

Predicting Differential Ion Mobility Behaviour *in silico* using Machine Learning

Supporting Information

Christian Ieritano,^{1,2} J. Larry Campbell,^{1,3,4} and W. Scott Hopkins^{1,2,3,5*}

1. Department of Chemistry, University of Waterloo, 200 University Avenue West, Waterloo, Ontario, N2L 3G1, Canada. * E-mail: shopkins@uwaterloo.ca

2. Waterloo Institute for Nanotechnology, University of 200 University Avenue West, Waterloo, Ontario, N2L 3G1, Canada

3. WaterMine Innovation, Inc., Waterloo, Ontario, N0B 2T0, Canada

4. Bedrock Scientific Inc., Milton, Ontario, L6T 6J9, Canada

5. Centre for Eye and Vision Research, Hong Kong Science Park, New Territories, 999077, Hong Kong.

Keywords: ion mobility spectrometry, differential mobility spectrometry, FAIMS, Random Forest Regression

Table of Contents

Supporting Figures and Tables	S3
Supplementary Section S1 – CCS calculation workflow	S8
Supplementary Section S2 – Dispersion plots from a validation set composed of Type A, B, and C ions	S10
References	S31

Supporting Figures and Tables

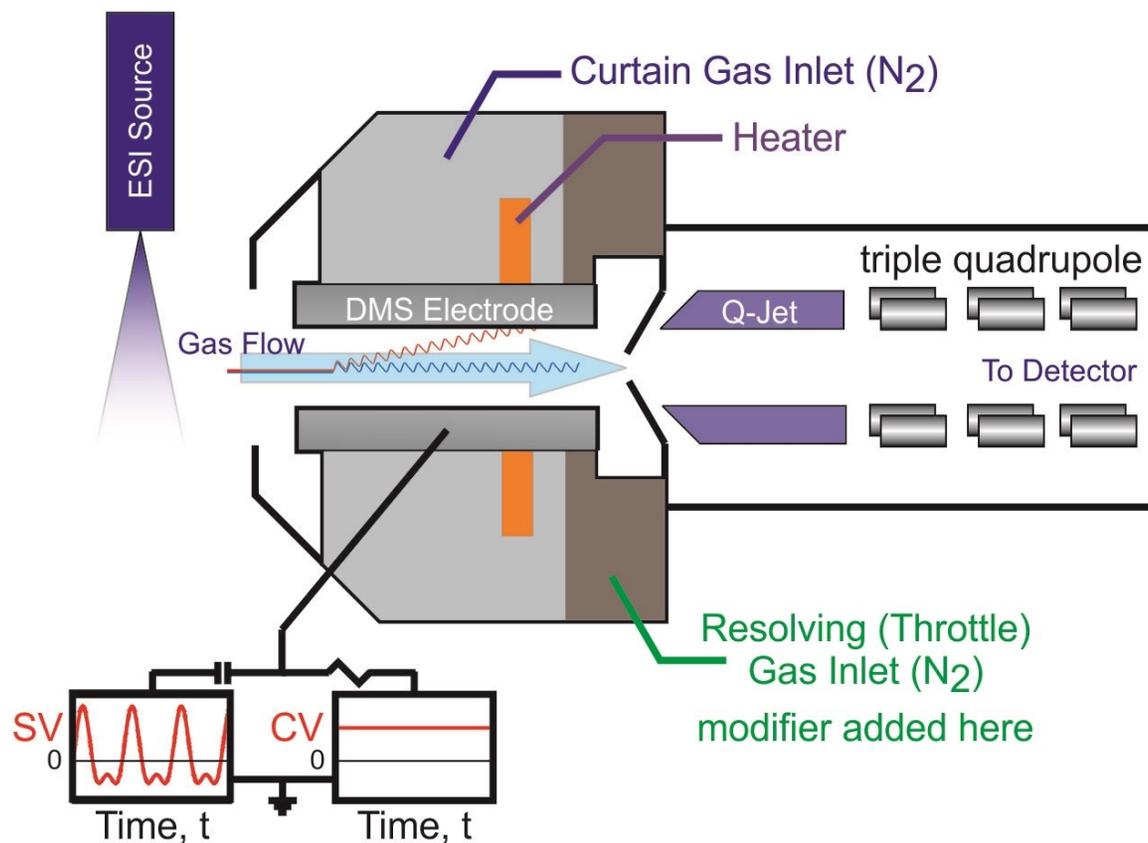


Figure S1. The basic layout of a differential mobility spectrometry (DMS) cell consists of two-planar electrodes separated by a 1 mm gap. Ions are introduced to the DMS cell by ESI and carried through the device with the flow of the N₂ carrier gas. Application of the oscillating separation field, denoted the separation voltage (SV), induces off-axis motion of the analytes that is determined by their field dependent mobility. Ion trajectories are stabilized by application of a static direct-current compensation voltage (CV). The magnitude of the CV applied is correlated with the differential mobility of the ion, enabling spatial resolution of analytes as they are directed towards the exit orifice of the DMS cell to the mass spectrometer. Figure adapted from Ref. 1.

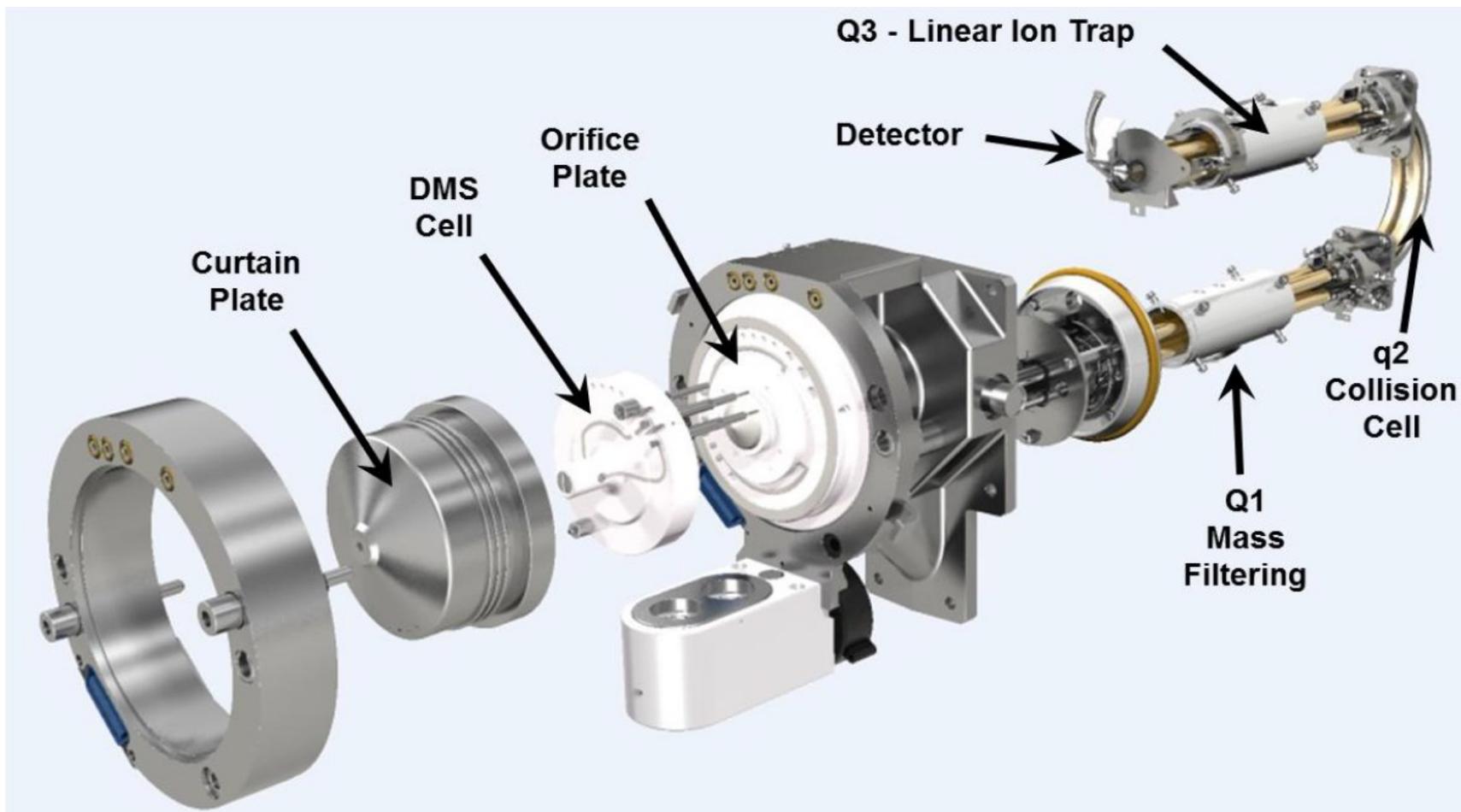


Figure S2. SelexION system coupled to the QTRAP 5500 (SCIEX) hybrid linear ion trap triple-quadrupole mass spectrometer.

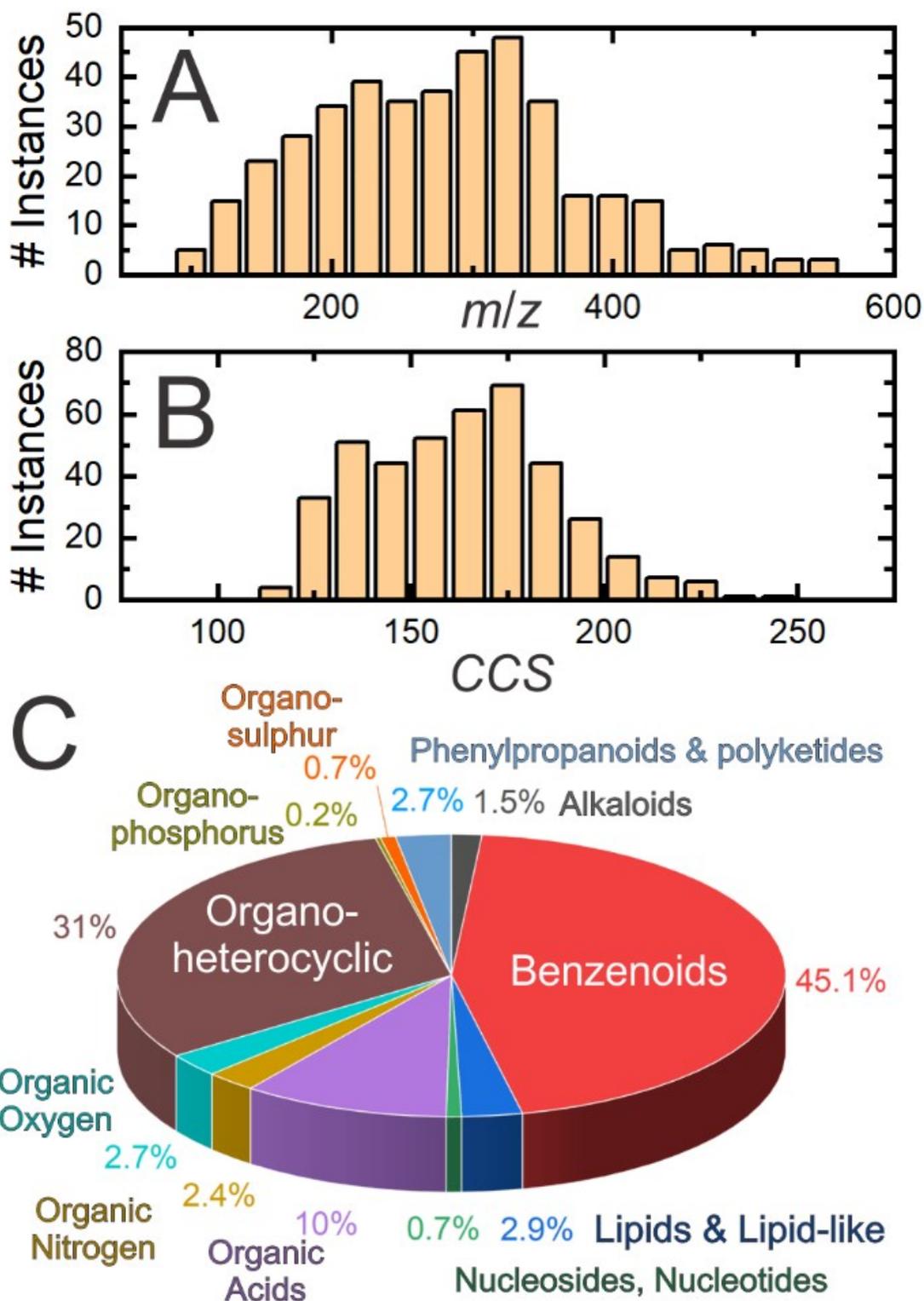


Figure S3. (A) m/z , (B) collision cross section, and (C) molecular superclasses as identified by Classyfire² distributions for the 409 compounds used in this study.

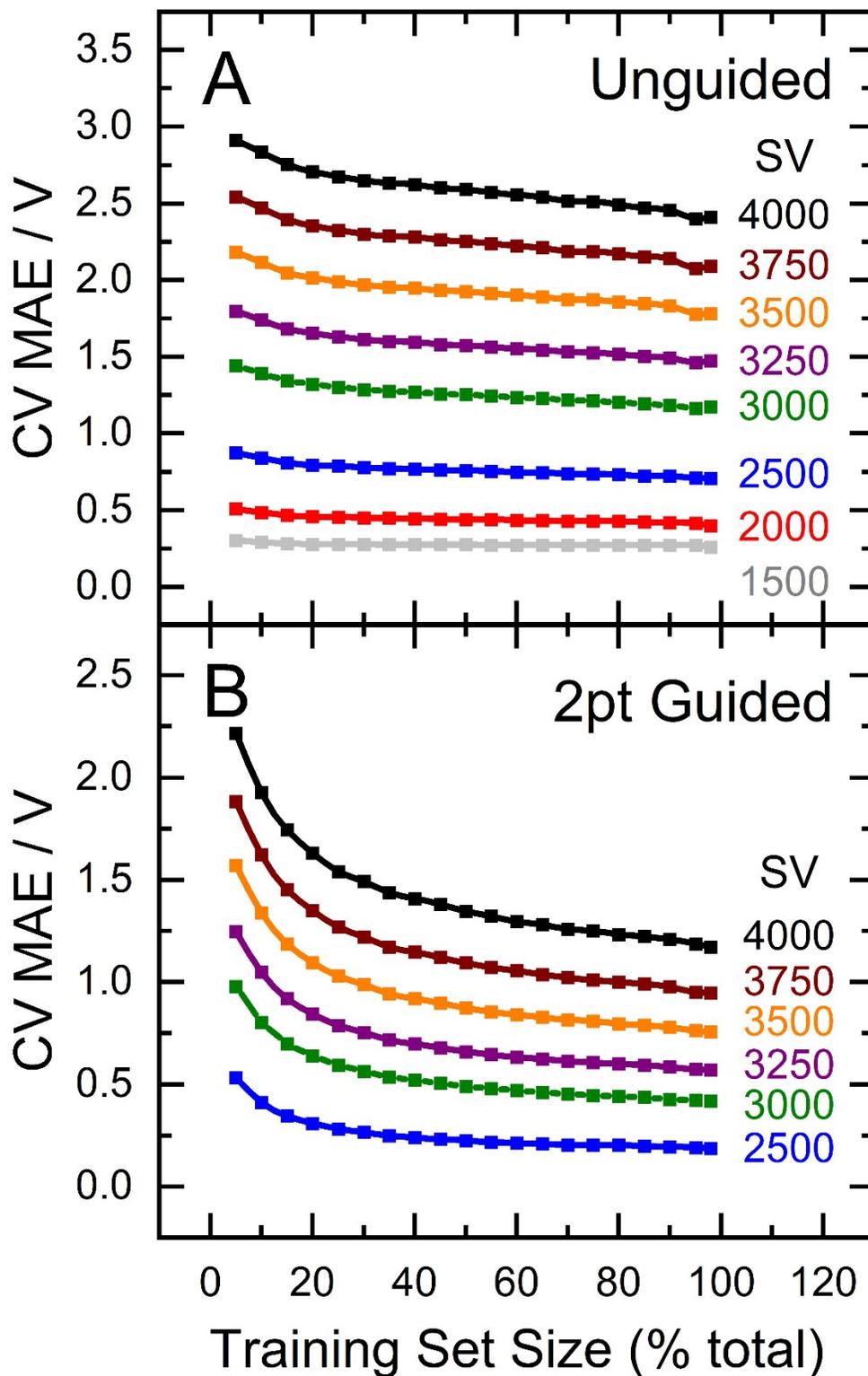


Figure S4. Learning curves for the prediction of CV and each SV sampled using the unguided (A) and 2 point guided (B) random-forest approach.

Table S1. DMS-MS experimental parameters as controlled using Analyst 1.7. N₂ was used as source gas, the curtain gas in the DMS cell, and as the collision gas for data acquisition in MRM mode. MRM transitions are provided in the Excel sheet accompanying this manuscript.

Parameter	Setting
Source	
Curtain Gas (CUR)	20.0
Ion spray voltage (IS)	5500
Source Temperature (TEM)	50
Nebulizing Gas Pressure (GS1)	30.0
Auxiliary Gas Pressure (GS2)	10.0
Compound	
Declustering Potential (DP)	100
Entrance Potential (EP)	10
Collision Energy (CE)	Varies
Collision Cell Exit Potential (CXP)	15
DMS	
DMS Temperature (DT)	Low
Modifier (MD)	None
Separation Voltage (SV)	Varies
Compensation Voltage	Varies
DMS Offset (DMO)	-3.0
DMS Resolution Enhancement (DR)	Off
Resolution – Quad 1	
Ion Energy 1 (IE1)	1.1
Q1 Resolution	Unit
Resolution – Quad 3	
Ion Energy 3 (IE3)	1.8
Q1 Resolution	Unit
Detector	
CEM (CEM)	2100.0

Supplementary Section S1 – CCS calculation workflow

All ions used in this study were products by electrospray ionization in positive mode. Thus, computational investigations of the ions considered in this study were either protonated molecules (*i.e.*, $[M + H]^+$) or those with permanent positive charges (*e.g.*, choline). All possible protonation sites were explored if a molecule exhibited more than one gas-phase basic site. In cases where protonation sites or molecular conformations could not be immediately obvious based on chemical intuition, the potential energy surface (PES) of the molecular ion was mapped using a basin-hopping (BH)^{3–7} algorithm interfaced with Gaussian 16 (Version C.01).⁸

Optimizations in the BH search were conducted using the Universal Force Field (UFF),⁹ which was supplemented with partial charges for a ‘guess structure’ of the parent ion as calculated at the B3LYP-D3/6-31+G(d,p) level of theory.^{10,11} Atomic partial charges were generated based on the electrostatic potential of the ion according to the Merz–Singh–Kollman (MK) partition scheme.^{12,13} Conformational searches within the BH code were done by randomly distorting all rotatable dihedral angles by $-10^\circ \leq \Phi \leq 10^\circ$. In total, 5000 to 20000 structures were sampled depending on the size of the ion in question since larger molecules require greater exploration of the PES to capture all relevant local minima.

Typically, the BH routine would identify 5 – 100 low-energy conformers ($< 75 \text{ kJ mol}^{-1}$) for each protonated isomer. Candidate structures were then carried forward for pre-optimization at the semi-empirical PM7 level of theory,^{14,15} and subsequently sorted based on energy and cosine similarities. For details on cosine similarities, see ref. ¹⁶. Unique conformers within 50 kJ mol^{-1} of the lowest energy PM7 geometry were carried forward for full optimization using density functional theory (DFT) at the B3LYP-D3/6-31++G(d,p) level of theory. Normal mode analyses were conducted to

verify that each structure corresponded to a minimum on the PES and to calculate the thermochemistry (enthalpy, entropy, and Gibbs energy) for each species.

The MobCal-MPI¹⁷ code was used to calculate ion-neutral CCS in N₂ (Ω_{N_2}) for all DFT isomers. All CCS calculations involved 10 complete cycles of mobility calculations that used 48 points of velocity integration and 512 points of impact parameter integration. Calculated Ω_{N_2} values are reported as average values with statistical errors assessed from the 10 cycles of calculation. For ions exhibiting multiple low-energy conformations or prototropic isomers, a Boltzmann-weighted CCS is reported based on the standard Gibbs corrected energies (T = 298.15 K) determined by DFT. The total error for a corresponding Boltzmann-weighted CCS was calculated from the standard errors for the Boltzmann-weighted CCSs of the low-energy isomers. Cartesian coordinates of each isomer alongside calculated thermochemical data, pertinent MobCal-MPI files, and calculated CCSs are available in the dedicated section with the supporting information.

Supplementary Section S2 – Dispersion plots from a validation set composed of Type A, B, and C ions

The following dispersion curves correspond to the 69th randomly generated validation set produced by training the random-forest framework with 95% of the dataset. Full details and outputs all other validation sets are provided with the GitHub release at <https://github.com/HopkinsLaboratory>. Predicted and experimental dispersion curves for the 69th set are shown below in Figures S5 – S25.

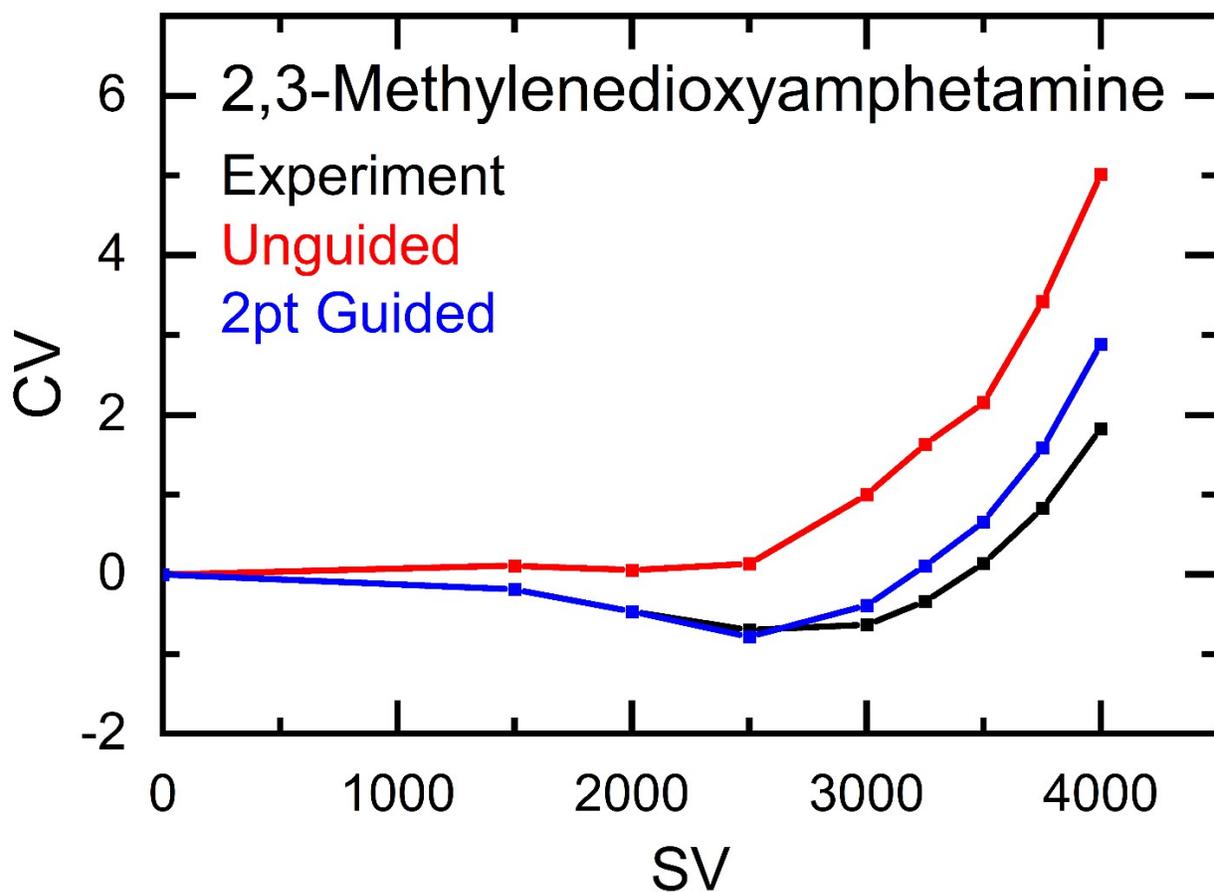


Figure S5. Dispersion curves of 2,3-methylenedioxyamphetamine as predicted by the unguided (red) and two-point guided (blue) ML-approaches compared to the experimental dispersion curve (black).

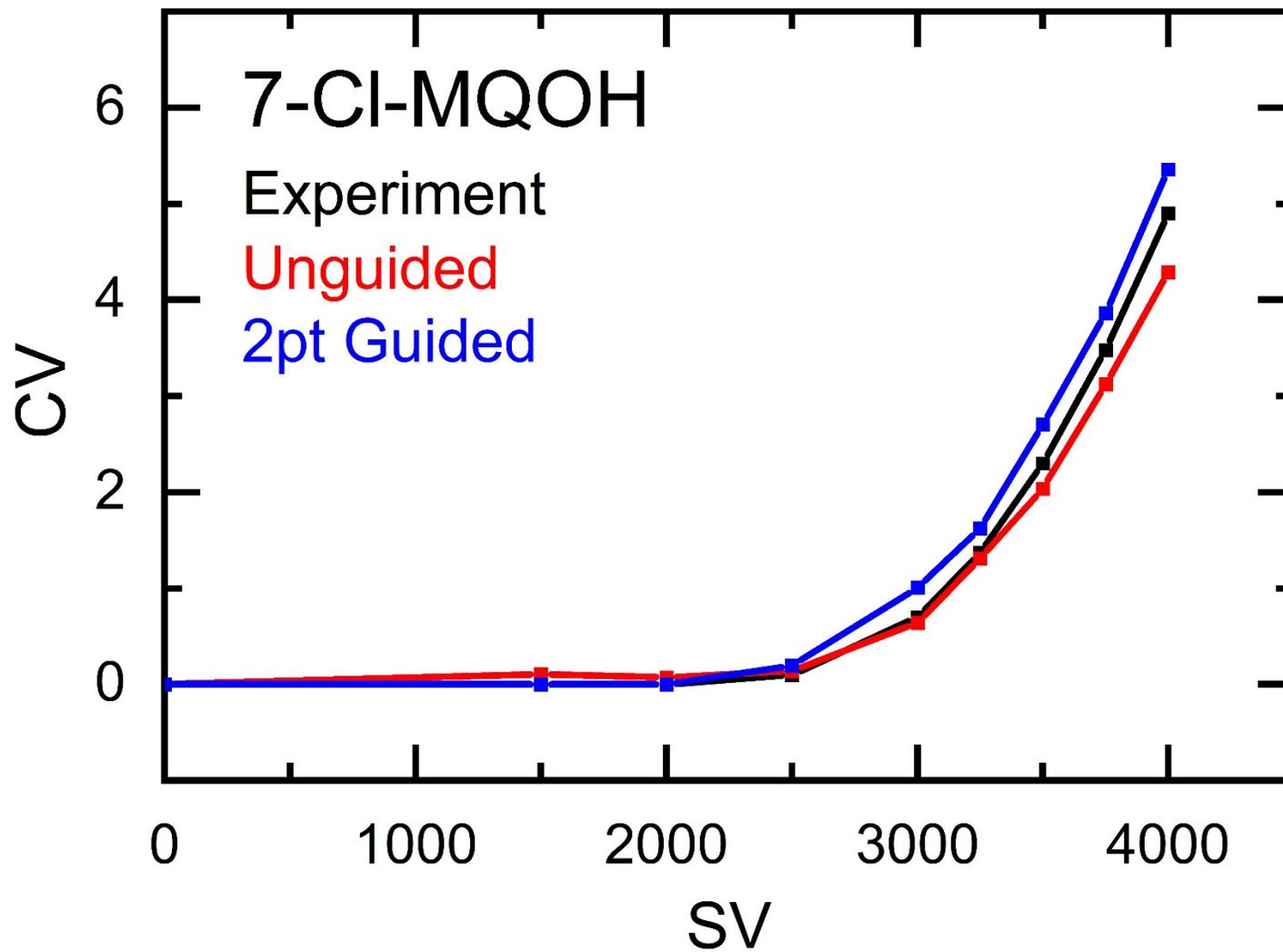


Figure S6. Dispersion curves of 7-chloro-2-methylquinolin-8-ol as predicted by the unguided (red) and two-point guided (blue) ML-approaches compared to the experimental dispersion curve (black).

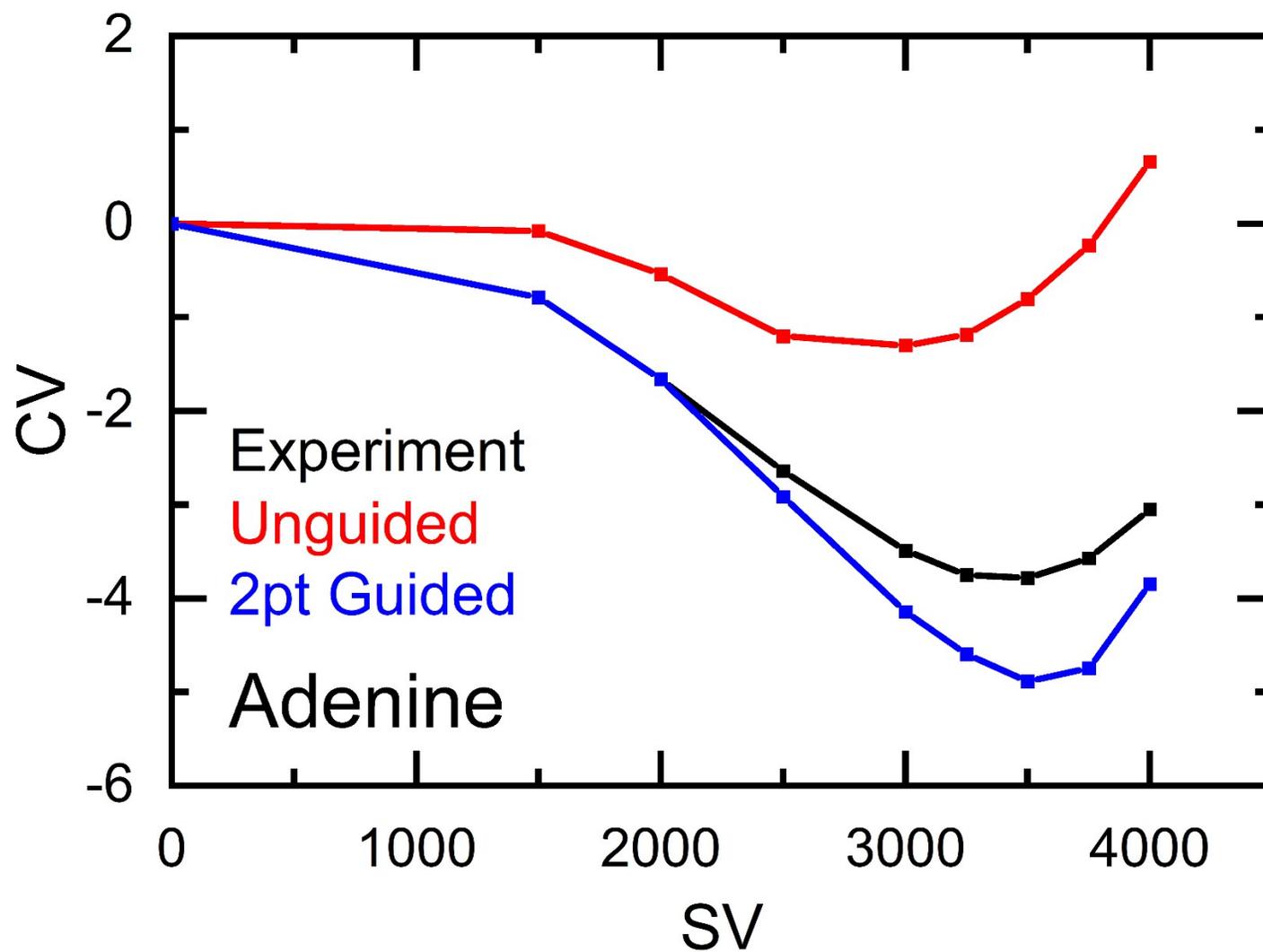


Figure S7. Dispersion curves of adenine as predicted by the unguided (red) and two-point guided (blue) ML-approaches compared to the experimental dispersion curve (black).

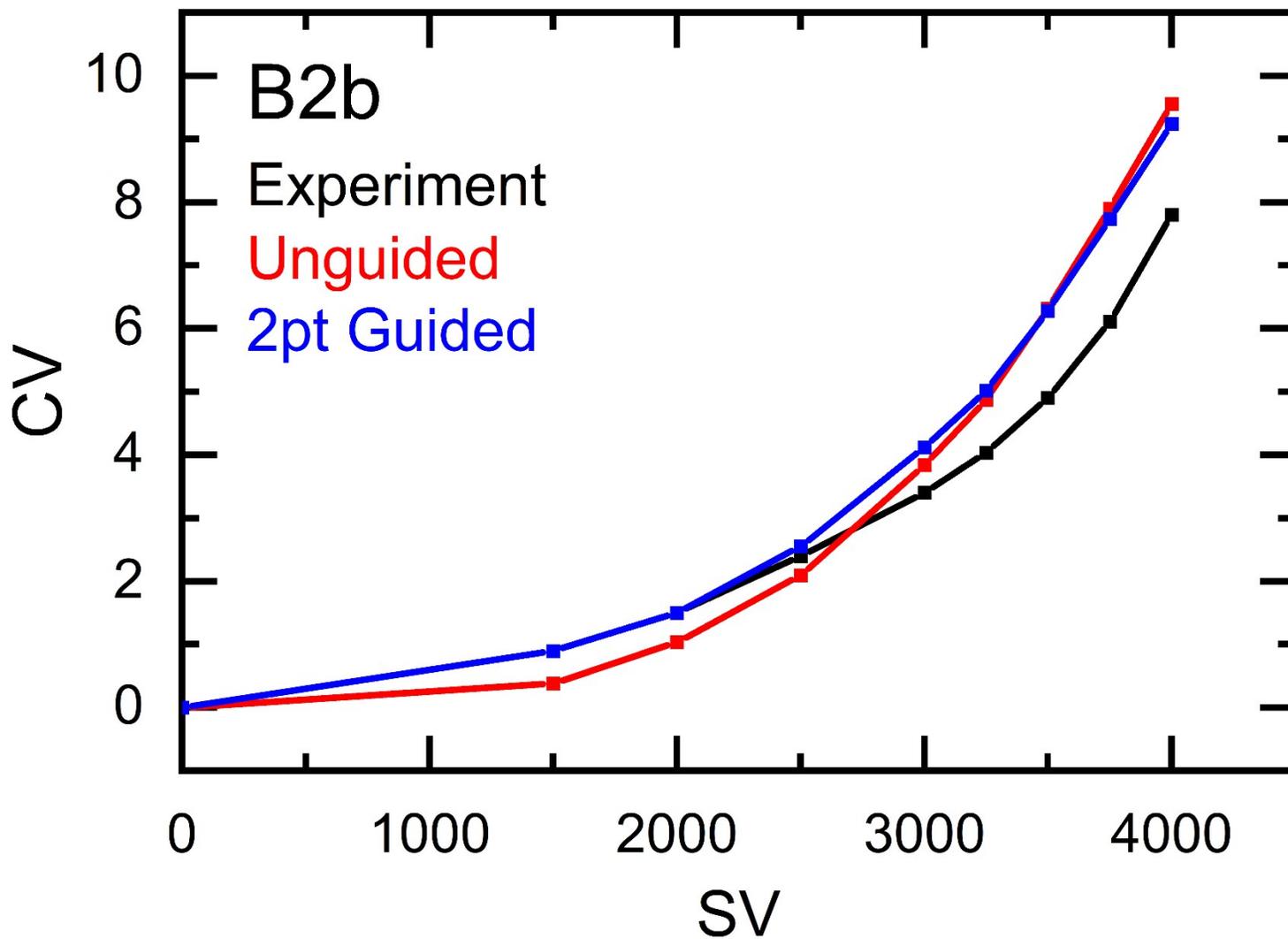


Figure S8. Dispersion curves of B2b (ethyl (2-(piperidine-1-carbonyl)phenyl)carbamate) as predicted by the unguided (red) and two-point guided (blue) ML-approaches compared to the experimental dispersion curve (black).

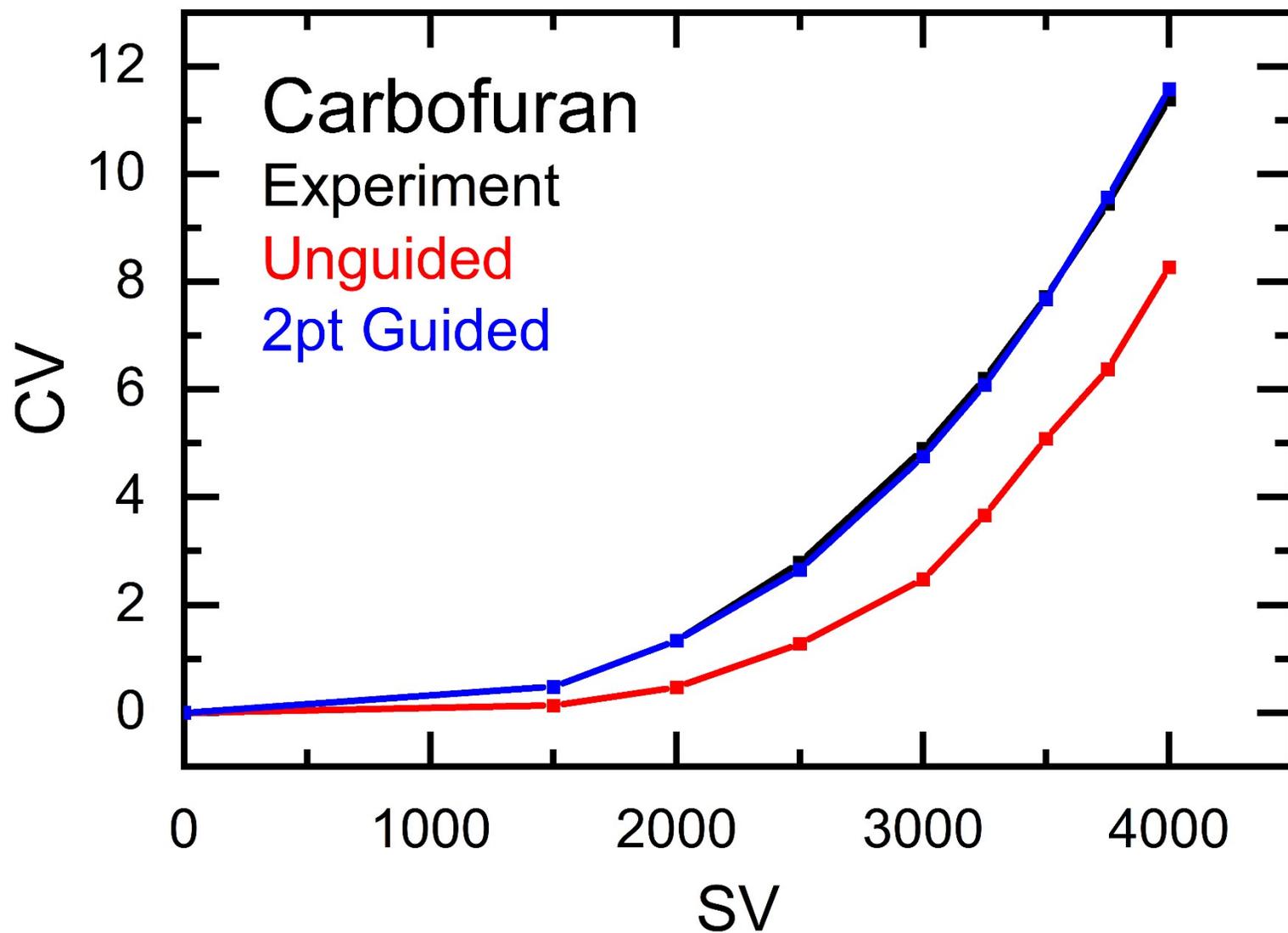


Figure S9. Dispersion curves of carbofuran as predicted by the unguided (red) and two-point guided (blue) ML-approaches compared to the experimental dispersion curve (black).

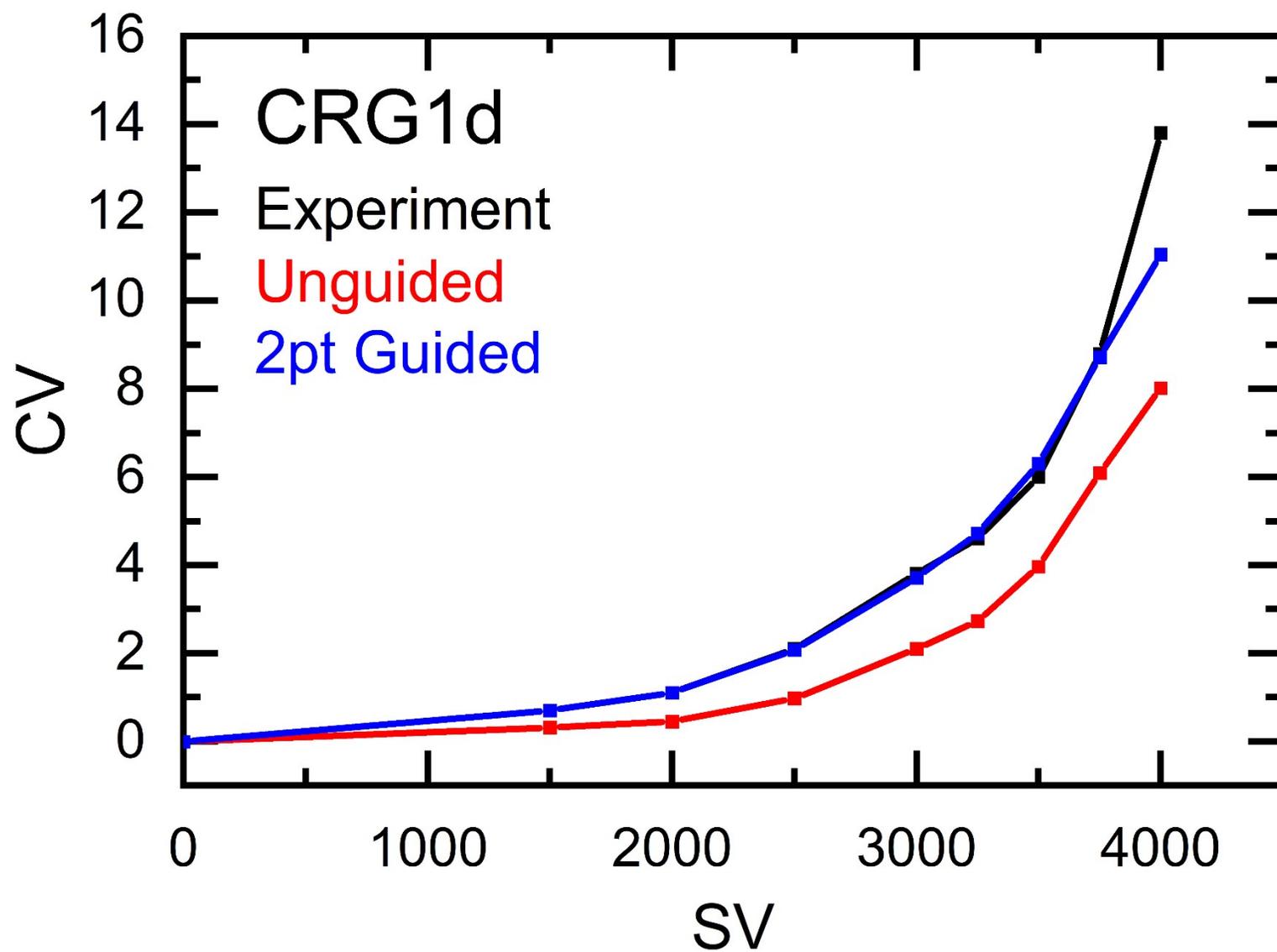


Figure S10. Dispersion curves of CRG1d (N-(pyridin-2-yl)acrylamide) as predicted by the unguided (red) and two-point guided (blue) ML-approaches compared to the experimental dispersion curve (black).

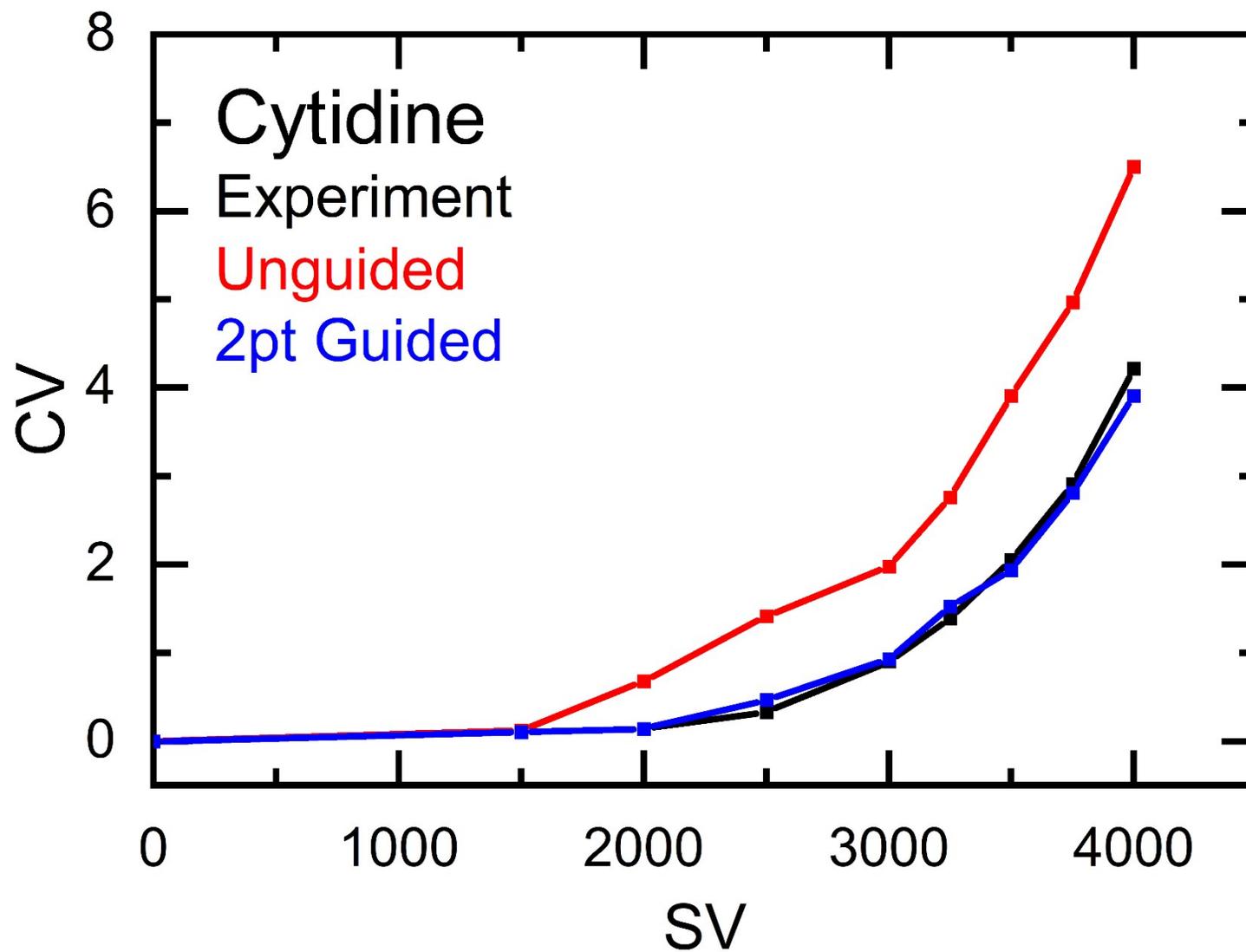


Figure S11. Dispersion curves of cytidine as predicted by the unguided (red) and two-point guided (blue) ML-approaches compared to the experimental dispersion curve (black).

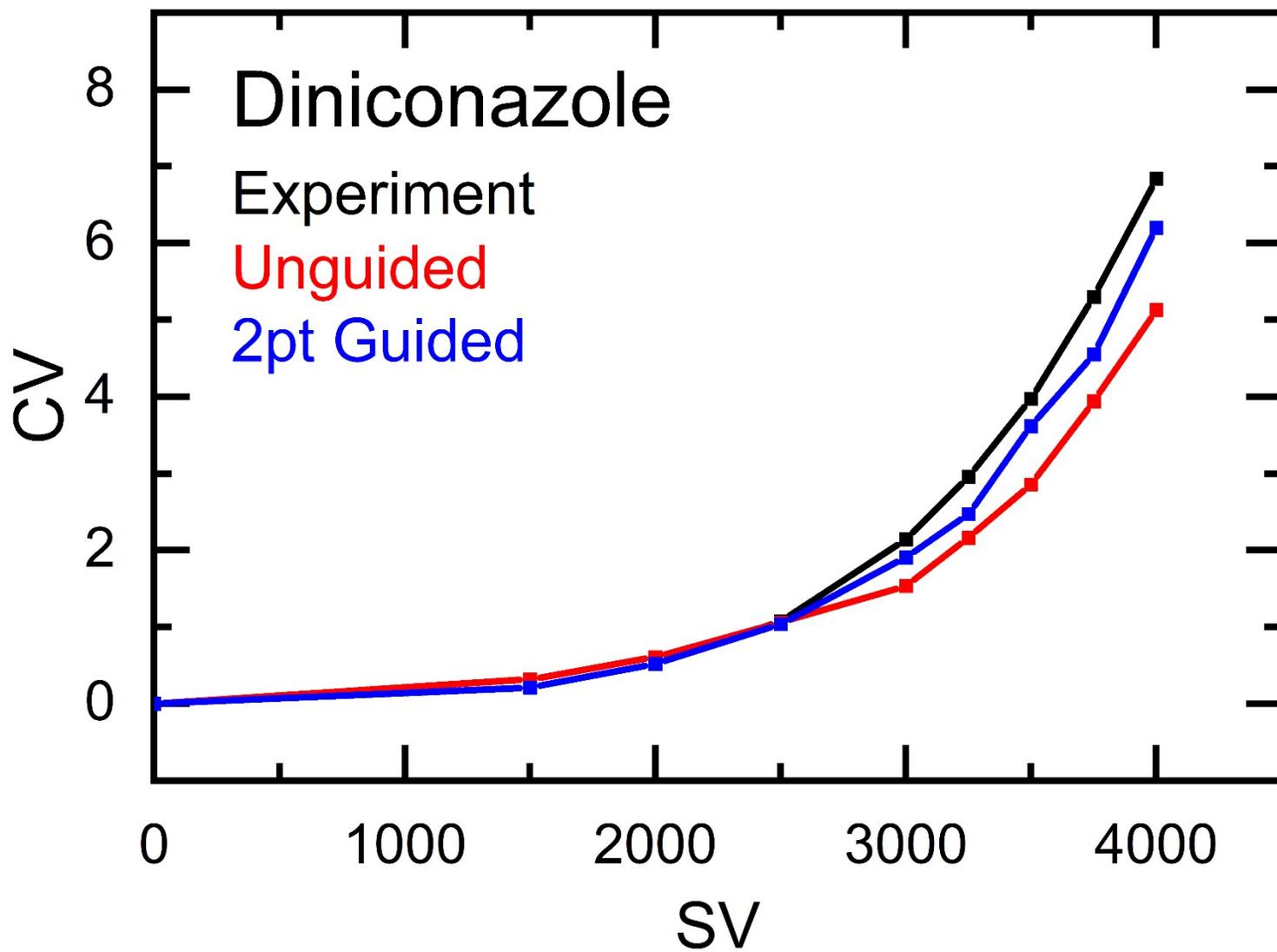


Figure S12. Dispersion curves of diniconazole as predicted by the unguided (red) and two-point guided (blue) ML-approaches compared to the experimental dispersion curve (black).

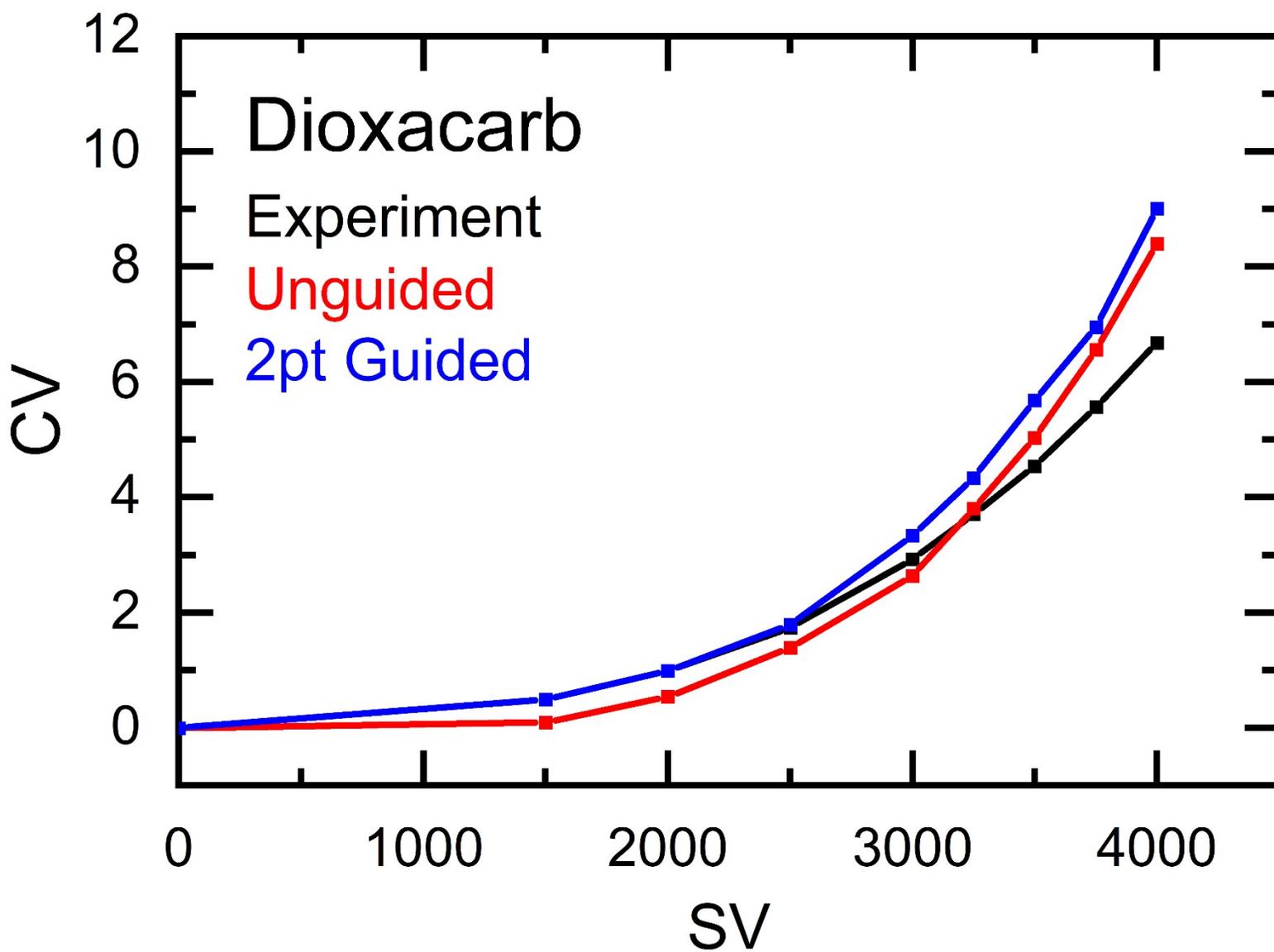


Figure S13. Dispersion curves of dioxacarb as predicted by the unguided (red) and two-point guided (blue) ML-approaches compared to the experimental dispersion curve (black).

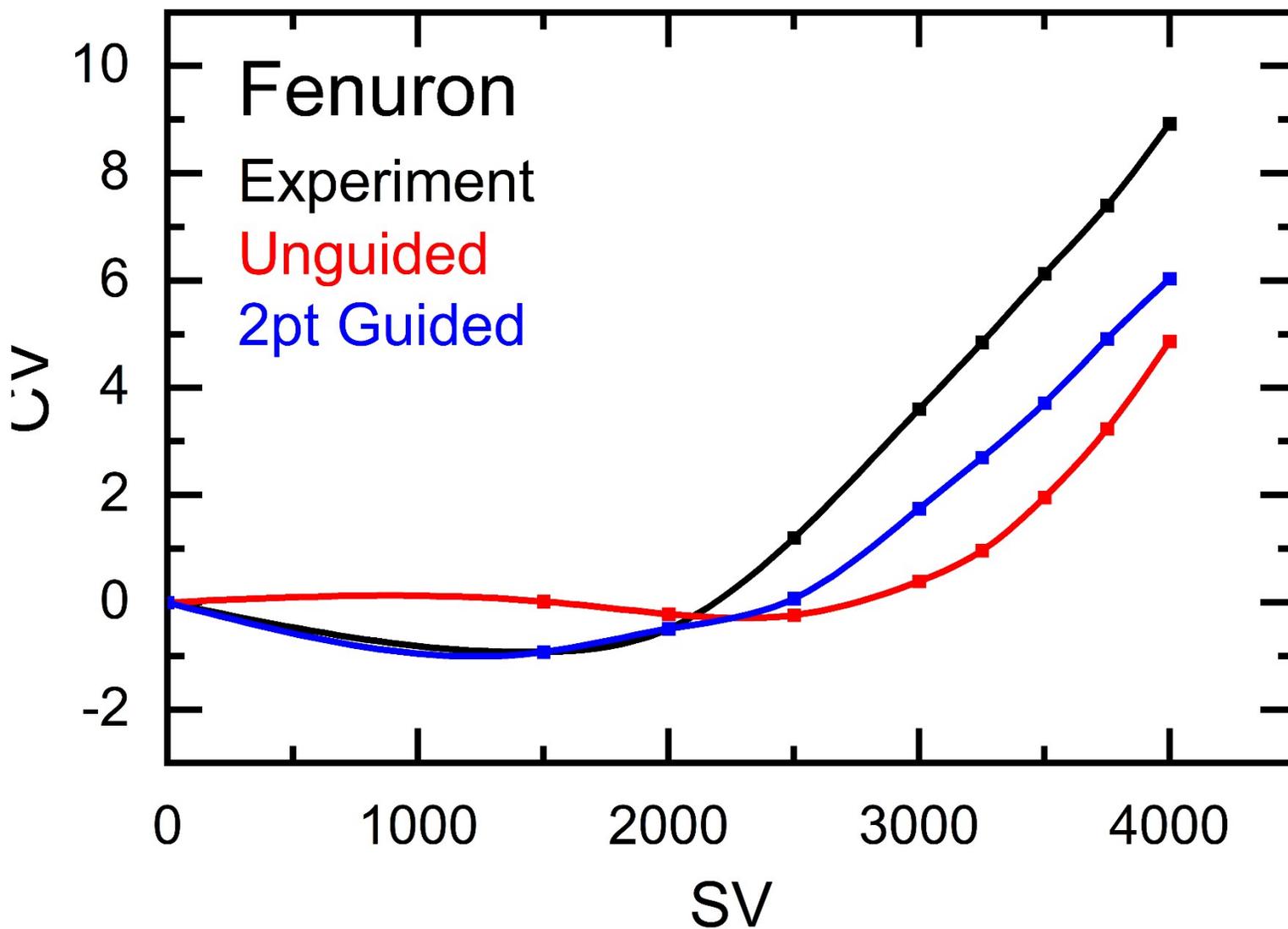


Figure S14. Dispersion curves of fenuron as predicted by the unguided (red) and two-point guided (blue) ML-approaches compared to the experimental dispersion curve (black).

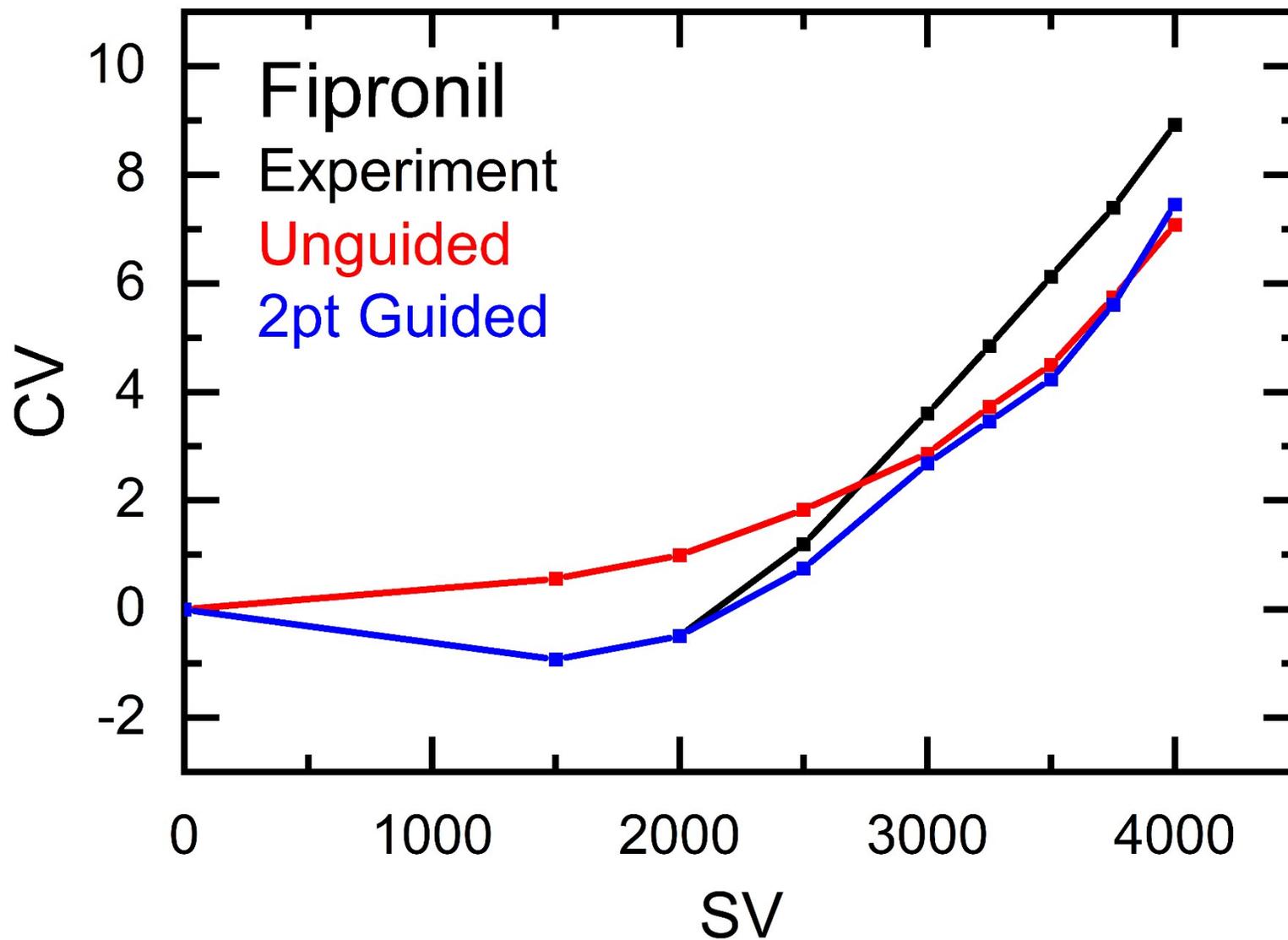


Figure S15. Dispersion curves of fipronil as predicted by the unguided (red) and two-point guided (blue) ML-approaches compared to the experimental dispersion curve (black).

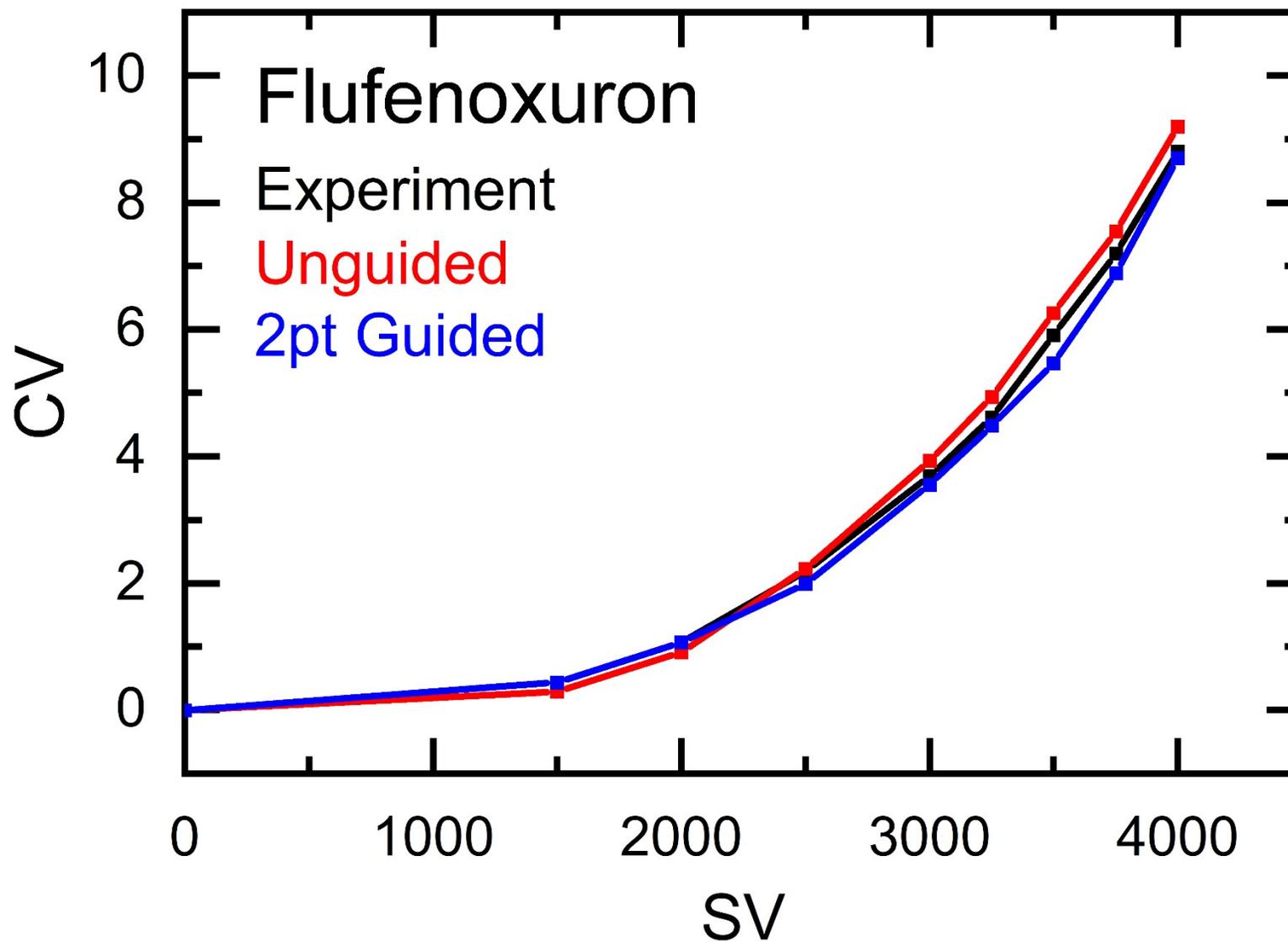


Figure S16. Dispersion curves of flufenoxuron as predicted by the unguided (red) and two-point guided (blue) ML-approaches compared to the experimental dispersion curve (black).

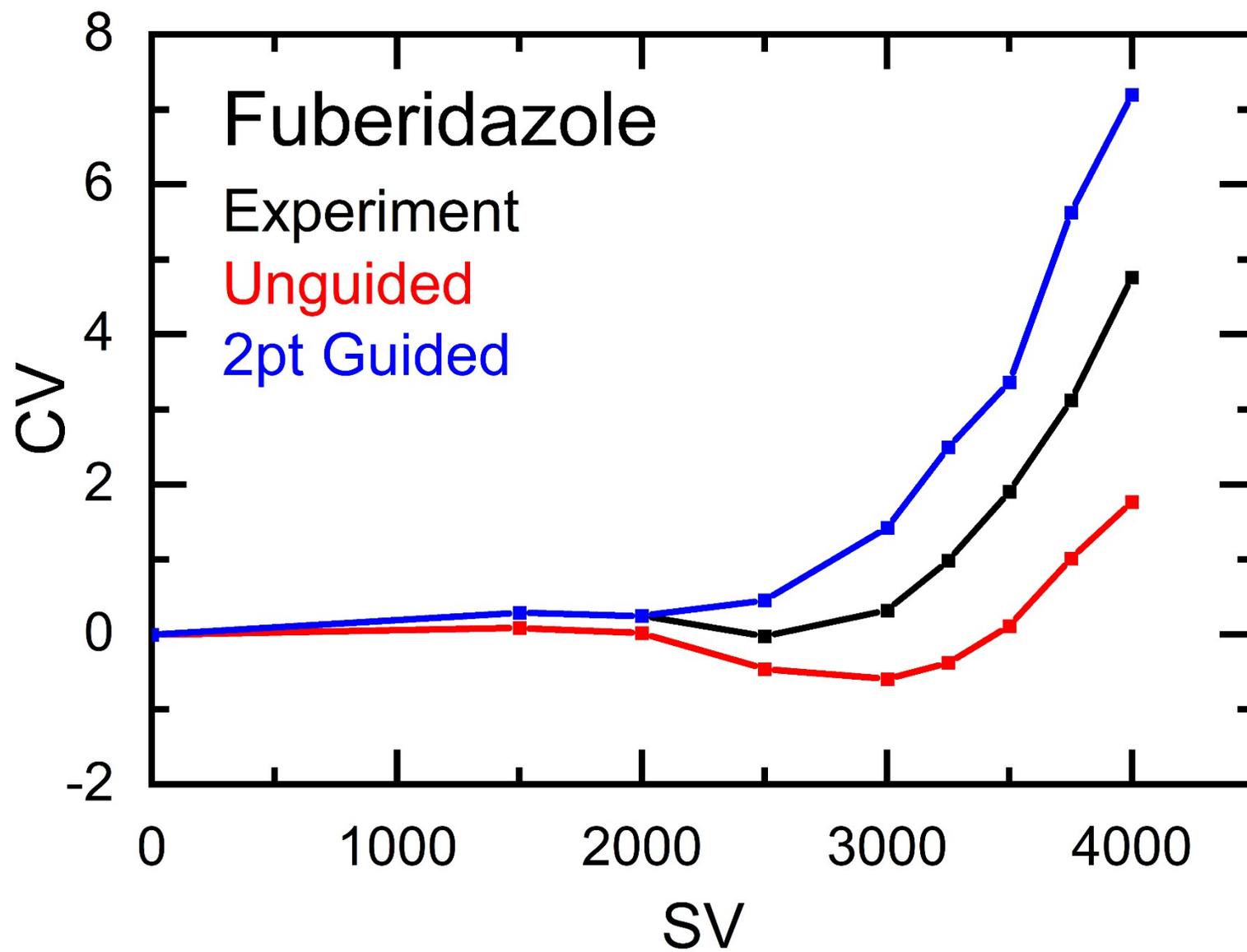


Figure S17. Dispersion curves of fuberidazole as predicted by the unguided (red) and two-point guided (blue) ML-approaches compared to the experimental dispersion curve (black).

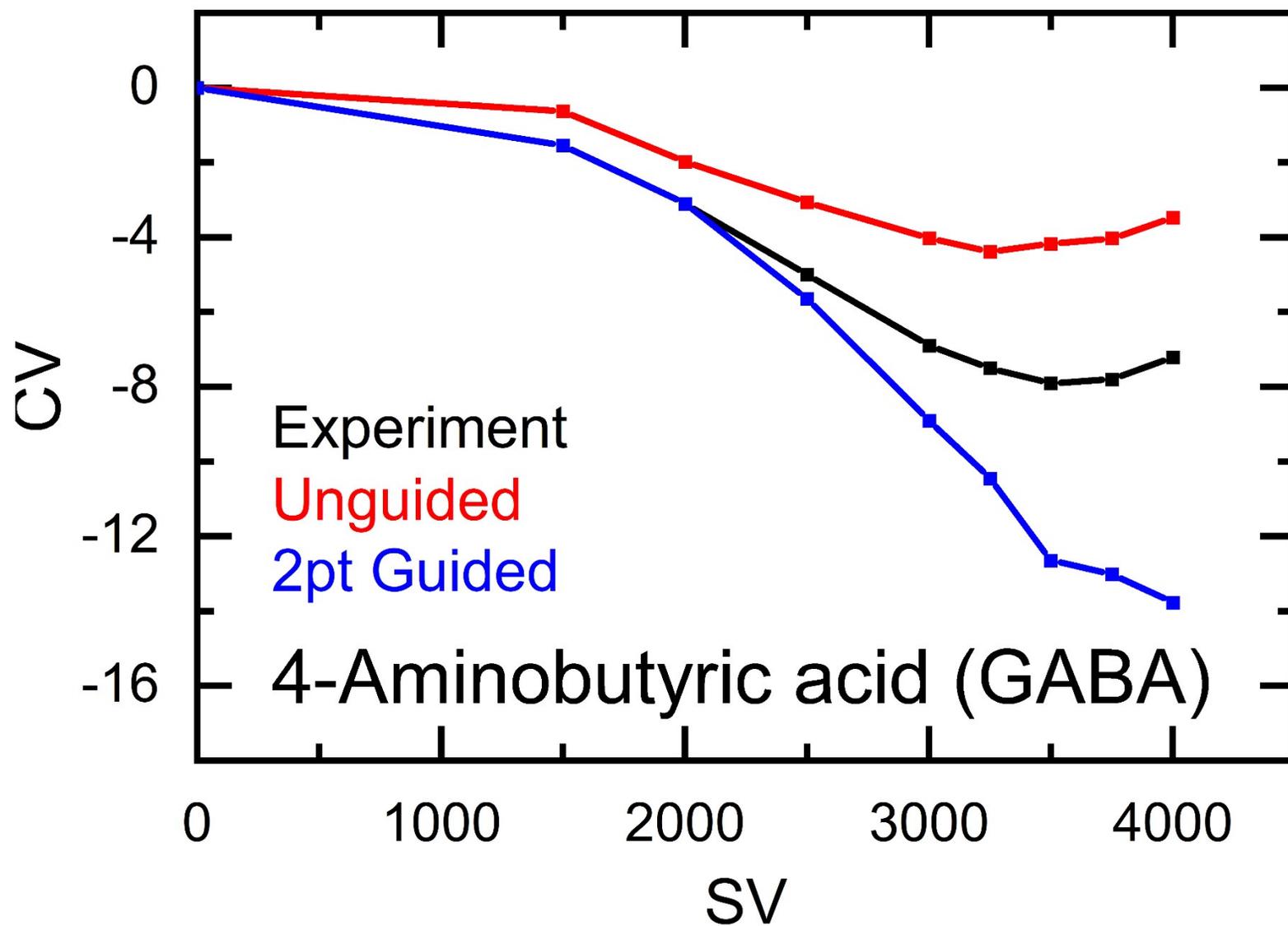


Figure S18. Dispersion curves of 4-aminobutyric acid (GABA) as predicted by the unguided (red) and two-point guided (blue) ML-approaches compared to the experimental dispersion curve (black).

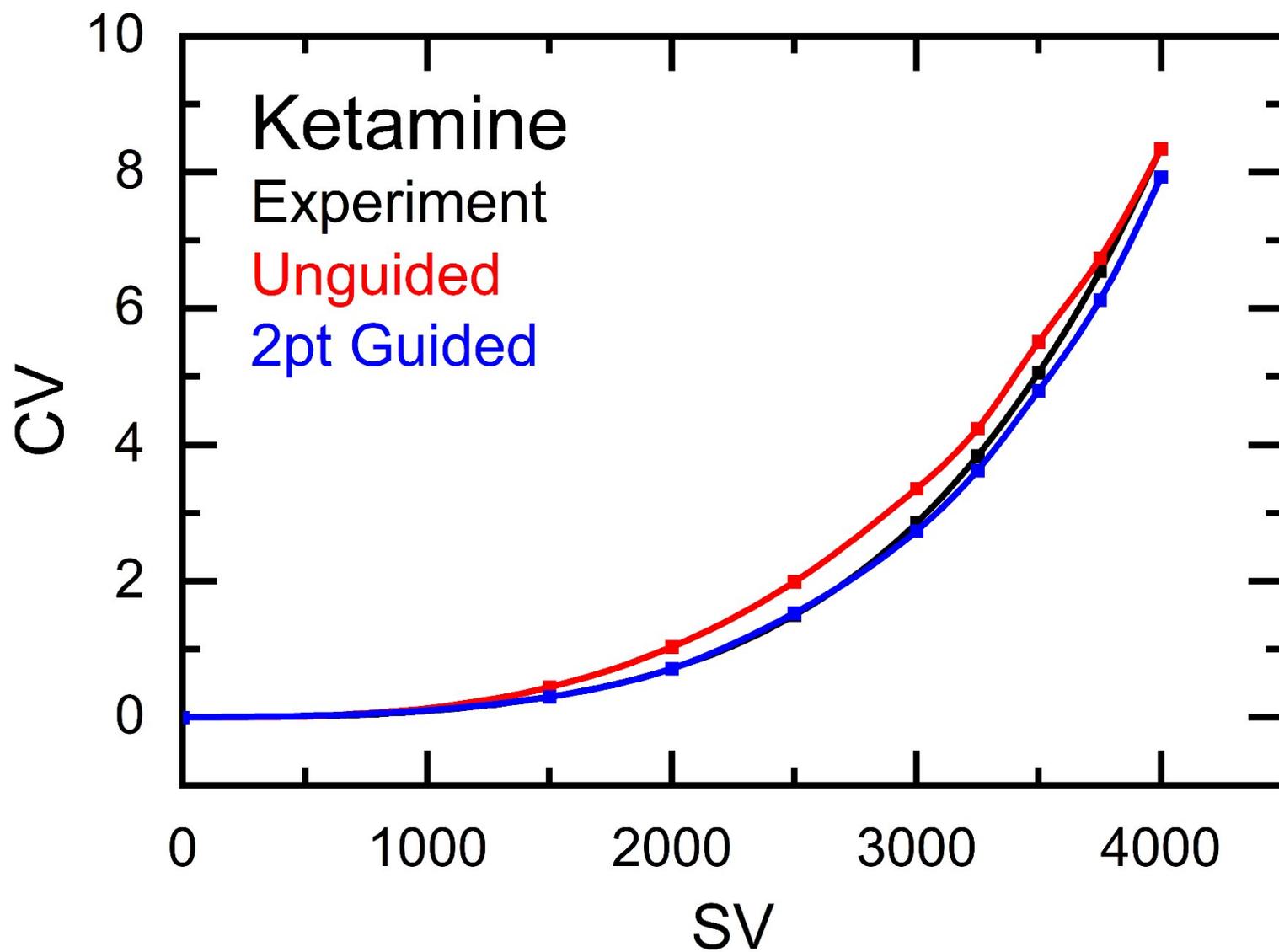


Figure S19. Dispersion curves of ketamine as predicted by the unguided (red) and two-point guided (blue) ML-approaches compared to the experimental dispersion curve (black).

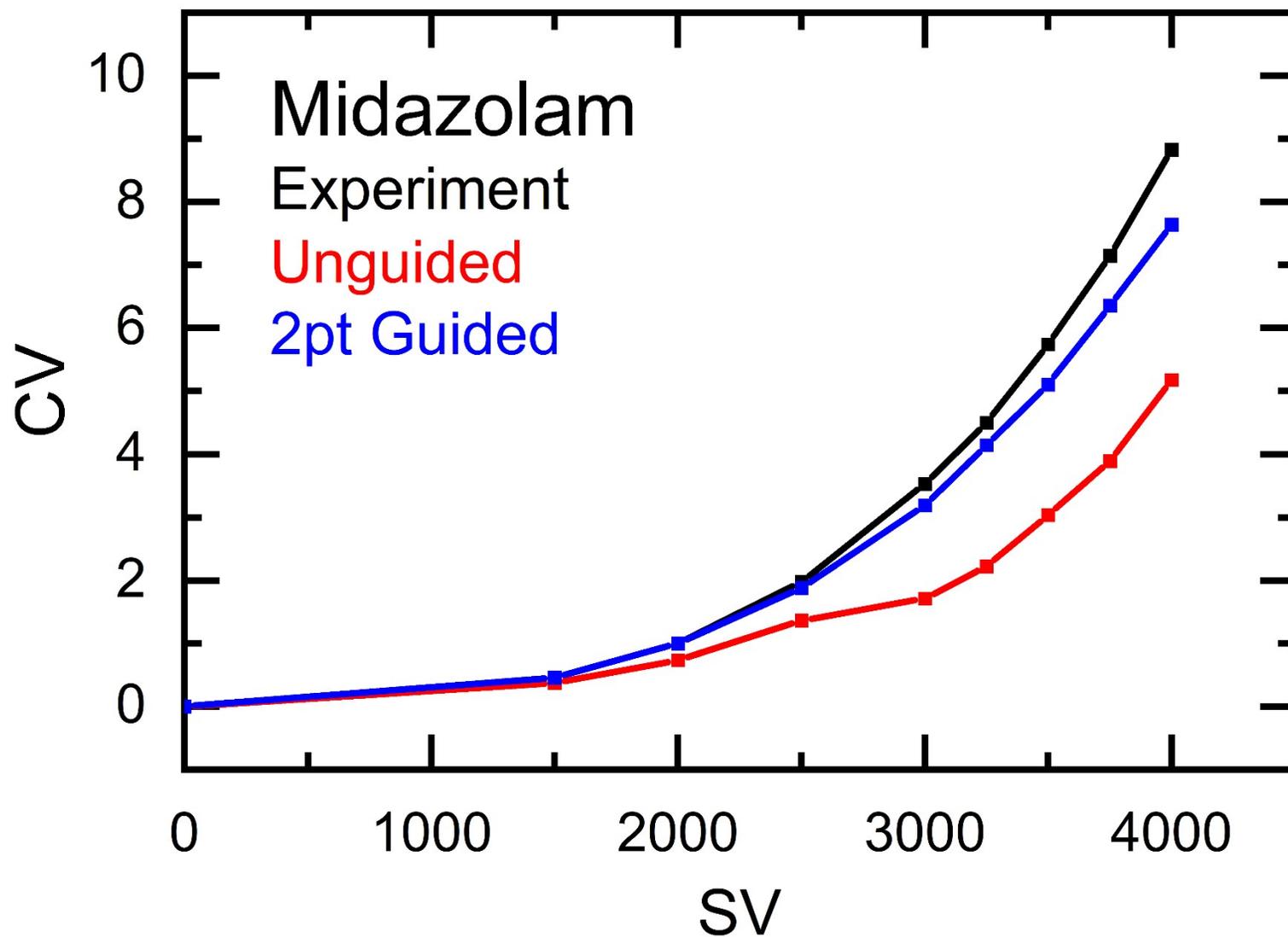


Figure S20. Dispersion curves of midazolam as predicted by the unguided (red) and two-point guided (blue) ML-approaches compared to the experimental dispersion curve (black).

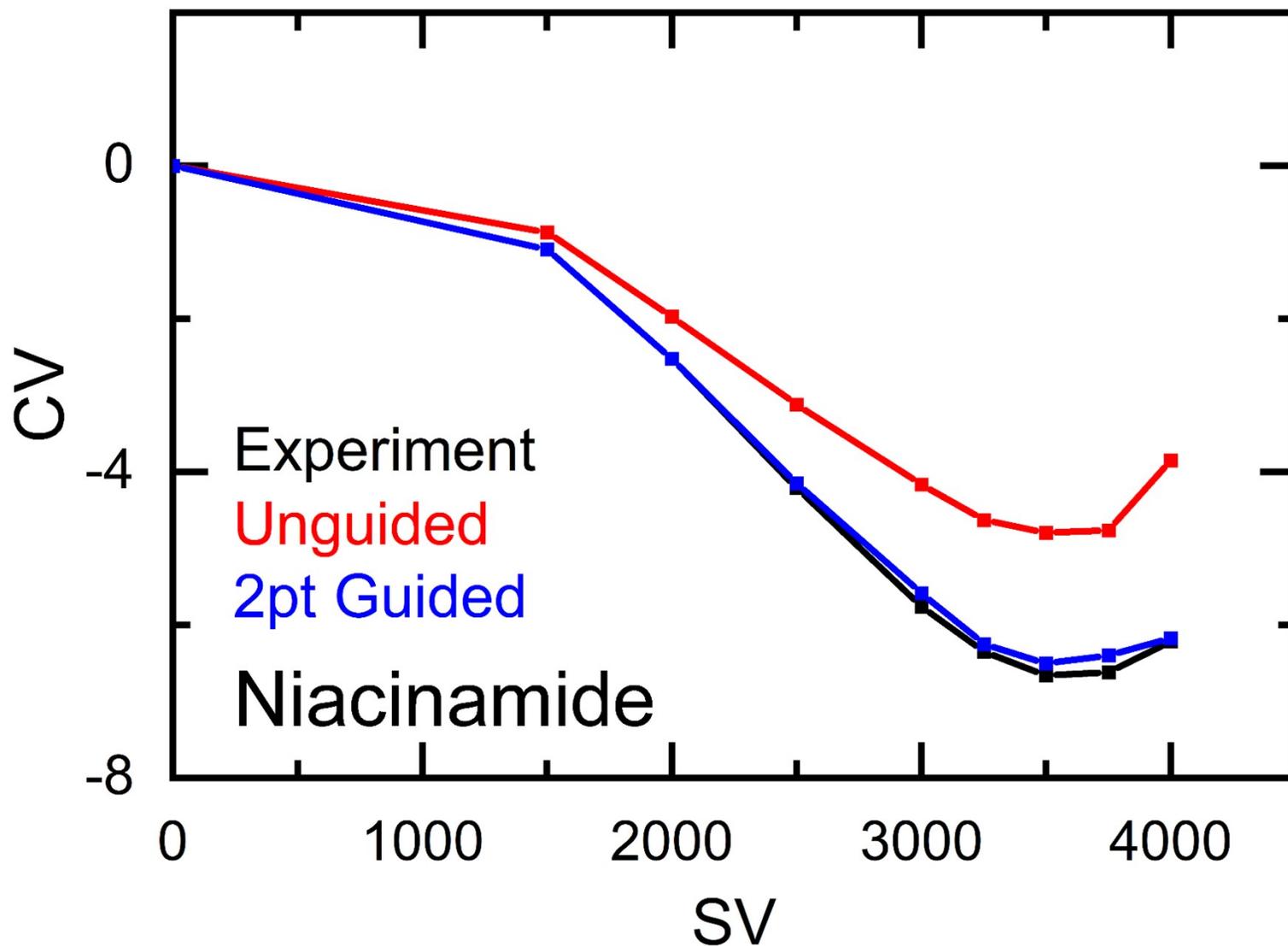


Figure S21. Dispersion curves of niacinamide (also known as nicotinamide) as predicted by the unguided (red) and two-point guided (blue) ML-approaches compared to the experimental dispersion curve (black).

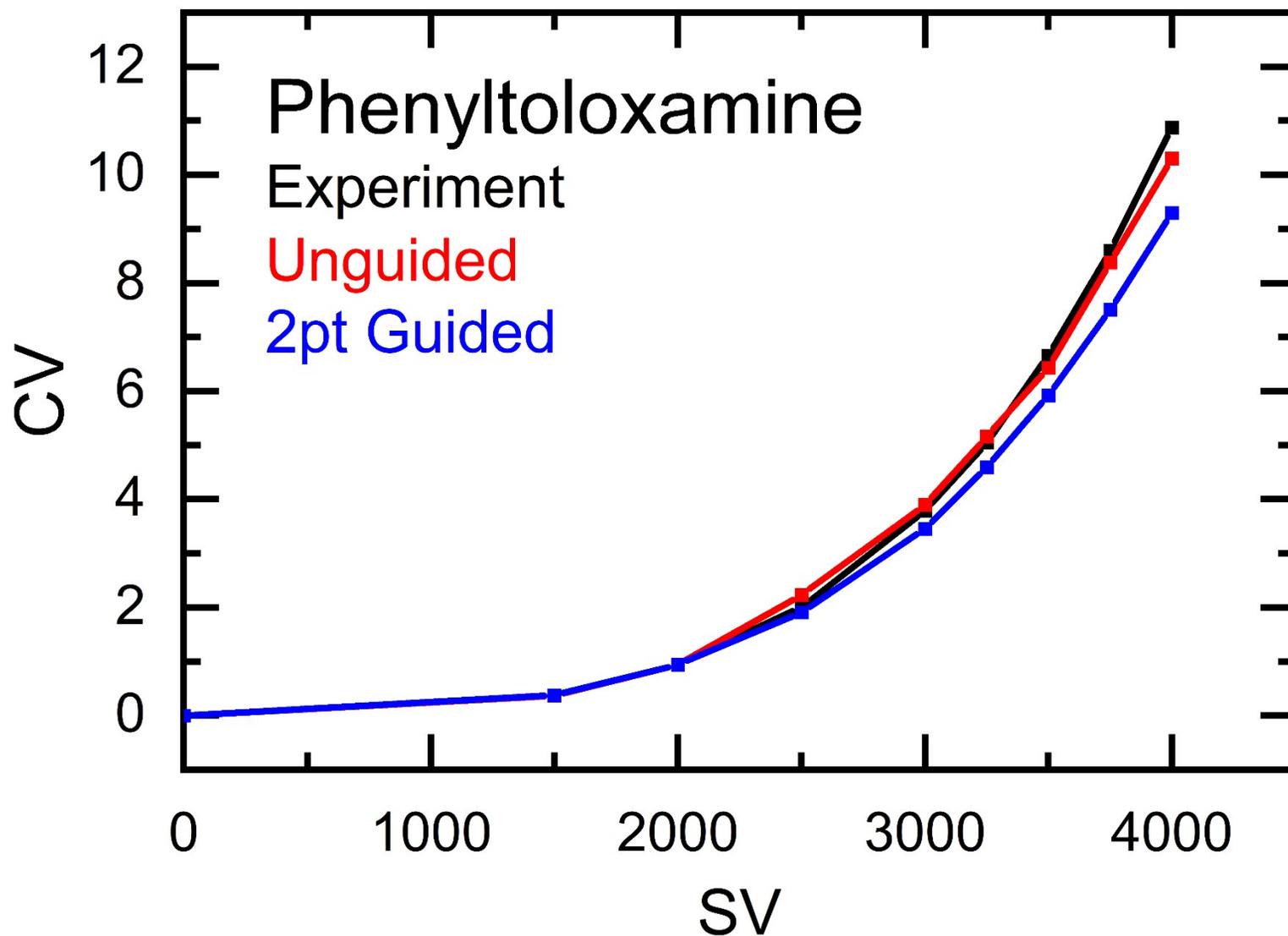


Figure S22. Dispersion curves of phenyltoloxamine as predicted by the unguided (red) and two-point guided (blue) ML-approaches compared to the experimental dispersion curve (black).

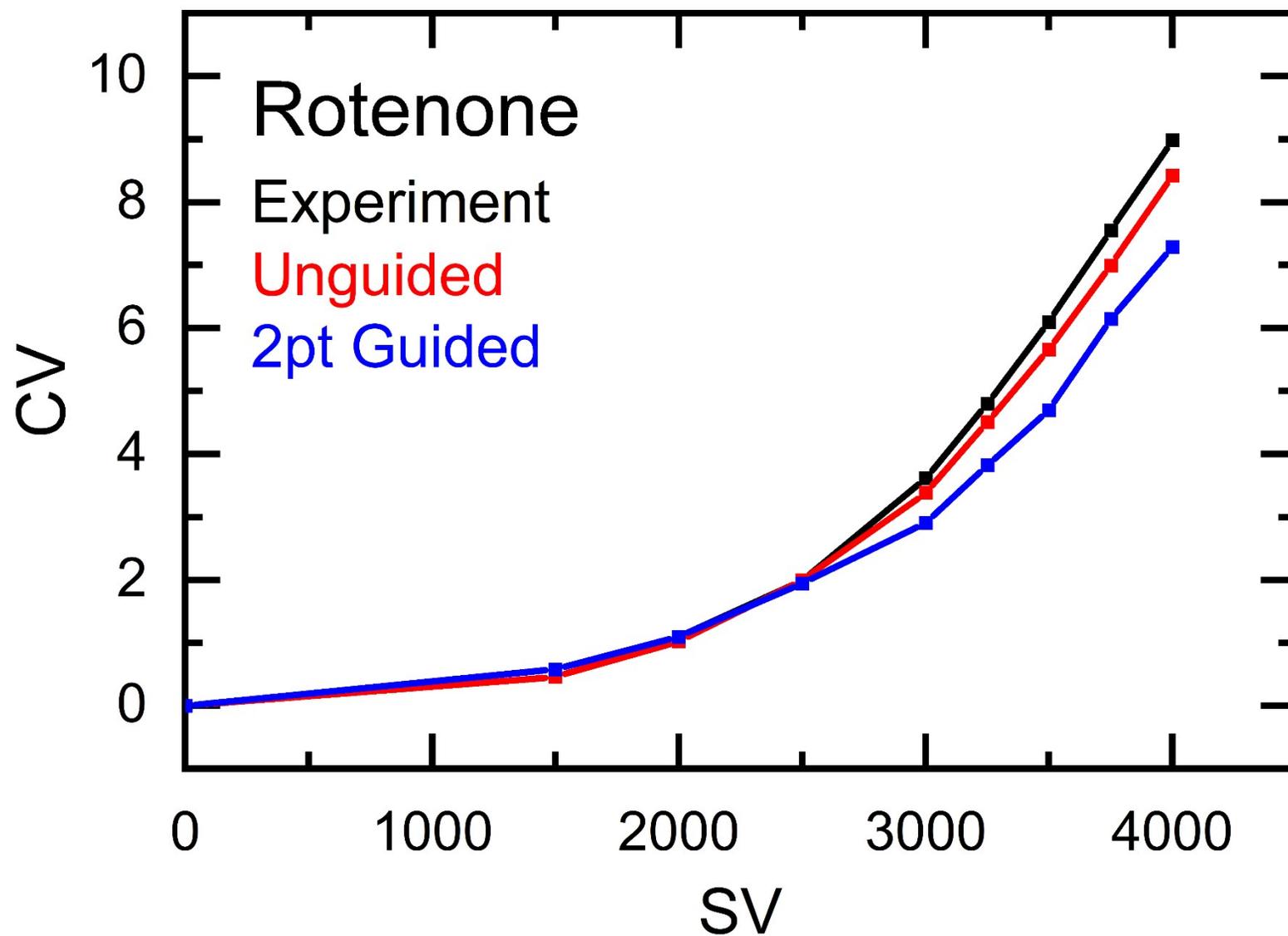


Figure S23. Dispersion curves of rotenone as predicted by the unguided (red) and two-point guided (blue) ML-approaches compared to the experimental dispersion curve (black).

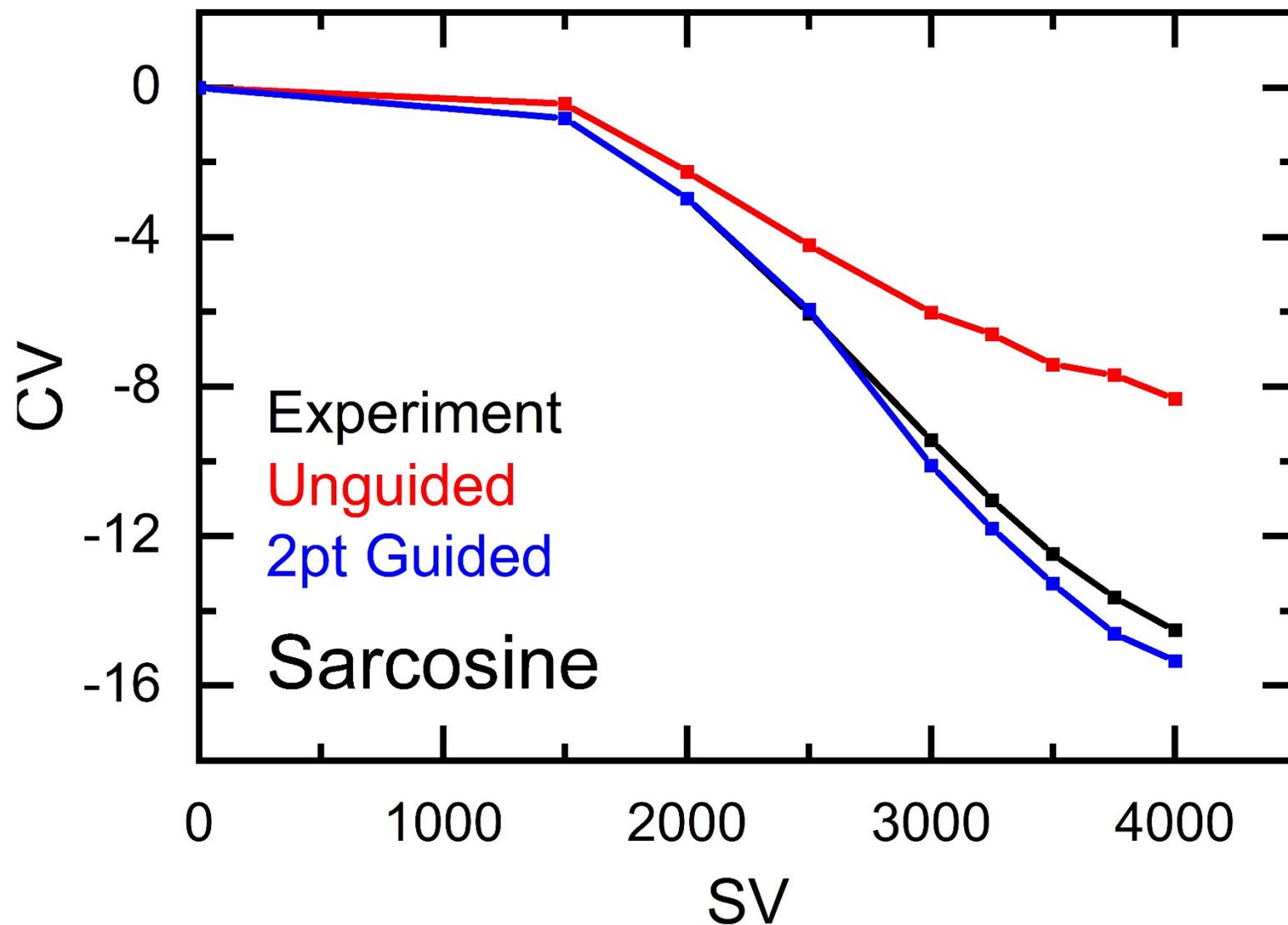


Figure S24. Dispersion curves of sarcosine as predicted by the unguided (red) and two-point guided (blue) ML-approaches compared to the experimental dispersion curve (black).

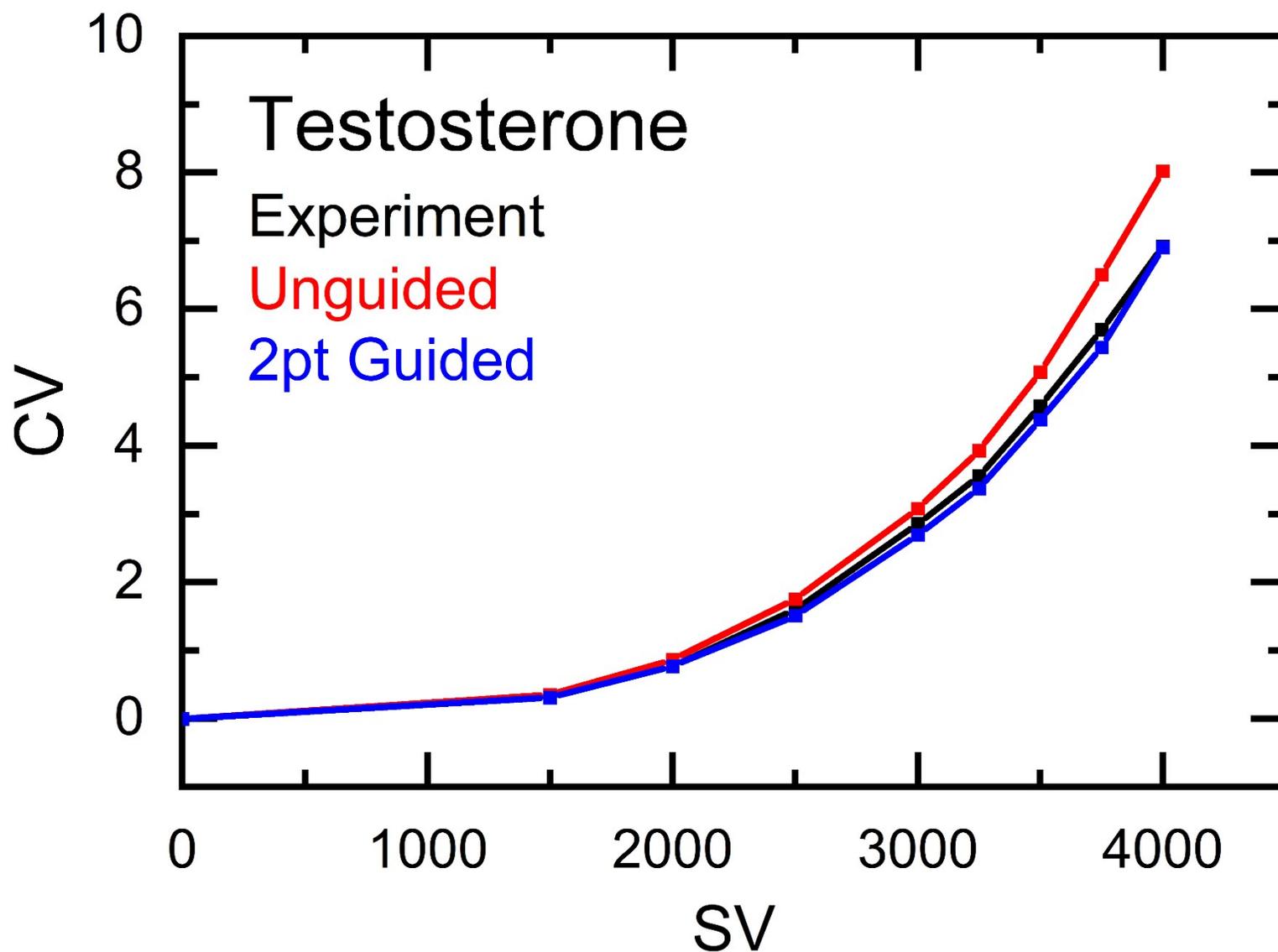


Figure S25. Dispersion curves of sarcosine as predicted by the unguided (red) and two-point guided (blue) ML-approaches compared to the experimental dispersion curve (black).

References

- (1) Coughlan, N. J. A.; Liu, C.; Lecours, M. J.; Campbell, J. L.; Hopkins, W. S. Preferential Ion Microsolvation in Mixed-Modifier Environments Observed Using Differential Mobility Spectrometry. *J. Am. Soc. Mass Spectrom.* **2019**, *30* (11), 2222–2227. <https://doi.org/10.1007/s13361-019-02332-1>.
- (2) Djoumbou Feunang, Y.; Eisner, R.; Knox, C.; Chepelev, L.; Hastings, J.; Owen, G.; Fahy, E.; Steinbeck, C.; Subramanian, S.; Bolton, E.; Greiner, R.; Wishart, D. S. ClassyFire: Automated Chemical Classification with a Comprehensive, Computable Taxonomy. *J. Cheminform.* **2016**, *8* (1), 1–20. <https://doi.org/10.1186/s13321-016-0174-y>.
- (3) Wales, D. J.; Doye, J. P. K. Global Optimization by Basin-Hopping and the Lowest Energy Structures of Lennard-Jones Clusters Containing up to 110 Atoms. *J. Phys. Chem. A* **1997**, *101* (28), 5111–5116. <https://doi.org/10.1021/jp970984n>.
- (4) Wales, D. *Energy Landscapes*; Cambridge University Press, 2004. <https://doi.org/10.1017/cbo9780511721724>.
- (5) Ieritano, C.; Featherstone, J.; Carr, P. J. J.; Marta, R. A.; Loire, E.; McMahon, T. B.; Hopkins, W. S. The Structures and Properties of Anionic Tryptophan Complexes. *Phys. Chem. Chem. Phys.* **2018**, *20* (41), 26532–26541. <https://doi.org/10.1039/c8cp04533j>.
- (6) Hopkins, W. S.; Marta, R. A.; McMahon, T. B. Proton-Bound 3-Cyanophenylalanine Trimethylamine Clusters: Isomer-Specific Fragmentation Pathways and Evidence of Gas-Phase Zwitterions. *J. Phys. Chem. A* **2013**, *117* (41), 10714–10718. <https://doi.org/10.1021/jp407766j>.

- (7) Scott Hopkins, W.; Marta, R. A.; Steinmetz, V.; McMahon, T. B. Mode-Specific Fragmentation of Amino Acid-Containing Clusters. *Phys. Chem. Chem. Phys.* **2015**, *17* (43), 28548–28555. <https://doi.org/10.1039/c5cp03517a>.
- (8) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; V. Barone, B.; Mennucci, G.; Petersson, A. Gaussian 09. Gaussian Incorporated: Wallingford, CT 2009.
- (9) Rappé, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. UFF, a Full Periodic Table Force Field for Molecular Mechanics and Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **1992**, *114* (25), 10024–10035. <https://doi.org/10.1021/ja00051a040>.
- (10) Becke, A. D. Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior. *Phys. Rev. A* **1988**, *38* (6), 3098–3100. <https://doi.org/10.1103/PhysRevA.38.3098>.
- (11) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B* **1988**, *37* (2), 785–789. <https://doi.org/10.1103/PhysRevB.37.785>.
- (12) Besler, B. H.; Merz, K. M.; Kollman, P. A. Atomic Charges Derived from Semiempirical Methods. *J. Comput. Chem.* **1990**, *11* (4), 431–439. <https://doi.org/10.1002/jcc.540110404>.
- (13) Singh, U. C.; Kollman, P. A. An Approach to Computing Electrostatic Charges for Molecules. *J. Comput. Chem.* **1984**, *5* (2), 129–145. <https://doi.org/10.1002/jcc.540050204>.
- (14) Stewart, J. J. P. Optimization of Parameters for Semiempirical Methods VI: More Modifications to the NDDO Approximations and Re-Optimization of Parameters. *J. Mol. Model.* **2013**, *19* (1), 1–32. <https://doi.org/10.1007/s00894-012-1667-x>.

- (15) Throssel, K.; Frisch, M. J. Evaluation and Improvement of Semiempirical Methods I: PM7R8: A Variant of PM7 with Numerically Stable Hydrogen Bonding Corrections. *Manuscr. Prep.*
- (16) Zhou, C.; Ieritano, C.; Hopkins, W. S. Augmenting Basin-Hopping With Techniques From Unsupervised Machine Learning: Applications in Spectroscopy and Ion Mobility. *Front. Chem.* **2019**, *7*, 519. <https://doi.org/10.3389/fchem.2019.00519>.
- (17) Ieritano, C.; Crouse, J.; Campbell, J. L.; Hopkins, W. S. A Parallelized Molecular Collision Cross Section Package with Optimized Accuracy and Efficiency. *Analyst* **2019**, *144* (5), 1660–1670. <https://doi.org/10.1039/c8an02150c>.