J Comput Aided Mol Des (2012) 26:1195–1205
DOI 10.1007/s10822-012-9610-x

Close intramolecular sulfur–oxygen contacts: modified force field parameters for improved conformation generation

Dmitry Lupyan · Yuriy A. Abramov · Woody Sherman

Received: 1 May 2012 / Accepted: 20 September 2012 / Published online: 6 October 2012

Abstract The Cambridge Structural Database (CSD) offers an excellent data source to study small molecule conformations and molecular interactions. We have analyzed 130 small molecules from the CSD containing an intramolecular sulfur–oxygen distance less than the sum of their van der Waals (vdW) radii. Close S···O distances are observed in several important medicinal chemistry motifs (e.g. a carbonyl oxygen connected by a carbon or nitrogen linker to a sulfur) and are not treated well with existing parameters in the MMFFs or OPLS_2005 force fields, resulting in suboptimal geometries and energetics. In this work, we develop modified parameters for the OPLS_2005 force field to better treat this specific interaction in order to generate conformations close to those found in the CSD structures. We use a combination of refitting a force field torsional parameter, adding a specific atom pair vdW term, and attenuating the electrostatic interactions to obtain an improvement in the accuracy of geometry minimizations and conformational searches for these molecules. Specifically, in a conformational search 58 % of the cases produced a conformation less than 0.25 Å from the CSD crystal conformation with the modified OPLS force field parameters developed in this work. In contrast, 25 and 37 % produced a conformation less than 0.25 Å with the MMFFs and OPLS_2005 force fields, respectively. As an application of the new parameters, we generated conformations for the tyrosine kinase inhibitor axitinib (trade name Inlyta) that could be correctly repacked into three observed polymorphic structures, which was not possible with conformations generated using MMFFs or OPLS_2005. The improved parameters can be mapped directly onto physical characteristics of the systems that are treated inadequately with the molecular mechanics force fields used in this study and potentially other force fields as well.

Keywords Force field · Conformational analysis · OPLS · Small molecule crystal structure · Computational crystal structure prediction

D. Lupyan · W. Sherman (✉)
Schrodinger Inc., 120 West Forty-Fifth Street, 17th Floor, New York, NY 10036, USA
e-mail: Woody.Sherman@schrodinger.com

Y. A. Abramov (✉)
Pfizer Global Research and Development, Eastern Point Road, Groton, CT 06340, USA
e-mail: Yuriy.A.Abramov@pfizer.com

Introduction

The Cambridge Structural Database (CSD) [1] offers an excellent data source to study small molecule conformations and molecular interactions. The high resolution of CSD structures allows the atomic coordinates to be placed without biases from restraints or force fields, thereby providing accurate molecular geometries that are governed by the true energetics of the system. These geometries can be used for building knowledge-based potentials, filtering computationally generated structures, or fitting force fields. In this work, we focus on a specific molecular interaction (close intramolecular sulfur–oxygen contacts) and use the CSD data to improve the OPLS_2005 force field for treating this specific interaction. We then apply the improved parameters to generate conformations for a drug candidate that is known to exhibit multiple polymorphic forms, which is an important area of application for high-quality molecular geometries. Indeed, polymorphism of the crystalline state of pharmaceutical compounds is a common phenomenon [2] that can impact the performance of a

drug in terms of solubility and bioavailability [3–5], chemical and physical stability [6, 7], and mechanical properties [8]. Undesired polymorphs have resulted in at least two cases of drugs being suspended or pulled from the market [9–11], indicating the need for tools to aid in polymorph prediction and characterization.

While drug discovery applications such as molecular docking or pharmacophore modeling typically consider a conformation with heavy atom RMSD of 1.0–2.0 Å to the bioactive conformation as acceptable, polymorph prediction tends to be much more sensitive to molecular geometries, with good predictions typically requiring accuracies in the range of 0.25–0.50 Å. This stringent requirement arises because most currently available computational crystal structure prediction (CSP) methods rely heavily on accurate input conformations, as the internal degrees of freedom are held rigid during the initial crystal packing calculations. Given the sensitivity to accurate conformations, proper selection of the starting molecular conformations for CSP is of great importance [12–15].

Close intramolecular contacts between a carbonyl oxygen and a sulfur (S···O) were previously demonstrated to display attractive interactions both in proteins [16] and in small molecules [17–21]. The weak S···O attraction is due to quantum mechanical interactions that can present challenges to model with molecular mechanics force fields [22–24], thereby warranting special treatment. For example, most force fields use either an arithmetic mean [25, 26], geometric mean [27–30], or an augmented variant [31–33] for the vdW mixing rules, which ensures that the vdW parameters for an interaction pair will lie somewhere between the parameters for the two individual atoms. The vdW combining rules are adequate in most cases, but when quantum mechanical effects involving dispersion interactions and significant orbital overlaps allow for closer contacts than the combination of the individual parameters would permit, then the traditional mixing rules will be insufficient to correctly describe the pairwise potential. Modifying either or both of the individual vdW parameters may fix the specific interaction of interest but will alter the interactions with other atoms in the system, like water and protein atoms, thereby introducing errors to interactions that were already well parameterized. For example, simply scaling down the vdW radius parameter for sulfurs and/or carbonyl oxygens would change interactions of those atoms with water and thereby alter the solvation free energy for molecules containing those groups. Given that the OPLS force field has been parameterized to reproduce solvation free energies and previous work has shown the accuracy of such predictions across a broad range of functional groups that included carbonyl oxygens and sulfurs [34, 35], it would not be reasonable to simply scale van der Waals parameters without a significant revision of the functional form of the force field.

Adding to the challenge, conformational search methods typically generate small molecule conformations in the gas phase, thus ignoring the effects of crystal packing. Crystal packing forces can have an effect on the geometry of a molecule, perturbing it from the gas phase minimum [36–38]. In cases where crystal packing is significant, even using quantum mechanics at a high level of theory and large basis set it would not be possible to accurately reproduce the crystal conformation using a local or global potential energy minimum without explicitly considering the crystal environment. This suggests that either crystal packing be considered explicitly during the conformation generation or an ensemble of conformations around local minima is needed even when a highly accurate energy model is available.

The primary goal of this work is to provide a practical solution for improving the accuracy of conformations generated for small molecules containing close S···O interactions. As will be shown in this work, there are a number of force field modifications that can be made to improve the reproduction of small molecule crystal conformations for this class of interactions. While the new parameters developed here might not represent a general improvement to the interaction potential between sulfur and oxygen atoms in the OPLS_2005 force field, it helps address a specific problem that is important in computational crystal structure prediction. Namely, for most computational crystal packing programs it is necessary to pre-generate conformations prior to the crystal packing calculations and the close intramolecular S···O contacts from the CSD could not be reproduced with the existing force fields we tried.

Here, we study two heavily used small molecule force fields, MMFFs [32, 33] and OPLS_2005 [35, 39], and make modifications to the latter to improve performance on molecules with close S···O contacts. We focus specifically on molecules with the motif SCxC=O (where "x" is either nitrogen or carbon, and sulfur is divalent), which is a common pattern that we have found in 8 % of the molecules in commercially available chemical databases. In addition, there are three approved small molecule drugs that contain the SCCC=O motif: cefoxitin, cilastatin, and benzylpenicilloyl polylysine (http://www.drugbank.ca). First, we characterize the molecules with close S···O contacts using three different levels of theory: molecular mechanics force fields, a semi-empirical method, and quantum mechanics. We find that gas phase quantum mechanics calculations best reproduce the CSD geometries whereas molecular mechanics and semi-emperical calculations perform much worse. Then, we show that conformations using the OPLS_2005 force field can be improved

with new parameters that better describe the close S···O interactions. Finally, we test the applicability of the modified OPLS_2005 force field on a pharmaceutically relevant molecule, axitinib (trade name Inlyta), developed at Pfizer to treat cancer by targeting the vascular endothelial growth factor (VEGF) [40]. Axitinib is a particularly interesting compound for this study because it can form five anhydrous unsolvated polymorphs [41, 42], with the S···O distance ranging from 2.8 to 3.4 Å. We show that the modified force field parameters allow for the generation of conformations that can be accurately repacked into three out of four single conformation polymorphs, including the most stable form of axitinib (form XLI). This level of accuracy was previously not possible with conformations of axitinib generated with MMFFs or the unmodified OPLS_2005 force field.

Materials and methods

In this work we study the ability of different levels of theory (molecular mechanics force field, semi-empirical, and quantum mechanics) to reproduce the conformations of molecules from the CSD [1] containing close sulfur–oxygen (S···O) interactions. We study two heavily used small molecule force fields, MMFFs [32, 33] and OPLS_2005 [35, 39], and make modifications to the latter to improve the ability to generate structures with low RMSD relative to small molecule crystal structures. The OPLS force field has a standard functional form, with bond distances and angles treated using a harmonic potential and torsions treated with a sinusoidal potential. Non-bonded interactions are treated with a 6–12 Lennard-Jones potential for the van der Waals (vdW) energy and Coulomb's law for the electrostatics with fixed atomic point charges. The functional form is similar to that in other force fields such as CHARMM [43] and Amber [44]. The MMFFs force field [32, 33] was developed by Halgren at Merck and is based on the earlier MM3 force field developed by Allinger [45]. Bonds, angle, torsions, and electrostatics in MMFFs are treated similarly to OPLS. However, the vdW potential takes the form of a 6-9 potential, which has the effect of softening the repulsive interaction at short inter-atomic distances. We also study the semi-empirical method PM3 and quantum mechanics with the DFT level of theory and B3LYP-MM hybrid functional, which has been shown to accurately account for dispersion interactions with minimal computational overhead compared with standard DFT functionals [46].

A total of 309 neutral organic molecules with the SCxC=O motif and R-factor below 0.1 were selected from the CSD database version 5.30 [1]. From this set, 202 molecules displayed close S···O intramolecular nonbonding contacts, defined as the S···O distance being shorter than sum of S and O van der Waals radii (3.32 Å). Further filtering resulted in a subset of 132 molecules that displayed a planar arrangement of S–C and C=O bonds (an absolute value of S–C–C–O pseudo-torsion angle did not exceed 20 degrees), which was the motif of greatest interest to us in this study. Finally, molecules with more than 30 heavy atoms or those containing two or more SCxC=O motifs were filtered out, resulting in a final count of 130 molecules. The Supporting Information contains a table with the 130 molecules and their associated properties, such as S···O distance, SCxC=O angle, molecular weight, and formal charge.

To characterize the S···O interaction, we first performed gas phase minimizations of the crystal structure conformations using a single copy of each of the 130 CSD molecules described above. Force field minimizations were performed with MacroModel [47] using a maximum of 500 steps and the Polak-Ribiere Conjugate Gradient (PRCG) method, which tends to perform well on small molecules and is the default in MacroModel for molecules with less than 500 atoms. Semi-empirical minimizations were performed with the PM3 method [48, 49] available within Schrödinger Suite 2010. Quantum mechanics minimizations were performed using Jaguar (Schrödinger, Inc., Portland, OR) with the B3LYP-MM/6-31G** method and level of theory [46]. The B3LYP-MM method has been parameterized on a large data set of CCSD(T) nonbonding interactions energies and has been shown to be an efficient and effective way to account for dispersion interactions.

A two-step minimization scheme was used for both MM and QM geometry optimizations. The first step allowed for a full minimization of hydrogen atoms while the heavy atoms were restrained with a high harmonic potential (50 kcal/mol/Å^2), which helps correct systematically shortened covalent bonds to hydrogens that have been observed in X-ray diffraction analysis [50]. Next, a full minimization was run without any restraints. The initial step was found to be necessary in order to get reasonable initial hydrogen positions before the full molecule was allowed to relax and resulted in lower RMSD of the atomic coordinates to the initial crystal structure as compared to a full minimization without this step.

After characterizing the molecules with the minimization methods described above, a conformational search was performed on each molecule using the force fields to determine the ability to predict accurate conformations. Conformational searches were performed with MacroModel using the LMCS (Low Mode Conformational Search) method in global search mode and all conformations within 5.0 kcal/mol from lowest energy conformation were retained [51]. The LMCS method follows the low-mode vector, but every 2 Å it applies a few steps of steepest descent minimization to relieve any strain due to distorted bond lengths and bond angles introduced by the

move. LMCS is an efficient search method that has the advantage that ring structures and variable torsion angles do not have to be specified. LMCS works by exploring the low-frequency eigenvectors of the system, which are expected to follow "soft" degrees of freedom, such as torsions.

To fit the torsional potential for the modified OPLS_2005 force field, we scanned the torsion of interest at 30° increments and used quantum mechanics (QM) to compute the reference potential. QM geometry optimizations were run with the B3LYP/6-31G** method and basis set to obtain minimized geometries using the program Jaguar. Single point energies were then computed for each geometry at the localized MP2 (LMP2) level of theory with the cc-pVTZ(-f) basis set. Parameters for the torsion of interest were then generated using internally developed nonlinear curve fitting code to minimize the energetic deviation from the quantum mechanical torsional potential. For the work here, a substructure of axitinib (without the ylethenyl-indazol fragment) was used to parameterize the CA–CA–S–CA torsional potential, where "S" is a sulfur atom and "CA" is an SP2 aromatic carbon atom. The torsion parameter was fitted with a single Fourier coefficient (V2), which does not capture the local minima in each of the two peaks that appear on the QM potential energy surface. However, for this work our focus was on reproducing the location of the minima and the relative barrier heights as opposed to the subtleties around the peaks, since the relevant structures for conformation generation in the context of crystal structure predictions will be relatively low in energy. The torsional parameters added in this work are shown in Table 1.

To allow for closer S···O contact distances, we created a special pair potential for the vdW energy between a carbonyl oxygen and a divalent sulfur. Figure 1 shows the shape of multiple vdW curves with different sigma values, where sigma defines the point that the vdW potential becomes positive. Smaller sigma values that allow closer contacts are shifted to the left. We selected a value of $\sigma = 2.50$ Å for the S···O vdW interaction because it is associated with a minimum interaction energy at an atomic separation of 2.72 Å, consistent with the median value of

Table 1 CA–CA–S–CA torsion fitting parameters for axitinib

Force field	Torsion definition	V2 term	V4 term
OPLS_2005	XX–CA–S–CA	0.5	0.5
$OPLS_{S\cdots O}$	CA–CA–S–CA	1.84	0.0

The OPLS_2005 force field describes this torsion using a wildcard, denoted by "XX". "CA" is an atom type for an SP2 aromatic carbon atom, originally described by Jorgensen and Severance [58]. "S" is for the sulfur. $OPLS_{S\cdots O}$ denotes the OPLS_2005 parameter set containing the new torsion type

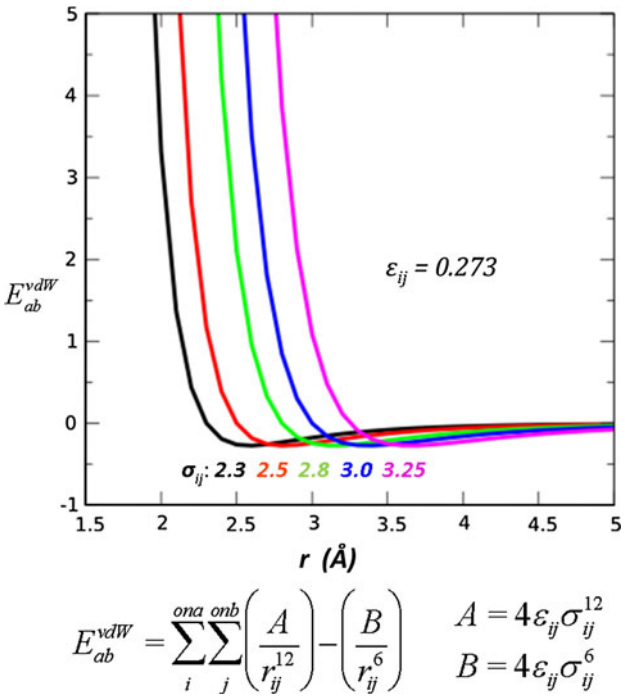

Fig. 1 Van der Waals (vdW) interaction energy profile between divalent sulfur and carbonyl oxygen atoms in kcal/mol. Using a Lennard-Jones 6–12 potential, the vdW energy between atoms i and j in OPLS_2005 can be expressed in terms of the parameters σ_{ij} and ε_{ij}. The total energy (E_{ab}^{vdW}) is the sum over all pairwise nonbonded interactions. Shown here are curves for values of σ ranging from 2.3 to 3.25 Å, with the latter value (*magenta curve*) being the default for OPLS_2005. To reproduce the most probable distance observed in the CSD between the sulfur and oxygen, a σ value of 2.5 Å was chosen for this work. The ε parameter was not altered from the default OPLS_2005 value of 0.273 kcal/mol

2.72 ± 0.11 Å for the S···O distance of the molecules in this study. In order to not deteriorate the interaction energy surface of sulfur or oxygen with other atoms, the sigma value described above was only applied to S···O interactions that involved a carbonyl oxygen and a divalent sulfur. This insures that parameters previously tuned to reproduce solvation free energies, condensed phase properties, and quantum mechanical potentials would be unaffected. The specific atom pair vdW potential was implemented within the MacroModel program [47] using an approach similar to the NBFIX option available in CHARMM [52]. This allows for explicit control over the vdW parameters for specific pairwise interactions without altering the form of the potential for these atoms interacting with other atoms. This is particularly important for the case of S···O interactions presented here, since these atoms are common in medicinally relevant molecules and we do not want to change the force field parameters for their interactions with anything else.

To improve the Coulomb interactions, we attenuated the electrostatic energies using a distance-dependent dielectric

constant. Because sulfur and oxygen both have negative partial atomic charges in the fixed charge force field representations in MMFFs and OPLS_2005, there is a natural repulsion between them. While reversal of the charge sign on one atom would create an attractive force, it would be inconsistent with the quantum mechanical electrostatic potential profile for sulfur and oxygen. Therefore, attenuating the polar interactions using a distance dependent dielectric allowed for the sulfur and oxygen repulsion to be less unfavorable while maintaining their assigned partial atomic charges. Combining each of the above enhancements we produced a modified OPLS force field, termed $OPLS_{S\cdots O}$, which allows for more accurate conformations to be generated for small molecules with close S···O interactions without a priori knowledge of the crystal packing.

Finally, we used the Polymorph Predictor module in Materials Studio 5.5 (Accelrys Software Inc.) to predict crystal forms of axitinib in the corresponding space groups using the MacroModel conformations described above as input. "Fine" quality settings were selected for the Polymorph Predictor sampling and the COMPASS force field was used [53]. In the first step of predictions, a Monte Carlo simulated annealing packing algorithm generates starting structures, treating the molecule as a rigid unit. Next, a geometry minimization of each structure is performed, optimizing unit cell parameters and relaxing molecular geometries, while imposing the space-group symmetry constraints. Finally, clustering is applied to all minimized structures based on interatomic distances to reduce the set to 500 diverse structures that are scored with the COMPASS force field and ranked by total energy. Since the COMPASS force field does not include the new S···O interaction parameters developed in this work, the results rely strongly on the starting molecular conformations. In addition, since the internal energies of the axitinib crystallographic conformations are not well described by the COMPASS force field, the calculated total energies reflect only a preliminary ranking of the polymorphic forms. A postprocessing of the generated crystal structures at a higher level of theory would be required for a reliable ranking of the axitinib polymorphs and will be the focus of future work.

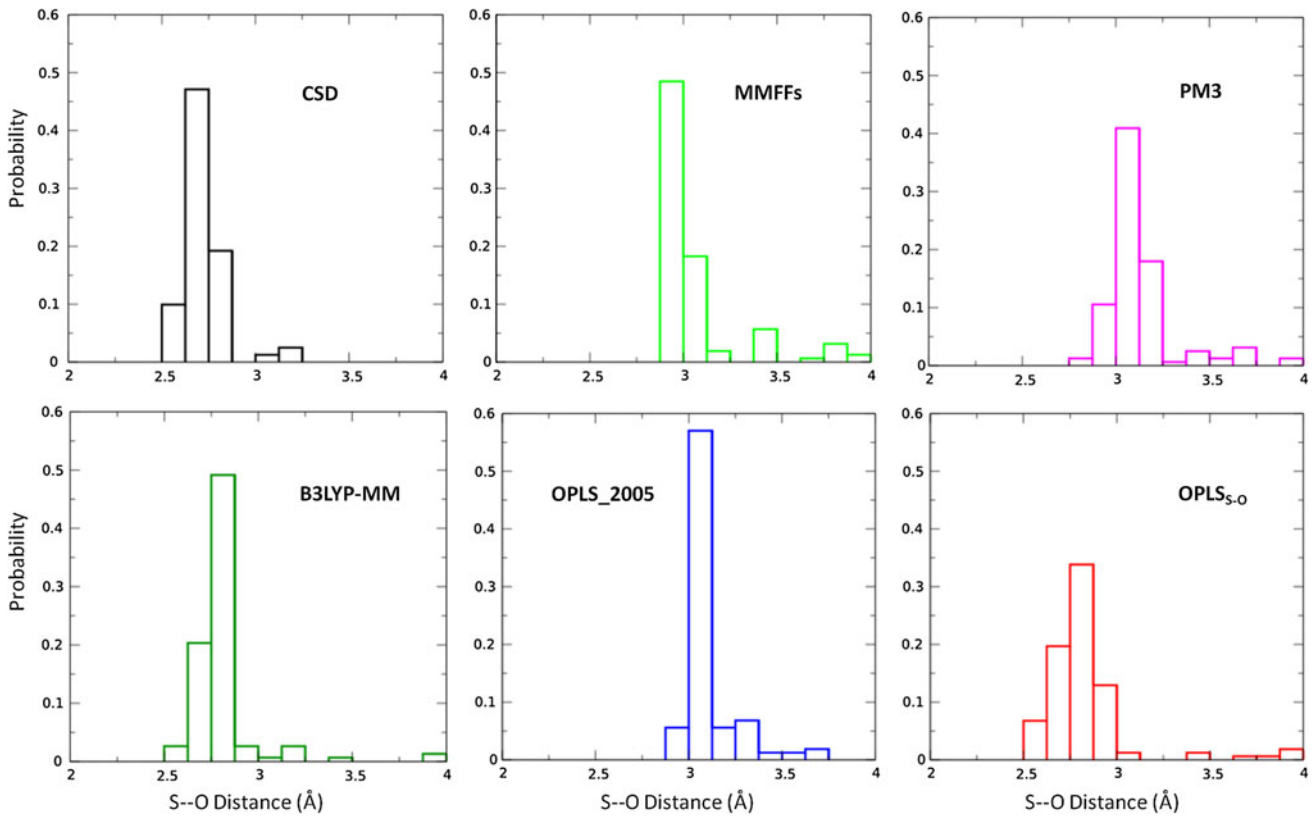

Fig. 2 Distribution of S···O distances in 0.125 Å bins for the 130 molecules in this study. Distributions are shown for the original Cambridge Structural Database (CSD) conformations as well as the geometries obtained through minimizing with various levels of theories, including molecular mechanics (MMFFs, OPLS_2005, and $OPLS_{S\cdots O}$), semi-empirical (PM3), and quantum mechanics (B3LYP-MM). $OPLS_{S\cdots O}$ denotes the OPLS_2005 force field with additional parameters developed in this work. The B3LYP-MM and $OPLS_{S\cdots O}$ distributions are in close agreement and also similar to the dist from the CSD

Results and discussion

The small molecules studied in this work were obtained from the Cambridge Structural Database (CSD) [1] and each contains a close sulfur–oxygen (S···O) contact with the SCxC=O chemical motif where a sulfur and oxygen are connected by 4 bonds and the "x" could be either a carbon or nitrogen atom. The distance between the sulfur and oxygen does not exceed the sum of van der Waals radii (3.32 Å) for any of the molecules. The 130 CSD molecules that satisfied these criteria were analyzed and minimized using the MMFFs and OPLS_2005 force fields as well as the semi-empirical method PM3 and quantum mechanics with the B3LYP-MM/6-31G** method and basis set. Figure 2 shows the distribution of S···O distances in the 130 CSD molecules. The peak in the distribution is at 2.75 Å, which is considerably shorter than the 3.32-Å distance for the minimum energy of the vdW potential in the OPLS force field, thereby resulting in a large energetic penalty due to very steep repulsive vdW potential. Minimizing each of the 130 molecules with the different methods produces the distributions shown in Fig. 2. While there are some molecules with an S···O distance less than 3.0 Å using the force fields, none of the molecules approach the 2.75-Å maximum probability distance observed in the CSD conformations. Interestingly, the semi-empirical method (PM3) performs as poorly as the molecular mechanics force fields. On the other hand, the quantum mechanics method (B3LYP-MM) produces S···O distances much more consistent with the observed geometries from the CSD.

In addition to the S···O distances, we were also interested in the planarity of the angle between the C–S and C=O bond vectors. The distribution from the CSD is shown in Fig. 3, along with the structures from the MMFFs, OPLS_2005, PM3, and B3LYP-MM minimizations. As was the case for the distances, a difference is observed in the distributions between the CSD conformations and the minimized force field conformations. While the distributions do overlap, the CSD conformations are peaked much more strongly at angles close to planar (i.e. 0°). The OPLS_2005 and MMFFs distributions are much broader, with many molecules adopting angles greater than 20°. In fact, 38.6 % of the molecules have an angle greater than 20° for the OPLS_2005 minimized structures and 19.1 % for MMFFs whereas by construction of the dataset none of the CSD structures have an angle greater than 20°. The PM3 semi-empirical method shows 14.2 % of the molecules greater than 20°. Similar to the distance analysis, the quantum

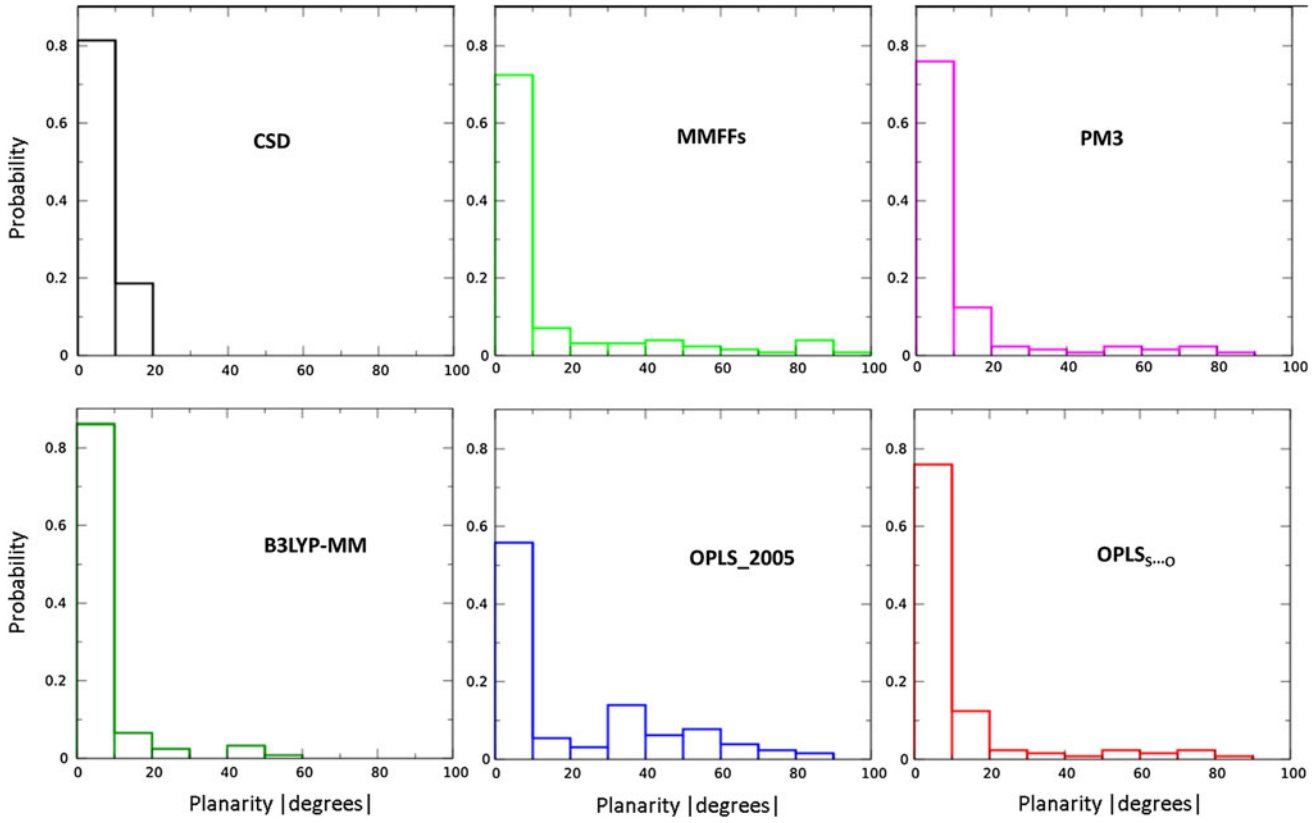

Fig. 3 Distribution of SCxC=O motif angle (deviation from planarity between the S–C bond and C=O bond) for the 130 CSD molecules. The geometry of each molecule was energy-minimized in vacuo using different levels of theory and force field parameters. See Fig. 2 for a description of the different plots

mechanics structures are closer to the crystal, with only 4.4 % of the molecules adopting planarity greater than 20°.

While the geometries from quantum mechanics minimizations are good, the calculations are computationally expensive and would be difficult to run routinely in drug discovery projects where hundreds or thousands of conformations need to be analyzed for each molecule. Therefore, we modified the OPLS_2005 force field by altering the torsional parameters, vdW treatment, and electrostatics (see "Materials and methods") in order to more accurately reproduce the quantum mechanical geometries. The parameter modifications included fitting of the torsional profile to a quantum mechanically-derived energy surface, scaling the vdW potential to match the S···O distance profile observed in the CSD, and attenuating the electrostatic potential using a distance-dependent dielectric. Figure 4a shows the substructure used for the CA–CA–S–CA torsion angle parameterization and the associated potential energy surface. Figure 4b shows the improved agreement between the modified OPLS force field (called $OPLS_{S\cdots O}$) and quantum mechanics relative to OPLS_2005. As seen in Figs. 2 and 3, the distribution with the modified force field is closer to the CSD distribution than the OPLS_2005, MMFFs, or PM3 distributions. The average S···O distance for $OPLS_{S\cdots O}$ is 2.84 ± 0.37 Å and only 11.5 % of the angles are greater than 20°, which is slightly higher than the quantum mechanics results but much better than OPLS_2005 or MMFFs. We also observed an improvement in the measured RMSD of minimized molecules using $OPLS_{S\cdots O}$. Figure 5 shows box plots of the RMSD distribution for the 130 molecules with $OPLS_{S\cdots O}$ and the other methods studied here.

Figure 6 shows an example structure from the CSD (refcode: FOKWEI; compound name: ethyl 3-(benzoylamino)-3-(isopropylthio)acrylate) that illustrates the problem observed with OPLS_2005 and MMFFs for most of the molecules in this study. The experimental CSD structure exhibits an S···O distance of 2.77 Å and planarity of 7.2°. The structures minimized with MMFFs and OPLS_2005 have S···O distances at 3.09 and 3.05 Å with the planarity of the motif at 12.9° and 32.2°, respectively. The heavy-atom RMSD with respect to CSD structure is 0.75 Å for the MMFFs-minimized structure and 0.61 Å for OPLS_2005-minimized structure. The PM3 method performs better than the force fields in this case, with S···O distance of 2.93 Å, planarity of 8.8°, and RMSD of 0.37 Å from the CSD structure. The B3LYP-MM/6-31G** optimized geometry is much closer to the crystal, with S···O distance of 2.75 Å, planarity of 3.2°, and RMSD of 0.11 Å. Finally, the modified $OPLS_{S\cdots O}$ force field parameters developed in this work produces an S···O distance of 2.83 Å, a planarity of 6.1°, and an RMSD of 0.24 Å. These results are better than the force fields (MMFFs and OPLS_2005) and between the PM3 and quantum mechanics calculations.

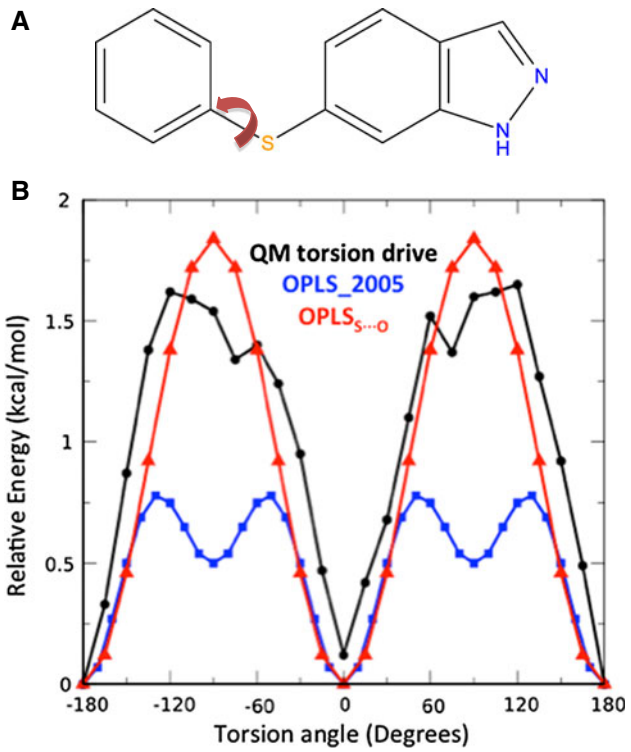

Fig. 4 **a** Substructure of axitinib used to fit the new torsion parameter (see "Materials and methods" for details). **b** Potential energy surface for the CA–CA–S–CA torsion angle in the SCCC = O motif. The QM energy surface is in *black circles*, OPLS_2005 is in *blue squares*, and refitted $OPLS_{S\cdots O}$ is in *red triangles*

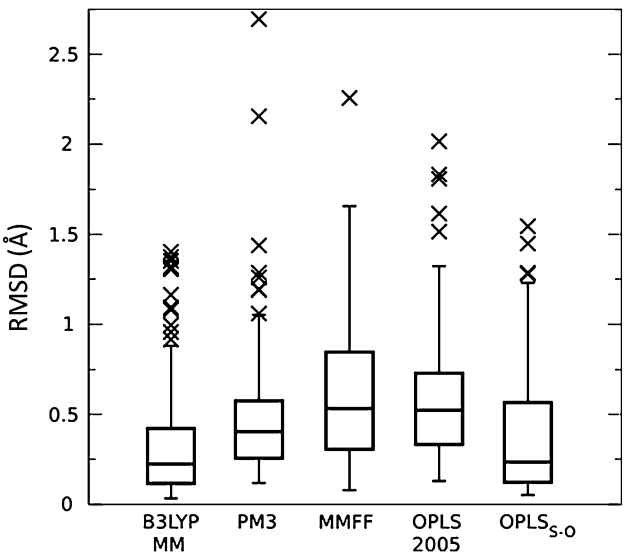

Fig. 5 RMSD *box plots* of the 130 test set structures using different levels of theory for geometry minimization. The median RMSD is denoted as a *horizontal line* through *each box*. The *bottom* and the *top* of the *box* mark the *lower* and *upper* quartile, respectively. The whiskers extend to 1.5*IQR (interquartile range) and the molecules with RMSD larger than that are the outliers, shown with '×' marks

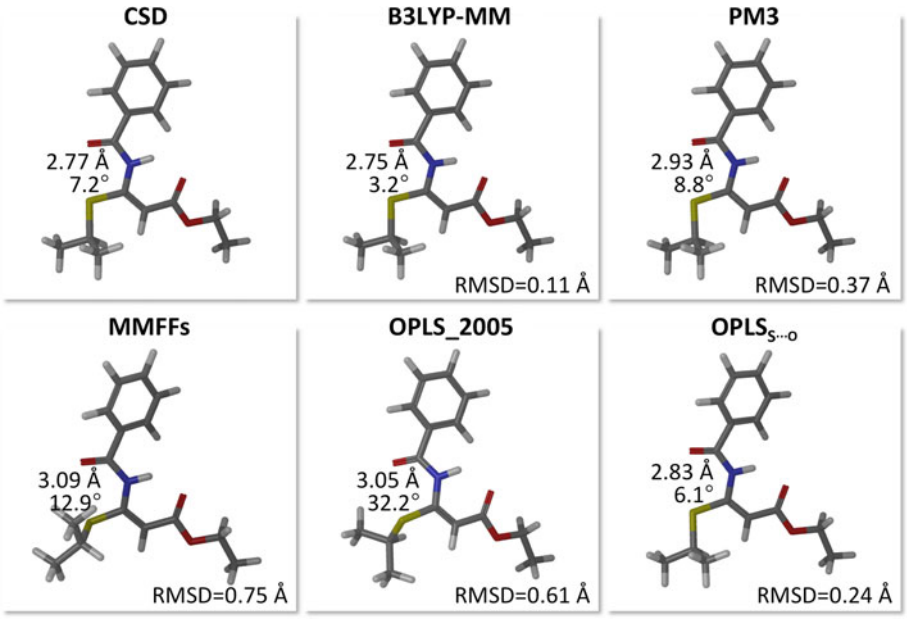

Fig. 6 Minimum energy structures of a CSD molecule (CSD refcode: FOKWEI) with different levels of theory. Numerical values to the *left* of each molecule show the S···O distance and SCNC=O motif planarity. RMSD with respect to the CSD conformation is shown to the *bottom right* of each structure

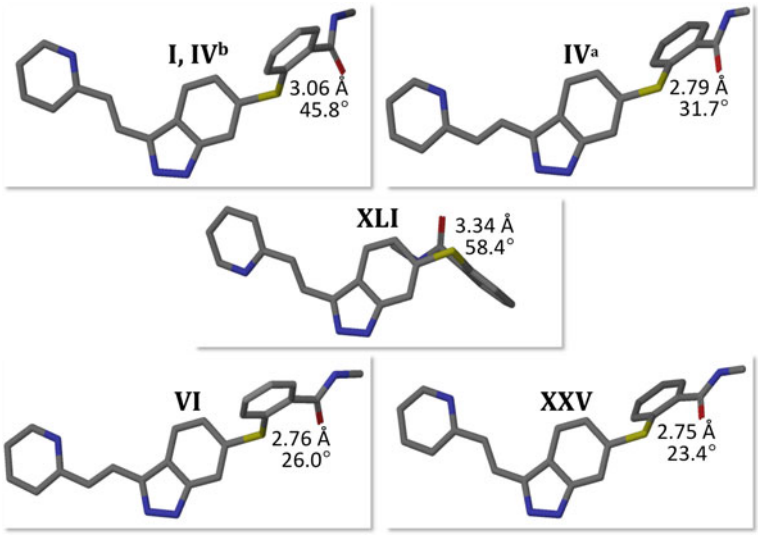

Fig. 7 Five experimentally observed crystallographic conformations of axitinib[33]. The CSD refcodes for forms I, IV, VI, XXV, and XLI are VUSDIX06, VUSDIX05, VUSDIX03, VUSDIX, and VUSDIX04, respectively. S···O distance and SCCC=O motif planarity are shown for each conformation. Form IV has two molecules in the asymmetric unit cell. Form IV conformation "b" (IV^{b}) is very close to the conformation of form I (RMSD < 0.5 Å) and has similar values for the distance and angle reported

To test the applicability of the modified force field to generate accurate conformations for crystal structure prediction on a pharmaceutically relevant molecule, we studied axitinib, developed at Pfizer to treat cancer by targeting the vascular endothelial growth factor (VEGF). Axitinib is a particularly interesting compound from the perspective of studying small molecule crystal structures because it forms five anhydrous polymorphs [41, 42]. Figure 7 shows the conformations of axitinib in the five polymorph structures and Table 2 shows the S···O distance and SCxC=O planarity. In addition, Table 3 shows the RMSD matrix between all 5 structures, highlighting that they are all different conformations. Each of these has a close S···O distance, ranging from 2.8 to 3.4 Å and the planarity ranges from 23° to 58°, with the largest value of distance and angle corresponding to the lowest energy crystal form. This is a good test for the modified force field because the angle is outside the range we observed in the 130-molecule CSD training set and therefore, if the S···O distance parameter was overfit, it would become apparent here.

Table 2 also shows the ability of different force fields to generate low RMSD axitinib conformations compared with the crystal structures. We find that each of the five observed crystal conformations is reproduced using the new OPLS$_{S\cdots O}$ force field. The average heavy-atom RMSD for the closest structure to each of the five crystal conformations is 0.33 Å and the maximum RMSD is 0.54 Å with OPLS$_{S\cdots O}$. This is in contrast to the results obtained using MMFFs and OPLS_2005, where the average RMSD is 0.94 Å and 0.80 Å, and the maximum is 1.38 and 1.18 Å, respectively. We find

Table 2 Tabulated properties for axitinib conformations

	CSD		B3LYP-MM			MMFFs			OPLS_2005			$OPLS_{S\cdots O}$		
Form	Dist	Angle	Dist	Angle	RMSD	Dist	Angle	RMSD	Dist	Angle	RMSD	Dist	Angle	RMSD
XXV	2.75	23.4	2.87	28.4	0.30	3.04	37.4	0.94	3.36	31.7	1.18	2.77	21.9	0.31
VI	2.76	26.0	2.82	31.1	0.29	3.21	41.5	1.38	3.40	38.0	1.11	2.77	27.8	0.21
XLI	3.34	58.4	3.36	59.2	0.24	3.27	66.8	0.73	3.21	45.9	0.56	3.15	57.4	0.35
IV[a]	2.79	31.7	2.85	32.8	0.42	3.12	39.4	0.47	3.27	44.2	0.60	2.86	30.1	0.54
I,IV[b]	3.06	45.8	3.01	39.2	0.26	3.37	42.5	1.21	3.25	53.3	0.54	3.02	46.8	0.25

"Dist" is the S···O distances (in Å) and "Angle" is the planarity of the SCxC=O motif (in degrees). Values are shown for the CSD crystal structure conformation and the minimized geometries for each of the methods. RMSD values (in Å) are shown are with respect to CSD structures

Refer to Fig. 7 caption for superscripts "a" and "b"

Table 3 Pairwise RMSD of five axitinib crystal conformations (in Å)

	XXV	VI	XLI	IV[a]	I, IV[b]
XXV	0				
VI	2.90	0			
XLI	3.06	4.12	0		
IV[a]	2.75	0.86	3.99	0	
I, IV[b]	2.63	1.13	3.71	0.62	0

Refer to Fig. 7 caption for superscripts "a" and "b"

that the $OPLS_{S\cdots O}$ conformational search geometries are almost as accurate as the quantum mechanics geometry minimizations of the CSD crystal structures, which have an average RMSD of 0.21 Å and a maximum of 0.37 Å.

Finally, the ultimate objective of this work is to generate conformations that can be repacked in the correct crystal form. To do this, a limited crystal structure prediction study was performed on axitinib, as follows. First, we took the structures from the conformational search using each force field with the lowest RMSD to the crystal forms I, VI, XXV, and XLI. These conformations were used as input for the Polymorph Predictor in attempt to reproduce the crystal forms in the corresponding space groups. Form IV was not considered in this limited study because it has two molecules in the asymmetric unit cell (termed IV[a] and IV[b]), which makes the computational crystal structure predictions substantially more challenging and time consuming.

The $OPLS_{S\cdots O}$ conformations were correctly packed to reproduce three out of four forms: forms I, VI and XLI. In fact, each of the predicted forms was ranked among the top few solutions from the 500 generated polymorphs as ranked by COMPASS force field. The predicted crystal forms displayed a good overlay with the experimental crystal structures (see Fig. 8), with RMSDs of 0.66, 0.54, and 0.47 Å for the overlay of 10 molecules in the crystal. In contrast, the crystal structure predictions based on conformations generated by the MMFFs and OPLS_2005 force fields failed to reproduce the crystal structures of forms I and XLI within any of the top 500 solutions. Only form VI was reproduced starting from the MMFFs conformation, resulting in an RMSD of 0.54 Å for the overlay of 10 molecules.

Due to the high level of computational time required, a reranking of the generated crystal structures at a higher level of theory (e.g. DFT with the dispersion energy corrections [54, 55]) was not considered in the current limited study. For the same reason, a full crystal structure prediction using the Polymorph Predictor program considering multiple space-groups for each starting conformation was not preformed. A complete study of the crystal structure

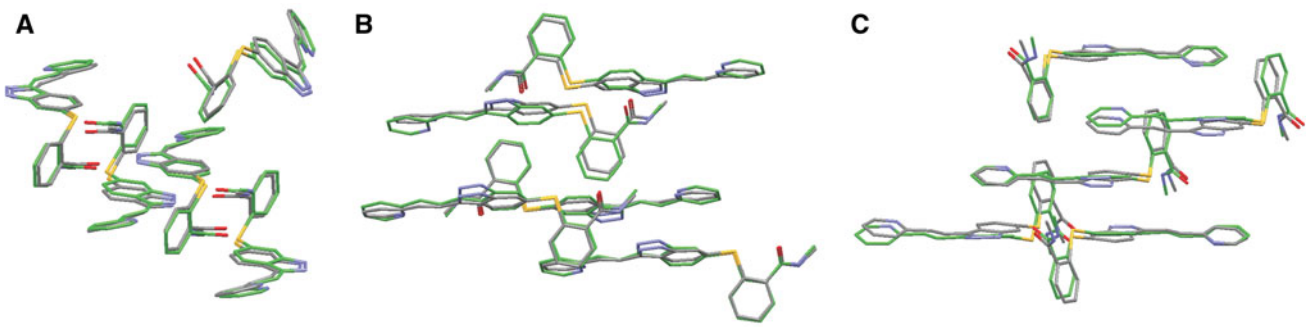

Fig. 8 Structural overlay of CSD (*green*) and predicted (*gray*) crystal packing geometries of axitinib using conformations generated using the $OPLS_{S\cdots O}$ force field parameters developed in this work. Structures of axitinib forms XLI (**a**), VI (**b**), and I (**c**) are shown

prediction of axitinib polymorphs is intended to be part of a future publication.

Conclusion

In this work, we have studied a specific type of interaction (close intramolecular S···O contacts) and how the OPLS_2005 force field can be modified to more accurately reproduce the geometries observed in the Cambridge Structural Database (CSD). We showed that modification of three force field parameters (torsion, vdW, and electrostatics) was sufficient to improve the results of energy minimizations and conformational searching. Furthermore, we showed that accurate conformations of an important pharmaceutical compound, axitinib, which contains a close S···O contact, could be generated with the modified force field parameters. Finally, these conformations could be correctly packed using crystal structure prediction (CSP) software to predict three out of four experimental crystal forms with one molecule in the asymmetric unit cell, which was not possible with the conformations from the OPLS_2005 and MMFFs force fields studied here.

While the results presented in this work are encouraging, they by no means represent a general solution to improving force fields for generating accurate molecular geometries. The new parameters were developed specifically for compounds with close S···O interactions. Other challenging structural motifs will require additional parameterization. However, due to the special atom pair construction of the vdW term described in this work, it is possible to develop new parameters that can address specific deficiencies in the force field without altering the general force field parameters for the atoms involved in the interactions of interest. It is likely that many other challenging interaction motifs exist that will require special parameterization. For example, fluorine interactions can be difficult to model with traditional force fields [56, 57] and could likely be improved by the approach taken in this work. Other interaction motifs, such as aromatic C–H hydrogen bonds and π–π interactions could also benefit from this approach, although one must take caution to not overparameterize the model. It is particularly important for any new terms to have a basis in the underlying physical chemistry in order for them to be robust and transferable.

The ultimate test for the success of a new set of force field parameters is the performance in real-world cases. The generation of conformations sufficiently accurate to reproduce repacking of the three polymorphs of axitinib presented here is only anecdotal, but encouraging. To test the general applicability of this approach it is necessary to expand to a much larger set of molecules with many different structural motifs. We are currently working on this and will report the results in a future publication.

Supporting information

A table with the 130 molecules and their associated properties, such as S···O distance, SCxC=O angle, molecular weight, and formal charge is shown. This material is available free of charge via the Internet at http://pubs.acs.org.

Acknowledgments We thank Wolfgang Damm and John Shelley for implementing the NBFIX functionality within the Schrodinger Suite. We also thank Ed Harder for helpful discussions regarding force fields and for comments on the manuscript.

References

1. Allen F (2002) Acta Crystallogr Sect B 58(3 Part 1):380
2. Bernstein J (2002) Polymorphism in molecular crystals, vol 14. Oxford University Press, USA
3. Abramov YA, Pencheva K (2010) Thermodynamics and relative solubility prediction of polymorphic systems. In: am Ende DJ (ed) Chemical engineering in the pharmaceutical industry: R&D to Manufacturing. Wiley, Hoboken, NJ
4. Kobayashi Y, Ito S, Itai S, Yamamoto K (2000) Int J Pharm 193(2):137
5. Brittain HG (2009) Polymorphism in pharmaceutical solids. Informa Healthcare, New York
6. Singhal D, Curatolo W (2004) Adv Drug Deliv Rev 56(3):335
7. Crowley KJ, Zografi G (2002) J Pharm Sci 91(2):492
8. Beyer T, Day GM, Price SL (2001) J Am Chem Soc 123(21):5086
9. Bauer J, Spanton S, Henry R, Quick J, Dziki W, Porter W, Morris J (2001) Pharm Res 18(6):859
10. Kempf DJ, Marsh KC, Denissen JF, McDonald E, Vasavanonda S, Flentge CA, Green BE, Fino L, Park CH, Kong XP (1995) Proc Nat Acad Sci 92(7):2484
11. Rascol O, Perez-Lloret S (2009) Expert Opin Pharmacother 10(4):677
12. Abramov YA, Zell M, Krzyzaniak JF (2010) Toward a rational solvent selection for conformational polymorph screening. In: am Ende DJ (ed) Chemical engineering in the pharmaceutical industry: R&D to manufacturing. Wiley, Hoboken, NJ
13. Ouvrard C, Price SL (2004) Cryst Growth Des 4(6):1119
14. Cooper TG, Hejczyk KE, Jones W, Day GM (2008) J Chem Theory Comput 4(10):1795
15. Day G, Motherwell W, Jones W (2007) Phys Chem Chem Phys 9(14):1693
16. Iwaoka M, Takemoto S, Okada M, Tomoda S (2002) Bull Chem Soc Jpn 75(7):1611
17. Burling FT, Goldstein BM (1992) J Am Chem Soc 114(7):2313
18. Senger S, Chan C, Convery MA, Hubbard JA, Shah GP, Watson NS, Young RJ (2007) Bioorg Med Chem Lett 17(10):2931
19. Senger S, Convery MA, Chan C, Watson NS (2006) Bioorg Med Chem Lett 16(22):5731
20. Brameld KA, Kuhn B, Reuter DC, Stahl M (2008) J Chem Inf Model 48(1):1
21. Reiter LA, Jones CS, Brissette WH, McCurdy SP, Abramov YA, Bordner J, DiCapua FM, Munchhof MJ, Rescek DM, Samardjiev IJ (2008) Bioorg Med Chem Lett 18(9):3000

22. Kucsman A, Kapovits I (1985) Non-bonded sulfur–oxygen interaction in organic sulfur compounds. In: Bernardi F, Csizmadia IG, Mangini A (eds) Organic sulfur chemistry: theoretical and experimental advances. Elsevier, Amsterdam
23. Nagao Y, Hirata T, Goto S, Sano S, Kakehi A, Iizuka K, Shiro M (1998) J Am Chem Soc 120(13):3104
24. Wu S, Greer A (2000) J Org Chem 65(16):4883
25. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983) J Comput Chem 4:187
26. Brooks BR, Brooks C III, Mackerell A Jr, Nilsson L, Petrella R, Roux B, Won Y, Archontis G, Bartels C, Boresch S (2009) J Comp Chem 30(10):1545
27. Jorgensen WL, Tirado-Rives J (1988) J Am Chem Soc 110:1657
28. Pranata J, Wierschke SG, Jorgensen WL (1991) J Am Chem Soc 113:2810
29. Weiner SJ, Kollman PA, Case DA, Singh UC, Ghio C, Alagona G, Profeta J, S., Weiner P (1984) J Am Chem Soc 106:765
30. Weiner SJ, Kollman PA, Nguyen DT, Case DA (1986) J Comput Chem 7:230
31. Halgren TA (1992) J Am Chem Soc 114(20):7827
32. Halgren TA (1996) J Comput Chem 17:520
33. Halgren TA (1996) J Comput Chem 17:490
34. Shivakumar D, Harder E, Damm W, Friesner RA, Sherman W (2012) J Chem Theory Comput 8(8):2553
35. Shivakumar D, Williams J, Wu Y, Damm W, Shelley J, Sherman W (2010) J Chem Theory Comput 6(5):1509
36. Dauber P, Hagler AT (1980) Acc Chem Res 13(4):105
37. Brock CP, Minton RP (1989) J Am Chem Soc 111(13):4586
38. Buntine MA, Hall VJ, Kosovel FJ, Tiekink ERT (1998) J Phys Chem A 102(14):2472
39. Jorgensen WL, Maxwell DS, Tirado-Rives J (1996) J Am Chem Soc 118(45):11225
40. Cohen EEW, Rosen LS, Vokes EE, Kies MS, Forastiere AA, Worden FP, Kane MA, Sherman E, Kim S, Bycott P (2008) J Clin Oncol 26(29):4708
41. Campeta AM, Chekal BP, Abramov YA, Meenan PA, Henson MJ, Shi B, Singer RA, Horspool KR (2010) J Pharm Sci 99(9):3874
42. Chekal BP, Campeta AM, Abramov YA, Feeder N, Glynn PP, McLaughlin RW, Meenan PA, Singer RA (2009) Org Process Res Dev 13(6):1327
43. MacKerell AD, Bashford D, Bellott, 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, Wiorkiewicz-Kuczera J, Yin D, Karplus M (1998) J Phys Chem B 102(18):3586
44. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA (1995) J Am Chem Soc 117(19):5179
45. Allinger NL, Yuh YH, Lii JH (1989) J Am Chem Soc 111(23):8551
46. Schneebeli ST, Bochevarov AD, Friesner RA (2011) J Chem Theory Comput 7(3):658
47. Mohamadi F, Richards NGJ, Guida WC, Liskamp R, Lipton M, Caufield C, Chang G, Hendrickson T, Still WC (1990) J Comp Chem 11(4):440
48. Stewart JJP (1989) J Comp Chem 10(2):209
49. Stewart JJP (1989) J Comp Chem 10(2):221
50. Speakman JC (1997) Molecular structure by diffraction methods. The Chemical Society, London
51. Kolossváry I, Guida WC (1996) J Am Chem Soc 118(21):5011
52. Baker CM, Lopes PEM, Zhu X, Roux B, MacKerell AD (2010) J Chem Theory Comput 6(4):1181
53. Sun H, Ren P, Fried J (1998) Comp Theor Poly Sci 8(1–2):229
54. Neumann MA, Perrin MA (2005) J Phys Chem B 109(32):15531
55. Abramov YA (2011) J Phys Chem A 115(45):12809
56. Baker RJ, Colavita PE, Murphy DM, Platts JA, Wallis JD (2011) J Phys Chem A 116(5):1435
57. Jorgensen WL, Schyman P (2012) J Chem Theory Comput. doi: 10.1021/ct300180w
58. Jorgensen WL, Severance DL (1990) J Am Chem Soc 112(12):4768