
3.2.1	<i>MolTK Coding Style Guidelines</i>	7
4	BIBLIOGRAPHY	9

List of Tables

No table of figures entries found.

List of Figures

No table of figures entries found.

List of Examples

1 Introduction to MolTK

1.1 Overview

1.1.1 What is MolTK?

There are three faces of MolTK; in order from easiest-to-use to most-powerful:

1. MolTK is a computer application that allows the user to view and align molecular sequences and structures.
2. Moltk is a Python programming language module that allows the user to align and otherwise compute on molecular sequences and structures.
3. Moltk is a C++ API that allow the programmer to compute with molecular sequences and structures with high efficiency.

1.1.2 Design Goals of MolTK

- Python programming language interactive environment for sequence/structure alignment that is just as easy to use as our (aging) dedicated alignment tool SEQUOIA.
- Flexible architecture that makes it easy to experiment with custom alignment methods and scoring systems.
- Consistent, well documented API for both Python and C++ programmers.
- Units-aware quantity type system. Thus alignment scores are not just numbers, they are information quantities with units of "bits". Atomic coordinates are not just numbers x, y, z, but are vector quantities with units of nanometers. Units-aware quantity types are an important part of scientific computing hygiene:
 - Converting a "quantity" to a raw number requires a "unit" to express the quantity in. This requires the user to pay attention to units at precisely the

moment when knowing the unit is most important. The rest of the time is "just works", even if you are wrong about what the current units are!

- This sort of type safety in scientific computing might help prevent errors such as, say, crashing \$100 million orbiters into planets.
- Type safety: Adding a length to a volume makes no sense, and will result in an error. Dividing a length by a time results in a velocity.
- Next-generation molecular sequence/structure viewer that adheres to our user interface principles.

1.2 Installing MolTK

1.2.1 Installing the MolTK Application

1.2.2 Installing the MolTK Python Module

1.2.3 Installing the MolTK C++ SDK

1.2.4 Building MolTK from Source Code

1.3 Software License

Copyright (C) 2011 Christopher M. Bruns

This program is free software; you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation; either version 2 of the License, or (at your option) any later version.

This program is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License along with this program; if not, write to the Free Software Foundation, Inc., 51 Franklin Street, Fifth Floor, Boston, MA 02110-1301 USA.

Commercial users should ask about our dual licensing model.

For questions contact: cmbruns@rotatingpenguin.com

1.4 Resources

Browse the MolTK home page at:

<http://rotatingpenguin.com/moltk/>

Download MolTK at:

<http://code.google.com/p/moltk/downloads/list>

Find questions and answers in the moltk-users forum at:

<http://groups.google.com/group/moltk-users>

Report issues and get the source code at:

<http://code.google.com/p/moltk/>

Study the MolTK Python API at:

http://www.rotatingpenguin.com/moltk/api_python/python_api.html

Study the MolTK C++ API at:

http://www.rotatingpenguin.com/moltk/api_cxx/cxx_api.html

2 The MolTK Application

3 Programming MolTK

3.1 Python

3.1.1 Tutorial

3.1.2 MolTK for SEQUOIA Users

3.2 C++

3.2.1 MolTK Coding Style Guidelines

Purpose: To provide a consistent look for both python and C++ MolTK code.

3.2.1.1 Indentation

Use four space characters per indent. No tabs.

3.2.1.2 Class names

Use CapitalizedWords

Use all caps for abbreviations, e.g. PDBStructure.

Class names should be nouns.

3.2.1.3 Macro, enum, and constant names

CAPITALIZED_WITH_UNDERSCORES

3.2.1.4 Method and function names

Use lower_case_with_underscores.

Begin method and function names with a verb.

3.2.1.5 Attribute and member names

lower_case_with_underscores

adjective_noun or noun

b_variable_name for boolean values

3.2.1.6 Namespace, package, and module names

lower case, single word, short names

3.2.1.7 Method parameters

lower_case_with_underscores

Remember that method parameter names are more important in python than in C++, because python allows named parameter use. In all C++ header files, every parameter to every exposed method should have an understandable name.

4 Bibliography

Dayhoff, M. O., 1969. *Atlas of Protein Sequence and Structure*. Silver Spring, MD: National Biomedical Research Foundation.

Edgar, R. C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32(5), pp. 1792-1797.

Gusfield, D., 1997. *Algorithms on Strings, Trees, and Sequences*. Cambridge: Cambridge University Press.

Henikoff, S. & Henikoff, J. G., 1992. Amino acid substitution matrices from protein blocks. *Proceedings of the National Academy of Sciences*, Volume 89, pp. 10915-10919.

Needleman, S. & Wunsch, C. D., 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology*, Volume 48, pp. 443-453.