Symmetrisation of the AMBER and CHARMM force fields

Edyta Małolepsza, Birgit Strodel, Mey Khalili,

Semen Trygubenko, Szilard Fejer and David J. Wales*

University Chemical Laboratories, Lensfield Road, Cambridge CB2 1EW, UK

June 6, 2009

*corresponding author: dw34@cam.ac.uk

Abstract

The AMBER and CHARMM force fields are analysed from the viewpoint of the permutational symmetry for feasible exchanges of chemically equivalent atoms and groups in amino and nucleic acids. In each case we propose schemes for symmetrising the potentials, which greatly facilitate the bookkeeping associated with constructing kinetic transition networks via geometry optimisation.

1 Introduction

The potential energy is usually the first property calculated in any computational study. Depending on the system size and computational resources available, different approaches can be applied. For small systems with up to a few hundred atoms *ab initio* calculations are possible. For larger molecules, such as proteins and nucleic acids, the number of atoms increases significantly and more empirical approaches are necessary. The same situation occurs when studying the interactions of molecules, or for chemical reactions involving more than a few dozen atoms, especially when the system is immersed in an explicit solvent. Hence there is great interest in approximate methods that facilitate examination of large systems.

The potential energy of a system in a molecular mechanics or empirical force field approach is calculated as a function of covalent and non-covalent terms. The covalent component is a sum of contributions from molecular bonds, angles, and dihedral angles, while the non-covalent component usually describes the electrostatic and van der Waals interactions. The exact formula differs between different force fields. For biomolecular applications there are several popular packages offering both sets of force fields for calculating the energy and molecular simulation modules for molecular dynamics (MD) and Monte Carlo (MC) simulations. Here we consider two of these force fields, namely AMBER9[1–3] and CHARMM.[4–7]

Independent of methodology, the force field should fulfil some basic physical requirements. In this paper we focus on the symmetry of the Hamiltonian and the consequences for the force field. The potential should be invariant to overall translation, rotation and to the permutation

of chemically equivalent atoms. The first two requirements are generally fulfilled in grid-free implementations. Grid-based approaches, such as these based on the Ewald summation, can lead to small changes in the energy corresponding to overall translation and rotation,[8] but here we focus on the permutational symmetry. In particular, we show how both the CHARMM and AMBER potentials can be symmetrised so that accessible permutational isomers have the same energy. Although the energy differences involved between isomers are small, and probably of little consequence in MC and MD simulations, they cause problems for the book-keeping required in approaches based on transition networks built from stationary points.[9–12] Typical energy differences between alternative permutational isomers range between 0.001 and 0.02 kcal/mol per residue for tightly converged minima, with corresponding changes in bond and dihedral angles of around a degree or less. However, for CHARMM we observed for some residues larger energy differences of more than 1 kcal/mol between alternative permutational isomers, with changes in bond and dihedral angles of up to 5 degrees.

In constructing databases of local minima and transition states from the potential energy surface we typically converge the energy and geometry very tightly, so that permutational isomers can be identified unambiguously. The unsymmetrical terms in the potential energy are significantly larger than the threshold we use to identify isomers, and the geometry difference between structures is also outside tolerance. These effects are very undesirable in approaches based on geometry optimisation, because they introduce an artificial complexity into the potential energy surface. However, with some minor adjustments to the potential exact permutational symmetry can easily be restored, as described below.

2 Permutational isomers

To introduce the problems posed by permutational isomers we will discuss the amino acids alanine, valine and phenylalanine, which are shown in Fig. 1. Any rotation of the methyl group in alanine by 120° around the local three–fold axis should generate an equivalent structure. Such rotations correspond to permutations of hydrogen atoms, e.g. H1⇒H2, H2⇒H3,

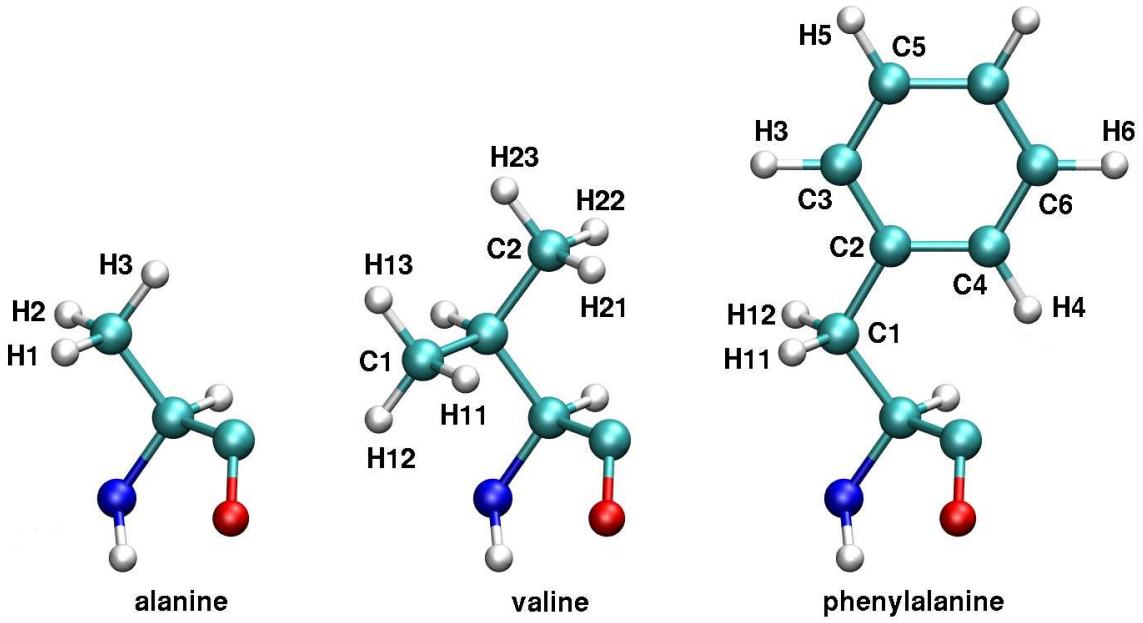


FIGURE 1: Structures of alanine, valine and phenylalanine.

H3⇒H1. For valine one can consider permutations of the hydrogen atoms in the two methyl groups separately along with permutations of both methyl groups, where eight atoms change places at the same time: C1⇔C2, H11⇔H21, H12⇔H22, H13⇔H23. Besides the permutation of hydrogen atoms in the methylene group (H11⇔H12), phenylalanine has another permutational isomer where the phenyl ring rotates by 180° around the C1–C2 bond. In this process eight atoms change places: C3⇔C4, C5⇔C6, H3⇔H4, H5⇔H6.

In general, the Hamiltonian is invariant to any permutation–inversion operation involving the complete nuclear permutation inversion group.[13,14] However, for standard biomolecular force fields most of these operations are prevented through the choice of harmonic bond length constraints. The feasible operations that we need to consider for biomolecular force fields are:

(1) exchange of hydrogen atoms in any methyl or methylene group,

(2) exchange of equivalent methyl groups in valine and leucine,

(3) permutation of hydrogen atoms in the NH_3 group for any N-terminal amino acid,

(4) permutation of hydrogen atoms in the NH_2 group for asparagine, glutamine and arginine,

(5) permutation of NH_2 groups in arginine,

(6) permutation of oxygen atoms in the carboxyl group for aspartate and glutamate,

(7) permutation of oxygen atoms in the carboxyl group for any C-terminal amino acid,

(8) rotation of the phenyl ring by 180° in phenylalanine and tyrosine,

(9) permutation of hydrogen atoms in the NH_2 group of adenine, cytosine, guanine.

Both the CHARMM and AMBER force fields can be systematically refined to support these symmetries, as described in the following sections.

3 The CHARMM force field

3.1 Origin of symmetry–breaking in the potential

For the united-atom force field CHARMM19[5] the possibility of exchanging hydrogen atoms in methyl or methylene groups does not exist. Since the united-atom potential for nucleic acids is not recommended[7] we did not consider case (9) for CHARMM19. For case (8) the standard CHARMM19 potential is symmetrical. However, for the exchange of equivalent methyl groups in valine and leucine the CHARMM19 potential was found to be unsymmetrical, and investigation showed that this effect is caused by the energy contributions from the dihedral angle torsions and the improper torsions. For the dihedral angle torsion around the C_β–C_γ bond in leucine and the C_α–C_β bond in valine only one dihedral angle is used in each case for the calculation of the dihedral angle potential,[4]

$$E_\phi = |k_\phi| - k_\phi \cos(n\phi)\,, \tag{1}$$

where the force constant $k_\phi = 1.6\,\mathrm{kcal\,mol^{-1}}$ and the periodicity $n = 3$. Inspection of the CHARMM19 topology file reveals that for leucine this dihedral angle is CA–CB–CG–CD2 and for valine it is N–CA–CB–CG1, where the CHARMM19 notation for the atom names is illustrated in Fig. 2. If the bond angle between the two methyl groups is exactly 120°,

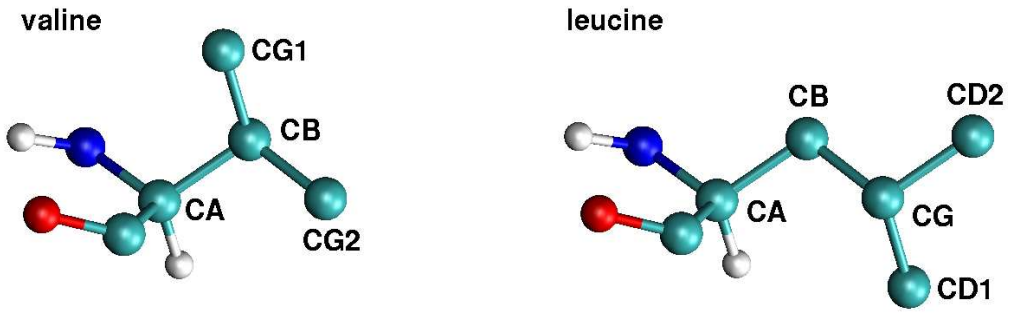


FIGURE 2: Structures of valine and leucine without nonpolar hydrogens (united-atom model) and with the atom names used in the CHARMM19[5] topology file.

the dihedral angle potential would be the same if one used instead the dihedral angles CA–CB–CG–CD1 for leucine and N–CA–CB–CG2 for valine, i.e. $E_\phi = E_{\phi,1} = E_{\phi,2}$, where $E_{\phi,i}$ denotes the potential energy for the dihedral angle terminating at CDi for leucine and CGi for valine, respectively. Thus, in this special case the dihedral angle potential is symmetrical for the exchange of both methyl groups. However, in most cases the bond angle between the two methyl groups will deviate from 120°, so that the threefold symmetry assumed for the dihedral angle potential in question is not fulfilled, and $E_{\phi,1} \neq E_{\phi,2}$. Thus the CHARMM19 dihedral angle potential is not symmetrical with respect to the permutation of the methyl groups in leucine and valine. Our solution to this problem is to use both dihedral angles CA–CB–CG–CD1 and CA–CB–CG–CD2 for leucine and N–CA–CB–CG1 and N–CA–CB–CG2 for valine, respectively, in an averaged torsional potential $E_\phi = \frac{1}{2}(E_{\phi,1} + E_{\phi,2})$. Of course, our objective is to symmetrise the potential without introducing any change that might require a full reparameterisation, or cause a significant computational overhead.

The above approach resolves the symmetry breaking of the dihedral angle potential for leucine and valine, but not the terms originating from the improper torsion potential,[4]

$$E_\omega = k_\omega(\omega - \omega_0)^2\,, \tag{2}$$

where in this case the improper torsion angle ω refers to the angle between the planes CG,CD2,CD1 and CD2,CD1,CB in leucine, and CB,CG2,CG1 and CG2,CG1,CA in valine, re-

spectively (see Fig. 2), where $\omega_0 = 35.26°$. In the CHARMM19 potential this improper torsion term is needed to prevent inversion about the tetrahedral centre without an explicit hydrogen,[4] i.e. about CG in leucine and about CB in valine. After exchanging the two methyl groups, the magnitude of this improper torsion angle stays the same but it changes in sign. The latter condition has caused problems in our geometry optimisation framework, since $\omega_0 \neq 0$, so that E_ω rises sharply after the permutation of the methyl groups in minimum energy geometries. Our solution to this problem is to use $\omega + \omega_0$ in (2) if ω is negative. Since this harmonic improper torsion potential has a force constant of $k_\omega = 55.0\,\mathrm{kcal\,mol^{-1}}$, large deviations from the equilibrium geometry are not allowed, and the criterion $\omega < 0$ is a safe indication that both methyl groups in leucine or valine have been permuted.

For the permutation of the hydrogen atoms of the NH_2 group we found symmetry–breaking for asparagine and glutamine, but not for arginine. The origin of the problem for asparagine and glutamine is the same as outlined above for the permutation of the methyl groups in leucine and valine: for both amino acids only one dihedral angle is used for the calculation of E_ϕ for the torsion around the CG–ND2 bond in asparagine and around the CD–NE2 bond in glutamine, respectively (see Fig. 3). Hence the potential is unsymmetrical if the amide group is not planar. This problem can be solved if one uses both dihedral angles, e.g., CB–CG–ND2–HD21 and CB–CG–ND2–HD22 for asparagine, and calculates $E_\phi = \frac{1}{2}(E_{\phi,1} + E_{\phi,2})$. In the case of arginine this approach was already implemented in the CHARMM19 topology file for the torsions around the CZ–NH1 and CZ–NH2 bonds (see Fig. 3). On the other hand, for the torsion around the NE–CZ bond in arginine only one dihedral angle was used, namely CD–NE–CZ–NH1, causing the CHARMM19 potential to be unsymmetrical for the permutation of the NH_2 groups. Again, the addition of the other dihedral angle, CD–NE–CZ–NH2, and calculation of the averaged potential E_ϕ produces a symmetrical potential.

For the permutation of the oxygen atoms in the carboxyl group in aspartate and glutamate the original CHARMM19 potential is already symmetrical, even though only one dihedral angle is used in the calculation of the dihedral angle potential (1) for the torsion around the

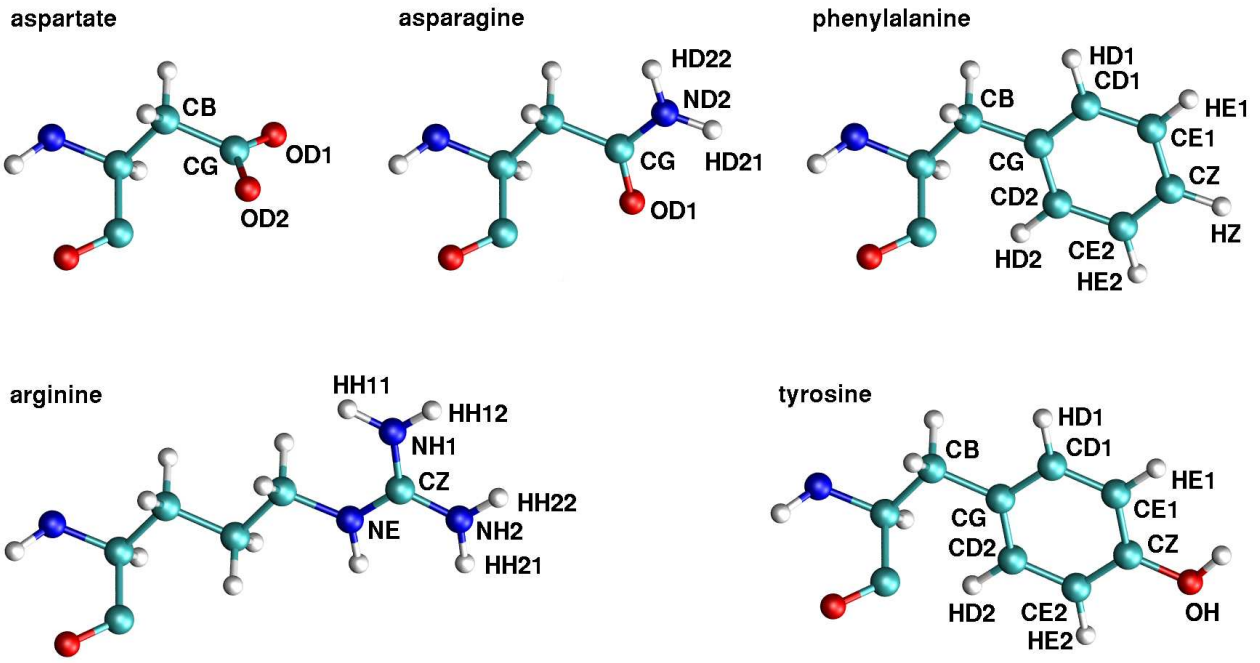


FIGURE 3: Structures of aspartate, asparagine, phenylalanine, arginine and tyrosine with the atom names used in CHARMM and AMBER for the definition of dihedral and improper torsion angles. Since the structures of glutamate and glutamine differ only by one additional methylene group compared to aspartate and asparagine, respectively, they are not shown. The atoms in glutamate and glutamine are named CB and CG for the methylene carbon atoms, CG for the carbonyl atom, and all other atom names are the same as in aspartate and asparagine.

CB–CG bond in aspartate and and the CG–CD bond in glutamate (see Fig. 3). Inspection of the parameter file revealed that the force constant assigned to these dihedral angle torsions is zero, which explains the symmetry. The corresponding force constant in the CHARMM22 potential[6] is 0.05 kcal mol^{-1}. In CHARMM19 there is also a zero force constant for the torsion around the bond between the C_α and the carbonyl carbon atoms in any C-terminal amino acid. Hence, the CHARMM19 dihedral angle potential is also symmetrical for the exchange of the two C-terminal oxygen atoms, which are denoted OT1 and OT2 in the CTER patch residue in the CHARMM19 topology file.[4,5] However, the improper torsion potential, which is needed to maintain planarity about the carboxyl carbon, is unsymmetrical for the permutation of OT1 and OT2. This situation could be changed by reordering the atoms defining the improper

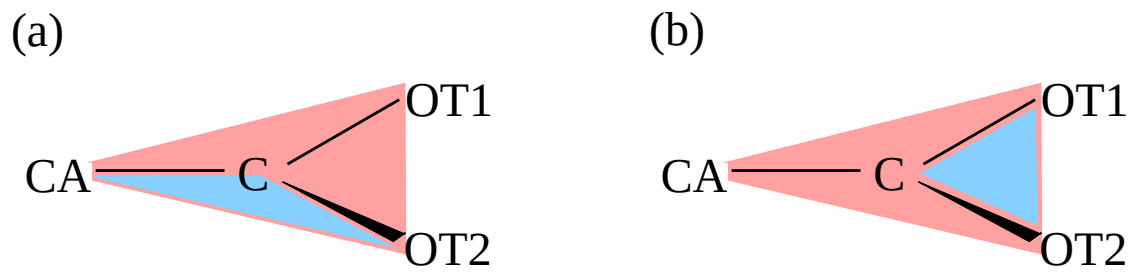


FIGURE 4: Definition of the improper torsion angle ω for the C-terminal carboxyl group in CHARMM19. In (a) ω is defined as the torsion angle $\phi_{\mathrm{C,CA,OT2,OT1}}$, or equivalently, the angle between the planes C,CA,OT2 and CA,OT1,OT2. In (b) ω is defined as the torsion angle $\phi_{\mathrm{C,OT1,OT2,CA}}$, or equivalently. the angle between the planes C,OT1,OT2 and OT1,OT2,CA. If $\omega \neq 0$, as illustrated here, in (a) ω will change in magnitude after the permutation of OT1 and OT2 whereas in (b) ω will only change its sign.

torsion angle in question. The original order was C–CA–OT2–OT1, which refers to the angle between the planes C,CA,OT2 and CA,OT2,OT1. The CHARMM convention in the definition of improper torsion angles is to list the central atom in the first position, while no rule exists for how to order the other three atoms. Thus six possibilities exist for the definition of an improper torsion angle. In the definition C–CA–OT2–OT1, the two oxygen atoms are treated differently, since only one of them, in this case OT2, is used for the determination of the first plane. If OT1 and OT2 are exchanged, the plane C,CA,OT2 will change as well, leading to a change in ω, unless $\omega = 0$, as illustrated in Fig. 4(a). In general ω will deviate from zero. If one changes the atom order to C–OT1–OT2–CA or C–OT2–OT1–CA, the oxygen atoms are treated equivalently because both are used in the definition of the two planes, as shown in Fig. 4(b). Permuting OT1 and OT2 now causes ω to change in sign but not magnitude, which does not affect the improper torsion potential (2) because $\omega_0 = 0$. Hence, from the six possible definitions for an improper torsion angle involving two permutable atoms, only the two definitions that use these two atoms for the determination of both planes in the calculation of ω lead to a symmetrical improper torsion potential for the permutational isomers.

The all-atom force fields CHARMM22 for proteins[6] and CHARMM27 for nucleic acids[7,15]

TABLE 1: Comparison of original and proposed reordering of atoms for improper torsion angles in the all-atom CHARMM22 and CHARMM27 force fields. The names of the atoms are indicated in Figs. 3 and 6.

residue	original order	proposed reordering
aspartate	CG–CB–OD2–OD1	CG–OD1–OD2–CB
glutamate	CD–CG–OE2–OE1	CD–OE1–OE2–CG
terminal COO^-	C–CA–OT2–OT1	C–OT1–OT2–CA
adenine	N6–C6–H61–H62	N6–H61–H62–C6

were found to be unsymmetrical only in cases where improper torsion angles are involved in the exchange of chemically equivalent atoms. For the CHARMM force fields released from CHARMM22 onward the bond and dihedral angles are in general not listed explicitly in the topology files. Instead they are automatically generated from the bond list via the 'AUTO-generate ANGL DIHE' command. Here, the complete set of dihedral angles is generated; for example, three dihedral angles are created if the fourth atom defining the dihedral angle is a hydrogen atom of a methyl group. This approach ensures that the CHARMM22 dihedral angle potential is invariant to any permutation–inversion operation. Our procedure to symmetrise the CHARMM19 dihedral angle potential is therefore analogous to the parameterisation procedure used for the CHARMM22 force field.[6]

The remaining symmetry breaking arises from improper torsion angles that are involved in the exchange of chemically equivalent atoms, namely for the permutation of hydrogen atoms in the NH_2 groups of asparagine and glutamine, the permutation of oxygen atoms in the carboxyl group of aspartate, glutamate and each terminal CO_2^-, and the permutation of hydrogen atoms in the NH_2 group of adenine, cytosine and guanine. Symmetrisation is easily achieved for permutation of oxygen atoms in the carboxyl groups and permutation of hydrogen atoms in the amino group of adenine by reordering the atoms that define the improper torsion angle

in question, as explained above and shown in Table 1. For the permutation of the hydrogen atoms in the NH_2 group of asparagine, glutamine, cytosine and guanine the two atoms in question are parameterised differently in the CHARMM22 potential to reflect the different chemical environments. For instance, in asparagine and glutamine one of the hydrogen atoms is *cis* and the other *trans* to the carbonyl oxygen atom (see Fig. 3), giving rise to a higher partial charge for the *cis* hydrogen atom. Hence these two hydrogen atoms should not be exchanged and one should exclude this possibility when sampling the conformational space. The alternative local minima that would otherwise result are not physically meaningful.

3.2 Symmetrisation of CHARMM

For the CHARMM19 force field we defined additional dihedral angles as described above for leucine, valine, asparagine, glutamine and arginine, and added them to the `toph19.inp` topology file. These modifications are associated with a change in the calculation of the dihedral angle potential to $E_\phi = \frac{1}{2}(E_{\phi,1} + E_{\phi,2})$. Since our aim was to avoid changing the CHARMM source code, we halved the force constants k_ϕ for the dihedral torsions in question. To this end we added the following parameters for the dihedral angle potential in the `param19.inp` CHARMM19 parameter file: $k_\phi = 0.8\,\mathrm{kcal\,mol^{-1}}$, $n = 3$ and $\phi_0 = 0$ for CH1E–CH2E–CH1E–CH3E in leucine and for NH1–CH1E–CH1E–CH3E in valine, which was derived from $k_\phi = 1.6\,\mathrm{kcal\,mol^{-1}}$, $n = 3$ and $\phi_0 = 0$ for X–CH2E–CH1E–X and X–CH1E–CH1E–X, respectively. For asparagine and glutamine we added $k_\phi = 4.1\,\mathrm{kcal\,mol^{-1}}$, $n = 2$ and $\phi_0 = 180°$ for the CH2E–C–NH2–H dihedral angles, which was derived from X–C–NH2–X with $k_\phi = 8.2\,\mathrm{kcal\,mol^{-1}}$, $n = 2$ and $\phi_0 = 180°$. The X–C–NH1–X torsion has the same parameters, so that CH2E–NH1–C–NC2 with $k_\phi = 4.1\,\mathrm{kcal\,mol^{-1}}$, $n = 2$ and $\phi_0 = 180°$ was added for the symmetrisation of the CHARMM19 potential for the exchange of the NH_2 groups in arginine. The improper torsion angle in CTER was reordered to C–OT1–OT2–CA in the `toph19.inp` CHARMM19 topology file. For the CHARMM22 force field we reordered the improper torsion angles as listed in Table 1 for aspartate, glutamate and CTER in the `top_all22_prot.inp`

topology file. In the `top_all27_na.rtf` CHARMM27 topology file the improper torsion angle for adenine was modified according to Table 1.

To implement the symmetry of the improper torsion potential for leucine and valine in CHARMM19 it was necessary to change the CHARMM source code to define $E_\omega = k_\omega(\omega + \omega_0)^2$ if $\omega < 0$ and $E_\omega = k_\omega(\omega - \omega_0)^2$ if $\omega \geq 0$. These changes were implemented in the `eintern.src` file for the standard energy routines and in `enefscal.src` for the fast energy and force calculations. These files can be downloaded together with the modified CHARMM19 topology and parameter files, `toph19_perm.inp` and `param19_perm.inp`, the all-atom topology files `top_all22_prot_perm.inp` for proteins and `top_all27_na_perm.rtf` for nucleic acids from URL `http://www-wales.ch.cam.ac.uk/software.html`.

4 The AMBER force field

We have considered the AMBER force fields ff99,[16] ff02[17] and the ff03[18,19] for all-atom representations and ff03ua for a united-atom representation. To check if these potentials are symmetrical with respect to feasible exchanges of identical atoms or groups we generated all possible permutational isomers for all amino and nucleic acids available in the AMBER libraries (including N- and C-terminal residues) and compared their energies. The AMBER force fields involving all-atom representations are symmetrical for the exchange of hydrogen atoms in all methyl and methylene groups, the exchange of equivalent methyl groups in valine and leucine, as well as permutation of hydrogen atoms in the NH_3 group for any N-terminal amino acid. However, for cases (4) to (9) listed in the Introduction, the energies of the permutational isomers are different for each of these three force fields.

For the united-atom force field ff03ua cases (1) and (9) do not apply. The force field is symmetrical only for permutations mentioned in case (2) in the Introduction, which means that, compared to the all-atom force fields, there is additional broken symmetry for the permutation of hydrogen atoms in the amino group of N-terminal amino acids.

4.1 Origin of symmetry breaking in the potential

To investigate the symmetry breaking we examined the potential energy formula used in the AMBER force field:[20]

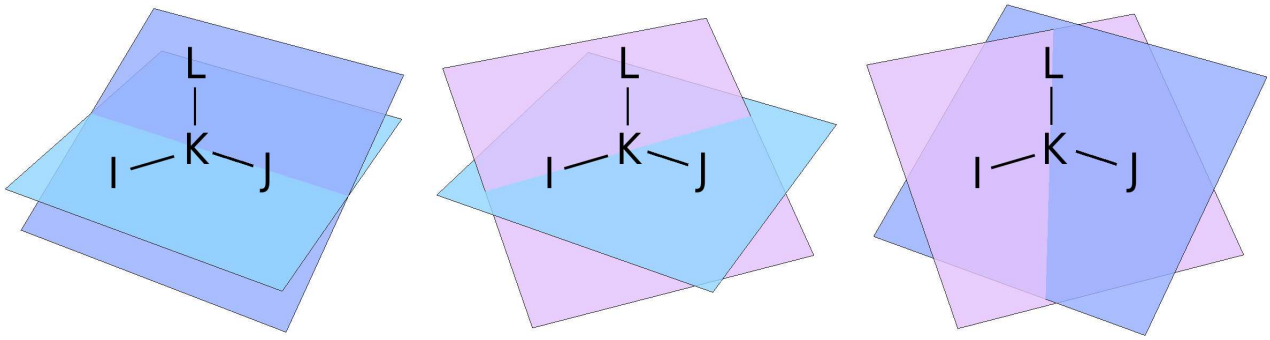
$$E_{\text{total}} = E_{\text{bonds}} + E_{\text{angles}} + E_{\text{dihedrals}} + E_{\text{electrostatic}} + E_{\text{van der Waals}}, \tag{3}$$

where the bond and angle terms are harmonic functions centred on equilibrium values, dihedral contributions are combinations of trigonometric functions fitted to *ab initio* calculations of the energy barrier for rotations, $E_{\text{electrostatic}}$ is a sum of electrostatic interactions for atomic charges, and the last term represents non-bonded interactions described by van der Waals potentials. The AMBER force field can also contain polarisation terms,[21] but these were not considered in the present work.

Closer examination of the permutational isomers showed that in all-atom force fields the improper torsions are responsible for lack of symmetry in the force field. Improper torsions are a subset of the dihedral angles where, instead of considering a chain of four atoms, one atom is connected to three others. The dihedral or improper torsion is calculated as the angle between the planes defined by atoms IJK and JKL.

The same energy formula is used for the dihedral and improper angle components. To calculate this value one needs to provide only the names of four atoms forming a particular angle (I, J, K, L on Fig. 5). In the case of the dihedral angle there are two possible orders for the atoms: IJKL or LKJI; in both cases we obtain the same value, because the same triple of atoms defines the two planes in question, namely IJK (or its permuted version KJI) and JKL (or LKJ). In the case of an improper torsion a different approach is applied: "the convention for an improper torsion named I–J–K–L is that the out-of-plane centre is listed in the third position and the order of the other three is determined alphabetically by atom type, and by atom number (i.e. their order in the molecule) when atom types are identical".[22] Basically, atom K needs to occupy the third position, while atoms I, J and L may be ordered in six different ways, which leads to three different definitions of an improper angle (there are three

FIGURE 5: Three possible definitions of improper torsion I–J–K–L as angle between planes IJK–JKL (left), IJK–IKL (middle), or JKL–IKL (right).



sets of possible planes: IJK–JKL, IJK–IKL, JKL–IKL, as shown in Fig. 5). If the atom types of two atoms are the same (as in all amino and nucleic acids containing the group NH_2 or in amino acids containing the group COO^-), it is sufficient to permute these atoms (this means changing their indices) to obtain different values for the improper torsion and hence for the potential energy. Also, one of the atoms I, J, L is connected to another atom of the side chain, while two others are connected only to atom K, which must also be taken into account.

For each amino and nucleic acid all possible atomic orderings for every improper torsion were tested. In the case of aspartate, glutamate, asparagine, glutamine, terminal COO^-, adenine, cytosine and guanine there are six combinations for each residue, while for arginine the number of combinations increases to 6^3 (arginine needs four impropers to define the geometry, but only three of them are crucial for the permutational isomers). The most time consuming amino acids were phenylalanine and tyrosine, where it was necessary to examine six impropers, and there were 6^6 combinations to test. In this case we have identified more than one symmetrical potential, but because they are equivalent we can choose any of them.

In the case of aspartate (see Fig. 3) the original order of the atoms in the improper torsion, set to enforce planarity of the carboxyl group, is CB–OD1–CG–OD2. However, to obtain a symmetrical potential this order has to be changed to OD1–CB–CG–OD2. A similar change (swapping the first and second atoms) should also be applied to glutamate and all

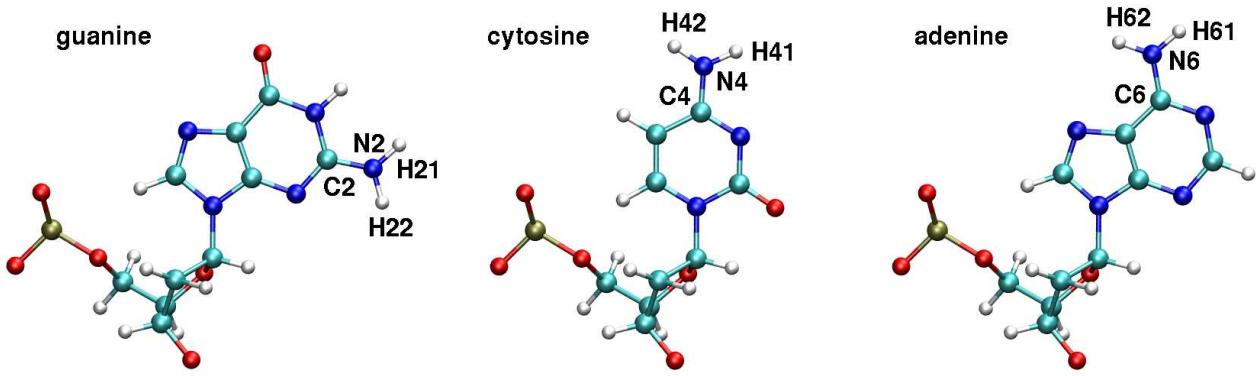


FIGURE 6: Structures of guanine, cytosine and adenine with the names of the atoms involved in the improper torsions.

C-terminal carboxyl groups (Table 2). Amino acids like arginine, phenylalanine and tyrosine need more than one change. Table 2 contains a complete set of definitions for the improper torsions, which guarantees that each of the all-atom AMBER force fields considered is invariant to a feasible permutation of any chemically equivalent atoms or groups.

The united-atom force field in AMBER exhibits the same symmetry problems as the all-atom force fields (cases (4) to (8) listed in the Introduction) along with one further issue: case (3), which is related to permutation of hydrogen atoms in the NH_3 group for N-terminal amino acids. Cases (4) to (8) can easily be corrected by the arrangements of improper torsions described above for the all-atom force fields. All improper torsions are defined in the same way in both types of force field and can be reordered using the same rules except for one improper in tyrosine (see subscripts in Tab. 2).

For the N-terminal amino group the problem is caused by the type definition for the nitrogen atom. In N-terminal NH_3 groups the nitrogen atom is usually described as sp^3 hybridised, while for planar NH_2 groups, as found in neutral arginine or lysine, the nitrogen corresponds to sp^2 hybridisation. In AMBER there are several different types defined for nitrogen atoms, depending on the environment. For planar amino groups the type is called N ("sp2 nitrogen in amide groups", quotation taken from the file `parm99.dat`), and for tetrahedral amino groups type N3 is used ("sp3 N for charged amino groups"). In the original force field the type of

the nitrogen atom in NH_3 is N instead of N3. This choice leads to usage of four impropers to describe the geometry of the amino group and no dihedral, instead of only dihedrals as in the case of N3. The energy formula is the same for all impropers, $E = 1 + \cos(2\phi - \pi)$, and has three minima at 0, π and 2π, while the expected values of these impropers are about $2/3\pi$ and $4/3\pi$ for the nonplanar NH_3 group.

Rearrangement of the atoms defining impropers did not provide a symmetrical potential for NH_3. However, adding two additional impropers to the four existing ones produces permutational isomers with the same energy. Table 3 presents the original and proposed impropers. Since two additional impropers are the same as the fourth improper, their magnitude needs to be scaled by 1/3.

The other way of symmetrising the potential for the NH_3 group is simply to change the type of the nitrogen atom into N3 by replacing N with N3 in the file `uni_aminont03.lib` for every amino acid. This change also requires adding angular force constants for the N3-CT-H1 angle (the angle between N3, C_α and the hydrogen attached to C_α), which are absent in the original data files. One can assume that the values should be the same as for the angle N-CT-H1, especially that all angles X–CT–H1, where X is a heavy atom, have the same force constants.

4.2 Symmetrisation of AMBER

There are two possibilities for symmetrising the all-atom potential. One is to change the source code of the LEaP program (a module from the AMBER package that generates topology files) to force the desired order of the atoms. The other is a change of definitions for improper torsions in the existing topology file. The topology file is necessary in any work with AMBER, and contains all the required information about the structure of the given system, i.e. the force field parameters for each atom, bond, angle, dihedral angle, etc., as well as information about which atoms form bonds, angles, dihedral angles, etc. Changing the source code is undesirable because many subroutines start from ordering atoms according to the rule mentioned above

(to speed up the LEaP) and therefore we concentrated on the topology file. We prepared a python script `perm-prmtop.py` that reads the topology file, finds unsymmetrised improper torsions, changes the order of appropriate atoms, and finally writes the new topology file, which is ready to work with AMBER. The script is available for download from our group web page `http://www-wales.ch.cam.ac.uk/software.html` or on request.

In the case of the united-atom force field one needs to redefine some improper torsions, as for the all-atom force fields, and additionally symmetrise the N-terminal amino groups for each amino acid. Since one improper torsion out of all is defined in different way than in the all-atom potential, separate files `perm-prmtop.py` are necessary for each kind of the force field. The latter operation can be achieved either by adding two additional improper torsions to the four in the original force field and rescaling three of them by a factor 1/3, or by changing the type of the nitrogen atom in the amino group from N to N3 in the file `uni_aminont03.lib`, and adding one additional angular force constant.

5 Conclusions

In the present work we have examined different AMBER and CHARMM force fields from the viewpoint of permutational symmetry. In each case the energies of permutational isomers were slightly different and symmetrisation of the potentials was desirable for our analysis of the potential energy surface.

In the case of the all-atom CHARMM and AMBER force fields, reordering of the atoms that define improper torsions is sufficient to obtain symmetrical potentials. For CHARMM, the changes can be implemented by reordering the atoms defining the affected improper torsion angles in the topology files. For AMBER, these changes can be implemented using a python script that reads a standard topology file and creates a new file containing the symmetric potential. The AMBER source code remains unchanged.

The united-atom CHARMM19 force field can be symmetrised with two sets of changes: additional dihedral angle terms in the energy formula, together with changes of force constants

(changes in the parameter and topology files), and a revised formula for the improper torsion potential, which needs to be implemented in the source code of the CHARMM. For the united-atom force field ff03ua in AMBER the same changes in atom orders defining improper torsions need to be applied as for the all-atom force fields, as well as either adding two impropers to the energy formula or changing the type of the nitrogen atom in N-terminal amino acids.

The overall properties of the force fields are conserved by these changes and no reparametrization is needed. All the files and scripts required for symmetrisation of these potentials are available for download from URL `http://www-wales.ch.cam.ac.uk/software.html`.

Acknowledgements

We gratefully acknowledge funding from BBSRC, and helpful comments on the original manuscript from Prof. Martin Karplus, Prof. David Case and Prof. C. L. Brooks III.

TABLE 2: Comparison of original and proposed reordering of atoms for improper torsions in the all-atom and united-atom AMBER force fields. The names of the atoms are indicated in Fig. 3 and Fig. 6. Subscripts $^{all-atom}$ and $^{united-atom}$ indicate the all-atom and united-atom force fields, respectively; otherwise it applies to both types of force fields.

name of residue	original order	proposed reordering
aspartate	CB–OD1–CG–OD2	OD1–CB–CG–OD2
glutamate	CG–OE1–CD–OE2	OE1–CG–CD–OE2
asparagine	CG–HD21–ND2–HD22	HD21–CG–ND2–HD22
glutamine	CD–HE21–NE2–HE22	HE21–CD–NE2–HE22
terminal COO^-	CA–O–C–OXT	O–CA–C–OXT
arginine	CZ–HH21–NH2–HH22	HH21–CZ–NH2–HH22
	CZ–HH11–NH1–HH12	HH11-CZ–NH1–HH12
	NE–NH1–CZ–NH2	NH1–NE–CZ–NH2
phenylalanine	CZ–CD2–CE2–HE2	CZ–HE2–CE2–CD2
	CE1–CE2–CZ–HZ	CE2–HZ–CZ–CE1
	CD1–CZ–CE1–HE1	CD1–HE1–CE1–CZ
	CG–CE1–CD1–HD1	HD1–CE1–CD1–CG
	CD1–CD2–CG–CB	CD2–CB–CG–CD1
tyrosine	CZ–CD2–CE2–HE2	CZ–HE2–CE2–CD2
	CD1–CZ–CE1–HE1	CD1–HE1–CE1–CZ
	CG–CE1–CD1–HD1	HD1–CE1–CD1–CG
	CE1–CE2–CZ–OH	CE2–OH–CZ–CE1
	CD1–CD2–CG–$CB^{all-atom}$	CD2–CB–CG–CD1
	CB–CD1–CG–$CD2^{united-atom}$	CD2–CB–CG–CD1
adenine (DA,RA)*	C6–H61–N6–$H62^{all-atom}$	H61–C6–N6–H62
cytosine (DC,RC)*	C4–H41–N4–$H42^{all-atom}$	H41–C4–N4–H42
guanine (DG,RG)*	C2–H21–N2–$H22^{all-atom}$	H21–C2–N2–H22

* D (deoxyribose) and R (ribose) units have the same order of atoms in improper torsions

TABLE 3: Comparison of original and proposed improper torsions in the united-atom AMBER force field for the N-terminal amino group NH_3.

original impropers	proposed impropers
H2–N–CA–H1	H2–N–CA–H1
H3–N–CA–H1	H3–N–CA–H1
H3–N–CA–H2	H3–N–CA–H2
H3–N–H2–H1	H3–N–H2–H1
	H2–N–H3–H1
	H1–N–H2–H3

References

[1] D. Case, T. Darden, T. C. III, C. Simmerling, J. Wang, R. Duke, R. Luo, K. Merz, D. Pearlman, M. Crowley, R. Walker, W. Zhang, B. Wang, S. Hayik, A. Roitberg, G. Seabra, K. Wong, F. Paesani, X. Wu, S. Brozell, V. Tsui, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, P. Beroza, D. Mathews, C. Schafmeister, W. Ross and P. K. (2006), AMBER 9, University of California (2006).

[2] D. Pearlman, D. Case, J. Caldwell, W. Ross, T. C. III, S. DeBolt, D. Ferguson, G. Seibel, and P. Kollman., Comp. Phys. Commun. **91**, 1 (1995).

[3] D. Case, T. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Jr., A. Onufriev, C. Simmerling, B. Wang and R. Woods, J. Computat. Chem. **26**, 1668 (2005).

[4] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. S. and M. Karplus, J. Comp. Chem. **4**, 187 (1983).

[5] E. Neria, S. Fischer and M. Karplus, J. Chem. Phys. **105**, 1902 (1996).

[6] A. D. MacKerell, Jr., D. Bashford, M. Bellott, J. R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin and M. Karplus, J. Phys. Chem. B **102**, 3586 (1998).

[7] N. Foloppe and A. D. MacKerell, Jr., J. Comput. Chem. **21**, 86 (2000).

[8] B. Strodel and D. J. Wales, J. Chem. Theor. Comput. **4**, 657672 (2008).

[9] F. Noé and S. Fischer, Curr. Op. Struct. Biol. **18**, 154 (2008).

[10] D. J. Wales, Mol. Phys. **100**, 3285 (2002).

[11] D. J. Wales, *Energy Landscapes*, Cambridge University Press, Cambridge (2003).

[12] D. J. Wales, Mol. Phys. **102**, 891 (2004).

[13] H. C. Longuet-Higgins, Mol. Phys. **6**, 445 (1963).

[14] P. R. Bunker and P. Jensen, *Molecular Symmetry and Spectroscopy*, NRC Press (Canada) (1998).

[15] A. D. MacKerell, Jr. and N. Banavali, J. Comput. Chem. **21**, 105 (2000).

[16] J. Wang, P. Cieplak and P. Kollman, J. Computat. Chem. **21**, 1049 (2000).

[17] P. Cieplak, J. Caldwell and P. Kollman, J. Computat. Chem. **22**, 1048 (2001).

[18] Y. Duan, C. Wu, S. Chowdhury, M. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo and T. Lee, J. Computat. Chem. **24**, 1999 (2003).

[19] M. Lee and Y. Duan, Proteins **55**, 620 (2004).

[20] S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta and P. Weiner, J. Am. Chem. Soc. **106**, 765 (1984).

[21] P. Cieplak, J. Caldwell and P. Kollman, J. Computat. Chem. **22**, 1048 (2001).

[22] A. Dejoux, P. Cieplak, N. Hannick, G. Moyna and F.-Y. Dupradeau, J. Mol. Model. **7**, 422 (2001).