MOPAC: An open-source semiempirical molecular orbital program

Jonathan E. Moussa [1][¶] **and James J. P. Stewart**[2]

1 Molecular Sciences Software Institute, Virginia Tech, Blacksburg, VA 24060, United States ROR **2** Stewart Computational Chemistry, Colorado Springs, CO 80921, United States ¶ Corresponding author

Summary

The Molecular Orbital PACkage (MOPAC) is a Fortran program that calculates chemical and physical properties of molecules, crystals, and nanostructures. MOPAC primarily operates as a command-line program that takes an input file defining a molecule by its approximate atomic coordinates and produces an output file containing the molecule's heat of formation and optimized coordinates alongside other useful properties. Keywords can then be added to the top of the input file to adjust MOPAC's behavior and request the use of other features. MOPAC is similar in function to *ab initio* quantum chemistry software, but it uses semiempirical models such as PM7 (Stewart, 2012) to reduce computational costs substantially while preserving as much accuracy as possible.

Statement of need

In the broader context of chemistry, physics, and materials science, the computer simulation of atomistic systems is intended as a partner to experiment – to help interpret experimental results, to extend their reach beyond what can be observed directly, and even to replace experiments altogether in some cases. Much of the rationale for this is motivated by the reduction of costs, that some physical properties ought to be fundamentally cheaper to understand and predict through computer simulations rather than by experiment. However, an unfortunate reality is that *ab initio* quantum chemistry has extremely high computational costs that grow very rapidly with molecular size, which has long been an impediment to this partnership with experiment. Historically, semiempirical models of thermochemistry were developed to mitigate these costs while preserving enough quantum mechanical structure to retain a high degree of model transferability. Semiempirical models of thermochemistry were pioneered by Pople (Pople, 1953; Pople et al., 1965) and later championed by Dewar (Dewar, 1975), culminating in the development of the Modified Neglect of Diatomic Overlap (MNDO) model form (Dewar & Thiel, 1977), the Austin Model 1 (AM1) parameterization (Dewar et al., 1985), and the MOPAC program (Stewart, 1983). Historically, MOPAC has been the development platform for the MNDO-family semiempirical models, but these models are also implemented in other software such as Gaussian (Frisch et al., 2016), CP2K (Kühne et al., 2020), Sparrow (Bosia et al., 2023), and ULYSSES (Menezes & Popowicz, 2022).

In the four decades since MOPAC was first released, the computing power of personal computers has increased a million fold. As a result, the practical cost considerations of quantum chemistry have changed significantly. Practical *ab initio* quantum chemistry calculations were once limited to Hartree-Fock calculations in a minimal atomic-orbital basis set and small molecules with a few atoms, but now density functional theory (DFT) calculations in moderately-sized basis sets (split valence and polarization) of large molecules with hundreds of atoms are routine. Relative to such routine DFT calculations, semiempirical calculations with MOPAC are roughly

a thousand times faster but half as accurate. Because of its distinct balance between cost and accuracy, semiempirical quantum chemistry software such as MOPAC is well-suited for simulation tasks that are highly sensitive to cost such as interactive simulations for chemical exploration and education (Weymuth & Reiher, 2021) and high-throughput virtual screening of molecules (Pyzer-Knapp et al., 2015). More generally, it is useful to perform calculations at a semiempirical level to estimate results and check for problems before committing to a much more expensive calculation at an *ab initio* level. For some expensive applications such as protein modeling, only the semiempirical level may be affordable to quantum chemistry users who do not have access to supercomputers. Semiempirical quantum mechanical models continue to be a middle ground in atomistic simulation between *ab initio* quantum mechanics and molecular mechanics based on classical force fields (Moussa, 2025b).

Features

The original 1990 software paper for MOPAC (Stewart, 1990) contains a detailed description of its features and architecture. At the time of its publication, MOPAC was public-domain software that was distributed by the Quantum Chemistry Program Exchange (QCPE) (Boyd, 2013). In 1993, MOPAC was acquired by Fujitsu Limited and became commercial, closed-source software until its open-source release in 2022. During its nearly three decades as commercial software and its recent open-source existence, many new features have been developed for MOPAC. While it was originally developed for thermochemistry calculations of organic molecules in vacuum containing chemical elements without valence d orbitals, it has since been extended to solids (Stewart, 2000) and molecules in solution (Klamt & Schüürmann, 1993), most elements of the periodic table (Stewart, 2007), and more physical properties such as electronic spectroscopy (Gieseking, 2021).

A major focus of the last two decades of MOPAC development has been the modeling of proteins and enzymes. Proteins usually contain hundreds or thousands of atoms, which are challenging to study with quantum chemistry software because of their rapid growth of costs with molecular size. The MOZYME solver (Stewart, 1996) can perform calculations in a localized molecular orbital (LMO) basis, which has costs that grow only linearly with molecular size. Most protein work is based on structures deposited in the Protein Data Bank (PDB) (Burley et al., 2017), and MOPAC can handle PDB structure files and add missing hydrogen atoms that are not reliably resolved by X-ray crystallography. To study proton transfer reactions in enzymes, MOZYME is able to locate transition states near an active site within a larger enzyme structure (Stewart, 2017). While most of the semiempirical models in MOPAC are intended for general use, it also has a model that is specifically optimized for biomolecular applications such as proteins and enzymes (Stewart & Stewart, 2023).

As an open-source project, the focus of MOPAC development has shifted to increasing its accessibility for important use cases. For example, high-throughput applications of MOPAC have been limited by its need for disk-based input and output on computers with a large number of compute cores but a slow file system. Since the 23.0.0 release of MOPAC, the most commonly used functionality of MOPAC is now available through an application programming interface (API) that avoids all disk usage and captures all essential state information from MOPAC in the input and output data of its API calls. This API has C bindings to avoid the application binary interface (ABI) incompatibility problems of Fortran, which allows it to serve as a base API layer for a Python API wrapper (Moussa, 2025a) and increases the accessibility of MOPAC to the Python software ecosystem.

Acknowledgements

References

Bosia, F., Zheng, P., Vaucher, A., Weymuth, T., Dral, P. O., & Reiher, M. (2023). Ultra-fast semi-empirical quantum chemistry for high-throughput computational campaigns with sparrow. *The Journal of Chemical Physics*, *158*(5), 054118. https://doi.org/10.1063/5.0136404

Boyd, D. B. (2013). Quantum chemistry program exchange, facilitator of theoretical and computational chemistry in pre-internet history. In *Pioneers of quantum chemistry* (pp. 221–273). American Chemical Society. https://doi.org/10.1021/bk-2013-1122.ch008

Burley, S. K., Berman, H. M., Kleywegt, G. J., Markley, J. L., Nakamura, H., & Velankar, S. (2017). Protein data bank (PDB): The single global macromolecular structure archive. In *Protein crystallography: Methods and protocols* (pp. 627–641). Springer New York. https://doi.org/10.1007/978-1-4939-7000-1_26

Dewar, M. J. S. (1975). Quantum organic chemistry. *Science*, *187*, 1037–1044. https://doi.org/10.1126/science.187.4181.1037

Dewar, M. J. S., & Thiel, W. (1977). Ground states of molecules. 38. The MNDO method. Approximations and parameters. *Journal of the American Chemical Society*, *99*, 4899–4907. https://doi.org/10.1021/ja00457a004

Dewar, M. J. S., Zoebisch, E. G., Healy, E. F., & Stewart, J. J. P. (1985). Development and use of quantum mechanical molecular models. 76. AM1: A new general purpose quantum mechanical molecular model. *Journal of the American Chemical Society*, *107*, 3902–3909. https://doi.org/10.1021/ja00299a024

Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Scalmani, G., Barone, V., Petersson, G. A., Nakatsuji, H., Li, X., Caricato, M., Marenich, A. V., Bloino, J., Janesko, B. G., Gomperts, R., Mennucci, B., Hratchian, H. P., Ortiz, J. V., … Fox, D. J. (2016). *Gaussian 16 rev. c.01*.

Gieseking, R. L. M. (2021). A new release of MOPAC incorporating the INDO/s semiempirical model with CI excited states. *Journal of Computational Chemistry*, *42*, 365–378. https://doi.org/10.1002/jcc.26455

Klamt, A., & Schüürmann, G. (1993). COSMO: A new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *Journal of the Chemical Society, Perkin Transactions 2*, 799–805. https://doi.org/10.1039/P29930000799

Kühne, T. D., Iannuzzi, M., Del Ben, M., Rybkin, V. V., Seewald, P., Stein, F., Laino, T., Khaliullin, R. Z., Schütt, O., Schiffmann, F., Golze, D., Wilhelm, J., Chulkov, S., Bani-Hashemian, M. H., Weber, V., Borštnik, U., Taillefumier, M., Jakobovits, A. S., Lazzaro, A., … Hutter, J. (2020). CP2K: An electronic structure and molecular dynamics software package - quickstep: Efficient and accurate electronic structure calculations. *The Journal of Chemical Physics*, *152*(19), 194103. https://doi.org/10.1063/5.0007045

Menezes, F., & Popowicz, G. M. (2022). ULYSSES: An efficient and easy to use semiempirical library for c++. *Journal of Chemical Information and Modeling*, *62*(16), 3685–3694. https://doi.org/10.1021/acs.jcim.2c00757

Moussa, J. E. (2025a). Mopactools. In *GitHub repository*. GitHub. https://github.com/openmopac/mopactools

Moussa, J. E. (2025b). The enduring relevance of semiempirical quantum mechanics. *The Journal of Physical Chemistry A*, *129*, 8465–8477. https://doi.org/10.1021/acs.jpca.5c03425

Pople, J. A. (1953). Electron interaction in unsaturated hydrocarbons. *Transactions of the Faraday Society*, *49*, 1375–1385. https://doi.org/10.1039/TF9534901375

Pople, J. A., Santry, D. P., & Segal, G. A. (1965). Approximate self-consistent molecular orbital theory. I. Invariant procedures. *The Journal of Chemical Physics*, *43*, S129–S135. https://doi.org/10.1063/1.1701475

Pyzer-Knapp, E. O., Suh, C., Gómez-Bombarelli, R., Aguilera-Iparraguirre, J., & Aspuru-Guzik, A. (2015). What is high-throughput virtual screening? A perspective from organic materials discovery. *Annual Review of Materials Research*, *45*, 195–216. https://doi.org/10.1146/annurev-matsci-070214-020823

Stewart, J. J. P. (1983). MOPAC, QCPE program #455. *Quantum Chemistry Program Exchange Bulletin*, *3*, 43.

Stewart, J. J. P. (1990). MOPAC: A semiempirical molecular orbital program. *Journal of Computer-Aided Molecular Design*, *4*, 1–103. https://doi.org/10.1007/BF00128336

Stewart, J. J. P. (1996). Application of localized molecular orbitals to the solution of semiempirical self-consistent field equations. *International Journal of Quantum Chemistry*, *58*, 133–146. https://doi.org/10.1002/(SICI)1097-461X(1996)58:2%3C133::AID-QUA2%3E3.0.CO;2-Z

Stewart, J. J. P. (2000). A practical method for modeling solids using semiempirical methods. *Journal of Molecular Structure*, *556*, 59–67. https://doi.org/10.1016/S0022-2860(00)00651-7

Stewart, J. J. P. (2007). Optimization of parameters for semiempirical methods v: Modification of NDDO approximations and application to 70 elements. *Journal of Molecular Modeling*, *13*, 1173–1213. https://doi.org/10.1007/s00894-007-0233-4

Stewart, J. J. P. (2012). Optimization of parameters for semiempirical methods VI: More modifications to the NDDO approximations and re-optimization of parameters. *Journal of Molecular Modeling*, *19*, 1–32. https://doi.org/10.1007/s00894-012-1667-x

Stewart, J. J. P. (2017). An investigation into the applicability of the semiempirical method PM7 for modeling the catalytic mechanism in the enzyme chymotrypsin. *Journal of Molecular Modeling*, *23*, 154. https://doi.org/10.1007/s00894-017-3326-8

Stewart, J. J. P., & Stewart, A. C. (2023). A semiempirical method optimized for modeling proteins. *Journal of Molecular Modeling*, *29*, 284. https://doi.org/10.1007/s00894-023-05695-1

Weymuth, T., & Reiher, M. (2021). Immersive interactive quantum mechanics for teaching and learning chemistry. *Chimia*, *75*, 45–49. https://doi.org/10.2533/chimia.2021.45