ORIGINAL PAPER

Automated conformational energy fitting for force-field development

Olgun Guvench · Alexander D. MacKerell Jr.

Abstract We present a general conformational-energy fitting procedure based on Monte Carlo simulated annealing (MCSA) for application in the development of molecular mechanics force fields. Starting with a target potential energy surface and an unparametrized molecular mechanics potential energy surface, an optimized set of either dihedral or grid-based correction map (CMAP) parameters is produced that minimizes the root mean squared error *RMSE* between the parametrized and targeted energies. The fitting is done using an MCSA search in parameter space and consistently converges to the same *RMSE* irrespective of the randomized parameters used to seed the search. Any number of dihedral parameters can be simultaneously parametrized, allowing for fitting to multi-dimensional potential energy scans. Fitting options for dihedral parameters include non-uniform weighting of the target data, constraining multiple optimized parameters to the same value, constraining parameters to be no greater than a user-specified maximum value, including all or only a subset of multiplicities defining the dihedral Fourier series, and optimization of phase angles in addition to force constants. The dihedral parameter fitting algorithm's performance is characterized through multi-dimensional fitting of cyclohexane, tetrahydropyran, and hexopyranose monosaccharide energetics, with the latter case having a 30-dimensional parameter space. The CMAP fitting is applied in the context of polypeptides, and is used to develop a parametrization that simultaneously captures the φ,ψ energetics of the alanine dipeptide and the alanine tetrapep-tide. Because the dihedral energy term is common to many force fields, we have implemented the dihedral-fitting algorithm in the portable Python scripting language and have made it freely available as "fit_dihedral.py" for download at http://mackerell.umaryland.edu.

Keywords Empirical force field · Molecular modeling · Dihedral parameter · Torsion parameter · CMAP

Introduction

The continued increase in the speed of computers and the ease-of-use of computational chemistry software has led to the widespread application of computational methods to chemical and biological problems. Molecular mechanics (MM) force fields are now a routinely used computational tool in the study of biological systems such as proteins, nucleic acids, carbohydrates, and lipids, and highly optimized force-fields are available for these types of systems [1]. However, force fields for condensed-phase simulations of small molecules of medicinal interest lag behind, primarily because the wide span of chemical space requires that a very large number of parameters need to be developed in order to support the simulation of arbitrary chemical entities of medicinal interest. Whereas the nonbonded parameters for small molecules can often be well-assigned either by manual inspection and analogy to previously parametrized molecules or in an automated fashion [2–4], the bonded parameters can pose more difficulty as the probability of having an unparametrized connectivity in the molecule of interest becomes successively larger for bonds (two consecutive atoms), valence angles (three consecutive atoms), and dihedral or torsion angles (four consecutive atoms).

O. Guvench · A. D. MacKerell Jr. (✉)
Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland,
20 Penn St., HSF II-629, Baltimore, MD 21201, USA
e-mail: alex@outerbanks.umaryland.edu

The ease-of-use of many quantum mechanical (QM) software packages and the wide availability of computing power has made QM geometry optimization accessible to most interested researchers, and missing bond and angle parameters can be quickly developed by taking parameters for a similar connectivity and manually adjusting them to match the QM data. However, the development of dihedral parameters to match QM conformational scans is a more difficult task because multiple conformational geometries and energies must be simultaneously fit. Additionally, because multi-dimensional QM scans are now tractable and drug-like molecules often contain more than one dihedral degree-of-freedom about rotatable bonds, simultaneous fitting of multiple dihedral parameters becomes a desirable goal. Such a task would benefit from a general, well-characterized, and easy-to-use automated dihedral parametrization software, such as the one we present here. The methodology we describe and characterize also includes conformational energy fitting to the CMAP grid-based cross-map energy term [5, 6], which was originally introduced in the context of the CHARMM protein force field [7] to further refine the energetics of the protein polypeptide backbone (i.e., φ,ψ or Ramachandran energy surface) relative to using only dihedral energies.

Fitting approach

The dihedral angle energy of a molecule in a given conformation, E_{dihedral}, in an MM force-field representation is commonly determined by

$$E_{\text{dihedral}} = \sum_{j}^{\text{dihedrals}} \sum_{n}^{\text{multiplicities}} K_{j,n}\left[1 + \cos\left(n\chi_j - \sigma_{j,n}\right)\right]. \quad (1)$$

The dihedral angle χ_j is defined by the bonded sequence of atoms 1–2–3–4, and the sum over multiplicities n is a Fourier series with coefficients of $K_{j,n}$ and phase angles $\sigma_{j,n}$. Thus, in principle, it is possible to reproduce any periodic function, and therefore any rotational energy profile, using Eq. 1.

Typical target data for fitting are the energies from QM adiabatic potential energy scans. For a hypothetical four-atom molecule with connectivity 1–2–3–4, a relaxed potential energy scan would be done for rotation about bond 2–3 in the QM representation. The same scan would then be done in the MM representation using all force field terms (bonds, angles, van der Waals, electrostatics, etc.) but with the $K_{j,n}$ of the dihedrals being optimized set to 0, and a difference potential would be calculated by subtraction of the QM and MM energy profiles. This difference potential would be fit exactly using Eq. 1 and the resulting $K_{j,n}$ and $\sigma_{j,n}$ would be the dihedral parameters.

In practice, the series in Eq. 1 is often truncated at $n=3$, reflecting the 3-fold nature of the energy profile for rotation around the bond connecting two sp^3-hybridized centers. The $n=1$ and $n=2$ terms are useful for reproducing local minima and barriers of different heights, as well as for capturing the energetics of scans in which one or both of the central two atoms is sp^2-hybridized. Additionally, the $n=6$ term can be useful in the case where the two central atoms are sp^2- and sp^3-hybridized, respectively, and the former has two and the latter has three identical substituents, as in e.g., benzyl sulfonate. Also, in practice, the phase angle $\sigma_{j,n}$ is typically constrained to be either 0° or 180°. This preserves the symmetry of the function about $\chi_j=0$, which in turn ensures that a molecule and its mirror image have the exact same dihedral energy for the same $K_{j,n}$ and $\sigma_{j,n}$. A final comment about Eq. 1 is that changing the sign of $K_{j,n}$ yields the same-shaped curve as changing $\sigma_{j,n}$ from 0° to 180°. Thus, a common convention in force-field parameters is to constrain all $K_{j,n}$ to be non-negative and allow for $\sigma_{j,n}$ values of both 0° and 180°.

The truncation of the dihedral Fourier series and the constraint of symmetry about $\chi_j=0$ mean that arbitrary difference-potentials cannot be fit exactly. In place of an exact fit, a target function that measures the difference between the QM and MM energies is optimized. A common target function is the root-mean squared error RMSE between these energies,

$$RMSE = \sqrt{\frac{\sum_{i}\left(E_i^{\text{QM}} - E_i^{\text{MM}} + c\right)^2}{\sum_{i} 1}}, \quad (2)$$

where the sum is over all conformations i of the molecule in the scan, E_i^{QM} is the QM energy of conformation i, E_i^{MM} is the total MM energy, including the energy of the dihedrals for which the parameters are being optimized (Eq. 1), and c is a constant that vertically aligns the data as the optimization proceeds and is defined by

$$\frac{\partial RMSE}{\partial c} = 0. \quad (3)$$

Any number of methods can be used to optimize RMSE. In the case where only the $K_{j,n}$ and not the $\sigma_{j,n}$ are to be fit, the problem can be expressed as a linear system of equations, with the $K_{j,n}$ being the coefficients to be solved for. In such a case, this "least squares" approach is the most direct method and gives the optimal solution, and has been applied to dipeptide conformational energetics [8]. Other possibilities, which allow for fitting of the $\sigma_{j,n}$, include

systematic searching of parameter space [9], self-consistent iteration [10], genetic algorithms [9, 11], and the use of simplex, Fletcher-Powell, and Newton-Raphson minimizers [12, 13].

Here we present details of a general dihedral parameter fitting algorithm that uses Metropolis Monte Carlo [14] to optimize *RMSE*. In addition to being an efficient way of searching parameter space, the Monte Carlo method allows for the introduction of various additional options into the optimization process. One such option is a constraint on the maximum value of the $K_{j,n}$ so as to produce physically reasonable parameters. Another is the equivalencing of $K_{j,n}$ for two or more different values of j. Equivalencing is useful in the case of a linear molecule like 1–2–3–4–5, in which a two-dimensional scan consisting of rotations about bonds 2–3 and 3–4 is used as target data and one wishes to enforce $K_{1-2-3-4,n}=K_{2-3-4-5,n}$. Weighting of different conformations can be done by extending Eq. 2 to

$$RMSE = \sqrt{\frac{\sum_i w_i \left(E_i^{QM} - E_i^{MM} + c\right)^2}{\sum_i w_i}}, \qquad (4)$$

where w_i is a weight factor for conformation i and can, for example, be used to favor more accurate fitting of low-energy conformations while sacrificing the fit of high-energy ones. Simultaneous fitting of multiple $K_{j,n}$ to data having a multiple number of dimensions is readily done and, as shown subsequently, *RMSE* convergence is achievable even for very high dimensionality. Finally, the phase angles $\sigma_{j,n}$ can be allowed to vary as part of the optimization process.

The above approach can also be extended to the parametrization of the grid-based cross map term (CMAP) employed in the CHARMM protein force field to accurately reproduce the conformational energetics of the polypeptide backbone [5, 6]. The CMAP energy is a function of two dihedral angles simultaneously. In the case of the polypeptide backbone, the CMAP energy is a function of the backbone dihedrals φ and ψ. The CMAP parameters are simply the difference potential energies between the QM and MM dipeptide surfaces calculated at 15° increments of φ and ψ, and an interpolation function is used for calculating the CMAP energies for off-grid φ/ψ values. This approach can almost exactly reproduce the target QM surface for the alanine dipeptide. However, exact reproduction of all energies is not possible if, for example, in addition to the dipeptide conformational energies, tetrapeptide conformational energies are also targeted.

In order to include both dipeptide and tetrapeptide conformational energies in the MCSA fitting, the quantity $RMSE_{CMAP}$ is targeted:

$$RMSE_{CMAP} = \frac{\left(w_{dipeptide} * RMSE_{dipeptide}\right) + \left(w_{tetrapeptide} * RMSE_{tetrapeptide}\right)}{w_{dipeptide} + w_{tetrapeptide}}$$
$$(5)$$

Here, $RMSE_{dipeptide}$ and $RMSE_{tetrapeptide}$ are defined independently by Eq. 4 for the dipeptide and tetrapeptide data. Importantly, the constant c in Eq. 4 varies independently for the two sets of data. The weight factors $w_{dipeptide}$ and $w_{tetrapeptide}$ can be chosen to bias the fit toward either the dipeptide or tetrapeptide data.

Computational details

QM energies were computed at the MP2/cc-pVTZ//MP2/6-31G(d) level for pyranose monosaccharides, and at the MP2/cc-pVTZ level for cyclohexane and tetrahydropyran [15–17]. Alanine dipeptide and tetrapeptide energies were at the RI-MP2/cc-pVTZ//MP2/6-31G(d) level [18–20]. MP2 and RI-MP2 data used in dihedral parameter fitting were from relaxed potential energy scans calculated using the Gaussian03 [21] and Q-Chem [22] software packages, respectively. MM energies were those from MM-optimized geometries (gradient $<10^{-3}$ kcal mol^{-1} Å^{-1}) with an infinite nonbonded cutoff and harmonic restraints with a force constants of 1,000 kcal mol^{-1} degree^{-2} on all dihedral angles that have parameters to be fit, and were computed using the CHARMM software [23] and the steepest descent [24] and conjugate gradient [25] optimizers implemented therein. Parameters for the MM calculations were the CHARMM22 all-atom protein set [7], additive CHARMM parameters for cyclohexane and ethers [26], and parameters under development for the monosaccharides (parameter set "combo*" in [27]).

Input for the dihedral fitting program consists of a file containing the QM energies, another file containing the MM energies calculated with the dihedral parameters to be fit set to 0, and a separate file for each unparametrized dihedral angle containing the values of that dihedral. The list of data in each file must represent the same ordering of conformations, and the first line in each of the dihedral angle files contains the four atom-types for that dihedral. Based on these atom types, the program automatically equivalences the parameters for all dihedrals that consist of the same atoms types. Thus, in fitting a two-dimensional scan of the $C_1C_2C_3C_4$ and $C_2C_3C_4C_5$ dihedrals in *n*-pentane, the parameters of the two dihedrals would automatically be constrained to be the same if the atom types were specified such that $C_1=C_5$ and $C_2=C_3=C_4$. The user can choose to further equivalence any other dihedrals

even though they may have different atom types, and also decide what multiplicities (n in Eq. 1) should be used for each equivalenced group. If non-uniform weighting of the conformations is desired, a file containing the weight-factor w_i for each conformation i is read prior to starting the Monte Carlo search, allowing, for example, the application of Boltzmann weighting to all points. Two possible temperature schemes are available for the Monte Carlo procedure, a constant-temperature or a simulated-annealing [28] protocol with an exponential-cooling schedule of

$$T_m = T_0 \exp(-4m/m_{\max}) \qquad (6)$$

where T_0 is the starting temperature, m is the current Monte Carlo step number, $m_{\max}$ is the maximum number of Monte Carlo steps, and T_m is the temperature at step m. The difference in energy, ΔE, between step m and step $m-1$ used in the Metropolis exchange criterion is

$$\Delta E = RMSE_m - RMSE_{m-1} \qquad (7)$$

and $RMSE$ at every step is recalculated using Eq. 4 and the constraint Eq. 3.

The implementation of the dihedral fitting, called "fit_dihedral.py", is in the Python scripting language (http://www.python.org) and uses only the "math", "random", "string", and "sys" libraries, which are a standard part of the Python distribution. fit_dihedral.py is freely available for download at http://mackerell.umaryland.edu.

The CMAP fitting also is implemented in Python, but writes out a new CMAP parameter file after each Monte Carlo step and calls the CHARMM program to calculate the energy, in contrast to fit_dihedral.py, in which Eq. 1 is implemented directly in Python and makes no calls to external programs.

Results and discussion

Equivalencing: cyclohexane

Parametrization of the ring C–C–C–C dihedrals in cyclohexane is an illustrative example of the automatic equivalencing of dihedral parameters. The energy surface for rotation about the χ_1 dihedral is complicated due to a barrier-crossing at $\chi_1 = -15°$ as the molecule goes from the global energy-minimum boat conformation to the local energy-minimum chair conformation. During the $\chi_1 = C_1C_2C_3C_4$ scan from $-100°$ to $+100°$, the five other C-C-C-C χ dihedrals also undergo changes in value. In the case of χ_2 and χ_6, this change spans over $100°$ as the molecule goes from a twist-boat conformation to a chair conforma-
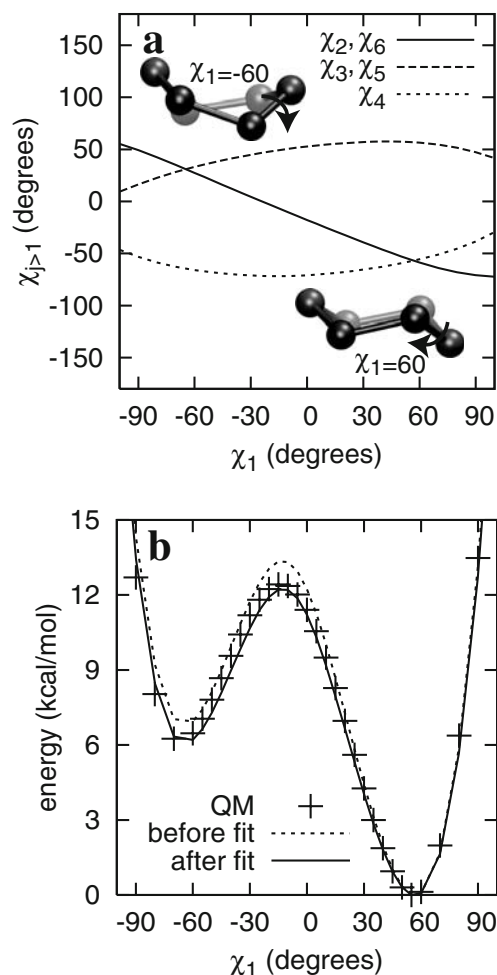

Fig. 1 Cyclohexane χ_2, χ_3, χ_4, χ_5, and χ_6 C–C–C–C dihedrals (**a**) and energy (**b**) during a relaxed potential energy scan of χ_1. $\chi_1 = C_1C_2C_3C_4$. Quantum mechanics (QM): MP2/cc-pVTZ (+), molecular mechanics (MM) before fit: $K=0.0$ kcal mol^{-1} (*dotted line*), and MM after fit: $K=0.19$ kcal mol^{-1}, $\sigma=180°$ (*solid line*)

tion (Fig. 1a). Because each of the six χ dihedrals is composed of the same atom types, the algorithm automatically constrains the $K_{j,n}$ to be the same, where j runs from 1 to 6. Using only the $n=3$ multiplicity and constraining $K_{j,3}$ to be in the range $[-3:3]$ kcal mol^{-1}, five 5,000-step simulated-annealing runs were seeded with random K values in this range. Since only a single multiplicity is used and equivalencing is in effect, only a single parameter K is being varied to optimize $RMSE$ as a function of K and the values of the six simultaneously changing χ dihedrals. From a starting temperature of $T_0 = 1,000$ K, all five runs converge to the same parameter value of 0.19 kcal mol^{-1} with a phase angle of $180°$, and reduce the $RMSE$ of 0.53 kcal mol^{-1} for $K=0$ kcal mol^{-1} to an $RMSE$ of 0.38 kcal mol^{-1} for $K=0.19$ kcal mol^{-1}, resulting in correction of the barrier height and chair conformation ($\chi_1 = 60°$) energies, which were both too high by nearly 1 kcal mol^{-1} prior to parametrization (Fig. 1b).

Simultaneous fitting: tetrahydropyran

Tetrahydropyran, in which one of cyclohexane's methylene groups is replaced by a ring ether, presents a case of simultaneous fitting. Taking the C–C–C–C dihedral parameter from cyclohexane leaves two pairs of equivalenced dihedrals, $C_1O_1C_5C_4/C_5O_1C_1C_2$ (χ_1/χ_3) and $O_1C_1C_2C_3/O_1C_5C_4C_3$ (χ_2/χ_4), to be fit. QM scans of χ_1 and of χ_2 show that, like cyclohexane, the other χ values in the system vary simultaneously as these two are scanned (Fig. 2). Inputting χ_1, χ_2, χ_3, and χ_4 leads to automatic equivalencing of χ_1 to χ_3 and χ_2 to χ_4 based on atom type, and the corresponding K are simultaneously optimized. Using the same optimization protocol as for cyclohexane ($n=3$ multiplicity, K constrained to be in the range [−3:3] kcal mol^{-1}, and five 5,000-step simulated-annealing runs seeded with random K values in this range) leads to convergence to the same $RMSE$ and nearly-identical K values in each of five annealing runs. The final optimized

Fig. 2 Variation in the tetrahydropyran χ_1, χ_2, χ_3, and χ_4 dihedrals during relaxed potential energy scans of χ_1 (**a**) and χ_2 (**b**). $\chi_1 = C_1O_1C_5C_4$, $\chi_2 = O_1C_1C_2C_3$, $\chi_3 = C_5O_1C_1C_2$, and $\chi_4 = O_1C_5C_4C_3$

values in the five runs are $K_{1,3} = K_{3,3} = 0.19$, 0.20, 0.21, 0.20, or 0.19 kcal mol^{-1} and $K_{2,3} = K_{4,3} = 0.33$, 0.31, 0.33, 0.31, or 0.31 kcal mol^{-1}, and the respective phase angles are 0° and 180°. Using only 3-fold parameters leads to a modest reduction of $RMSE$ from 0.98 kcal mol^{-1} to 0.92 kcal mol^{-1}, reflecting the good agreement with the target data prior to fitting. There are nonetheless specific conformations that benefit from the optimization process, in particular conformations with $\chi_1 = 40$, 50, $\chi_2 = 10$, 20, and $\chi_{C_1C_2C_3C_4} = -50$, −40, −30, which in the respective scans of these dihedrals come to match the QM energies post-optimization, compared to over-estimation of these energies by up to 1.1 kcal mol^{-1} prior to optimization (Fig. 3, Table 1).

Fitting in multi-dimensional parameter space: pyranose monosaccharides

Fragment-based approaches to parameter development divide the molecule of interest into smaller fragments, thereby reducing the number of atoms as well as the number of dihedral degrees-of-freedom and making QM relaxed potential energy scans tractable. However, increasing computer power has made direct QM scans of many-atom molecules with multiple dihedral degrees-of-freedom possible. Dihedral parameters derived from these more complicated scans are preferable to relying on the transferability of dihedral parameters from smaller fragments since dihedral parameters are critically important to the conformational energetics of flexible molecules.

An illustrative case of the importance of QM scans of the complete molecule vs smaller fragments is the diastereomers of the hexopyranose form of the monosaccharide glucose. A chirality change at C_1 converts α-D-glucose to β-D-glucose, while similar changes at the C_2, C_3, and C_4 positions yield α-D-mannose, α-D-allose, and α-D-galactose, respectively (Fig. 4). These changes place the hydroxyl groups in differing local chemical environments, which cannot be captured using a fragment-based approach, for example by using cyclohexanol as the model compound, because of the extensive number of intramolecular hydrogen bonds between hydroxyls in the full monosaccharides. Additionally, rotation of the $O_5C_5C_6O_6$ dihedral and the C_6 hydroxyl allow for hydrogen-bonding of this "exocyclic" hydroxyl with the O_5 ring ether or the C_4 hydroxyl, posing a further complication to a fragment-based approach.

To characterize the MCSA fitting algorithm, we apply it to fitting the energetics of 1,860 hexopyranose conformations comprising hydroxyl, exocyclic group, and ring deformation scans of a variety of glucopyranose diastereomers (Table 2). The resultant parameters will be applicable to the various diastereomers so that, for example,
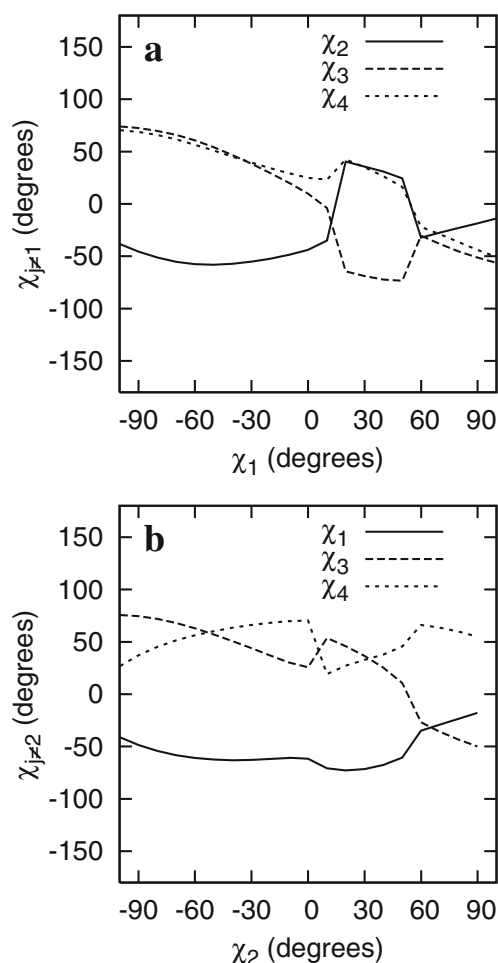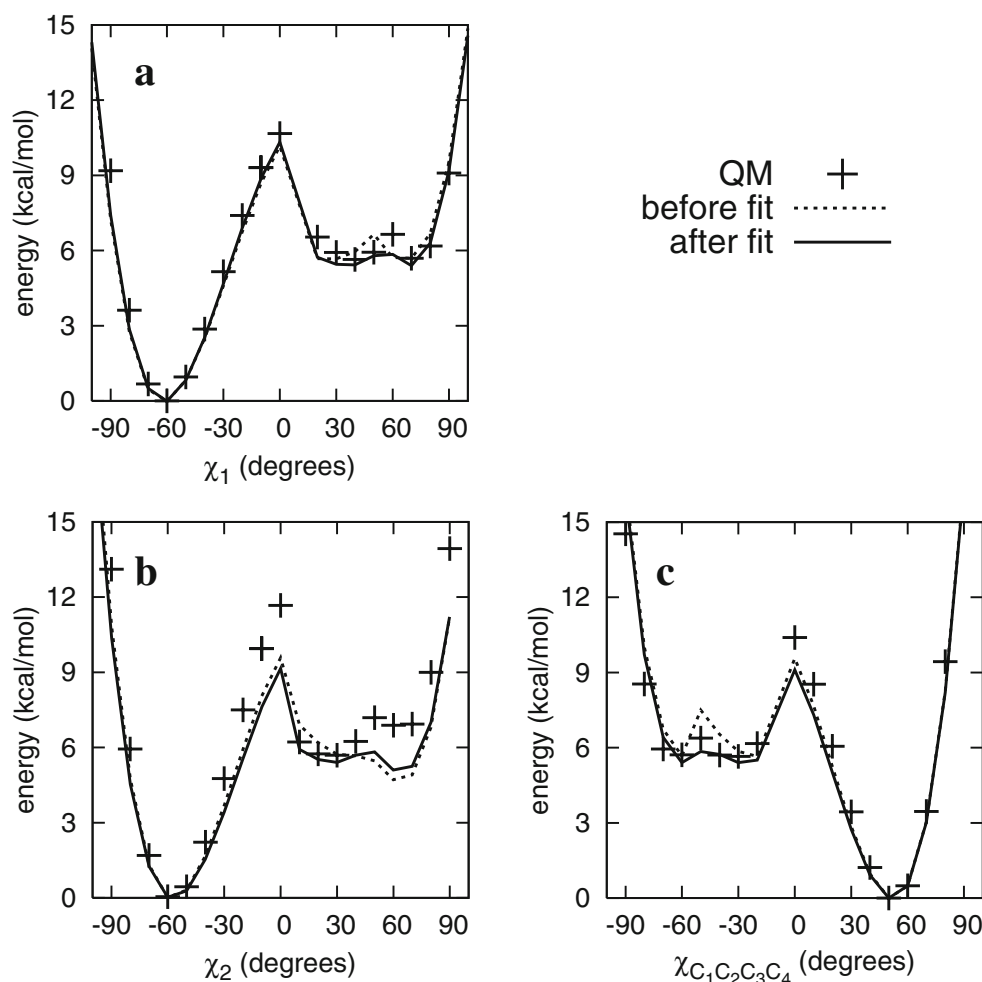
Fig. 3 Tetrahydropyran χ_1 ($C_1O_1C_5C_4$) (**a**), χ_2 ($O_1C_1C_2C_3$) (**b**), and $\chi_{C_1C_2C_3C_4}$ (**c**) relaxed potential energy scans. QM: MP2/cc-pVTZ (+), MM before fit: $K_{1,3}=K_{3,3}=K_{2,3}=K_{4,3}=$ 0 kcal mol^{-1} (*dotted line*), MM after fit: $K_{1,3}=K_{3,3}=$ 0.20 kcal mol^{-1}, $K_{2,3}=K_{4,3}=$ 0.31 kcal mol^{-1}, $\sigma_{1,3}=\sigma_{3,3}=0°$, and $\sigma_{2,3}=\sigma_{4,3}=180°$ (*solid line*)

glucose, galactose, and mannose will have the identical parameter set. Transferring existing dihedral parameters from alkanes, tetrahydropyran, and ethylene glycol still leaves undetermined the parameters for 13 hexopyranose dihedrals (hydroxyl rotation: $H_1O_1C_1C_2$, $H_1O_1C_1O_5$,

$H_2O_2C_2C_1$, $H_2O_2C_2C_3$, $H_3O_3C_3C_2$, $H_3O_3C_3C_4$, $H_4O_4C_4C_3$, $H_4O_4C_4C_5$, $C_5C_6O_6H_6$; ring deformation: $O_5C_1C_2O_2$, $O_1C_1O_5C_5$, $O_5C_5C_4O_4$; exocyclic-group rotation: $O_5C_5C_6O_6$). Automatic equivalencing based on atom-type ($H_2O_2C_2C_3 = H_3O_3C_3C_2 = H_3O_3C_3C_4 = H_4O_4C_4C_3$) reduces this to ten unique dihedrals, and allowing for $n=$ 1, 2, 3 multiplicity for each of these means $10 \times 3 = 30$ dihedral parameters to be simultaneously parametrized, with K values constrained as previously to the range $[-3:3]$ kcal mol^{-1}. Thus, this example represents an extreme case of fitting in multi-dimensional parameter space.

In contrast to the much simpler cases of cyclohexane and tetrahydropyran, in which the optimal parameters were determined in the first several hundred steps of 5,000-step exponential-cooling Monte Carlo runs, the 30-dimensional fit to the pyranose energetics shows much slower convergence behavior. Using exponential cooling (Fig. 5a), the maximum number of steps must be set to 50,000 in order to consistently converge to the same *RMSE* in each of ten Monte Carlo runs, while runs of 500 or 5,000 steps are insufficient (Fig. 5c,e). Using a constant-temperature scheme at 35 K (Fig. 5b), which yields a Monte Carlo

Table 1 Relative energies of selected tetrahydropyran conformers before and after dihedral parameter optimization. *QM* Quantum mecahnics, *MM* molecular mechanics

χ scan[a]	χ value (degrees)	Energy (kcal mol^{-1}) QM[b]	MM before fit	MM after fit
χ_1	40	5.64	5.95	5.43
χ_1	50	5.93	6.64	5.79
χ_2	10	6.22	6.90	5.92
χ_2	20	5.75	6.19	5.52
$\chi_{C_1C_2C_3C_4}$	−50	6.38	7.50	5.84
$\chi_{C_1C_2C_3C_4}$	−40	5.71	6.53	5.73
$\chi_{C_1C_2C_3C_4}$	−30	5.65	5.85	5.40

[a] Illustrated in Fig. 3
[b] MP2/cc-pVTZ

Fig. 4 Epimers of α-D-glucose. Starting from the upper-right molecule and going clockwise around α-D-glucose (center molecule), a chirality change at C_1 yields β-D-glucose, at C_2 yields α-D-mannose, at C_3 yields α-D-allose, and at C_4 yields α-D-galactose. Carbon atoms are numbered per the IUPAC convention. The ring-ether oxygen is O_5, and hydroxyl atom numbers are the same as the carbon atom to which they are attached
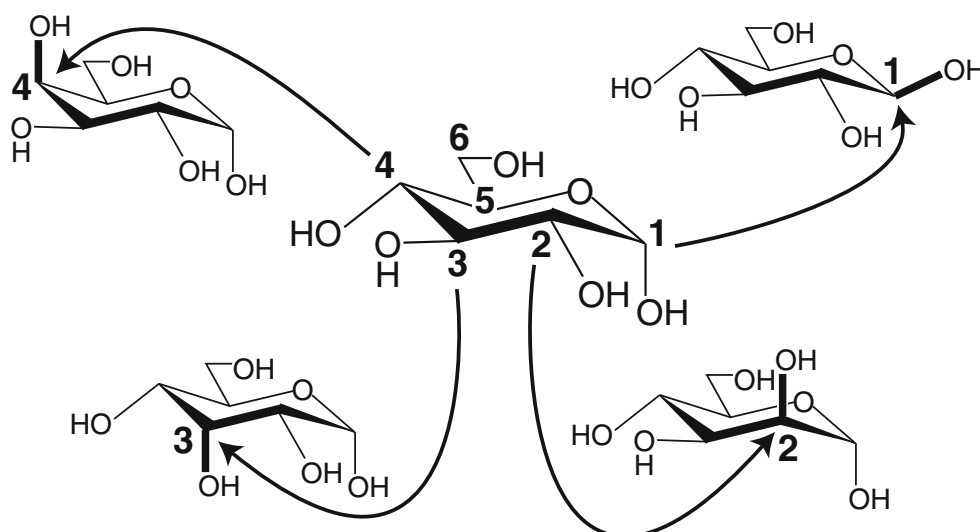
acceptance ratio of 0.2 to 0.3, the results are the same in that consistent convergence is seen only for runs of 50,000 steps (Fig. 5d,f). The advantage of the MCSA with exponential cooling as opposed to constant-temperature Monte Carlo is that the user does not have to take a trial-and-error approach to finding a temperature that yields a reasonable acceptance ratio.

While the search problem is much more difficult in this example compared to the prior cases, it is nonetheless possible to achieve converged RMSE results when simultaneously fitting 30 dihedral parameters. Another contrast with the simpler systems is that, though converged behavior is achieved with respect to RMSE, the parameters themselves show significant variability, both in the magnitude of

each $K_{j,n}$ as well as whether the associated $\sigma_{j,n}$ is 0° or 180°. For example, in the ten independent 50,000-step MCSA runs that converge to RMSE spanning only a 0.09 kcal mol^{-1} window (1.74 to 1.83 kcal mol^{-1}, Fig. 5c and e), the $n=3$ term for the $H_1O_1C_1O_5$ dihedral takes on parameter values ranging from $K=3.00$ kcal mol^{-1}, $\sigma=0°$ to $K=0.53$ kcal mol^{-1}, $\sigma=180°$ (i.e., $K=-0.53$ kcal mol^{-1}, $\sigma=0°$). Likewise, the $n=1$ term for the $O_1C_1O_5C_5$ dihedral takes on parameter values spanning the range $K=2.74$ kcal mol^{-1} down to $K=-1.83$ kcal mol^{-1}. The complete set of K values for each of the ten runs is listed in Table 3, along with the corresponding RMSE for each run. The standard deviations in the fit parameters K are as large as 1.27 kcal mol^{-1}, in stark contrast to the standard deviation of 0.03 kcal mol^{-1} for the RMSE of the ten independent runs. Thus, parameter space for such a complicated case is populated with multiple minima in different regions of the parameter space, with each minimum having a near-identical RMSE.

Weighted fitting: pyranose monosaccharide ring deformation

The dataset of hexopyranose conformations is populated mostly with ring conformations in the $^4C_1$ chair form (Fig. 4). In order to properly capture the energetics of chair-to-boat conversion, which are influenced by the $O_5C_1C_2O_2$, $O_1C_1O_5C_5$, and $O_5C_5C_4O_4$ dihedral parameters, scans of these ring dihedrals were included in the fit (Table 2, "ring" and "all"). The number of non-chair conformations resulting from these scans is dwarfed by the number of chair conformations from the other scans. As a result, this data set is weighted heavily toward the development of optimized parameters that reproduce chair energetics at the potential expense of boat energetics. This indeed does turn out to be the case in practice for β-D-galactopyranose,

Table 2 List and number of pyranose monosaccharide conformations used as target data for dihedral fitting

Monosaccharide	Type of conformational scan[a]	Number of conformations
α-Altrose	All	450
β-Altrose	All	450
α-Glucose	All	354
β-Glucose	All	330
α-Galactose	Exocyclic + ring	90
β-Galactose	Exocyclic + ring	90
β-Mannose	C2 hydroxyl	24
α-Talose	C3 hydroxyl	24
β-Gulose	C4 hydroxyl	24
β-Idose	C3 hydroxyl	24
Total		1,860

[a] *Exocyclic* $O_5C_5C_6O_6$ exocyclic dihedral; *ring* $O_5C_1C_2O_2$, $O_1C_1O_5C_5$, and $O_5C_5C_4O_4$ ring dihedrals; *all* exocyclic + ring + C1, C2, C3, C4, and C6 hydroxyls

Fig. 5 Fitting results for the hexopyranoses in the exponential cooling (**a**) or constant temperature (**b**) schemes. Root-mean squared error *RMSE* for ten runs as a function of the maximum number of steps and of exponential cooling (**c**) or constant temperature (**d**) Monte Carlo. Progress of the best-sampled *RMSE* vs the Monte Carlo step for ten runs using exponential cooling (**e**) or constant temperature (**d**). *Horizontal lines* in **c**–**f** denote *RMSE* with *K* to be optimized all set to 0 kcal mol^{-1}
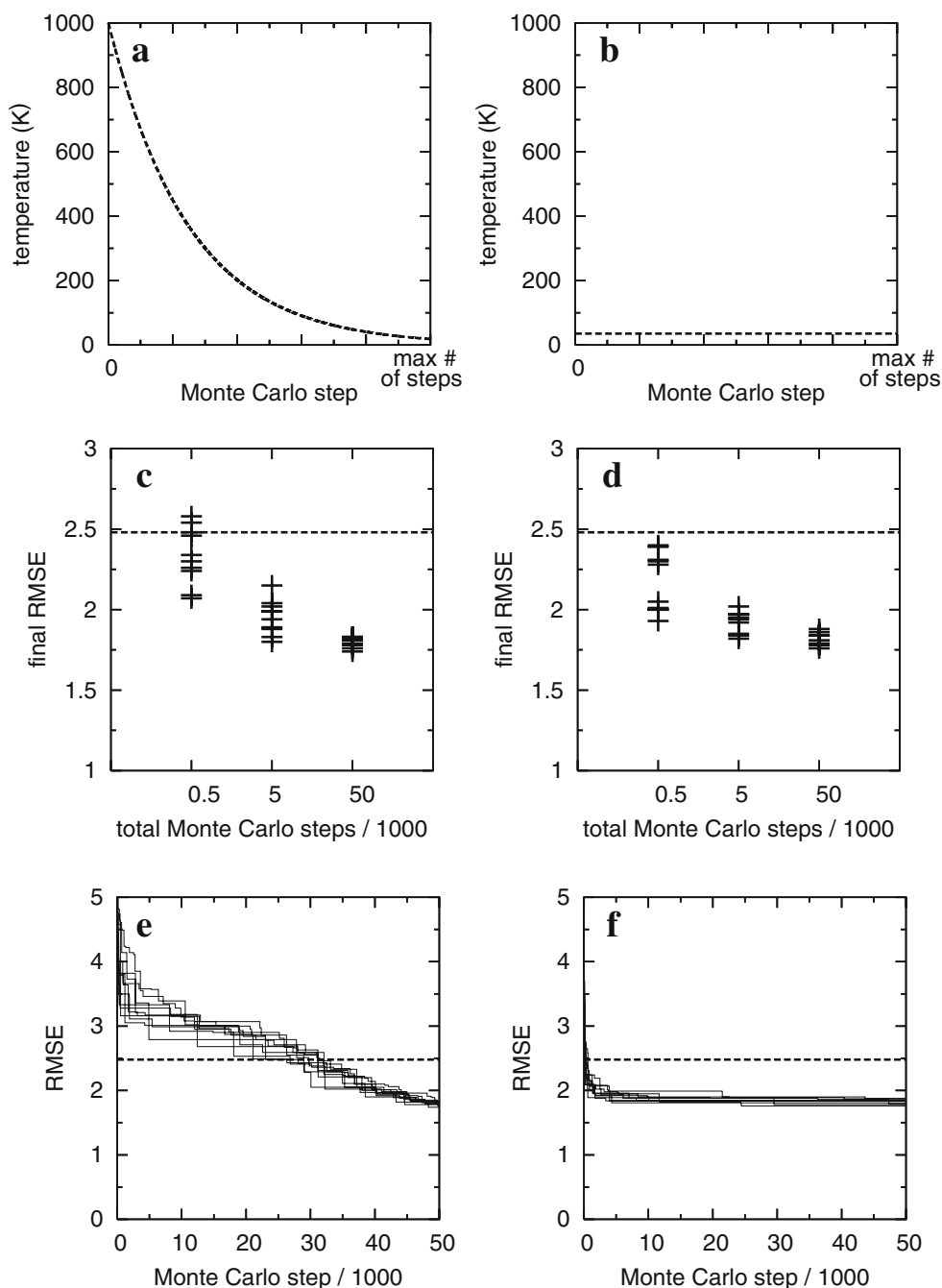
where the $O_5C_1C_2O_2$ scan is qualitatively incorrect relative to the QM in the absence of weighting, with the chair and boat conformations being isoenergetic instead of separated by 5 kcal mol^{-1} (Fig. 6).

The problem of under-represented conformations can be corrected simply by increasing the weight factors w_i for these conformations (Eq. 4). The choice of weight factors is an empirical task and may take several iterations of choosing different w_i values to get the desired results. In the present example, applying a weight factor of 5 to conformations in the scan with dihedral values of 75° to 150° is sufficient to correct their under-representation and

achieve dramatic improvement in chair vs boat energetics (Fig. 6). As with the unweighted fitting, exponential cooling over 50,000 Monte Carlo steps is sufficient to converge the *RMSE* of ten independent runs. The *RMSE* of the best unweighted fit, 1.74 kcal mol^{-1}, increases negligibly to 1.78 kcal mol^{-1} with this weighting scheme. Weight factors must be applied judiciously so as to balance the effect of the increased weighting of some conformations on the energies of the other conformations. If a sizable minority of conformations is given large weight factors, the resultant near-exact fitting of the energetics of their respective conformations will come at the expense of the
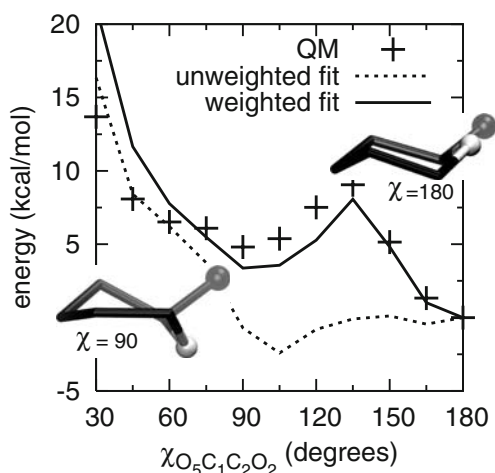
Fig. 6 Effect of weighting conformations in a ring-deformation scan of β-galactopyranose during fitting to 1,860 hexopyranose (Table 2) MP2/cc-pVTZ//MP2/6-31G(d) (QM, +) conformational energies. The weighted fit had w_i values (Eq. 4) of 5 for $75° \leq \leq 150°$ in this scan and w_i values of 1 for the other 1,854 conformations. Hydrogens and hydroxyl groups have been omitted from the graphics for clarity, and O_5 and C_6 are displayed as *white* and *black spheres*, respectively

energetics of the rest of the conformations. For example, Boltzmann weighting based on the target QM energies often yields accurate parametrization of low-energy conformations. However, Boltzmann weighting can also cause inaccurate energies for conformations located at or near high-energy barriers, which in turn will, for example, compromise the barrier-crossing transition rates in molecular dynamics simulations. In our experience, w_i values of less than or equal to five are typically appropriate for the under-represented conformations, assuming w_i values of unity for the rest of the target data.

Fitting phase angles in addition to force constants: pyranose monosaccharide exocyclic rotation

The previous two examples involved fitting exclusively the D forms of hexopyranose monosaccharides. Nonetheless, the parameters from those unweighted and weighted fits are transferable to the L forms of these sugars since the phase angles $\sigma_{j,n}$ were constrained to 0°/180° so as to preserve the symmetry of Eq. 1 about $\chi_j=0°$. It is possible to further refine the parameters by removing this constraint. Of course, the resultant increase in accuracy comes at the expense of decreased transferability of the parameters. In particular, a pair of enantiomers will require unique phase angles $\sigma_{j,n}$.

Taking the parameter set from the prior unweighted fit to 1,860 hexopyranose conformational energies, we reoptimized just the $O_5C_5C_6O_6$ dihedral parameters that determine the energetics of exocyclic-group rotation by allowing for $-180° \leq \sigma_{j,n} \leq 180°$ and using β-D-galactopyranose $O_5C_5C_6O_6$ scans as target data. The additional degrees-of-

freedom afforded by variability in the $\sigma_{j,n}$ yield a significant improvement in the force field's ability to reproduce the QM target data (Fig. 7). This 6-dimensional fitting ($j=O_5C_5C_6O_6$, $n=1$, 2, 3, both $K_{j,n}$ and $\sigma_{j,n}$ allowed to vary), like the low-dimension cyclohexane and tetrahydropyran fits, showed convergence both in *RMSE* as well as the actual values of all of the parameters in multiple 5,000-step exponential cooling runs seeded with random parameter values and started at 1,000 K.

Breaking the symmetry of Eq. 1 by allowing phase-angle variability means that the L enantiomers of these β-D-galactopyranose conformations will have different energies using this parameter set, which is chemically incorrect. Thus, increased accuracy comes at the cost of decreased generality. Additionally, non-zero phase angles introduce singularities into the derivatives of the dihedral energy, and
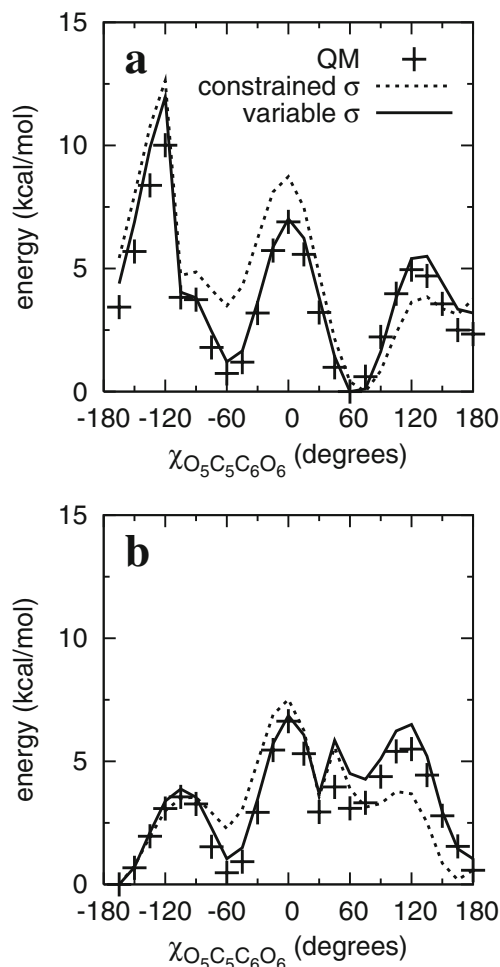

Fig. 7 a, b β-D-Galactopyranose $O_5C_5C_6O_6$ dihedral parameter fitting with constraints (constrained σ, *dotted line*) and without constraints (variable σ, *solid line*) on the phase angles. MP2/cc-pVTZ//MP2/6-31G(d) (QM, +) data for fitting were from both forward (**a**) and backward (**b**) scans of the dihedral, and differ because different intramolecular interactions are formed depending on the scan's direction (i.e., incrementing χ vs decrementing χ)

computer code must take this into account [29]. If possible, it is preferable to improve the fit through the use of non-uniform weighting of conformations instead of removing the constraints on the phase-angle parameters. Nonetheless, there may be particular instances when allowing variable phase angles is desirable, especially in biological systems where often only one enantiomer is found (e.g., amino acids [11] or nucleic acids [30]) or is relevant (e.g., chiral drugs). In such instances, the MCSA fitting approach is able to obtain converged optimization of both the $\sigma_{j,n}$ and the $K_{j,n}$.

Grid-based correction map

The CHARMM all-atom force field functional form was recently extended so as to better reproduce the conformational energetics of the polypeptide backbone in proteins. The extension involved the introduction of a new energy term, CMAP, which is a grid-based energy correction map and is a function of the backbone φ/ψ angles [5, 6]. Just as the dihedral energy term in Eq. 1 seeks to reproduce the difference energy between the MM surface with the target dihedrals set to zero and the QM surface as a function of the dihedral angle χ, the CMAP energy term reproduces the

difference energy between the QM and MM surfaces as a function of the φ and ψ angles simultaneously. That is, $E_{CMAP}=f(\varphi,\Psi)$ where $f(\varphi,\Psi)$ is constructed by two-dimensional bi-cubic interpolation through grid points located in φ/ψ space [5]. These grid points are evenly placed at 15° increments of φ and ψ, and each grid point has associated with it a difference energy. The difference energies at these grid points are the CMAP parameters.

Using the CMAP energy term, it is possible to exactly reproduce any difference energy as a function of φ/ψ. Thus, in the case of, e.g., alanine dipeptide, the entire adiabatic QM φ/ψ surface can be reproduced by the MM model, which is not the case using only dihedral terms for φ and ψ, as previously discussed [6]. In practice, based on data from protein crystal simulations, it was found that an empirical adjustment to the exact QM dipeptide surface was required to better capture the conformational properties of polypeptides [6].

In an effort to further refine the current CMAP parametrization [6], we have adapted the described MCSA fitting protocol to simultaneously fit CMAP parameters to alanine dipeptide and alanine tetrapeptide relative conformational energies. The target QM dipeptide and tetrapep-

Fig. 8 a–d Simultaneous alanine tetrapeptide and dipeptide energetic fitting. a φ/ψ angles in 51 alanine tetrapeptide conformations (x) and the corresponding CMAP grid points (squares). b Alanine tetrapeptide energies at the RI-MP2/cc-pVTZ//MP2/6-31G(d) level (QM, ×), the MM representation using CMAP to directly reproduce the alanine dipeptide QM surface (before fitting, dotted line), and the MM representation using a CMAP simultaneously fit to alanine dipeptide and tetrapeptide (after fitting, solid line). c Alanine dipeptide surface in the MM representation using alanine dipeptide fit CMAP parameters. d Alanine dipeptide surface in the MM representation after simultaneous fitting of CMAP to alanine dipeptide and alanine tetrapeptide QM relative energies. Contours are every 1 kcal mol⁻¹ and the data have been aligned so that the global minimum is at 0 kcal mol⁻¹
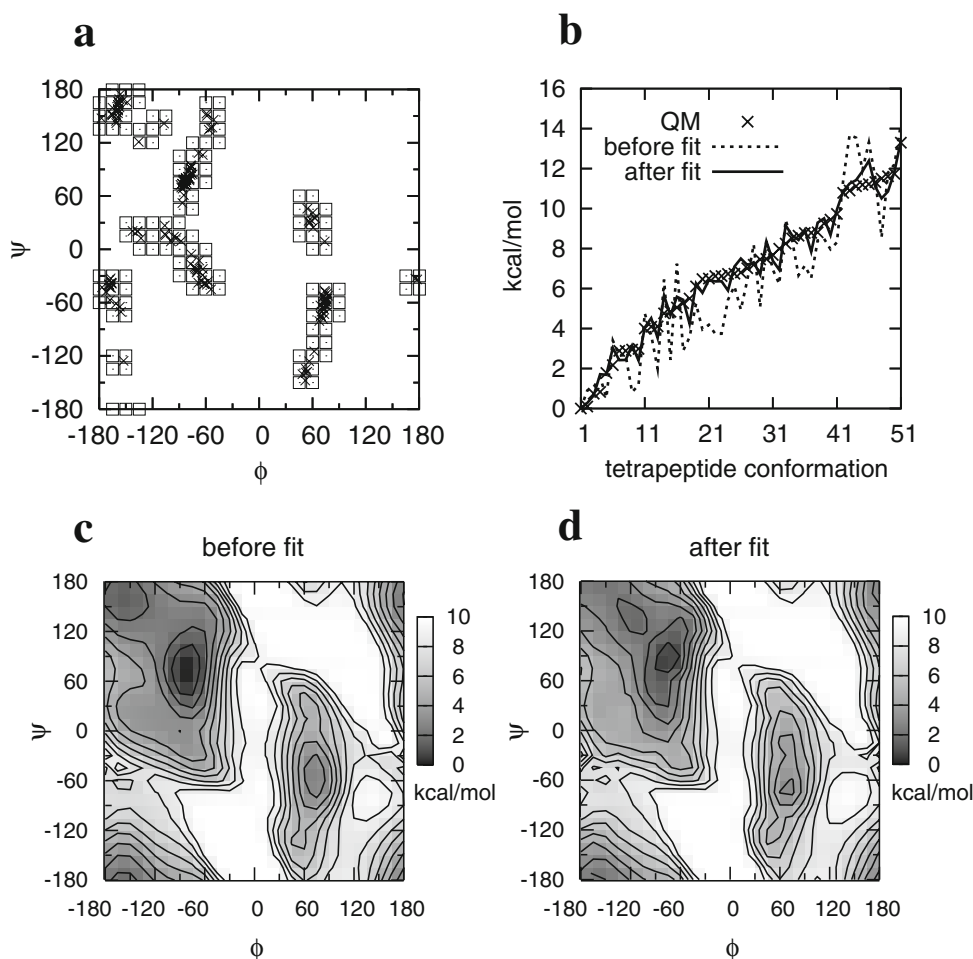
tide data were single-point RI-MP2/cc-pVTZ energies calculated from MP2/6-31G(d)-optimized geometries. While the dipeptide data consisted of the entire φ/ψ surface, the tetrapeptide data consisted of 51 structurally distinct conformations derived by clustering conformations sampled by MM molecular dynamics simulations [31]. For the purposes of fitting, the dipeptide and tetrapeptide data were given equal weighting ($w_{dipeptide}=w_{tetrapeptide}$ in Eq. 5), and the CMAP parameters (i.e., offset energies at the grid points) within a ±2 kcal mol⁻¹ window were sampled. The starting CMAP parameters were those that exactly reproduced the dipeptide surface such that $RMSE_{dipeptide}=0$.

The 153 φ/ψ values sampled in the 51 alanine tertrapeptide conformations are illustrated in Fig. 8a and populate all regions of φ/ψ space observed in high-quality protein crystal structures [32]. Figure 8a also shows the CMAP grid points whose parameters were allowed to vary by ±2 kcal mol⁻¹ during the fitting. To retain the

smoothness of the surface, whenever one of these parameters was changed by δE, the parameters of all the adjacent grid points were changed by $0.5 * \delta E$. Adjacent grid points not part of the set shown in Fig. 8a were allowed to vary at most by ±1 kcal mol⁻¹ relative to their starting values. With the starting CMAP parameters, $RMSE_{tetrapeptide}=1.53$ kcal mol⁻¹, which reflects the considerable scatter in the MM tetrapeptide energies relative to the QM target energies, and the common occurrence of errors as large as 2 kcal mol⁻¹ (Fig. 8b). In contrast, after MCSA fitting to combined tetrapeptide and dipeptide QM relative energies, the tetrapeptide MM energies are greatly improved, with most errors reduced to less than 0.5 kcal mol⁻¹ (Fig. 8b) and a final $RMSE_{tetrapeptide}$ of 0.57 kcal mol⁻¹.

The improvement in the tetrapeptide energies results from three subtle changes to the starting (i.e., exact QM) alanine dipeptide surface (Fig. 8c,d). First, the local minimum at φ/ψ=−165°/165° in the extended backbone

Table 3 Force constants from ten independent Monte Carlo simulated annealing (MCSA) fitting runs on the pyranose monosaccharides

Dihedral	Multiplicity	MCSA run										Average	Standard deviation
		1	2	3	4	5	6	7	8	9	10		
$C_5C_6O_6H_6$	1	1.16[a]	1.01	0.78	0.09	0.86	0.69	0.69	-0.45	-0.05	-0.27	0.45	0.57
	2	0.76	1.03	1.60	0.49	0.30	0.04	0.02	0.02	-0.03	0.40	0.46	0.53
	3	0.49	0.27	0.08	0.11	0.51	0.61	0.53	0.43	0.28	0.29	0.36	0.18
$H_1O_1C_1C_2$	1	-0.55	1.15	0.26	0.29	-0.09	-0.01	0.33	0.68	0.06	0.32	0.24	0.46
	2	-0.20	0.30	0.95	-0.11	0.30	0.80	0.64	0.85	-0.57	0.69	0.37	0.51
	3	-1.66	-0.18	-1.32	-2.60	-1.02	-1.53	0.07	-2.13	-2.04	0.47	-1.19	1.02
$H_1O_1C_1O_5$	1	0.68	1.61	1.59	1.97	1.99	1.33	2.57	2.37	1.70	2.27	1.81	0.55
	2	1.58	2.26	1.99	-0.01	2.58	0.67	1.53	0.83	0.99	1.14	1.36	0.79
	3	2.17	-0.53	1.31	2.64	1.41	2.82	0.31	3.00	1.42	0.25	1.48	1.20
$H_2O_2C_2C_1$	1	-0.66	0.21	-0.90	-0.87	-0.74	-0.05	-0.82	0.31	0.43	0.52	-0.26	0.59
	2	0.40	0.76	0.66	-0.42	-0.14	0.69	0.55	0.61	-0.51	-0.51	0.21	0.54
	3	0.40	0.93	-0.77	-0.15	-0.62	-0.02	0.48	-0.25	0.57	-0.66	-0.01	0.58
$H_2O_2C_2C_3$	1	0.02	0.20	-0.42	-0.63	-0.29	-0.11	-0.45	0.10	-0.73	0.17	-0.21	0.34
	2	0.39	0.30	0.21	-0.45	0.11	0.39	0.48	0.17	-0.69	-0.01	0.09	0.38
	3	-0.75	-0.27	-0.02	-0.21	-0.02	0.11	-0.43	-0.22	-0.40	-0.29	-0.25	0.25
$H_4O_4C_4C_5$	1	0.28	0.78	-0.08	0.15	-1.34	0.18	-1.35	-0.44	-0.08	-0.58	-0.25	0.69
	2	0.40	0.44	0.28	0.10	-0.14	1.09	-0.27	-0.06	0.31	-0.20	0.20	0.41
	3	0.20	-0.87	-0.43	-0.29	-0.66	0.01	0.52	0.15	-0.67	-0.18	-0.22	0.45
$O_5C_1C_2O_2$	1	-1.75	-2.08	-2.12	-2.46	-2.75	-2.36	-2.25	-1.29	-1.74	-3.00	-2.18	0.51
	2	-0.28	0.39	0.26	-0.04	-0.59	-0.14	0.08	1.22	-0.19	-0.33	0.04	0.50
	3	0.65	-0.68	-1.53	-0.27	-0.12	1.05	0.08	1.53	-0.95	-1.85	-0.21	1.09
$O_1C_1O_5C_5$	1	0.76	0.57	-0.87	0.43	0.81	-1.83	2.74	-0.09	0.75	-1.08	0.22	1.27
	2	1.32	1.08	0.40	1.20	1.77	-0.74	2.85	0.88	1.82	1.06	1.16	0.94
	3	-1.81	-1.28	-1.09	-1.75	-0.26	-1.39	0.53	-1.74	-0.69	-2.11	-1.16	0.81
$O_5C_5C_4O_4$	1	-1.85	-2.51	-1.78	-0.57	-0.12	-1.92	0.23	-2.16	-2.05	-1.28	-1.40	0.93
	2	-1.19	-0.12	-0.19	0.67	1.14	0.19	0.46	-0.09	-0.81	0.44	0.05	0.69
	3	0.43	0.14	0.85	1.60	0.45	0.02	0.82	-0.18	0.66	2.17	0.70	0.72
$O_5C_5C_6O_6$	1	-0.79	-1.25	-0.83	-1.11	0.19	-1.52	-1.30	-1.35	-1.12	-0.18	-0.93	0.55
	2	-0.02	0.14	0.70	0.62	0.83	0.48	0.45	0.77	0.65	1.40	0.60	0.39
	3	0.18	-0.31	-0.29	-0.42	-0.64	0.45	-0.33	-0.54	-0.46	-0.14	-0.25	0.33
RMSE		1.83	1.79	1.74	1.81	1.76	1.78	1.82	1.83	1.81	1.83	1.80	0.03

[a] All parameter and RMSE values are in kcal mol⁻¹

region of φ/ψ space has been shifted slightly to −120°/ 135°. Second, the local minimum at −150°/30° has been raised by ~1 kcal mol^{-1} and is no longer a local minimum. And third, the local minimum at 60°/−75° has been increased by less than 1 kcal mol^{-1} in energy. Qualitatively, the "before" and "after" dipeptide surfaces remain very similar, and the *RMSE* of the MCSA fit surface, $RMSE_{dipeptide}$, is 0.33 kcal mol^{-1} compared to a starting value of 0 kcal mol^{-1}.

One thousand MCSA steps were sufficient to achieve the improvements in the alanine tetrapeptide energies. The $RMSE_{CMAP}$ (Eq. 5) went from an initial value of 0.77 kcal mol^{-1} to a final value of 0.45 kcal mol^{-1}. Since changes to a single CMAP parameter affect only conformations with φ/ ψ values very close to that grid point, and therefore lead to small changes in $RMSE_{CMAP}$ and hence small ΔEs (Eq. 7), a starting temperature T_0 (Eq. 6) of 1 K gave good MC acceptance ratios during the annealing. The low-temperature annealing also makes the MCSA more akin to a minimization, which is appropriate given the fact that a single CMAP parameter affects only the energies of conformations nearby in φ/ψ space. This is in contrast to the dihedral parameter fitting, where changes in a dihedral parameter affect the energies of all conformations, necessitating a higher T_0 so as to not get trapped in local *RMSE* minima while searching parameter space.

Conclusions

We have presented and characterized an MCSA conformational-energy fitting algorithm for use in the development of molecular mechanics force fields. For the fitting of dihedral parameters, the algorithm consistently converges to the same optimized parameters, and therefore to the same value of the target function *RMSE* (Eq. 4), when the number of parameters to be fit is small. In a case of very high dimensionality (30 dihedral parameters to be fit), multiple MCSA runs also converge to a very narrow range of *RMSE*. However, the parameters that yield near-identical optimized *RMSE* can be qualitatively different, illustrating that there are multiple minima in dihedral parameter space that offer "best fits" to the target data. Extending the algorithm to the fitting of the CHARMM force-field CMAP term shows that the MCSA approach is also an effective way to develop CMAP parameters that find a balance between, in this case, alanine dipeptide and alanine tetrapeptide energetics. This approach to CMAP and multi-dimensional dihedral fitting is expected to prove useful in future applications such as parametrizing the energetics of the nucleic acid backbone, of glycosyl linkages in polysaccharides, and of flexible drug-like small-molecules.

References

1. MacKerell AD Jr (2004) J Comput Chem 25:1584–1604
2. Halgren TA (1996) J Comput Chem 17:490–519
3. Wang JM, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) J Comput Chem 25:1157–1174
4. Wang JM, Wang W, Kollman PA, Case DA (2006) J Mol Graphics Modell 25:247–260
5. MacKerell AD Jr, Feig M, Brooks CL III (2004) J Am Chem Soc 126:698–699
6. MacKerell AD Jr, Feig M, Brooks CL III (2004) J Comput Chem 25:1400–1415
7. MacKerell AD Jr, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiórkiewicz-Kuczera J, Yin D, Karplus M (1998) J Phys Chem B 102:3586–3616
8. Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL (2001) J Phys Chem B 105:6474–6487
9. Wang JM, Kollman PA (2001) J Comput Chem 22:1219–1228
10. Park S, Radmer RJ, Klein TE, Pande VS (2005) J Comput Chem 26:1612–1616
11. Okur A, Strockbine B, Hornak V, Simmerling C (2003) J Comput Chem 24:21–31
12. Maxwell DS, Tirado-Rives J (1994) Fitpar. Yale University, New Haven, CT
13. Norrby PO, Liljefors T (1998) J Comput Chem 19:1146–1166
14. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) J Chem Phys 21:1087–1092
15. Møller C, Plesset MS (1934) Phys Rev 46:618–622
16. Hariharan PC, Pople JA (1973) Theor Chim Acta 28:213–222
17. Woon DE, Dunning TH Jr (1993) J Chem Phys 98:1358–1371
18. Feyereisen M, Fitzgerald G, Komornicki A (1993) Chem Phys Lett 208:359–363
19. Weigend F, Haser M (1997) Theor Chem Acc 97:331–340
20. Distasio RA, Steele RP, Rhee YM, Shao YH, Head-Gordon M (2007) J Comput Chem 28:839–856
21. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomery JA, Vreven Jr T, Kudin KN, Burant JC, Millam JM, Iyengar SS, Tomasi J, Barone V, Mennucci B, Cossi M, Scalmani G, Rega N, Petersson GA, Nakatsuji H, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda K, Kitao O, Nakai H, Klene M, Li TW, Knox JE, Hratchian HP, Cross JB, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Ayala PY, Morokuma K, Voth GA, Salvador P, Dannenberg JJ, Zakrzewski VG, Dapprich S, Daniels AD, Strain MC, Farkas O, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Ortiz JV, Cui Q, Baboul AG, Clifford S, Cioslowski J, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Gonzalez C, Pople JA (2003) Gaussian 03. Gaussian Inc, Pittsburgh PA

22. Shao Y, Molnar LF, Jung Y, Kussmann J, Ochsenfeld C, Brown ST, Gilbert ATB, Slipchenko LV, Levchenko SV, O'Neill DP, DiStasio RA, Lochan RC, Wang T, Beran GJO, Besley NA, Herbert JM, Lin CY, Van Voorhis T, Chien SH, Sodt A, Steele RP, Rassolov VA, Maslen PE, Korambath PP, Adamson RD, Austin B, Baker J, Byrd EFC, Dachsel H, Doerksen RJ, Dreuw A, Dunietz BD, Dutoi AD, Furlani TR, Gwaltney SR, Heyden A, Hirata S, Hsu CP, Kedziora G, Khalliulin RZ, Klunzinger P, Lee AM, Lee MS, Liang W, Lotan I, Nair N, Peters B, Proynov EI, Pieniazek PA, Rhee YM, Ritchie J, Rosta E, Sherrill CD, Simmonett AC, Subotnik JE, Woodcock HL, Zhang W, Bell AT, Chakraborty AK, Chipman DM, Keil FJ, Warshel A, Hehre WJ, Schaefer HF, Kong J, Krylov AI, Gill PMW, Head-Gordon M (2006) Phys Chem Chem Phys 8:3172–3191
23. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983) J Comput Chem 4:187–217
24. Levitt M, Lifson S (1969) J Mol Biol 46:269–279
25. Fletcher R, Reeves C (1964) Comput J 7:149–154
26. Vorobyov I, Anisimov VM, Greene S, Venable RM, Moser A, Pastor RW, MacKerell AD Jr (2007) J Chem Theory Comput 3:1120–1133
27. Guvench O, Greene SN, Kamath G, Brady JW, Venable RM, Pastor RW, MacKerell AD Jr (2008) J Comput Chem (in press)
28. Kirkpatrick S, Gelatt CD, Vecchi MP (1983) Science 220:671–680
29. Blondel A, Karplus M (1996) J Comput Chem 17:1132–1141
30. Perez A, Marchan I, Svozil D, Sponer J, Cheatham TE, Laughton CA, Orozco M (2007) Biophys J 92:3817–3829
31. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C (2006) Proteins: Struct, Funct, Bioinf 65:712–725
32. Lovell SC, Davis IW, Adrendall WB, de Bakker PIW, Word JM, Prisant MG, Richardson JS, Richardson DC (2003) Proteins: Struct, Funct, Bioinf 50:437–450