Additive CHARMM Force Field for Pterins and Folates

Alexey Aleksandrov, Hannu Myllykallio, Jean-Christophe Lambry, Wenlu Yin, Elsa Balduzzi

Additive CHARMM Force Field for Pterins and Folates

Elsa Balduzzi[1], Wenlu Yin[1], Jean-Christophe Lambry[1], Hannu Myllykallio[1], and Alexey Aleksandrov[1*]

[1]Laboratoire d'Optique et Biosciences (CNRS UMR7645, INSERM U1182), Ecole Polytechnique, Institut polytechnique de Paris, F-91128 Palaiseau, France

*Corresponding author: alexey.aleksandrov@polytechnique.edu

Running title: CHARMM Force Field for Folates

Keywords: folic acid; tetrahydrofolate; tetrahydrobiopterin; pterin; CHARMM; CGenFF; force field; molecular dynamics;

ABSTRACT

Folates comprise a crucial class of biologically active compounds related to folic acid, playing a vital role in numerous enzymatic reactions. One-carbon metabolism, facilitated by the folate cofactor, supports numerous physiological processes, including biosynthesis, amino acid homeostasis, epigenetic maintenance, and redox defense. Folates share a common pterin heterocyclic ring structure capable of undergoing redox reactions and existing in various protonation states. This study aimed to derive molecular mechanics parameters compatible with the CHARMM36 all-atom additive force field for pterins and biologically important folates, including pterin, biopterin, and folic acid. Three redox forms were considered: oxidized, dihydrofolate, and tetrahydrofolate states. Across all protonation states, a total of 18 folates were parameterized. Partial charges were derived using the CHARMM force field parametrization protocol, based on targeting reference quantum mechanics monohydrate interactions, electrostatic potential, and dipole moment. Bonded terms were parameterized using one-dimensional adiabatic potential energy surface scans, and two-dimensional scans to parametrize in-ring torsions associated with the puckering states of dihydropterin and tetrahydropterin. The quality of the model was demonstrated through simulations of three protein complexes using optimized and initial parameters. These simulations underscored the significantly enhanced performance of the folate model developed in this study compared to the initial model without optimization in reproducing structural properties of folate-protein complexes. Overall, the presented molecular mechanics model will be valuable for modeling folates in various redox states and serve as a staring point for parameterizing other folate derivatives.

INTRODUCTION

Folates named for their abundant presence in dark green leafy vegetables form an important class of biologically active compounds related to folic acid (vitamin B_9).[1] As an important one-carbon source essential for maintaining normal cell growth, a deficiency in folic acid can lead to abnormalities in one-carbon metabolism, thereby triggering various chronic diseases and developmental disorders, such as Alzheimer's disease and autism.[2-3] Mammals cannot synthesize folic acid *de novo* and must rely on dietary intake for its supply.[1] Folates can consist of three distinct chemical moieties linked together, with the common part, pterin (2-amino-4-hydroxy-pteridine) heterocyclic ring. While pterin itself has no significant biological function, its derivative tetrahydrobiopterin (BH4) is an essential co-factor in the synthesis of serotonin, dopamine, norepinephrine, epinephrine, and melatonin.[4] BH4 serves as a cofactor required for an enzyme's activity of mainly hydroxylases.[5]

In addition to the fully oxidized form of folic acid, there are two other semireduced and fully reduced forms, dihydrofolate and tetrahydrofolate, respectively. Dihydrofolate is an important intermediate in the biological reduction processes, in particular in the biosynthesis of tetrahydrofolate, with its pteridine ring in a semi-reduced state.[6] However, it is inactive in one-carbon transfer reactions, thus lacking coenzyme activity.[7] Tetrahydrofolate is a fully reduced derivative of folic acid, serving as the coenzyme form of folate in various metabolic reactions.[6] In living organisms, tetrahydrofolate serves as a carrier of one-carbon units, where the nitrogen atoms at N5 and N10 positions in its molecular structure can bind to one-carbon units of three different oxidation states (methyl, formyl, and formate).[8] For instance, 5,10-methylenetetrahydrofolate, formed by the connection of N^5 to N^{10} of tetrahydrofolate by the methylene group,[9] serves as a one-carbon donor for thymidylate synthase, which reaction methylates dUMP to form dTMP, and is crucial for pyrimidine synthesis within organisms.[10-11] Additionally, 10-formyltetrahydrofolate, formed by the attachment of a formyl group at N10 of tetrahydrofolate,[12] participates in purine biosynthesis via the pentose phosphate pathway and it also undergoes formylation with methionyl-tRNA formyltransferase to generate fMet-tRNA, which is crucial for genetic material and protein synthesis.[13] In addition to three redox states, the pterin group of folates is characterized by a wide range of different protonation and tautomeric states that they can adopt in solvent and proteins.[14-16]

While the CHARMM force field model exists for a wide range of molecules including proteins, nicotinamide adenine dinucleotide phosphate (NADP) and flavins,[17-19] folates have not received any specific attention in context of force field development, despite their importance. Although CHARMM additive models can be generated using the CGenFF program, based on analogy to model compounds with existing parameters, such models may contain parameters transferred from molecules with poor analogy. This can undermine the quality of the results from simulation studies relying on such models. In the present study, we present a systematic development of a force field model for folate derivatives in important protonation and redox states. By using the standard CHARMM parametrization protocol, the force field is developed to be compatible with the additive CHARMM36 force field for proteins, flavins, NADP, and CGenFF force field for small molecules.[20] To summarize, the model developed in this work is suitable to investigate interactions of folates with a wide range of

proteins and can be used as a template to parametrize other folate derivatives.

RESULTS

Set of parametrized molecules

We parameterized three redox forms of folates, shown in Figure 1. For each redox form of folates the most important protonation states were considered at the physiological pH of 6.5. Specifically, the pterin group can be protonated at the N^1 or N^3 positions, with N^3 protonation being preferred.[16, 21] However, when bound to protein, the protein environment can shift the equilibrium toward N^1 protonation, as observed in folate receptors.[22] Additionally, the rare N^1,N^3-deprotonated form of the oxidized folates was parameterized. We considered the pterin group in the following molecules: biopterin (also known as sapropterin), 6-methyl-pterin and folic acid. The pterin group common to these molecules was first parameterized as a standalone molecule, and then the parameters were transferred to the pterin moiety of other molecules following the standard development protocol of the CHARMM force field.[20] Considering all protonation states, a total of 18 folates were parameterized in this work.

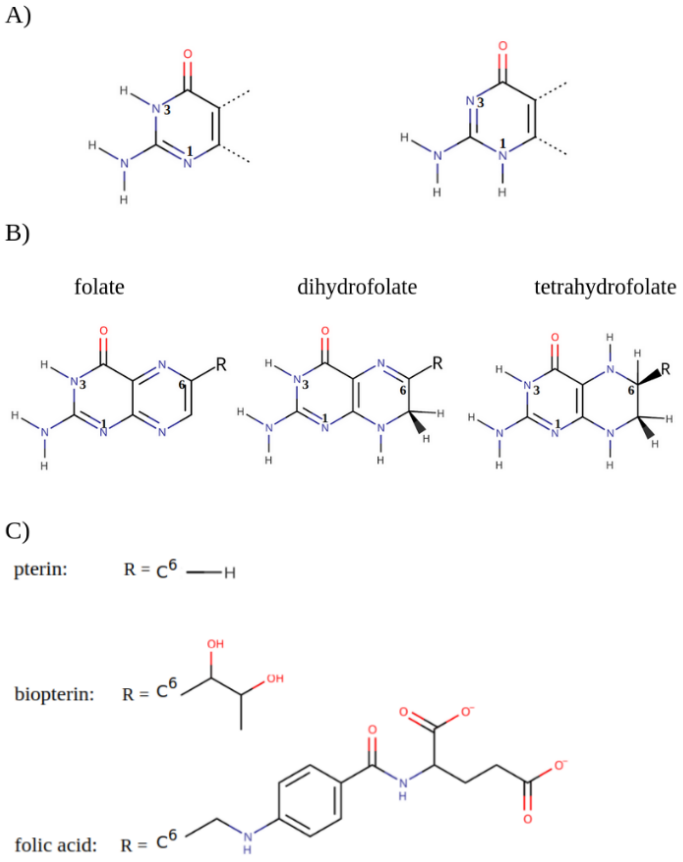

Figure 1. Protonation states A) and redox forms B) of the pterin group parameterized in this work. Molecular structures featuring the pterin group C) parameterized in this study.

Charge optimization

The CHARMM atomic charges were derived by targeting reference Quantum Mechanics (QM) data, including water-compound minimum interaction energies and geometries, the dipole moment magnitude and orientation, and electrostatic potential (ESP). Charge development was performed for 30 model compounds representing

different groups of folates. Table 1 summarizes the results of the charge development. Hereafter, we will refer to parameters provided by the CGenFF program as "initial", and parameters optimized in this work as "optimal".

Table 1. Statistics for partial charge development and agreement with target data for all model compounds used to parameterize folates.

Property	*N* points	RMSD optimal/initial	MAE optimal/initial
[a]norm of μ	28	0.09/2.33	0.05/1.83
[b]direction of μ	28	1.3/50.5	0.6/36.7
[c]water-solute E_{int}	867	0.62/1.84	0.46/1.32
[d]water-solute d_{min}	867	0.13/0.37	0.07/0.12
[e]φ_{elec}	30	2.26/7.84	2.20/7.05

[a]The magnitude of the dipole moment (μ) is given in Debye; [b]angle (°) between the *ab initio* and empirical dipole moment vectors; [c,d]probe water-model compound interactions; [e]RMSD between the *ab initio* and empirical ESP.

Table 1 presents the dipole moment statistics for 28 molecules; the dipole moment for the anionic form of pterin was excluded from the fitting process in accord with the standard CHARMM protocol. The symmetric pyrazine, having a zero dipole moment, was also excluded from the statistics, as trivial. The RMSD between the QM and MM dipole moments is 0.09 Debye with the optimized charges, compared to 2.33 Debye with the initial charges. The direction of the dipole moment also shows improvement relative to the QM moments. The root mean square angle between the MM and QM dipole moments is 50.5° with the CGenFF charges and 1.3° with the optimized charges. Overall, a good agreement was obtained for the dipole moment.

In accordance with the standard CHARMM optimization protocol, water probe interactions were given a high weight during the fitting process relative to ESP. For 30 molecules, a total of 867 individual water positions were computed, averaging 28.9 water interactions per model compound. The agreement between QM and CHARMM energies for probe water interactions is demonstrated in Figure 2. The interaction energies with water were significantly improved from 1.84 to 0.62 kcal·mol^{-1} with the initial to optimized charges. The water interaction distances were also improved, decreasing RMSD from 0.37 to 0.13 Å. Overall, the optimization of charges significantly improved water-compound interactions for all compounds. Interestingly, interactions computed with the optimized charges tend to slightly underestimate the QM interaction energies, which is reflected in the regression fit constant term of 0.22 kcal·mol^{-1}. This discrepancy can be attributed to a slight imbalance between different contributions to the cost function: empirically corrected interaction energies computed at the HF level and the dipole moment and ESP computed at the MP2 level in implicit solvent.

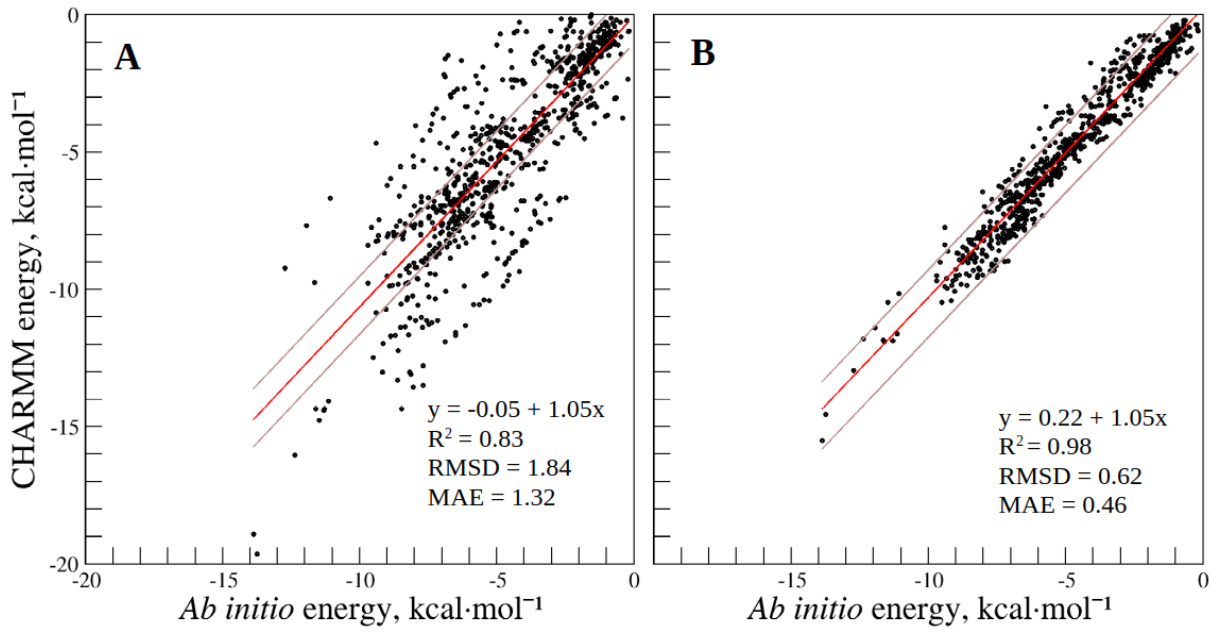

Figure 2. Comparison of QM and CHARMM water interaction energies for compound-water monohydrates computed with A) the initial charges and B) the optimal atomic charges. The red line represents the linear regression fit between QM and CHARMM data, while the diagonal gray lines indicate deviations of ±1.0 kcal·mol^{-1} from the regression line.

Electrostatic potential was included as an additional restraint during charge fitting to ensure a more accurate charge distribution in the model compounds, following previous studies.[23] In all cases, including the anionic model compound, the electrostatic potential was significantly improved relative to the initial values. The RMS deviation between MM and QM ESPs decreased from 7.84 kcal·mol^{-1}·e^{-1} with the initial charges to 2.26 kcal·mol^{-1}·e^{-1} with the optimized charges. To attain such improvement, the initial charges underwent optimization, resulting in an RMS deviation of 0.19*e* between the initial and optimal charges. Notably, we found that the charge assignment to the carbon atom C4A in the pterin ring structure often necessitated significant adjustment, with an RMS deviation of 0.45*e* observed between the initial and optimal charges. This particular carbon, positioned between the two rings, consistently incurred a high CGenFF penalty, indicating that such a molecular structure context has not been considered in CGenFF parameterization before. In contrast, hydrogens required the least adjustment, with an RMS deviation of just 0.04*e*, indicating that these atoms are typically well described by the CGenFF force field.

Within this work, we also investigated the recently developed CM5 model, derived from Hirshfeld population analysis, proposed for use in molecular mechanics.[24] Table S1 provides a comparison between results obtained with optimal charges and CM5 charges. As anticipated, CM5 charges do not reproduce well the overestimated dipole moment in implicit solvent, as the CM5 model was calibrated to match the dipole moment in vacuum. However, the direction of the dipole moment is accurately reproduced with CM5 charges, with an RMSD of just 2.5°, compared to 1.3° with optimal charges. Interactions with probe waters are better reproduced with CM5 charges compared to the initial charges, with an energy RMSD in comparison to reference QM data of

1.18 kcal·mol^{-1}, in contrast to 1.84 kcal·mol^{-1} with the initial charges. However, it is still inferior to the 0.62 kcal·mol^{-1} obtained with the optimized set of charges. Additionally, ESP is more accurately reproduced with CM5 charges than with the initial charges, with an RMSD with respect to reference QM ESP of 5.14 kcal·mol^{-1}·e^{-1}, compared to 7.84 kcal·mol^{-1}·e^{-1} with initial charges. Overall, although CM5 charges perform worse than the optimal charges for water interaction energies, dipole moment, and ESP, they emerge as an alternative to CGenFF as initial charges for charge optimization.

Optimization of bonded terms

All bonded terms, including stiff and soft terms, were parametrized based on reproducing QM potential energy surfaces (PESs), as in our previous studies.[25-27] Bonded parameters with a non-zero ParamChem penalty were considered for refinement. Seventeen molecules, including model compounds and folates, were created to parameterize the bonded terms in eighteen folates and their forms. The phase and multiplicity of dihedral angle terms were initially taken from the ParamChem guess. Additional multiplicities were considered for soft dihedrals to improve alignment with the QM reference data. The phase was allowed to vary between 0º and 180º, except for dihedral angles in conjugated systems, where the multiplicity was set to two and the phase to 180º.

The agreement between structural properties for equilibrium geometries computed with *ab initio* and empirical models, using both the initial and optimal sets of parameters, is summarized in Table 2. The RMS deviation between the *ab initio* and CHARMM-optimized Cartesian coordinates for all atoms, averaged over 17 model compounds, is 0.28 Å (standard deviation (SD): 0.17 Å) with the initial parameters and 0.19 Å (SD: 0.14 Å) with the optimal parameters. Table 2 also provides the agreement for bonds, valence angles, and torsions in equilibrium structures, for which corresponding bonded terms were considered for optimization.

Table 2. Comparison between empirical and *ab initio* optimized geometries for equilibrium structures.

Property	*N* values	[b]MAE	RMSD
		optimal/initial	optimal/initial
[a]RMSD (Å)	17	-	0.19/0.28
bond (Å)	4	0.001/0.009	0.002/0.012
angle (º)	26	0.8/3.2	1.9/4.6
dihedral (º)	99	2.4/5.3	5.1/8.2

[a]average RMS deviation between QM and MM optimized equilibrium structures for all atoms

For bonds, the RMS deviation between bond distances in QM and MM optimized structures is 0.012 Å with the initial parameters and 0.002 Å with the optimal parameters. For valence angles, the RMS deviation is 4.6º with the initial parameters and 1.9º with the optimal parameters. For torsions, the RMS deviation is 8.2º with the initial parameters and 5.1º with the optimal parameters. The mean absolute error (MAE) for torsions with the optimal model is 2.4º, which is three times higher than the MAE of 0.8º computed for valence angles. Overall, the optimized bonded parameters enable the CHARMM model to reproduce the QM geometries well.

Table 3 summarizes the agreement between CHARMM and QM energies of PES scans. The RMS deviation between QM and empirical energies for PES scans shows systematic improvement for all terms. For

instance, for valence angle terms, the energy RMSD is reduced from 0.33 kcal·mol^{-1} with the initial parameters to 0.13 kcal·mol^{-1} with the optimal parameters. Similarly, the energy (MAE) decreases from 0.22 kcal·mol^{-1} to 0.07 kcal·mol^{-1}. This represents a significant improvement, with both the energy MAE and RMSD for valence angles reduced by approximately three times. Comparable improvements are observed for bond terms and improper angles, with reductions in both MAE and RMSD by around three times.

Table 3. Comparison between empirical and *ab initio* energies of PES scans with the optimal and initial sets of parameters

Term	[a]*N* points	[c]RMSD optimal/initial	[d]MAE optimal/initial
bond	20	0.13/0.49	0.09/ 0.26
angle	115	0.13/0.33	0.07/0.22
stiff dihedral	260	0.26/ 0.51	0.11/0.32
rotatable dihedral	296	1.84/3.07	1.32/2.10
in-ring dihedral	1064	1.52/5.92	1.07/3.98
improper angle	15	0.07/0.54	0.04/0.36
[b]local minimum	103	2.59/4.38	1.86/3.23

[a]Number of PES conformations used to optimize bonded parameters; [b]energies of local minimum structures; [c,d]RMS deviation between QM and MM energies, and mean absolute error, respectively.

To fit dihedral terms associated with soft degrees of freedom, Fourier series with the minimum number of multiplicities needed to fit the energy profiles were sought to enhance parameter transferability. If a satisfactory agreement could not be achieved, additional multiplicities were considered. A total of 8 PES scans were performed for rotatable dihedrals across all model compounds. For these dihedral angles, where full rotation scans were conducted, the energy RMSD improved from 3.07 kcal·mol^{-1} with the initial parameters to 1.84 kcal·mol^{-1} with the optimized parameters, while the MAE improved from 2.10 kcal·mol^{-1} to 1.32 kcal·mol^{-1}. This represents a 60% improvement in the energy RMS deviation. This moderate improvement in comparison to the stiff degrees of freedom was expected, given the substantial impact of nonbond interactions on their PES. Rotatable dihedrals typically presented the greatest challenge, exhibiting the largest RMS deviations compared to QM data, in contrast to more rigid degrees of freedom.

Figures 3 demonstrates the results for fitting rotatable dihedrals in N-methylaniline. This model compound was used to fit the dihedral angle terms associated with the methylene link between the pterin group and p-aminobenzoyl. Figure 3 illustrates that the empirical PES computed with the initial parameters exhibits additional minima at -135° and 45°, which are not present in the reference QM PES. The barriers at ±90° are underestimated, with the empirical values at 2.5 kcal·mol^{-1} compared to the QM value of 4.8 kcal·mol^{-1}. In contrast, with the optimal parameters, the number of minima and the relative heights of the barriers separating these minima on the PES are well reproduced. However, the position of the QM barriers is not accurately reproduced with either the initial or optimal parameters. The QM barriers are located at -135° and 45°, not symmetrically relative to zero,

whereas the barriers with the empirical model are at ±90°, with no improvement upon fitting the dihedral parameters. Additional multiplicities did not correct the barrier positions. Nonetheless, this misalignment of barriers is not expected to significantly affect the equilibrium structures simulated in MD simulations.

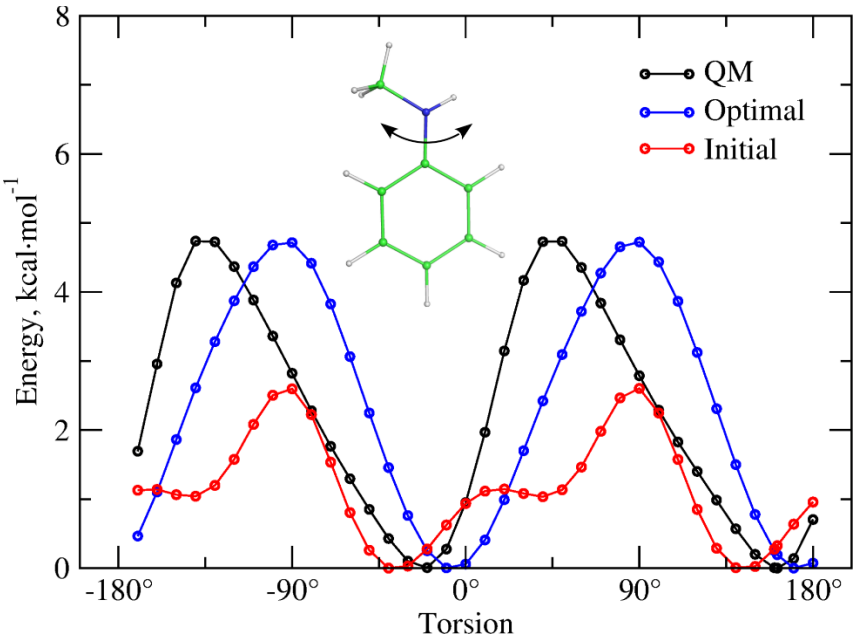

Figure 3. PES scans for N-methylaniline. The PES scan was performed for rotation around the bond depicted in the inset.

Figure 4 shows the PES scan for another dihedral associated with the same linkage, parametrized in the model compound 2-(phenylaminomethyl)tetrahydropyrazine. The QM PES reveals a very high energy barrier at 40° of 27.5 kcal·mol^{-1}. Since geometries with high energies (>10 kcal·mol^{-1}) are not considered during fitting, the height of this barrier only moderately improved. In contrast, the second barrier, which is 7 kcal·mol^{-1} in the QM PES, shows significant improvement in the empirical model, with the barrier reduced from 20.1 kcal·mol^{-1} to 12.5 kcal·mol^{-1}. This barrier is largely influenced by nonbonded interactions between the rings, which are challenging to reproduce accurately with the empirical model. The position of the minimum at 60° is well reproduced, and the position of the shallow minimum at higher energy around -135° is also accurately reproduced. With the optimal parameters, the energy of the second minimum is better matched with 1.0 kcal·mol^{-1} compared to 2.6 kcal·mol^{-1}, whereas with the initial parameters, it was 7.5 kcal·mol^{-1}. Additionally, the position of the minimum with the optimal parameters is at -145°, closer to the QM PES position of -135°.

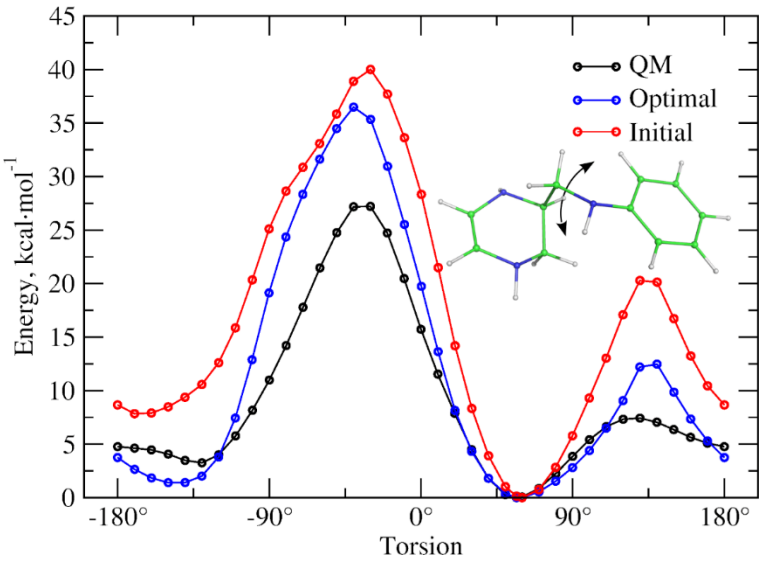

Figure 4. PES scans for 2-(phenylaminomethyl)tetrahydropyrazine. The PES scan was performed for rotation around the bond depicted in the inset.

Puckering states and local energy conformation minima

Soft dihedrals within ring structures typically lead to puckering states that a molecule can adopt in simulations, influenced by the protein and solvent environment. These puckering states can result in substantially different orientations of ring substituents, leading to different interactions with proteins and solvents. In this work, to fit dihedral terms associated with soft degrees of freedom in the ring structures of folates, two-dimensional PES scans were performed for in-ring dihedrals. Additionally, energy-minimum conformations identified through conformation searches were included in the fit. It is important to note that full rotation for in-ring dihedrals is not possible without breaking the ring structure. Therefore, we adopted a method where calculations are done on a grid, similar to rotatable dihedrals, but the exploration is limited in energy to prevent breaking the chemical structure.

The energy RMS deviation between empirical and QM energies for all PES scans of in-ring dihedrals improved from 5.92 kcal·mol^{-1} with the initial parameters to 1.52 kcal·mol^{-1} with the optimal parameters. The MAE also showed significant improvement, decreasing from 3.98 kcal·mol^{-1} with the initial parameters to 1.07 kcal·mol^{-1} with the optimal parameters. This improvement for in-ring dihedrals is more pronounced than for other rotatable dihedrals and stiff degrees of freedom. This can be attributed to the strains often present in ring structures, which are challenging to reproduce, and dihedral parameters associated with in-ring torsions tend to be less transferable.

Figure 5 compares the QM and empirical energies for 2D PES scans computed with the initial and optimal sets of parameters for 1,2-dihydropyrazine. This model compound was used to fit in-ring dihedral angle terms present in dihydrofolates. The ring conformation can be described by two dihedral angles: the rotation around the C6-C7 bond and the position of the HN hydrogen. The 2D PES demonstrates four stable conformations for 1,2-dihydropyrazine, as shown in Figure 5. These four conformations are practically equienergetic. The RMSD

between conformations I and II, and III and IV, is 0.9 Å and 1.0 Å, respectively; conformations I and III, and II and IV, are more similar, with an RMSD of 0.2 Å. The difference between conformations with smaller RMSD is mainly due to the orientation of the NH hydrogen. In contrast to the reference QM PES, the initial parameters yield only two stable conformations, which differ by the orientation of the HN8 hydrogen. The geometry of the molecule is planar, with the N8-C7-C6-N5 dihedral preferring an angle of 0°. With the optimal parameters, the empirical PES demonstrates four minima, and the N8-C7-C6-N5 dihedral has a preference for ±45°, in agreement with the QM PES. The position of the HN8 hydrogen is also well reproduced in these minima. Although the QM energies between -90° and 90° for the dihedral angle C4A-C8A-N8-HN8 are underestimated by the optimal MM parameters, this will not affect the rate of proton inversions around the N8 nitrogen center. These inversions are dominated by the transition through the ±180° angle, which is practically barrierless.

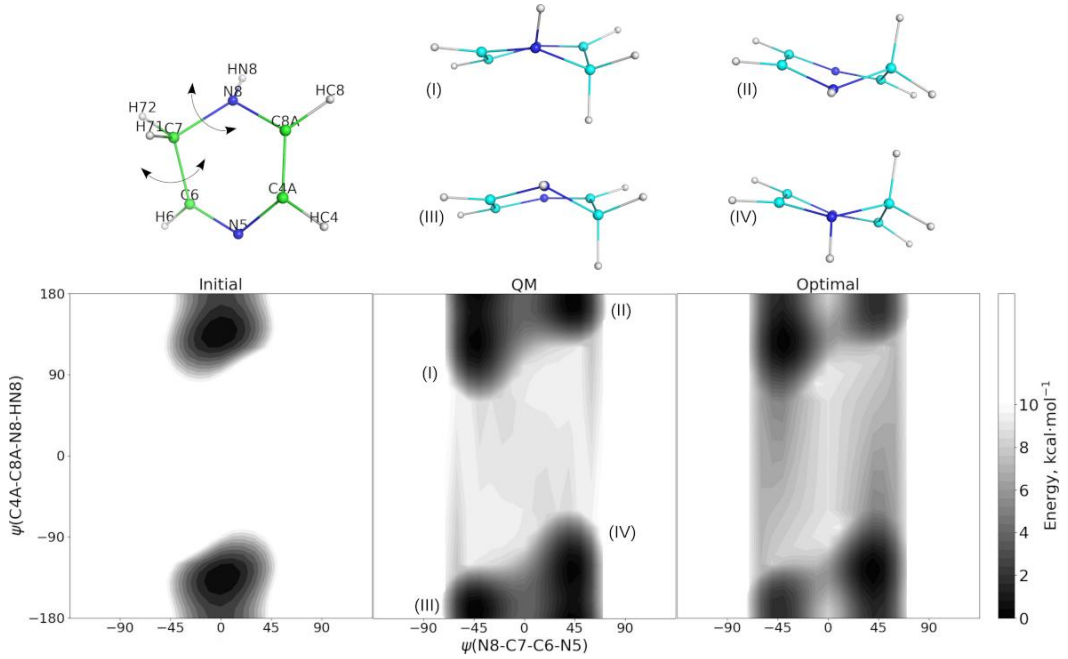

Figure 5. Two-dimensional PES scan for 1,2-dihydropyrazine. Upper panel: the two torsion angles for the PES scan are shown by the arrows; four minimum energy QM conformations. Bottom panel: 2D PES scans with initial and optimal sets of parameters along with the QM reference data.

In this study, local-energy minimum conformations were incorporated into the fitting process. A total of 103 stable conformations were identified through conformational analysis for the 17 model compounds parameterized in this work, averaging approximately 6 conformations per molecule. The energy RMSD for these conformations is 4.48 kcal·mol^{-1} with the initial parameters and 2.59 kcal·mol^{-1} with the optimal parameters. Reproducing the energies of stable conformations posed several challenges. Firstly, these energies are significantly influenced by non-bonded interactions, especially pronounced for electrostatic interactions in vacuum without solvent screening. Secondly, the use of existing parameters in the CGenFF force field, optimized for different molecular contexts, may not be fully suitable for the molecules parameterized in this study. Addressing the latter issue would necessitate fitting bonded parameters across multiple molecules simultaneously. However, such a

comprehensive refitting of existing CHARMM parameters is beyond the scope of the current study.

Molecular Dynamics simulations of protein complexes

To illustrate the quality of the model, MD simulations were conducted on protein complexes with folic acid and tetrahydrofolate. Three protein structures, ranging from high to medium resolution, were selected for these simulations. Additional information on the chosen protein structures can be found in Table S2. In these MD simulations, no restraints were imposed on protein or ligand atoms. MD simulations were performed with the optimal parameters and initial parameters for comparison. The folate rotatable dihedrals are given in Tables S3-S5, while the distances between heavy atoms participating in hydrogen bonds are given in Tables S6-S8.

The RMS deviations between simulation and experimental structures are provided in Table 4. The RMS deviation for protein backbone atoms within 10 Å of folates in all MD simulations ranges from 0.69 Å to 1.49 Å with the optimal parameters. Furthermore, the RMSD for folates after superimposing them onto the crystal structure based on the non-hydrogen atoms of folates is within the range of 1.11 Å to 1.61 Å, representing a significant improvement compared to simulations with the initial parameters. In simulations with the initial parameters, the RMS deviation is consistently larger for both the protein backbone and the folate ligands. With the optimal parameters, the RMS deviation for the folates, after superposition using the protein backbone within approximately 10 Å around the ligand, is notably improved. This RMS deviation is influenced by both the reorientation of the ligand and changes in its conformation. For instance, in the case of dihydrofolate reductase in complex with tetrahydrofolate (PDB reference code 6CW7[28]), the RMSD for THG is 3.54 (SD: 0.28) with the initial parameters, compared to 1.59 (SD: 0.26) with the optimized parameters. This demonstrates that the ligand undergoes significant conformational changes in simulations with the initial parameters.

Table 4. RMS deviation in molecular dynamics simulations in Angstroms (Å). Standard deviations are given in parenthesis.

Ligand	PDB ref. code	[a]Protein initial/optimal	[b]Ligand initial/optimal	[c]Ligand initial/optimal
Folic acid	4P3Q	1.71(0.30)/1.25(0.32)	1.84(0.31)/1.19(0.13)	3.31(0.64)/1.49(0.21)
Tetrahydrofolate	6CW7	1.49(0.15)/1.49(0.24)	2.55(0.21)/1.11(0.17)	3.54(0.28)/1.59(0.26)
Tetrahydrofolate	4O1F	0.82(0.12)/0.69(0.10)	1.91(0.38)/1.61(0.25)	5.87(1.62)/2.71(0.52)

[a]RMSD was computed for backbone atoms around 10 Å around the ligand after superposition on the experimental structure; [b]RMSD was computed for the heavy atoms of the ligand after superimposing the ligand; [c]RMSD was calculated for the ligand after superimposing the protein backbone atoms.

In simulations with folic acid and dihydrofolate reductase (DHFR, crystal structure 4P3Q[29]), we observe that interactions are significantly better reproduced with the optimal set of parameters. The interaction distances are given in Table S6. Specifically, the interaction between NH2 of Arg57 and oxygens of the glutamate part of folic acid is well maintained in MD simulations with the optimal parameters. The distance between Arg57 and the oxygens is 2.75 Å and 2.68 Å in simulations with the optimal parameters, compared to around 8 Å in simulations with the initial parameters. These interactions are stable in simulations with the optimal parameters, with RMS

deviation of around 0.1 Å from the average values. In contrast, simulations with the initial parameters show RMS deviation in the distances of around 2.0 Å, indicating that the interaction between Arg57 and folic acid is not stable and the distance between them increases over the 500 ns MD simulations. The dihedral angles of the link parameterized in this work given in Table S4 are better reproduced in the MD simulations with the optimal parameters. Specifically, the torsions N5-C6-C9-N10 and C6-C9-N10-C14 deviate by 62° and 64° from the values in the experimental structure 4P3Q, respectively. This discrepancy may explain the loss of interactions with Arg57.

The hydrogen bond interaction between Arg52 and OE2 of folic acid is not reproduced with either the initial or optimal set of parameters. However, a close examination of the crystal structure 4P3Q reveals that Arg52 is positioned on a hydrophobic surface and thus desolvated, which is unlikely for the cationic arginine. In other crystal structures, such as 4KJJ[30] and 4PDJ,[31] Arg52 is solvent-exposed and does not interact with OE2 of folic acid in contrast to the structure 4P3Q. This observation is consistent with the results from the present MD simulations.

The selected interaction distances and folate dihedral angles in simulations of the DHFR complex with tetrahydrofolate (PDB reference code 6CW7[28]), are given ion Tables S3 and S7. The N8 atom of THG interacts with the backbone group of Ile5. These interaction is well maintained in MD simulations with both the initial and optimal parameter sets. However, interactions between the glutamate part of THG and Arg57 are not preserved in simulations with the initial parameters, resulting in long distances between the carboxylic group of THG and the guanidinium nitrogens of Arg57. In contrast, with the optimal parameters, this interaction is preserved. This improvement can be attributed to better sampling of the dihedral angles involved in the link between the pterin ring and the benzene part. Specifically, the dihedral angle C6-C9-N10-C14 deviates by 33 degrees from the value observed in the experimental structure 6CW7 in simulations with the initial parameter set.

Additionally, simulations were performed for THG in complex with the methyltransferase component (PDB entry 4O1F[32]). Interaction distances, as given in Table S8, clearly demonstrate a significant improvement in reproducing experimental structural properties with optimal parameters. With the initial parameters, THG was unstable in the binding pocket and lost its initial hydrogen bonds. However, with the optimized parameters, these interactions were reasonably maintained. This improvement is reflected in the RMSD values for the ligand after superimposing the protein backbone on the experimental structure: 5.87 Å with the initial parameters and 2.71 Å with the optimized parameters. Despite these improvements, the torsions for the link between the pterin ring and the benzene part were poorly reproduced with both the initial and optimized parameters (Table S5). Analysis of the crystal structure reveal that the benzene and glutamate parts are mostly solvent-exposed, allowing for large conformational fluctuations, which is consistent with the poor electron density observed for these parts of the molecule in the crystal structure 4O1F. In contrast, the pterin head remains well anchored in the protein binding site during MD simulations.

Overall, the optimized folate model performs significantly better in molecular dynamics simulations,

demonstrating superior accuracy in reproducing interaction distances for hydrogen bonds and in capturing the conformations of folate and proteins.

Discussions

This study represents a systematic development of a force field model for folates, focusing on the most important protonation and redox states. The parameterization was chosen to be consistent with the standard methods used in developing the additive CHARMM force field, ensuring compatibility with other components in the CHARMM force field, including proteins, flavins, and NADP. Partial charges were parameterized to reproduce both QM ESP and interactions with individual water molecules, maintaining a balance between folate interactions with solvent and other system components, such as protein residues. Although the CGenFF program provided a good initial guess, the initial charges did not provide a satisfactory agreement for water interactions, ESP, and dipole moments, indicating that the molecular groups were not explicitly parameterized in CGenFF.

Additionally, we tested CM5 charges, which are based on Hirshfeld population analysis. Surprisingly, while CM5 charges performed worse than the optimal charges for water interaction energies, dipole moments, and ESP, they performed better than the initial CGenFF charges. This improvement is likely due to the absence of the molecular groups, such as pterin in the CGenFF. This suggests that CM5 charges could serve as an alternative to CGenFF as initial charges for future force field development.

A special emphasis was placed on the quality of bonded parameters, with all bonded parameters, including soft torsions and stiff harmonic terms, adjusted using computationally intensive PES scans. In this study, all parameters not previously existing in the CGenFF force field were considered for refinment. The bonded parameters were then optimized to reproduce PES energies and geometries. The model successfully reproduces the QM geometry for folates, and the PES for soft dihedrals. To parameterize dihedral terms in rings, which can adopt different puckering states, expensive 2D PES scans were performed for in-ring dihedral terms. This approach enabled the parameterization of dihydropterin and tetrahydropterin ring structures, leading to a more accurate reproduction of the puckering states of the pterin ring structure. These puckering states can be crucial for protein-folate interactions, as they modify the position of ring substituents.

Molecular dynamics simulations of protein complexes using both the initial and optimal parameters demonstrated the significant impact of force field parameter optimization. In all simulations with the optimal parameters, folates maintained interactions with the protein as observed in experimental structures, as well as preserved the ligand and protein conformation. In contrast, simulations with initial parameters showed significant structural deviations: the conformation of the folates deviated from the experimental structure, and protein-ligand interactions were not maintained.

Overall, the developed model should be valuable for modeling protein-folate complexes and can serve as a template for future force field development for other folate forms, such as methylene intermediates.

METHODS AND MATERIALS

Parametrization protocol

The functional form of the additive all-atom CHARMM form was adopted in this work.[33] The atom types were adopted from the CGenFF force field.[20] The CGenFF program was used to assign existing atomic types and to obtain initial guesses for the parameters of model compounds.[34-35] Bonded parameters that were assigned zero score (called "Penalty") by CGenFF, indicating prior optimization, were excluded from further optimization in this study. Lennard-Jones potential parameters were not considered for optimization. All QM optimizations were performed Gaussian 16,[36] employing the MP2/6-31G(d) model chemistry (MP2/6-311G(d) for anions) with default tight tolerances.

For molecules with multiple conformations, the conformation with the lowest energy, as predicted by conformational analysis, was used to determine partial charges. This conformational analysis was performed using the Conformer-Rotamer Ensemble Sampling Tool (CREST) version 2.12.[37] The CREST-generated conformations were subsequently optimized with Gaussian 16.[36] The resultant conformations underwent filtering based on their relative energies and root mean square (RMS) deviations between Cartesian coordinates. Conformations were considered identical if the RMS deviation was lower than 0.1 Å, and the absolute energy difference was less than 0.1 kcal·mol^{-1}. The RMS deviation between QM-optimized conformations was computed using the obrms tool in Open Babel.[38]

Determination of the intermolecular force field parameters

The intermolecular component of the total energy consists of Coulomb and Lennard–Jones terms. Following the additive CHARMM force field development, atomic partial charges were optimized targeting QM reference data. These reference data included interactions between the model compound and individual water molecules, the dipole moment for neutral molecules, and the electrostatic potential.[23]

Individual probe water molecules were positioned in idealized linear orientations to maximize interaction with the target site.[17] Various orientations of the water molecule around the interaction axis were considered: for polar atoms, the complex was calculated at every 45º or 90º rotation of the water probe, and only one or two orientations for non-polar atoms. Each water-compound structure was then optimized by adjusting the interaction distance to find the energy minimum for the water position, using the TIP3P model geometry for the water molecule. Consistent with the CHARMM force field parametrization protocol, calculations were done at the HF/6-31G(d) level,[17, 20] to preserve the balance in interactions between parameterized model compounds and the rest of CHARMM force field. In accordance with the standard CHARMM parametrization protocol,[17] the *ab initio* interaction energies were scaled by an empirical factor of 1.16 only for neutral polar compounds and the HF/6-31G(d) minimum interaction distance was corrected by subtracting 0.2 Å for all polar interactions in neutral compounds. The dipole moment was included in the reference data for neutral compounds during charge fitting.[39]

The dipole moment and electrostatic potential (ESP) were calculated in the implicit solvent at the MP2/6-31G(d) level, using the Polarizable Continuum Model (PCM) implemented in Gaussian 16.[40-41] During charge

optimization, symmetrical atoms were assigned identical charge values. Aliphatic group charges were not optimized, adhering to the standard CHARMM method, which assigns a charge of +0.09*e* to the aliphatic protons. Additionally, Charge Model 5 (CM5) charges, derived from Hirshfeld Population Analysis, were computed using Gaussian 16.[24]

Determination of the intermolecular force field parameters

Determination of bonded harmonic energy terms

Parameters for the bonded terms described by harmonic potentials, i.e. bonds, angles, improper dihedrals, and the stiff dihedrals were optimized using the previously developed protocol.[27] An adiabatic PES scan for each degrees of freedom that has adjustable parameters in the force field has been performed near the equilibrium value. All QM PES scans were conducted using the MP2/6-31G(d) model chemistry and MP2/6-311G(d) for anions. To limit the energy of deformed structures, the following method is used.[26] In a preliminary calculation using initial predefined values of distortions, the range for the scanned degree of freedom was estimated to ensure that the energy of the deformed structures remained below 2.0 kcal·mol^{-1}. For molecular mechanics (MM) calculations, a two-step procedure was employed.[27] First, the geometry was optimized with constraints only on rotatable dihedral angles. Equilibrium values for bonds, valence angles, improper dihedral, and dihedral angles from these MM-optimized conformations were then used for MM PES scans. These PES scans were performed around the MM equilibrium values with the same deformations used in the reference QM calculations. The cost function included the RMS energy difference between QM and MM structures, aiming to reproduce QM conformational flexibility. Two additional terms were introduced to reproduce structural properties and to accurately optimize force constants associated with valence angle terms that share the same atomic center.[27] To fit the reference QM data, an in-house program previously used to parameterize a large set of non-standard amino acids and flavins was employed.[25-26] At each optimization iteration of bonded parameters, PES scans were performed with the CHARMM program[42] using the updated set of MM parameters. The MM parameters were adjusted until the cost function could no longer be significantly reduced.

Optimization of flexible dihedral parameters

To parameterize terms associated with rotatable dihedrals, one-dimensional (1D) potential energy surface (PES) scans were performed on the torsions. Torsion angles were scanned in the range from -180° to 180° in 10° increments. During the 1D PES scans, only one torsion was varied while all other rotatable dihedrals were constrained to values from the minimum-energy geometry identified through the conformational analysis described above.

For dihedral terms associated with atoms in rings, which can exhibit different puckering states, a different procedure was used. Specifically, two-dimensional (2D) PES scans were performed along two torsions formed by shared atoms. These scans were conducted on a grid with 10° increments but were limited to the low-energy basin within 15 kcal·mol^{-1} of the minimum energy structure identified through conformational analysis. For MM calculations, each conformation was extracted from QM scans and minimized with a harmonic restraint force

constant of $5 \cdot 10^4$ kcal·mol·radian^{-2} on the target torsion(s). The MM dihedral parameters were optimized to minimize the deviation between the QM and MM surfaces, focusing on the lower energy regions. PES points with QM energy more than 10 kcal·mol^{-1} above the minimum energy were not considered in the optimization.

Molecular Dynamics simulations

The protein complexes used in this work are summarized in Table S2. High- to medium-resolution crystal structures were obtained from the Protein Data Bank (PDB). The simulations included all protein residues, water molecules present in the crystal structures, and folate ligands. Protonation states of histidines were assigned based on visual inspection and ideal stereochemistry, while PROPKA 3.0 was used to determine the protonation states of other residues. [43-44]

In addition to crystal waters, a cubic box of water was overlaid, and waters overlapping the protein, ligands, and crystal water molecules were removed. The size of the solvent box was chosen to maintain a minimum distance of 10 Å between any protein atom and the box's edge. Periodic boundary conditions were applied. Long-range electrostatic interactions were computed using the particle mesh Ewald method.[45] The appropriate number of chloride or potassium counterions were included to neutralize the system; additionally, chloride and potassium ions were added to achieve a final salt concentration of 0.15 M. A switching function acting from 9 Å was used to truncate all van der Waals interactions at the distance of 11 Å. Long range electrostatic forces were evaluated every 4 steps, while short-range non-bonded interactions were computed at each step. The system was first minimized with 5000 steps of minimization, with restraints imposed on the position of non-hydrogen atoms. MD simulations were performed at constant room temperature and pressure, after 150,000 steps of equilibration with restraints acting on heavy atoms gradually released. Constant pressure was maintained using the Berendsen pressure bath coupling[46] with the relaxation of 500 fs the compressibility parameter of liquid water. Constant temperate was maintained by simulating temperature coupling to a heat bath with a room temperature by correcting forces as implemented in the NAMD program.[47] The CHARMM36m force field was used for the protein[48-50] and the modified TIP3P model for water.[17, 51-52] The folates were modeled using the force field model specifically developed in this work. Calculations were done with the NAMD program,[47] with MD simulations of the protein complexes continuing for 500 nanoseconds. For the methyltransferase (PDB reference code 4O1F[32]), simulations were limited to 100 nanoseconds due to the significantly large system size. Additionally, MD simulations were conducted with folates modeled using the initial parameters to demonstrate the impact of optimization for the same simulation time.

Supporting Information

Tables with statistics for agreement with target data with the optimized and CM5 charges; with experimental protein structures used for MD simulations; with selected rotatable dihedrals in MD simulations; with selected distances in MD simulations of folates in protein complexes. CHARMM topology and parameter file for the folate model developed in this work.

Acknowledgements

This project was provided with computer and storage resources by GENCI at CINES thanks to the grants 2023-A0150814640 and 2023-AD010814466 on the supercomputer Adastra.

References

1. Saini, R. K.; Nile, S. H.; Keum, Y. S., Folates: Chemistry, analysis, occurrence, biofortification and bioavailability. *Food Res. Int.* **2016,** *89* (Pt 1), 1-13.
2. Hinterberger, M.; Fischer, P., Folate and Alzheimer: when time matters. *Journal of neural transmission (Vienna, Austria : 1996)* **2013,** *120* (1), 211-24.
3. Lyall, K.; Schmidt, R. J.; Hertz-Picciotto, I., Maternal lifestyle and environmental risk factors for autism spectrum disorders. *Int. J. Epidemiol.* **2014,** *43* (2), 443-64.
4. Latremoliere, A.; Costigan, M., GCH1, BH4 and pain. *Curr. Pharm. Biotechnol.* **2011,** *12* (10), 1728-41.
5. Kappock, T. J.; Caradonna, J. P., Pterin-Dependent Amino Acid Hydroxylases. *Chem. Rev.* **1996,** *96* (7), 2659-2756.
6. Cieslik, E.; Cieślik, I., Occurrence and significance of folic acid. *Pteridines* **2018,** *29*, 187-195.
7. Green, J. M.; Matthews, R. G., Folate Biosynthesis, Reduction, and Polyglutamylation and the Interconversion of Folate Derivatives. *EcoSal Plus* **2007,** *2* (2), 10.1128/ecosalplus.3.6.3.6.
8. Fox, J. T.; Stover, P. J., Folate-mediated one-carbon metabolism. *Vitam. Horm.* **2008,** *79*, 1-44.
9. Födinger, M.; Hörl, W. H.; Sunder-Plassmann, G., Molecular biology of 5,10-methylenetetrahydrofolate reductase. *Journal of nephrology* **2000,** *13* (1), 20-33.
10. Filée, J.; Becker, H. F.; Mellottee, L.; Eddine, R. Z.; Li, Z.; Yin, W.; Lambry, J. C.; Liebl, U.; Myllykallio, H., Bacterial origins of thymidylate metabolism in Asgard archaea and Eukarya. *Nature communications* **2023,** *14* (1), 838.
11. Karunaratne, K.; Luedtke, N.; Quinn, D. M.; Kohen, A., Flavin-dependent thymidylate synthase: N5 of flavin as a Methylene carrier. *Arch. Biochem. Biophys.* **2017,** *632*, 11-19.
12. Vickers, T. J.; Murta, S. M.; Mandell, M. A.; Beverley, S. M., The enzymes of the 10-formyl-tetrahydrofolate synthetic pathway are found exclusively in the cytosol of the trypanosomatid parasite Leishmania major. *Mol. Biochem. Parasitol.* **2009,** *166* (2), 142-52.
13. Denis, V.; Daignan-Fornier, B., Synthesis of glutamine, glycine and 10-formyl tetrahydrofolate is coregulated with purine biosynthesis in Saccharomyces cerevisiae. *Mol. Gen. Genet.* **1998,** *259* (3), 246-55.
14. Soniat, M.; Martin, C. B., Theoretical Study on the Relative Energies of Neutral Pterin Tautomers. *Pteridines* **2008,** *19* (1), 120-124.
15. Schwalbe, C. H.; Lowis, D. R.; Richards, W. G., Pterin 1H–3H tautomerism and its possible relevance to the binding of folate to dihydrofolate reductase. *J. Chem. Soc., Chem. Commun.* **1993,** (15), 1199-1200.
16. Gocheva, G.; Petkov, N.; Luri, A.; Iliev, S.; Ivanova, N.; Petkova, J.; Mitrev, Y.; Madjarova, G.; Ivanova, A., Tautomerism in folic acid: Combined molecular modelling and NMR study. *J. Mol. Liq.* **2019,** *292*, 111392.
17. Mackerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Smith, J.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M., An all-atom empirical potential for molecular modelling and dynamics study of proteins. *J. Phys. Chem. B* **1998,** *102*, 3586-3616.
18. MacKerell, A. D., Jr.; Feig, M.; Brooks, C. L., 3rd, Improved treatment of the protein backbone in empirical force fields. *J. Am. Chem. Soc.* **2004,** *126* (3), 698-9.
19. Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E.; Mittal, J.; Feig, M.; Mackerell, A. D., Jr., Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone φ, ψ and side-chain χ(1) and χ(2) dihedral angles. *J. Chem. Theory Comput.* **2012,** *8* (9), 3257-3273.
20. Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D., CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J Comp Chem* **2010,** *31* (4), 671-690.
21. Nekkanti, S.; Martin, C., Theoretical study on the relative energies of cationic pterin tautomers. *Pteridines* **2015,** *26*, 13-22.
22. Chen, C.; Ke, J.; Zhou, X. E.; Yi, W.; Brunzelle, J. S.; Li, J.; Yong, E.-L.; Xu, H. E.; Melcher, K., Structural basis for molecular recognition of folic acid by folate receptors. *Nature* **2013,** *500* (7463), 486-489.
23. Huang, L.; Roux, B., Automated Force Field Parameterization for Nonpolarizable and Polarizable Atomic Models Based on Ab Initio Target Data. *J. Chem. Theory Comput.* **2013,** *9* (8), 3543-3556.
24. Marenich, A. V.; Jerome, S. V.; Cramer, C. J.; Truhlar, D. G., Charge Model 5: An Extension of Hirshfeld Population Analysis for the Accurate Description of Molecular Interactions in Gaseous and Condensed Phases. *J. Chem. Theory Comput.* **2012,** *8* (2), 527-541.
25. Croitoru, A.; Park, S.-J.; Kumar, A.; Lee, J.; Im, W.; MacKerell, A. D., Jr.; Aleksandrov, A., Additive CHARMM36 Force Field for Nonstandard Amino Acids. *J. Chem. Theory Comput.* **2021,** *17* (6), 3554-3570.
26. Aleksandrov, A., A Molecular Mechanics Model for Flavins. *J. Comput. Chem.* **2019,** *40* (32), 2834-2842.
27. Croitoru, A.; Aleksandrov, A., Parametrization of Force Field Bonded Terms under Structural Inconsistency. *Journal of chemical information and modeling* **2022,** *62* (19), 4771-4782.
28. Cao, H.; Gao, M.; Zhou, H.; Skolnick, J., The crystal structure of a tetrahydrofolate-bound dihydrofolate reductase reveals the origin of slow product release. *Communications biology* **2018,** *1*, 226.
29. Keedy, D. A.; van den Bedem, H.; Sivak, D. A.; Petsko, G. A.; Ringe, D.; Wilson, M. A.; Fraser, J. S., Crystal cryocooling distorts conformational heterogeneity in a model Michaelis complex of DHFR. *Structure* **2014,** *22* (6), 899-910.
30. van den Bedem, H.; Bhabha, G.; Yang, K.; Wright, P. E.; Fraser, J. S., Automated identification of functional dynamic contact

networks from X-ray crystallography. *Nat. Methods* **2013,** *10* (9), 896-902.
31. Wan, Q.; Bennett, B. C.; Wilson, M. A.; Kovalevsky, A.; Langan, P.; Howell, E. E.; Dealwis, C., Toward resolving the catalytic mechanism of dihydrofolate reductase using neutron and ultrahigh-resolution X-ray crystallography. *Proc. Natl. Acad. Sci. U. S. A.* **2014,** *111* (51), 18225-30.
32. Sjuts, H.; Dunstan, M. S.; Fisher, K.; Leys, D., Structures of the methyltransferase component of Desulfitobacterium hafniense DCB-2 O-demethylase shed light on methyltetrahydrofolate formation. *Acta Crystallogr. D Biol. Crystallogr.* **2015,** *71* (Pt 9), 1900-8.
33. Vanommeslaeghe, K.; MacKerell, A. D., Jr., CHARMM additive and polarizable force fields for biophysics and computer-aided drug design. *Biochim. Biophys. Acta* **2015,** *1850* (5), 861-71.
34. Vanommeslaeghe, K.; MacKerell, A. D., Jr., Automation of the CHARMM General Force Field (CGenFF) I: bond perception and atom typing. *J. Chem. Inf. Model.* **2012,** *52* (12), 3144-54.
35. Vanommeslaeghe, K.; Raman, E. P.; MacKerell, A. D., Jr., Automation of the CHARMM General Force Field (CGenFF) II: assignment of bonded parameters and partial atomic charges. *Journal of chemical information and modeling* **2012,** *52* (12), 3155-68.
36. 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.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 16 Rev. C.01*, Wallingford, CT, 2016.
37. Pracht, P.; Bohle, F.; Grimme, S., Automated exploration of the low-energy chemical space with fast quantum chemical methods. *Phys. Chem. Chem. Phys.* **2020,** *22* (14), 7169-7192.
38. O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R., Open Babel: An open chemical toolbox. *Journal of Cheminformatics* **2011,** *3* (1), 33.
39. Xu, Y.; Vanommeslaeghe, K.; Aleksandrov, A.; MacKerell, A. D.; Lennart Nilsson, J. L., Additive CHARMM force field for naturally occurring modified ribonucleotides. *J. Comput. Chem.* **2016,** *37* (10), 896-912.
40. Tomasi, J.; Mennucci, B.; Cammi, R., Quantum Mechanical Continuum Solvation Models. *Chem. Rev.* **2005,** *105* (8), 2999-3094.
41. Scalmani, G.; Frisch, M. J., Continuous surface charge polarizable continuum models of solvation. I. General formalism. *J. Chem. Phys.* **2010,** *132* (11), 114110.
42. Brooks, B. R.; Brooks, C. L., 3rd; Mackerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M., CHARMM: the biomolecular simulation program. *J. Comput. Chem.* **2009,** *30* (10), 1545-614.
43. Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H., PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J. Chem. Theory Comput.* **2011,** *7* (2), 525-537.
44. Søndergaard, C. R.; Olsson, M. H. M.; Rostkowski, M.; Jensen, J. H., Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values. *J. Chem. Theory Comput.* **2011,** *7* (7), 2284-2295.
45. Darden, T., Treatment of Long-Range Forces and Potential. In *Computational Biochemistry & Biophysics*, Becker, O. M.; MacKerell, A. D., Jr; Roux, B.; Watanabe, M., Eds. Marcel Dekker, NY.: 2001.
46. Berendsen, H.; Postma, J.; van Gunsteren, W.; DiNola, A.; Haak, J., Molecular dynamics with coupling to an external bath. . *J Chem Phys.* **1984,** *811*, 3684-3690.
47. Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K., Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005,** *26* (16), 1781-802.
48. Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; MacKerell, A. D., Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ, ψ and Side-Chain $\chi1$ and $\chi2$ Dihedral Angles. *J. Chem. Theory Comput.* **2012,** *8* (9), 3257-73.
49. Huang, J.; MacKerell, A. D., Jr., CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. *J. Comput. Chem.* **2013,** *34* (25), 2135-45.
50. Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell, A. D., CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **2017,** *14* (1), 71-73.
51. Jorgensen, W.; Chandrasekar, J.; Madura, J.; Impey, R.; Klein, M., Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys* **1983,** *79*, 926-935.
52. Neria, E.; Fischer, S.; Karplus, M., Simulation of activation free energies in molecular systems. *J. Chem. Phys.* **1996,** *105*, 1902-1921.