

Reduction pathway of glutaredoxin 1 investigated with QM/MM molecular dynamics using a neural network correction

Cite as: J. Chem. Phys. **157**, 154104 (2022); <https://doi.org/10.1063/5.0123089>

Submitted: 28 August 2022 • Accepted: 26 September 2022 • Accepted Manuscript Online: 26 September 2022 • Published Online: 21 October 2022

 Julian Böser,  Tomáš Kubař,  Marcus Elstner, et al.

COLLECTIONS

Paper published as part of the special topic on [Modern Semiempirical Electronic Structure Methods](#)



View Online



Export Citation



CrossMark

ARTICLES YOU MAY BE INTERESTED IN

[Can the Hartree-Fock kinetic energy exceed the exact kinetic energy?](#)

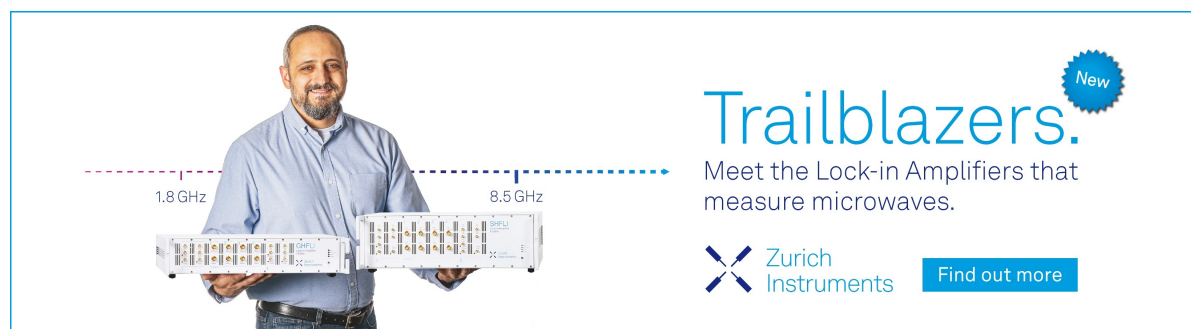
The Journal of Chemical Physics **157**, 154106 (2022); <https://doi.org/10.1063/5.0105684>


[Molecular-orbital-based machine learning for open-shell and multi-reference systems with kernel addition Gaussian process regression](#)

The Journal of Chemical Physics **157**, 154105 (2022); <https://doi.org/10.1063/5.0110886>


[Coupled local mode method for simulating vibrational spectroscopy](#)

The Journal of Chemical Physics **157**, 154103 (2022); <https://doi.org/10.1063/5.0119631>



Trailblazers. 

Meet the Lock-in Amplifiers that measure microwaves.

 Zurich Instruments [Find out more](#)

Reduction pathway of glutaredoxin 1 investigated with QM/MM molecular dynamics using a neural network correction

Cite as: J. Chem. Phys. 157, 154104 (2022); doi: 10.1063/5.0123089

Submitted: 28 August 2022 • Accepted: 26 September 2022 •

Published Online: 21 October 2022



View Online



Export Citation



CrossMark

Julian Böser,¹  Tomáš Kubař,¹  Marcus Elstner,^{1,2,a)}  and Denis Maag¹ 

AFFILIATIONS

¹Institute of Physical Chemistry, Karlsruhe Institute of Technology, 76131 Karlsruhe, Germany

²Institute of Biological Interfaces (IBG-2), Karlsruhe Institute of Technology, 76131 Karlsruhe, Germany

Note: This paper is part of the JCP Special Topic on Modern Semiempirical Electronic Structure Methods.

^{a)}Author to whom correspondence should be addressed: marcus.elstner@kit.edu

ABSTRACT

Glutaredoxins are small enzymes that catalyze the oxidation and reduction of protein disulfide bonds by the thiol–disulfide exchange mechanism. They have either one or two cysteines in their active site, resulting in different catalytic reaction cycles that have been investigated in many experimental studies. However, the exact mechanisms are not yet fully known, and to our knowledge, no theoretical studies have been performed to elucidate the underlying mechanism. In this study, we investigated a proposed mechanism for the reduction of the disulfide bond in the protein HMA4n by a mutated monothiol *Homo sapiens* glutaredoxin and the co-substrate glutathione. The catalytic cycle involves three successive thiol–disulfide exchanges that occur between the molecules. To estimate the regioselectivity of the different attacks, classical molecular dynamics simulations were performed and the trajectories analyzed regarding the sulfur–sulfur distances and the attack angles between the sulfurs. The free energy profile of each reaction was obtained with hybrid quantum mechanical/molecular mechanical metadynamics simulations. Since this required extensive phase space sampling, the semi-empirical density functional tight-binding method was used to describe the reactive cysteines. For an accurate description, we used specific reaction parameters fitted to B3LYP energies of the thiol–disulfide exchange and a machine learned energy correction that was trained on coupled-cluster single double perturbative triple [CCSD(T)] energies of thiol–disulfide exchanges. Our calculations show the same regioselectivity as observed in the experiment, and the obtained barrier heights are about 12 and 20 kcal/mol for the different reaction steps, which confirms the proposed pathway.

© 2022 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>). <https://doi.org/10.1063/5.0123089>

I. INTRODUCTION

Glutaredoxins (Grxs) are enzymes that are involved in many biological processes, such as cell signaling and oxidative stress response.^{1,2} They are responsible for the reduction and oxidation of disulfide bonds in proteins, and their catalytic efficiency depends largely on the cellular concentration of the co-substrates glutathione and glutathione disulfide (2GSH/GSSG), which can glutathionylate and deglutathionylate the enzyme or the protein substrate.³ Considering the number of cysteines in the active site, Grxs can be divided into two groups, dithiol and monothiol Grxs.⁴ Dithiol Grxs contain two enzymatically active cysteines (C-XX-C motif), while monothiol Grxs contain only one active cysteine (C-XX-S motif). The two groups operate by different mechanisms and by mutating

the second cysteine of dithiol Grxs to a serine, it has been demonstrated that the monothiol pathway is more efficient than the dithiol pathway.^{5–8} For example, Ukuwela *et al.* investigated the catalytic cycles of dithiol *Homo sapiens* HsGrx1 (C-PY-C motif) and a monothiol HsGrx1 mutant (C-PY-S motif) in which the C-terminal cysteine was mutated to a serine, compare Fig. 1(a).⁷ The metal binding domain HMA4n was used as substrate and GSH/GSSG as co-substrate, see Figs. 1(b) and 1(c). Based on the analysis of the intermediates by electrospray ionization mass spectrometry, they proposed a monothiol and a dithiol pathway. While monothiol HsGrx1 can only use the monothiol route, shown in Fig. 1(d), dithiol HsGrx1 can employ the monothiol and a dithiol route in parallel. In both catalytic cycles, a sequence of thiol–disulfide

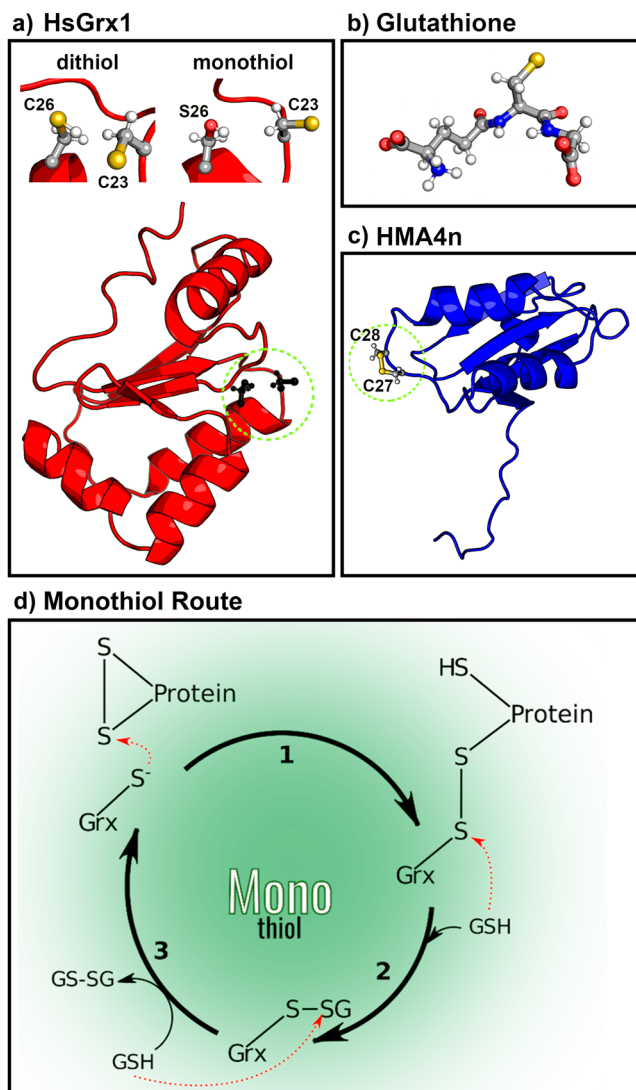


FIG. 1. (a) Wild-type human dithiol HsGrx1 (PDB ID: 1JHB¹¹) with two cysteines in the active site and monothiol HsGrx1 in which the C-terminal cysteine has been mutated to a serine. The N-terminal nucleophilic cysteines are shown in their anionic forms. (b) The co-substrate GSH (γ -L-glutamyl-L-cysteinylglycine tripeptide) in its anionic form GS^- . (c) The protein substrate HMA4n based on PDB ID: 2KKH.¹² The two oxidized cysteines forming a disulfide bond are highlighted. (d) Proposed monothiol route by Ukuwela *et al.* for the reduction of a protein disulfide bond by HsGrx1.⁷

exchanges occurs, i.e., a $\text{S}_{\text{N}}2$ reaction between a thiolate anion and two disulfide bonded sulfurs. The dithiol route requires an additional intramolecular attack of the Cys26 sulfur atom on the Cys23 sulfur atom and is, therefore, less efficient. The first step in both proposed pathways is the nucleophilic attack on the protein disulfide bond by the enzyme Grx. However, there are also other mechanistic models. For example, it has been proposed that GSH attacks the disulfide bond of the substrate in the first step, resulting in a glutathionylated protein $\text{P}(\text{SH})(\text{S-SG})$.^{3,9,10} In the next step, the GSH moiety is then attacked by Grx, resulting in the reduced substrate

$\text{P}(\text{SH})_2$ and the glutathionylated Grx enzyme $\text{Grx}(\text{S-SG})$. Thus, the exact mechanism is still under debate.

Computational studies could help identify the correct mechanism. With classical molecular dynamics (MD) simulations, the accessibility for different nucleophilic attacks can be determined, for example, for a direct attack of HsGrx1's sulfur atom $\text{S}_{\text{Grx}}^{23}$ on a disulfide bonded sulfur and, thus, the probability for the different reactions estimated. Moreover, the reaction barriers and reaction energies can be determined to assess whether or not the reactions are feasible. To this end, hybrid quantum mechanical/molecular mechanics (QM/MM) simulations of the systems can be employed, where the reactive part of the system is described quantum mechanically (QM) and the rest of the system with a classical force field.¹³ By generating thermodynamic ensembles, in which all relevant microstates of the systems are sufficiently sampled, the differences in free energy between the reactant and the product can be calculated and the correct pathways determined. However, if the two states are separated by a high barrier, the required sampling time for the transition most likely exceeds the time scale of *ab initio* and Density-Functional Theory (DFT) methods. Moreover, the configurational space of biomolecules exhibits many minima and saddle points due to the large number of degrees of freedom. Thus, many irrelevant areas of the high-dimensional free energy surface will be sampled or the system might get trapped in a local minimum from which it cannot escape during the simulation time. As an alternative, several methods have been developed that systematically search for a path connecting the reactant and product states. However, such a path will most likely contain many irrelevant saddle points or might not end up in the "true" product state, but rather a local minimum similar to the product state.

Hence, there is no way around sampling, and instead of *ab initio* or DFT methods, semi-empirical methods can be used, such as the Density Functional Tight-Binding (DFTB) approach. DFTB is derived from DFT with several well-defined approximations, which make DFTB calculations 2–3 orders of magnitude faster than DFT calculations, using medium sized basis sets.¹⁴ Extensive tests on organic molecules showed that for standard properties, such as geometries, reaction energies, and vibrational frequencies, a very good accuracy can be achieved, when an appropriate strategy for fitting E_{rep} is applied. On average, a slightly lower accuracy than DFT with medium sized to large basis sets can be obtained, however, being better than DFT-GGA with double-zeta basis sets.¹⁵ Also for reaction barriers, quite good agreement with full DFT calculations can be achieved.¹⁶ Although this quantitative error can be low on average, there are also cases where DFTB shows much larger deviation, which can be addressed as a qualitative error.¹⁷ It can have two reasons: (i) In the standard DFTB/3OB method, all tabulated integrals are computed from DFT using the Perdew–Burke–Ernzerhof (PBE) functional. Thus, DFTB has the same deficiencies as DFT, which are due to the well-known delocalization error. This error, e.g., shows up in the reaction between a thiolate anion and two disulfide bonded sulfurs.¹⁸ (ii) Furthermore, qualitative errors are related to the specific approximations inherent in DFTB, i.e., the usage of a minimal basis set and the integral approximations. Known examples are the proton affinities of nitrogen¹⁵ and the P–O bond.¹⁹ In these cases, the repulsive energy contribution has to be parameterized specifically for a certain bonding situation, which is called a specific reaction parameterization (SRP). The user has to decide with

the application at hand which parameterization to use—the standard parameters for a large class of molecules or the modified parameters for the specific application.

The accuracy that can be achieved with the standard DFTB parameterization is limited due to the fact that E_{rep} represents only a two-body interaction. Thus, the properties of linear bonds can be improved very much; however, bond and dihedral angles are not affected by E_{rep} and, therefore, cannot be improved. This was, for example, the reason for using a Δ ML approach to obtain accurate energies for the important dihedral angles in peptides where the rotational barriers are underestimated with DFTB due to the minimal basis set.²⁰ Since the Δ ML approach covers the many-body interaction, it can be used to correct both the qualitative and quantitative errors of the DFTB. This is because it combines both, the dependency of the potentials on the chemical environment (e.g., four-fold vs five-fold co-ordinated phosphorus) and the many-body nature of the repulsive energy, which are neglected in standard DFTB. At this point, we also note that the accuracy of semi-empirical QM/MM simulations can be improved by introducing an appropriate ML, perhaps a Δ ML correction to the electrostatic interactions between the QM and MM regions. Such methods are emerging, but appear to remain somewhat limited so far, for instance, incorporating the MM region directly in an NN potential has been applied to simple reactions,²¹ or the QM–MM interaction has been corrected in a short range.²² Since reliable Δ MLs of the QM–MM interactions for realistic systems, akin to the subject of this work, seem far from straightforward and in need of further developments, this work will use ML for the QM region and describe the QM–MM interactions on the original DFTB/MM level.

In a recent study, we corrected DFTB's description of the thiol-disulfide exchange, the reaction studied in this work, using the above mentioned approaches: the SRP and the Δ ML framework.¹⁷ A benchmark study by Neves *et al.* showed that proper inclusion of electron correlation is required to obtain the correct energetics and transition state geometry, where the three sulfur atoms are linearly aligned with short S–S bonds of about 2.4 Å.²³ However, with DFTB, the S–S bonds are too long, and the activation energy has an error of greater than 6 kcal/mol, due to errors inherited by the PBE functional. Thus, we performed an SRP, where we reparameterized the repulsive S–S pair potentials so that the energetics are consistent with the higher level QM methods, B3LYP and G3B3. This *ad hoc* workaround led to a good description of the thiol-disulfide exchange and yielded the correct transition state energy and geometry. In the Δ ML approach,^{20,24,25} we used a Behler–Parinello Artificial Neural Network^{26,27} (ANN) to learn the energy difference between DFTB and the higher level QM methods, B3LYP and coupled-cluster single double perturbative triple [CCSD(T)], for the thiol-disulfide exchange in the gas phase. The Δ ML scheme reproduced the energies and geometries of the thiol-disulfide exchange with *ab initio* accuracy, with a computational cost similar to that of a DFTB calculation. The ANN trained on gas-phase data could also be applied to thiol-disulfide exchange in water and to intramolecular thiol-disulfide exchanges in a protein, without loss of accuracy. Due to the low computational cost of the ANN, we were furthermore able to generate accurate multi-dimensional free energy profiles of the reactions from extensive molecular dynamics simulations and enhanced sampling methods.

In this work, we investigate the proposed catalytic reduction cycle of HMA4n by monothiol HsGrx1 with GSH as co-substrate.⁷ We perform classical MD simulations of the different systems and analyze the sulfur–sulfur distances and angles between the sulfurs to estimate the regioselectivity and accessibility for the various thiol-disulfide exchanges. The free energy profiles of the reactions are then obtained with extensive DFTB/MM metadynamics simulations using the specific reaction parameters and the Δ ML approach introduced in Ref. 17.

II. METHODS

A. Density-functional tight-binding

The semi-empirical Density-Functional Tight-Binding (DFTB) method is derived from the Density-Functional Theory (DFT) total energy. Expanding the Kohn–Sham total energy expression in a Taylor series up to the third order with respect to charge density fluctuations $\delta\rho(\mathbf{r}) = \rho(\mathbf{r}) - \rho^0(\mathbf{r})$ yields

$$\begin{aligned} E^{\text{DFTB}^3} &= E^0[\rho_0(\mathbf{r})] + E^1[\rho_0(\mathbf{r}), \delta\rho(\mathbf{r})] \\ &\quad + E^2[\rho_0(\mathbf{r}), (\delta\rho(\mathbf{r}))^2] + E^3[\rho_0(\mathbf{r}), (\delta\rho(\mathbf{r}))^3], \\ &= \frac{1}{2} \sum_{A,B} V_{AB}^{rep} + \sum_i^{\text{occ}} \sum_{\mu \in A, \nu \in B} c_{\mu}^i c_{\nu}^i H_{\mu\nu}^0 \\ &\quad + \frac{1}{2} \sum_{A,B} \gamma_{AB} \Delta q_A \Delta q_B + \frac{1}{3} \sum_{A,B} (\Delta q_A)^2 \Delta q_B \Gamma_{AB}. \end{aligned} \quad (1)$$

Truncation after the first-order term yields the original DFTB1 method.^{28,29} The zeroth-order term corresponds to the core–core repulsion and is approximated as a sum of repulsive two-body potentials V_{AB}^{rep} . These are fitted as spline functions to reference atomization energies of given molecular geometries.^{15,30} The first-order term is the electronic energy, approximated as the sum of occupied Kohn–Sham orbital energies. The orbitals are expanded in a confined minimal basis restricted to the valence electrons. In addition, a two-center approximation is used for the elements of the charge independent Hamiltonian matrix $H_{\mu\nu}^0$, which are precalculated with DFT using the PBE functional. Together with the corresponding overlap matrix, the Hamiltonian is stored in parameter files that are read during a DFTB calculation.

For systems that are sensitive to charge fluctuations, higher order terms have to be considered. Including the second- and third-order terms yields the self-consistent-charge DFTB methods, DFTB2³¹ and DFTB3.^{14,32} Here, the density fluctuations are approximated by charge monopoles that are represented by atomic point charges obtained from Mulliken population analysis. In the second order, the charge–charge interactions are described by the analytical γ_{AB} -function, which corresponds to a Coulomb interaction at large distances. At short distances ($A = B$), the on-site repulsion is accounted for by the Hubbard parameter U_A , which is related to the atomic hardness of an isolated atom. For a better description of charge fluctuations, the third-order term Γ_{AB} -function is introduced, which includes the charge derivative of the Hubbard parameter U_A^d .

Several parameter sets are available for DFTB, such as the 3OB set, which is recommended for organic and biological systems. However, some reactions are inaccurately described by this set, for example the thiol-disulfide exchange. In our recent work, we

improved the performance of DFTB for the thiol–disulfide exchange with a specific reaction parameterization (SRP).¹⁷ The sulfur–sulfur pair repulsive parameters in Eq. (1) were reparameterized such that the potential energy surface (PES) of the thiol–disulfide exchange obtained with DFTB agrees with the PES obtained with B3LYP using the def2-TZVPP basis set. A detailed description can be found in Ref. 17.

B. Artificial neural network correction for thiol–disulfide exchange

The accuracy of DFTB can also be improved with machine learning techniques, such as the Δ ML approach.^{20,24,25} In a recent study,¹⁷ we used a Behler–Parinello Artificial Neural Network^{26,27} (ANN) to learn the energy differences between DFTB/3OB and CCSD(T)/aug-cc-pVTZ for the thiol–disulfide exchange

$$\Delta E = E_{\text{CCSD(T)}} - E_{\text{DFTB}}. \quad (2)$$

In proteins, a thiol–disulfide exchange occurs between two disulfide-bonded cysteines and a deprotonated cysteine anion. A QM description of the three amino acids requires that the bonds between $C\alpha$ and $C\beta$ are capped with a hydrogen atom, resulting in a dimethyl disulfide and a methylthiolate anion. The ANN used these two molecules consisting of 15 atoms as input. The complete training set included 13 548 different structures of the molecules, which were generated with two different approaches. The first subset of 5112 different structures was obtained from an unrelaxed potential energy scan, where the sulfurs were linearly aligned, as described in Ref. 18. To obtain more non-linear structures, the intramolecular thiol–disulfide exchange in a small protein was enforced with QM/MM metadynamics and snapshots of the QM region were extracted and binned regarding the sulfur–sulfur distances. This yielded 8436 bins, and one structure from each bin was added to the second subset. To achieve a normalized distribution of the data, some data points were duplicated, resulting in a final dataset of 18 357 structures.

Since the Cartesian coordinates are not invariant regarding rotation, translation, and permutation of identical atoms, they were converted to a vector of radial and angular atom-centered symmetry function (ACSF) values. ACSFs are functions of the interatomic distances R_{ij} and angles θ_{ijk} between atoms within a cut-off sphere of radius R_c ,

$$G_i^{\text{rad}} = \sum_{\text{atoms } j \neq i}^{\text{within } R_c} e^{-\eta(R_{ij}-R_{s,m})^2},$$
$$G_i^{\text{ang}} = \sum_{\text{atoms } j,k \neq i}^{\text{within } R_c} \frac{2^{1-\zeta} (1 + \lambda \cos \theta_{ijk})^\zeta}{e^{\eta(R_{ij}^2 + R_{jk}^2 + R_{ik}^2)}}. \quad (3)$$

By varying the hyperparameters η , ζ , and λ , different function values can be generated for the same structure. The optimal combination of hyperparameters was obtained with a genetic algorithm. The ANN itself comprised two hidden layers for each atom, each consisting of 15 neurons, using the tanh as activation function. The final model featured a root-mean-square error (RMSE) of 0.62 kcal/mol.

In order to use the machine learned energy correction in quantum chemical calculations, the ANN was implemented in

DFTB+ as a new module that does not rely on external libraries. The forces acting on the atoms are obtained as

$$F_{\text{DFTB+}}^\alpha = F_{\text{DFTB}}^\alpha - \sum_i^{\text{atoms}} \sum_m^{\text{ACSF on atom } i} \frac{\partial E}{\partial G_{im}} \cdot \frac{\partial G_{im}}{\partial \alpha}, \quad (4)$$

where F_{DFTB} is the forces obtained from DFTB, α the Cartesian coordinates of the atoms, $\partial E/\partial G$ the derivatives of the ANN, and $\partial G/\partial \alpha$ the derivatives of the ACSFs. The modified version of DFTB+ is freely available on Github.³³ For more details about the ANN architecture, the training, and the implementation, please consult the original work in Ref. 17.

C. System setup and simulation details

1. Structural models

The structural model of HsGrx1 is based on a nuclear magnetic resonance (NMR) solution structure of dithiol HsGrx1 in the fully reduced form with PDB ID: 1JHB.¹¹ To generate the monothiol model, the C-terminal cysteine (Cys26) in the active site was mutated to a serine with PyMOL's Mutagenesis Wizard tool.³⁴ The sulfur atom of the enzymatically active Cys23 was prepared as an anion to enable the reaction with disulfide bonded sulfurs.

For the protein substrate HMA4n, the NMR solution structure with PDB ID: 2KKH was used.¹² The additional water molecules and the zinc ion in the structure were removed, and a disulfide bond between Cys27 and Cys28 was formed by manually reducing the distances between the sulfur atoms.

The sulfur atom of glutathione's cysteine was prepared as an anion (GS^-) to enable reactions with disulfide bonded sulfurs; the structure was generated with PyMOL Builder tool.³⁴ GS^- contains a γ -glutamyl residue, which is not included in the standard parameterization of the AMBER99SB-ILDN force field³⁵ used in this work, making it necessary to construct its topology and obtain the atomic charges. For the parameterization, the zwitterionic molecule γ -glutamyl-(N-methyl)amide zwitterion was geometry optimized at the level B3LYP/6-31G* using the polarizable continuum model (PCM) at 300 K. The application of PCM was necessary to prevent an undesired proton transfer from the amino and to the carboxyl group, which had occurred in pilot optimizations of the molecule in the gas phase. Then, the electrostatic potential induced in the surroundings of the molecule was obtained at the level HF/6-31G*, and was determined on four layers of points surrounding the molecule, starting at a distance of 1.4 times the Merz–Kollman radius of the respective nearest atom and spaced by 0.2 times the radius, with the density of 1 point per \AA^2 . (The Merz–Kollman radii are 1.2, 1.5, 1.5, and 1.4 \AA for H, C, N, and O, respectively.) The quantum chemical calculations were performed with GAUSSIAN 09 A.02.³⁶ The atomic charges were obtained with the two-stage restrained electrostatic potential fit,³⁷ as implemented in the Antechamber tool in the AmberTools suite.³⁸ The charges of the peptide backbone atoms were constrained to their respective values occurring in the standard AMBER force field. Finally, the capping group NH-CH_3 was removed, to yield the topology and parameter file for an N-terminal γ -glutamyl residue, termed GGL.

2. Classical simulations

All classical simulations were performed with GROMACS 2020.2³⁹ pathed with PLUMED 2.6.1.^{40,41} The AMBER99SB-ILDN³⁵

force field was used, periodic boundary conditions were set, and electrostatic Lennard-Jones interactions were calculated using a cut-off of 1 nm. Long-range electrostatic interactions were calculated by particle-mesh Ewald summation; the neighbor list was updated every ten MD steps. The leap-frog integrator was used with a time step of 2 fs. Initial velocities of the atoms were assigned from the Maxwell-Boltzmann distribution at 300 K. The temperature was maintained at 300 K by the Bussi thermostat⁴² with $\tau_T = 0.1$ ps, in the NVT- and NPT-ensemble. In NPT simulations, the pressure was kept constant with the Parrinello-Rahman barostat⁴³ at $p = 1$ bar and $\tau_p = 2.0$ ps.

For the first catalytic **step 1** shown in Fig. 1(b), monothiol HsGrx1 and HMA4n were placed together in a cubic simulation box. The distance between the molecules and the box was about 2 nm, which corresponds to a box size of $\sim 11 \times 11 \times 11$ nm³. The box was solvated with 36 436 TIP3P water molecules and electro-neutralized by adding four chloride atoms. The system was shortly equilibrated for 10 ns in a NVT ensemble and subsequently for 10 ns in a NPT ensemble, during which harmonic position restraints were applied to the heavy atoms with a force constant of $10\,000$ kJ mol⁻¹ nm⁻². A 500 ns NPT simulation followed, from which two structures were selected that showed a large difference in their root-mean-square deviation (RMSD) and radius of gyration. Each structure was then simulated for 500 ns after assigning new initial velocities from the Maxwell-Boltzmann distribution at 300 K. During the simulations, the distances between the sulfur atom of HsGrx1 and the sulfur atoms of HMA4n were restrained to values smaller than 9 Å with a force constant of $10\,000$ kJ mol⁻¹ nm⁻² to reduce the conformational sampling.

For the reaction **step 2**, a snapshot of the system containing the Grx-S-S-HMA4n intermediate was selected from a QM/MM metadynamics simulation of reaction 1. Due to the previous thiol-disulfide exchange, one of the sulfur atoms of HMA4n was deprotonated and thus carried a negative charge. To prepare the system for a nucleophilic attack by GS⁻, the following equilibration scheme was applied. First, the system was equilibrated with the deprotonated sulfur atom for 100 ns. Next, the sulfur atom was protonated, one chloride ion was added, and the system was equilibrated for 100 ns. Then the GS⁻ molecule was introduced, and two chloride ions were introduced into the system, which was simulated for another 100 ns. Two structures with high mutual RMSD and radius of gyration were selected from the obtained trajectory and, analogous to the procedure described in the previous paragraph, used as starting structure for two additional 500 ns long simulations with new initial velocities. The distances between the sulfur atom of GS⁻ and the sulfur atoms of the Grx-S-S-HMA4n intermediate were restrained to values smaller than 9 Å with a force constant of $50\,000$ kJ mol⁻¹ nm⁻².

For the third and final catalytic **step 3**, a structure of the HsGrx1-S-SG intermediate was taken from a QM/MM metadynamics simulation of reaction 2. The molecule was placed in a cubic box of $\sim 6 \times 6 \times 6$ nm³, solvated with 8126 TIP3P water molecules, electro-neutralized with one sodium ion and equilibrated for 100 ns. The nucleophile GS⁻ was then introduced, followed by another simulation of 100 ns, from which two starting structures were selected according to the same criteria described for reactions (1) and 2. Analogously, two 500 ns long simulations were performed with new initial velocities of the atoms, keeping the S-S distances

between GS⁻ and HsGrx1-S-SG below 6 Å with a force constant of $10\,000$ kJ mol⁻¹ nm⁻².

The 2×500 ns long trajectories per system were analyzed to estimate the regioselectivity of the possible thiol-disulfide exchanges. To this end, 2D histograms of the distances between the nucleophilic sulfur atoms and the two respective target sulfur atoms were obtained with a bin size of 0.1 Å.

3. QM/MM simulations

All QM/MM simulations were performed with a local GROMACS 2020³⁹ version patched with PLUMED 2.5.1^{40,41} and interfaced with a modified DFTB+ 19.1⁴⁴ in which the Δ ML correction was implemented. Analogous to previous studies,^{17,45} the QM regions of the respective systems comprised the side chains of the three cysteines for which a thiol-disulfide exchange was investigated. The link atom approach was used to cap the QM region with hydrogen atoms, which were placed along the C α and C β bonds at a fixed distance. The QM regions consisted of 15 atoms, described by the semi-empirical density functional theory method DFTB3, and two different sets of QM/MM simulations of the catalytic cycle were performed. The first set used the 3OB parameter set¹⁹ with a reparameterized S-S repulsive potential based on B3LYP/def2-TZVPP data.¹⁷ In the second set, the artificial neural network correction that learned the energy difference between DFTB with the unmodified 3OB parameter set and CCSD(T)/aug-cc-pVTZ level of theory was applied.¹⁷ The rest of the systems were described with the AMBER99SB-ILDN³⁵ force field and TIP3P water.

For each system, a starting structure was taken from one of the two respective classical 500 ns MD simulations that met an accessibility criterion for a possible nucleophilic attack. An attack was considered possible if the distance between the nucleophilic sulfur and the target was less than 5 Å, and the attack angle greater than 130°. The systems were equilibrated for 100 ns in an NPT ensemble with the same settings as for the classical simulations, with two changes. A time step of 0.5 fs was used, and the electrostatic interactions between the QM region and the MM system were scaled down by a factor of 0.75 to compensate for the missing polarization of the MM environment.^{46,47}

4. QM/MM metadynamics

The free energy profiles of the thiol-disulfide exchanges using the SRP and the Δ ML approach were obtained with QM/MM well-tempered multiple walker metadynamics.⁴⁸⁻⁵¹ A 2D setup was considered, where the S-S distances between the respective nucleophilic sulfur atoms S_{nuc} and the attacked sulfur atoms S_{ctr} were used as the first reaction coordinate, and the S-S distances between the attacked sulfur atoms S_{ctr} and the leaving sulfur atoms S_{lg} as the second reaction coordinate. Depending on the system, the S-S distances were restrained to values smaller than 10 Å with a force constant of $10\,000$ kJ mol⁻¹ nm⁻² to reduce conformational sampling. Moreover, the S_{nuc}-S_{ctr}-S_{lg} angles were restrained to values larger than 130° with a force constant of $100\,000$ kJ mol⁻¹ rad⁻¹. Either 16 or 24 walkers were used with a simulation time of at least 1.8 or 1.4 ns per Walker, respectively. This resulted in a total simulation time of at least 28.8 ns for each system. Gaussian potentials with a width of 0.2 Å and an initial height of 0.5 kJ mol⁻¹ were deposited and read every 500 fs. A bias factor of $\gamma = 20$ was used. Following the setup in Refs. 17 and 45, the sum of switching functions depending on the

three S–S distances was restrained to prevent bond breaking of the disulfide bond, while the sulfur anion S_{nuc} is too far away. Restraints were also applied to the coordination numbers of the C β atoms with their bonded hydrogen atoms to prevent deprotonation by the sulfur anion at short distances.

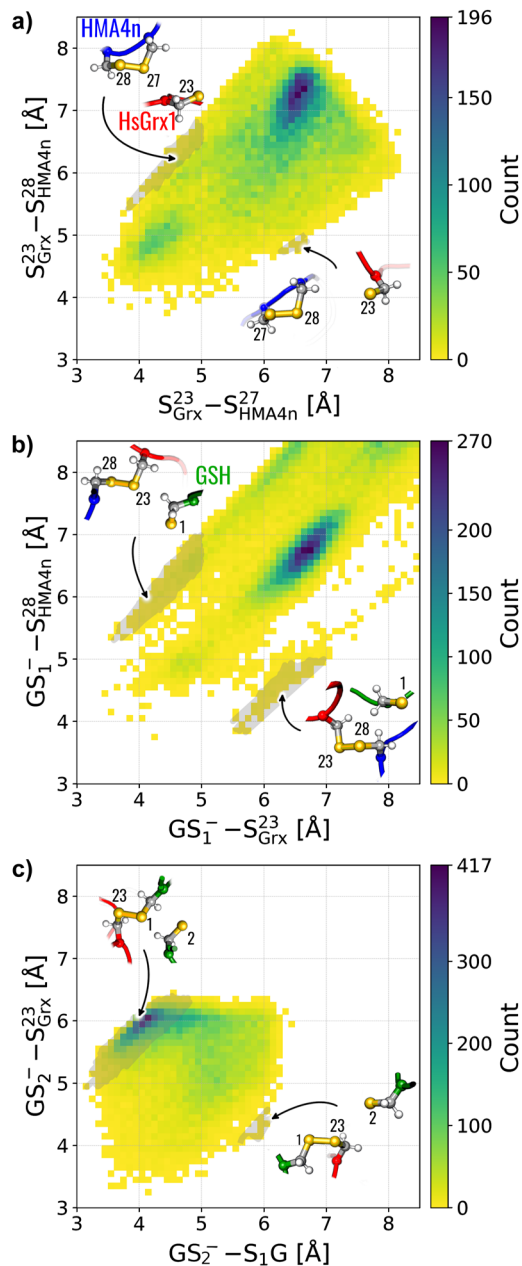


FIG. 2. Histograms of the sulfur–sulfur distances calculated from 2×500 ns long classical MD simulations of the respective systems of each reaction step. (a) Histogram of the $S_{\text{Grx}}^{23}-S_{\text{HMA4n}}^{27}$ and $S_{\text{Grx}}^{23}-S_{\text{HMA4n}}^{28}$ distances for **step 1**. (b) Histogram of the $GS_1^- - S_{\text{Grx}}^{23}$ and $GS_1^- - S_{\text{HMA4n}}^{28}$ distances for **step 2**. (c) Histogram of the $GS_2^- - S_1G$ and $GS_2^- - S_{\text{Grx}}^{23}$ distances for **step 3**. The gray shaded areas correspond to structures that meet the accessibility criterion.

III. RESULTS AND DISCUSSION

The proposed monothiol reduction cycle of HsGrx1 involves three catalytic steps; see Fig. 1(d).⁷ In each step, the respective nucleophilic sulfur anion S_{nuc} can attack one of the disulfide bonded sulfurs, resulting in two possible thiol–disulfide exchanges per step. The cycle starts with reaction 1, where the sulfur anion of HsGrx1 S_{Grx}^{23} can either attack sulfur atom S_{HMA4n}^{27} or S_{HMA4n}^{28} of the protein substrate HMA4n. In order to proceed, **step 2** requires an attack of the glutathione thiolate GS_1^- on the disulfide bonded S_{Grx}^{23} . However, the disulfide bonded sulfur atom of HMA4n (S_{HMA4n}^{27} or S_{HMA4n}^{28}) could also be attacked, which would then lead to a different pathway. In the final **step 3**, a second glutathione anion GS_2^- must attack the disulfide bonded glutathione GS_1 to form GSSG, but an attack on S_{Grx}^{23} is also conceivable, although this would lead to the same mixed disulfide molecule (HsGrx1–S–SG).

To investigate the regioselectivity, we split the calculations into two parts: In the first step, we estimated the accessibility of a certain reactant conformation using MD simulations. That is, we sampled the reactant conformations, which allowed to attack a specific sulfur atom and computed the probabilities of special sub-ensembles, which may support different reaction pathways. These simulations were performed with a classical force field because of the long times needed (500 ns) and also because sufficient accuracy is expected for the description of different protein conformations. In the second step, we use the conformations of these sub-ensembles to estimate the reaction barriers for the respective reaction pathways by means of QM/MM metadynamics simulations. Both steps together lead to an estimate of the regioselectivity.

A. Regioselectivity

To estimate the accessibility, we calculated 2D histograms of the distances between the respective nucleophilic sulfur atoms S_{nuc} and the potential nucleophilic targets S_{ctr} from classical MD simulations, compare Fig. 2. We then counted how many times S_{nuc} approaches the target sulfur atoms such that an attack complex can be formed. Only geometries where the $S_{\text{nuc}}-S_{\text{ctr}}$ distances were less than 5 Å and the angle between the three sulfur atoms $\angle(S_{\text{nuc}}S_{\text{ctr}}S_{\text{lig}}) \geq 130^\circ$ were considered. For each system, the accessibility for each attack was calculated, expressed as the percentage of the structures that met the criterion, see Table I.

TABLE I. Estimated accessibility of the different possible nucleophilic attacks in the catalytic reduction cycle of HsGrx1, based on the histograms shown in Fig. 2. An attack was considered feasible if the distance between the nucleophilic and the attacked sulfur atom ($S_{\text{nuc}}-S_{\text{ctr}}$) was less than 5 Å and the attack angle between the sulfurs $\angle(S_{\text{nuc}}S_{\text{ctr}}S_{\text{lig}}) \geq 130^\circ$. The accessibility is expressed as the sum of structures that met the criterion divided by the total number of counts.

Step	Attack	Accessibility (%)
1	Grx-S23 [−] → S27-HMA4n	0.54
	Grx-S23 [−] → S28-HMA4n	0.04
2	GS ₁ [−] → S23-Grx	1.30
	GS ₁ [−] → S28-HMA4n	0.39
3	GS ₂ [−] → GS ₁	28.04
	GS ₂ [−] → S23-Grx	0.04

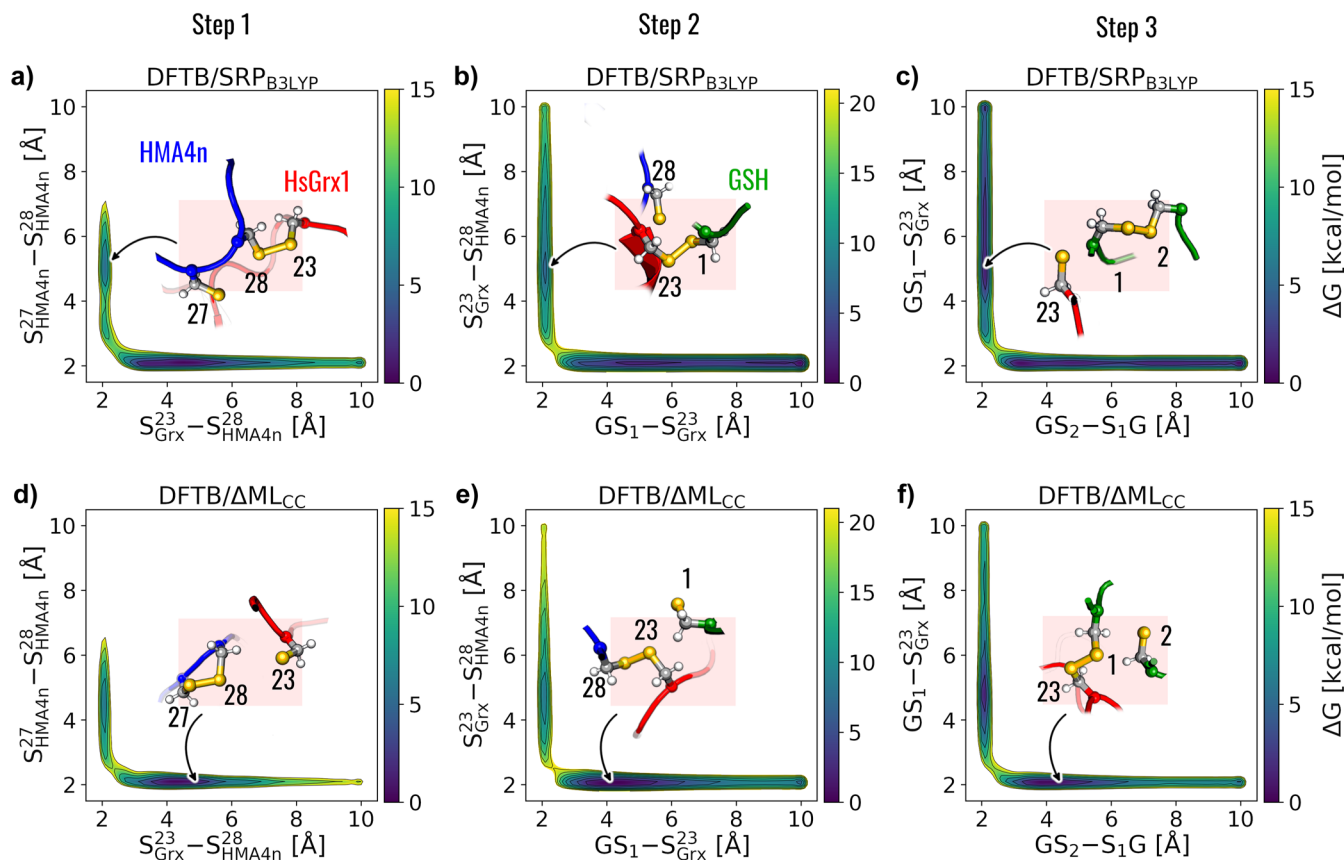


FIG. 3. Potentials of the mean force (PMFs) of the three consecutive thiol–disulfide exchanges in the monothiol route of HsGrx1 with HMA4n as substrate and GSH as co-substrate. PMFs of **steps 1 to 3** (left to right) obtained with DFTB using the SRP (a–c) and with DFTB using the Δ ML correction (d–f). Contour lines are drawn every 2.5 kcal/mol.

We find that in **step 1**, S_{HMA4n}^{27} might be attacked more often than S_{HMA4n}^{28} by S_{Grx}^{23} , compare Fig. 2(a). The estimated accessibilities are 0.54% and 0.04%, respectively. However, QM/MM metadynamics of this attack showed a very large barrier of more than 23 kcal/mol (Fig. S1), whereas the barrier height for an attack on S_{HMA4n}^{28} is much smaller with ~ 12 kcal/mol [Figs. 3(a) and 3(d)]. Therefore, S_{HMA4n}^{28} is considered as a nucleophilic target in the following; further details on this choice are provided in the [supplementary material](#).

In **step 2**, the mixed disulfide bond $S_{\text{Grx}}^{23}-S_{\text{HMA4n}}^{28}$ between the enzyme and the protein substrate is then attacked by GS_1^- ; see Fig. 2(b). In our classical simulations, GS_1^- can easily attack S_{Grx}^{23} with an accessibility of 1.30% compared to 0.39% for an attack on S_{HMA4n}^{28} . In the last **step**, another GS_2^- molecule attacks one sulfur atom of the newly formed disulfide bond $\text{GS}_1^- - S_{\text{Grx}}^{23}$, compare Fig. 2(c). An attack on GS_1^- is more likely to occur with an accessibility of 28.04% compared to an attack on S_{Grx}^{23} with 0.04%. The regioselectivity of **step 1** is not known from experiments, but our estimated regioselectivities for **step 2** and **step 3** are the same as those experimentally observed by Ukuwela *et al.*⁷ They find only small traces of the glutathionylated HMA4n, i.e., the less likely attack we find for step 2, and also only small traces of the glutathionylated HsGrx1 enzyme,

i.e., the product of **step 2** and the less likely attack we find for **step 3**. To further investigate the possible reaction pathways, free energy profiles of the three regioselective attacks were computed with QM/MM metadynamics simulations.

B. QM/MM metadynamics

Each attack was simulated twice, using either the SRP or the Δ ML energy correction. The obtained potentials of the mean force are shown in Fig. 3 and the reaction barriers in Table II. The barrier heights and reaction energies determined by the two methods agree very well, with differences ranging from 0.2 to 1.5 kcal/mol.

In **step 1**, the disulfide bond between $S_{\text{HMA4n}}^{27}-S_{\text{HMA4n}}^{28}$ is the global minimum on the free energy surface and the reduction product $S_{\text{Grx}}^{23}-S_{\text{HMA4n}}^{28}$ is a local minimum. Note that sulfur atom S_{HMA4n}^{27} is deprotonated in the product state due to the QM/MM metadynamics setup. Using DFTB/SRP or DFTB/ Δ ML as QM method, the heights of the barriers are $\Delta G_{\text{red}}^\ddagger = 11.3$ and 12.5 kcal/mol, respectively. The corresponding reaction energies are $\Delta G = +5.0$ and $+5.8$ kcal/mol and consequently, the barrier heights for the reverse oxidation reaction are $\Delta G_{\text{ox}}^\ddagger = 6.3$ and 6.7 kcal/mol, respectively.

TABLE II. Reaction barriers of the catalytic reactions between HsGrx1 (Grx), the protein substrate HMA4n (P), and GSH. The barriers were obtained with DFTB using the 3OB parameter set. Either the SRP or the Δ ML correction was employed. All values in kcal/mol.

Method	$\text{Grx-S}^- + \text{P-(S)}_2$ $\rightleftharpoons \text{Grx-S-S-P(S}^-)$	$\text{Grx-S-S-P(SH)} + \text{GS}^-$ $\rightleftharpoons \text{Grx-S-SG} + \text{P-(SH) (S}^-)$	$\text{Grx1-S-SG} + \text{GS}^-$ $\rightleftharpoons \text{Grx-S}^- + \text{GSSG}$
DFTB/SRP	11.3/6.3	18.6/12.1	12.1/11.1
DFTB/ Δ ML	12.5/6.7	20.1/12.1	12.4/11.4

In the next catalytic **step 2**, the global minimum corresponds to a disulfide bond between $\text{S}_{\text{Grx}}^{23} - \text{S}_{\text{HMA4n}}^{28}$ with $\text{S}_{\text{HMA4n}}^{27}$ being protonated in this setup. A nucleophilic attack of a glutathione anion on the sulfur atom of HsGrx1 leads to a disulfide bond $\text{GS}_1^- - \text{S}_{\text{Grx}}^{23}$, which corresponds to a local minimum on the free energy surface. The reaction energies are $\Delta G = +6.5$ and $+8.0$ kcal/mol with DFTB/SRP and DFTB/ Δ ML, respectively. The reaction barriers for the reduction and oxidation are $\Delta G_{\text{red}}^\ddagger = 18.6$ and 20.1 kcal/mol, respectively, and $\Delta G_{\text{ox}}^\ddagger = 12.1$ and 12.1 kcal/mol, respectively.

In the final **step 3**, a second glutathione anion attacks the first glutathione disulfide bonded to HsGrx1, forming a GSSG molecule, which is the local minimum on the free energy surface. With DFTB/SRP or DFTB/ Δ ML, the reaction energies are $\Delta G = 1.0$ kcal/mol, in each case. The barrier heights for the reduction are $\Delta G_{\text{red}}^\ddagger = 12.1$ and 12.4 kcal/mol, respectively, and for the oxidation $\Delta G_{\text{ox}}^\ddagger = 11.1$ and 11.4 kcal/mol, respectively.

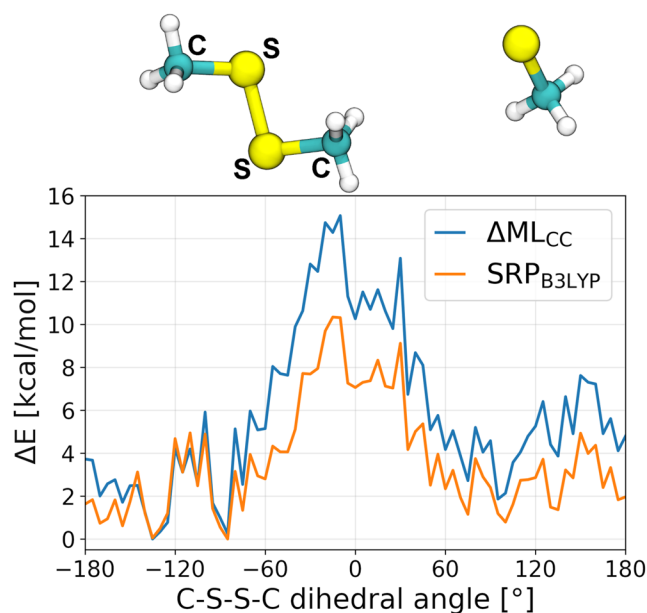
The free energy surfaces obtained with the two corrections differ not only in their energetics but also in their topography. With DFTB/ Δ ML, the minima are narrower, and the energy increases more steeply than with DFTB/SRP. For example, in **step 3**, the global minimum corresponding to the disulfide bond $\text{GS}_1^- - \text{S}_{\text{Grx}}^{23}$ spans $\text{GS}_2 - \text{S}_1\text{G}$ distances from 3.5 to 5.5 Å with DFTB/ Δ ML [horizontal minimum in Fig. 3(f)]. With DFTB/SRP, the minimum spans over distances from 3.5 to almost 10 Å [horizontal minimum in Fig. 3(c)]. These differences can be attributed to the treatment of many-body interactions. The modified repulsion potential E_{rep} of the SRP covers only two-body interactions, whereas the Δ ML accounts for many-body interactions. These become particularly important when the sulfur atoms are not linearly aligned, which can occur in our setup; the angles between the respective sulfur atoms can take values between 130° and 180° .

C. Evaluation of energy differences

We already provided a performance analysis of the machine-learning model in Ref. 17. Here, the previous analysis is expanded to provide further details, and also to compare the performance of the SRP approach with the Δ ML model. The generated potential energy surfaces and a discussion of the SRP accuracy are presented in the [supplementary material](#). This analysis revealed a considerably inaccurate description of the $-\text{S}-\text{S}-$ torsion with the SRP. The larger errors with DFTB/SRP can be attributed to the inaccurate description of the torsional potential between $\text{C}-\text{S}-\text{S}-\text{C}$ of the dimethyl disulfide. DFTB is known to underestimate rotational barriers due to the neglect of three-center and four-center interactions.^{20,52} Thus, we performed a potential energy scan of the $\text{C}-\text{S}-\text{S}-\text{C}$ dihedral

angle of the dimethyl disulfide in the gas phase with DFTB/SRP and DFTB/ Δ ML. The angle was scanned from -180° to 180° with 5° increments obtained from short 40 ps QM/MM simulations of the dimethyl disulfide and the methyl thiolate in the gas phase. The $\text{S}-\text{S}$ distances between the dimethyl disulfide and the methyl thiolate were ~ 7.0 Å. The obtained potential energy profile, the $\text{C}-\text{S}-\text{S}-\text{C}$ dihedral, is shown in Fig. 4. For a better comparability, the minimum at -135° was set to zero. Between -145° and -85° , the energies obtained with two corrections agree very well within 1 kcal/mol. However, for all other angles there are large energy differences up to 5.5 kcal/mol at -30° .

Thiol-disulfide exchange poses considerable challenges for computational approaches: first, the reaction barriers are highly dependent on the interaction with the environment, and the reaction in gas-phase is barrierless, while in solution or in protein environment, barriers of up over 20 kcal/mol can occur, as shown in this work and Ref. 17. Furthermore, the reaction is difficult to tackle for approximate electronic structure methods such as DFTB3, which is derived from DFT-GGA, as shown in detail in Ref. 23. This

**FIG. 4.** Potential energy profile of the $\text{C}-\text{S}-\text{S}-\text{C}$ dihedral angle of the dimethyl disulfide in the gas phase obtained with DFTB/SRP and DFTB/ Δ ML.

necessitates the augmentation of the delta learning approach. In combination with a semi-empirical method such as DFTB, this then allows extensive sampling of reaction barriers, needed to compute reliable free energy values. Since the correction is based on CCSD(T), the question about the absolute computational cost arises. Generating the CCSD(T) training data for 13 500 structures in Ref. 17 took in total ~12 node-days with a dual Xeon Silver 4214 setup, while the time to determine a single free energy barrier using DFTB-QM/MM metadynamics took ~12 days on the same node. In this work, we calculated a total of eight reaction barriers, i.e., the parameterization procedure was considerably faster than the actual computation of the free energies.

IV. CONCLUSION AND OUTLOOK

The mechanism of enzymatic disulfide bond reduction by glutaredoxins is still under debate, and several models have been proposed. In this work, we used classical and QM/MM simulations to investigate the reduction pathway proposed by Ukuwela *et al.*⁷ Based on their experiments, the reduction cycle should proceed over three consecutive regio-specific thiol–disulfide exchanges between the monothiol enzyme HsGrx1, the protein substrate HMA4n, and the co-substrate GSH. In our classical MD simulations, we find the same regioselectivity by means of an accessibility criterion from histograms of the S–S distances between the respective nucleophilic and the two target sulfur atoms. We then obtained the free energy profiles of each attack with QM/MM metadynamics and determined the barrier heights and reaction energies to analyze whether the enzymatic reactions were feasible or not.

This approach requires extensive phase space sampling, which poses a challenge for *ab initio* and DFT methods due to their high computational cost. As an alternative, minimum energy pathway algorithms can be employed or relaxed geometry optimization calculations for stationary points along a reaction coordinate performed, as done by Neves, Fernandes, and Ramos for glutathione disulfide reduction by protein disulfide isomerase.⁵³ However, such approaches usually neglect motions of the environment, which can lead to errors, and approximations for the entropy have to be used to obtain the Gibbs free energy. In contrast, adequate sampling accounts for all important configurational changes, and the Gibbs free energy is obtained directly. Hence, we used the semi-empirical DFTB method as a QM method, which is about 3 orders of magnitude faster than DFT-GGA, using moderately sized basis sets. However, DFTB fails to accurately describe the thiol–disulfide exchange reaction, as do most DFT-GGA functionals due to correlation effects.²³ As a workaround, we developed specific reaction parameters (SRP) and a machine learned energy correction Δ ML used in a recent study¹⁷ and used them here for the free energy calculations. The computational cost of the Δ ML correction is comparable to a DFTB calculation. With both corrections, the obtained barrier heights are about 12 kcal/mol for **steps 1** and **3**, and about 20 kcal/mol for **step 2**, which lie in the order of magnitude of typical enzymatic reactions. Therefore, our calculations indicate that disulfide bond reduction by HsGrx1 most likely occurs via the mechanism proposed by Ukuwela *et al.*

Still, there is an interesting question of the relative merit of the SRP and Δ ML approaches, as Δ ML offers a potential for higher

accuracy and greater flexibility for a somewhat increased computational cost. In our current work, there are minor differences in the obtained free energy profiles: The Δ ML correction yielded barriers that are up to 1.5 kcal/mol higher than those obtained with the SRP, and the minima are somewhat narrower than with the SRP. It would be important to know in what situations this agreement may break down, with the Δ ML approach possibly becoming superior. Comparing the corrections with their respective reference methods, B3LYP for the SRP and CCSD(T) for the Δ ML correction, showed that the SRP reproduces the reference energies at short S–S distances very well, but exhibits large errors at larger S–S distances. In addition, the SRP underestimates the torsional energy of the C–S–S–C dihedral angle between disulfide bonded molecules. The Δ ML correction, on the other hand, reproduces the reference energies for all given structures accurately. The incorrect DFTB3/3OB description of torsions –S–S– remains uncorrected and wrong with the SRP, but it is within the applicability area of the Δ ML correction. Thus, the reason for the good performance of the SRP in the current study seems to be the fact that only S–S distances vary during the reactions—for which the SRP could be and was parameterized—while no other more complex degrees of freedom such as –S–S– torsions play a role. In a more general case potentially involving those, PMFs obtained with the Δ ML correction would be considered more accurate.

Nevertheless, all investigated reactions are uphill, which is unexpected for enzymatic reactions. This can be attributed to the QM/MM setup, since we did not consider the subsequent protonation of the cysteinyl thiols S28 and S27 of the protein substrate HMA4n for technical reasons. The protonated HMA4n species are most likely lower in energy than the deprotonated ones, since most cysteinyl thiols are considered to have a pK_a of 8 or higher. In contrast, the enzymatically active cysteinylthiols of glutaredoxins typically have a $pK_a < 5$ and are therefore usually deprotonated and readily available to initiate a nucleophilic attack. Moreover, the direction of the catalytic cycle is controlled by the concentration of GSH. An excess of GSH leads to the reduction of the disulfide bond of the substrate and an excess of GSSG to oxidation. Considering all these facts, we can conclude that disulfide bond reduction by HsGrx1 most likely proceeds via the mechanism proposed by Ukuwela *et al.* In a future work, the regioselectivities and barrier heights for other mechanistic models can be studied and compared with the results presented here.

SUPPLEMENTARY MATERIAL

See the [supplementary material](#) for force field parameters of GSH, additional results on the regioselectivity and free energy surface for **step 1**, and additional performance analysis of the machine-learning model.

ACKNOWLEDGMENTS

This work was supported by the German Science Foundation (DFG) under Project No. GRK 2450 and further by the state of Baden-Württemberg through bwHPC and by the DFG through Project No. INST 40/467-1 FUGG (JUSTUS cluster).

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Julian Böser: Formal analysis (equal); Investigation (equal); Methodology (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). **Tomáš Kubař:** Formal analysis (supporting); Funding acquisition (supporting); Investigation (supporting); Methodology (equal); Writing – original draft (supporting); Writing – review & editing (supporting). **Marcus Elstner:** Formal analysis (supporting); Funding acquisition (equal); Methodology (equal); Project administration (equal); Resources (lead); Writing – original draft (supporting); Writing – review & editing (equal). **Denis Maag:** Formal analysis (equal); Investigation (equal); Methodology (equal); Project administration (lead); Visualization (equal); Writing – original draft (lead); Writing – review & editing (equal).

DATA AVAILABILITY

The data that support the findings of this study are available within the article and its [supplementary material](#).

REFERENCES

- 1 L. E. S. Netto, M. A. de Oliveira, C. A. Tairum, and J. F. da Silva Neto, *Free Radical Res.* **50**, 206 (2016).
- 2 F. T. Ogata, V. Branco, F. F. Vale, and L. Coppo, *Redox Biol.* **43**, 101975 (2021).
- 3 M. Deponte, *Biochim. Biophys. Acta, Gen. Subj.* **1830**, 3217 (2013).
- 4 C. H. Lillig, C. Berndt, and A. Holmgren, *Biochim. Biophys. Acta, Gen. Subj.* **1304**, 1780 (2008).
- 5 Y. Yang, S.-c. Jao, S. Nanduri, D. W. Starke, J. J. Mielay, and J. Qin, *Biochemistry* **37**, 17145 (1998).
- 6 A. A. Ukuwela, A. I. Bush, A. G. Wedd, and Z. Xiao, *Biochem. J.* **474**, 3799 (2017).
- 7 A. A. Ukuwela, A. I. Bush, A. G. Wedd, and Z. Xiao, *Chem. Sci.* **9**, 1173 (2018).
- 8 J. Zimmermann, J. Oestreicher, S. Hess, J. M. Herrmann, M. Deponte, and B. Morgan, *Redox Biol.* **36**, 101598 (2020).
- 9 M. M. Gallogly, D. W. Starke, and J. J. Mielay, *Antioxid. Redox Signaling* **11**, 1059 (2009).
- 10 C. H. Lillig and C. Berndt, *Antioxid. Redox Signaling* **18**, 1654 (2013).
- 11 C. Sun, M. J. Berardi, and J. H. Bushweller, *J. Mol. Biol.* **280**, 687 (1998).
- 12 M. Zimmermann, O. Clarke, J. M. Gulbis, D. W. Keizer, R. S. Jarvis, C. S. Cobbett, M. G. Hinds, Z. Xiao, and A. G. Wedd, *Biochemistry* **48**, 11640 (2009).
- 13 A. Warshel and M. Levitt, *J. Mol. Biol.* **103**, 227 (1976).
- 14 M. Gaus, Q. Cui, and M. Elstner, *J. Chem. Theory Comput.* **7**, 931 (2011).
- 15 M. Gaus, A. Goetz, and M. Elstner, *J. Chem. Theory Comput.* **9**, 338 (2013).
- 16 M. Gruden, L. Andjeklović, A. K. Jissy, S. Stepanović, M. Zlatar, Q. Cui, and M. Elstner, *J. Comput. Chem.* **38**, 2171 (2017).
- 17 C. L. Gómez-Flores, D. Maag, M. Kansari, V.-Q. Vuong, S. Irle, F. Gräter, T. Kubař, and M. Elstner, *J. Chem. Theory Comput.* **18**, 1213 (2022).
- 18 M. Putzu, F. Gräter, M. Elstner, and T. Kubař, *Phys. Chem. Chem. Phys.* **20**, 16222 (2018).
- 19 M. Gaus, X. Lu, M. Elstner, and Q. Cui, *J. Chem. Theory Comput.* **10**, 1518 (2014).
- 20 J. Zhu, V. Q. Vuong, B. G. Sumpter, and S. Irle, *MRS Commun.* **9**, 867 (2019).
- 21 L. Bösel, M. Thürlmann, and S. Riniker, *J. Chem. Theory Comput.* **17**, 2641 (2021).
- 22 T. J. Giese, J. Zeng, Ş. Ekesan, and D. M. York, *J. Chem. Theory Comput.* **18**, 4304 (2022).
- 23 R. P. P. Neves, P. A. Fernandes, A. J. C. Varandas, and M. J. Ramos, *J. Chem. Theory Comput.* **10**, 4842 (2014).
- 24 R. Ramakrishnan, P. O. Dral, M. Rupp, and O. A. von Lilienfeld, *J. Chem. Theory Comput.* **11**, 2087 (2015).
- 25 L. Shen and W. Yang, *J. Chem. Theory Comput.* **14**, 1442 (2018).
- 26 J. Behler and M. Parrinello, *Phys. Rev. Lett.* **98**, 146401 (2007).
- 27 J. Behler, *Chem. Rev.* **121**, 10037 (2021).
- 28 D. Porezag, T. Frauenheim, T. Köhler, G. Seifert, and R. Kaschner, *Phys. Rev. B* **51**, 12947 (1995).
- 29 G. Seifert, D. Porezag, and T. Frauenheim, *Int. J. Quantum Chem.* **58**, 185 (1996).
- 30 M. Gaus, C.-P. Chou, H. Witek, and M. Elstner, *J. Phys. Chem. A* **113**, 11866 (2009).
- 31 M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai, and G. Seifert, *Phys. Rev. B* **58**, 7260 (1998).
- 32 Y. Yang, H. Yu, D. York, Q. Cui, and M. Elstner, *J. Phys. Chem. A* **111**, 10861 (2007).
- 33 T. Kubař, <https://github.com/tomaskubar/dftbplus/tree/machine-learning>, 2021 (last accessed on 27 July 2022).
- 34 Schrödinger, LLC and W. DeLano, “PyMOL,” 2021.
- 35 K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, and D. E. Shaw, *Proteins: Struct., Funct., Bioinf.* **78**, 1950 (2010).
- 36 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, GAUSSIAN 09, Revision A.02, Gaussian, Inc., 2009.
- 37 C. I. Bayly, P. Cieplak, W. Cornell, and P. A. Kollman, *J. Phys. Chem.* **97**, 10269 (1993).
- 38 J. Wang, W. Wang, P. A. Kollman, and D. A. Case, *J. Mol. Graph. Model.* **25**, 247 (2006).
- 39 M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl, *SoftwareX* **1–2**, 19 (2015).
- 40 M. Bonomi, G. Bussi, C. Camilloni, G. A. Tribello, P. Banáš, A. Barducci, M. Bernetti, P. G. Bolhuis, S. Bottaro, D. Branduardi, R. Capelli, P. Carloni, M. Ceriotti, A. Cesari, H. Chen, W. Chen, F. Colizzi, S. De, M. De La Pierre, D. Donadio, V. Drobot, B. Ensing, A. L. Ferguson, M. Filizola, J. S. Fraser, H. Fu, P. Gasparotto, F. L. Gervasio, F. Giberti, A. Gil-Ley, T. Giorgino, G. T. Heller, G. M. Hocky, M. Iannuzzi, M. Invernizzi, K. E. Jelfs, A. Jussupow, E. Kirilin, A. Laio, V. Limongelli, K. Lindorff-Larsen, T. Löhner, F. Marinelli, L. Martin-Samos, M. Masetti, R. Meyer, A. Michaelides, C. Molteni, T. Morishita, M. Nava, C. Pisoni, E. Papaleo, M. Parrinello, J. Pfandner, P. Piaggi, G. Piccini, A. Pietropaolo, F. Pietrucci, S. Pipolo, D. Provasi, D. Quigley, P. Raiteri, S. Raniolo, J. Ryzewski, M. Salvalaglio, G. C. Sosso, V. Spiwok, J. Šponer, D. W. H. Swenson, P. Tiwary, O. Valsson, M. Vendruscolo, G. A. Voth, and A. White (PLUMED Consortium), *Nat. Methods* **16**, 670 (2019).
- 41 G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, and G. Bussi, *Comput. Phys. Commun.* **185**, 604 (2014); [arXiv:1310.0980](https://arxiv.org/abs/1310.0980).
- 42 G. Bussi, D. Donadio, and M. Parrinello, *J. Chem. Phys.* **126**, 014101 (2007).

- ⁴³S. Nosé, *Mol. Phys.* **52**, 255 (1984).
- ⁴⁴B. Hourahine, B. Aradi, V. Blum, F. Bonafé, A. Buccheri, C. Camacho, C. Cevallos, M. Y. Deshayé, T. Dumitrică, A. Dominguez, S. Ehlert, M. Elstner, T. van der Heide, J. Hermann, S. Irlé, J. J. Kranz, C. Köhler, T. Kowalczyk, T. Kubař, I. S. Lee, V. Lutsker, R. J. Maurer, S. K. Min, I. Mitchell, C. Negre, T. A. Niehaus, A. M. N. Niklasson, A. J. Page, A. Pecchia, G. Penazzi, M. P. Persson, J. Řezáč, C. G. Sánchez, M. Sternberg, M. Stöhr, F. Stuckenberg, A. Tkatchenko, V. W.-z. Yu, and T. Frauenheim, *J. Chem. Phys.* **152**, 124101 (2020).
- ⁴⁵D. Maag, M. Putzu, C. L. Gómez-Flores, F. Gräter, M. Elstner, and T. Kubař, *Phys. Chem. Chem. Phys.* **23**, 26366 (2021).
- ⁴⁶I. Leontyev and A. Stuchebrukhov, *Phys. Chem. Chem. Phys.* **13**, 2613 (2011).
- ⁴⁷B. J. Kirby and P. Jungwirth, *J. Phys. Chem. Lett.* **10**, 7531 (2019).
- ⁴⁸A. Laio and M. Parrinello, *Proc. Natl. Acad. Sci. U. S. A.* **99**, 12562 (2002).
- ⁴⁹A. Laio and F. L. Gervasio, *Rep. Prog. Phys.* **71**, 126601 (2008).
- ⁵⁰A. Barducci, G. Bussi, and M. Parrinello, *Phys. Rev. Lett.* **100**, 020603 (2008).
- ⁵¹P. Raiteri, A. Laio, F. L. Gervasio, C. Micheletti, and M. Parrinello, *J. Phys. Chem. B* **110**, 3533 (2006).
- ⁵²K. H. Lee, U. Schnupf, B. G. Sumpter, and S. Irlé, *ACS Omega* **3**, 16899 (2018).
- ⁵³R. P. P. Neves, P. A. Fernandes, and M. J. Ramos, *Proc. Natl. Acad. Sci. U. S. A.* **114**, E4724 (2017).