JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

Wolf$_2$Pack – Portal Based Atomistic Force-Field Development

Ottmar Krämer-Fuhrmann,[†] Jens Neisius,[†] Niklas Gehlen,[†] Dirk Reith,[†,#] and Karl N. Kirschner*[,†,‡]

[†]Department of Simulation Engineering, [‡]Department of Bioinformatics, Fraunhofer Institute for Algorithms and Scientific Computing SCAI, Schloss Birlinghoven, 53754 Sankt Augustin, Germany

Supporting Information

ABSTRACT: In this contribution we introduce the technical concept and implementation details concerning the front end of our force-field optimization workflow package for intramolecular degrees of freedom, called Wolf$_2$Pack. The package's design follows our belief that parameter optimization should be a user-driven, but program guided, workflow with specific modular tasks that reduce human errors and save time. Through this design, parameter optimization becomes more reliable and reproducible. Wolf$_2$Pack can integrate common force fields from different research areas, allowing the user to optimize balanced parameters; alternatively users can develop highly specialized force fields that suite their chemical systems. Included in the package's front end is a force-field and molecular database whose contents facilitate parameter optimization. Wolf$_2$Pack can be accessed at www.wolf2pack.com.

INTRODUCTION

Computational chemistry is a powerful tool for investigating and understanding chemical and biological systems at atomistic resolutions levels. An excellent example of its use is computer-aided drug design in pharmaceutical research.[1−4] Equally important insights can be achieved in the field of condensed-phase material design.[5−9] Due to the large size of many investigated systems, molecular mechanics (MM)-based techniques are an appealing approach for many researchers.

The accuracy and predictability of these techniques (e.g., molecular dynamics simulations, Monte Carlo simulations, docking) are directly dependent upon the force field employed and its optimized parameters. Clearly, there exist reliable force fields for basic chemical groups and their combination, and consequently for many chemical and biological systems (e.g., amino acids, nucleic acids, carbohydrates). These popular force fields are continually refined by a comparatively small percentage of research groups. Unfortunately, the force-field options available to researchers drop significantly when their model involves complex chemical groups, combination thereof, or "nonstandard" residues (e.g., natural products). To partially address this issue, we have developed a scientific **Wo**rkflow for **f**orce-field optimization **Pack**age (Wolf$_2$Pack) that allows nonforce-field experts the ability to optimize needed intramolecular parameters that are balanced with existing major force fields.

Scientifically, Wolf$_2$Pack targets the optimization of bonds, angles, and dihedral internal coordinates. By coupling to the R.E.D. algorithm[10,11] it also determines the partial atomic charges that are needed for describing Coulombic forces. Wolf$_2$Pack is unique because it guides the user throughout the entire parameter optimization process — from selecting the molecule that contains the internal coordinate of interest, to generating the quantum mechanics (QM) target energy curves, to the parameter optimization itself. Through this user-guided workflow, we emphasize that parametrization is done in a systematic manner, whose steps and results should be reproducible.

For the sake of completeness, force fields also require the optimization of intermolecular parameters. This task differs greatly from the strategy to optimize the intramolecular parameters. The main reasons are first, that experimental condensed phase or vapor−liquid equilibrium thermodynamic data are needed as target observables. Second, the target observables are isolated numbers in contrast to full functional angular curves. We address intermolecular parameters using gradient-based optimization strategies in a separate workflow-based modular tool called GROW.[12,13]

In addition to its support service in scientific investigations, Wolf$_2$Pack is meant to serve as a platform for researchers to participate in creating a force field and molecular database and for establishing further methodological developments. In this contribution, we focus on the technical implementation issues that led to Wolf$_2$Pack's current version. In a previous publication we outlined the scientific workflow that is the foundation for Wolf$_2$Pack's algorithm development.[14] One set of algorithms involves shell scripts that automate a specific task at each workflow step. A second set of algorithms involves the development of a user interface that communicates the researcher's desires to the shell scripts. While Wolf$_2$Pack can be used via line command, the development of a graphical user interface (GUI) was necessary to facilitate nonforce-field experts easy access to the parametrization workflow. Thus, Wolf$_2$Pack's front end is a Web interface formed using Java

Received: June 22, 2012
Published: March 1, 2013

   dx.doi.org/10.1021/ci300290g | J. Chem. Inf. Model. 2013, 53, 802−808

portlets,[15,16] making it accessible from any location that has an Internet connection.

The creation of tools for optimizing intramolecular force constants has previously been done by other research groups. Several groups have developed programs that are suited for their interests, such as TORSFIND,[17] Parmscan,[18] fit dihedrals,[19] ForceFit,[20] and AMBER's Antechamber.[21] More general and online programs for different aspects of parametrization include the R.E.D. Server[22] and ParamChem.[23−25] In general, the available programs specialize in one or two aspects (e.g., torsion parameters, partial atomic charges, atom typing) of building a complete force field and are program dependent. In contrast, our primary goal is to create a flexible, user-friendly, and user-driven package that addresses the complete parametrization scheme — from generating QM target observables to assigning atom types and partial atomic charges to optimizing internal coordinate parameters. Our secondary goal is to provide a force-field and molecular database whose contents facilitate parameter optimization. Such a program should provide force-field support for several MD packages. Thus, Wolf$_2$Pack is designed to interface with several existing packages (e.g., AMBER,[26] GROMACS[27]) allowing users to develop balanced parameters that are missing from the force fields within those packages.

A user will access Wolf$_2$Pack with specific goals in mind. These goals could include a) the optimization a missing parameters for a given force field, b) the development of a new specialized force field, or c) the validation of existing force-field parameters using QM target data. Having a single tool that enables the pursuit of these goals gives researchers more control of their modeling and simulations.

SCIENTIFIC WORKFLOW

Our first step in developing Wolf$_2$Pack was to generate a scientific workflow that represents our procedure for optimizing force-field parameters. These steps, and hence the components within the workflow, are composed of both linear and cyclic progression of actions. Scientific workflows offer the benefit that they

a) save time by automating certain tasks and reducing human error,

b) make force-field development almost deterministic and reproducible,

c) enable tasks execution in a distributed environment,

d) accommodate new ideas, algorithmic changes, and updates within each step, and finally

e) accelerate and transform the scientific-analysis process.[28−30]

The details of Wolf$_2$Pack's workflow has been previously published and essentially describe the back end.[14] Briefly, the workflow is composed of 13 global steps, as shown in Figure 1, with several steps composed of subworkflows.

This workflow was designed to be semiautomated, allowing the user to direct specific and important scientific decisions that ultimately effect the resulting parameters. An important example of this is the actual parameter optimization, which is a user-directed iterative process. While it is technically possible to fully automate the optimization process, doing so would inhibit the researcher from including his or her chemical intuition.

Once the scientific workflow has been conceptualized, the programming of modules that actualizes each step follows. The advantage of modular programming is that a large program can be separated into interlinked executables that can be independently written, updated, and maintained. New executables that represent new ideas within the workflow are also easily incorporated. While a command-line usage of these executables provide users with the greatest control, non-specialists usually prefer to use a GUI front end.

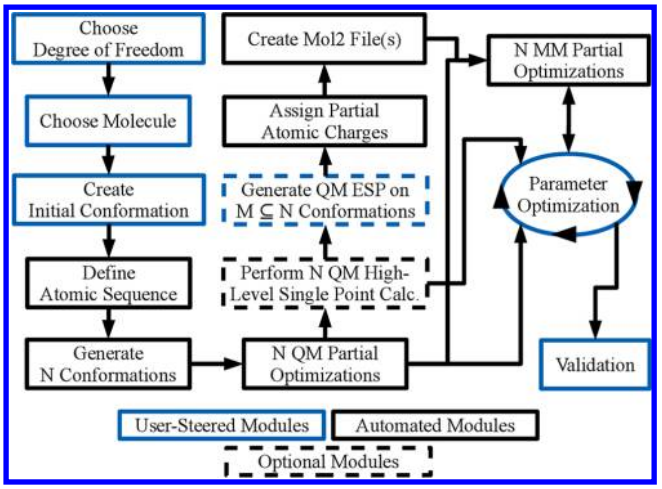

Figure 1. Conceptualization of Wolf$_2$Pack's scientific workflow.[14]

PORTAL CONCEPT

Twenty-first century scientific research is increasingly conducted on a global scale, where collaborations occur between groups separated by very long distances. Exchanging and sharing knowledge between groups occurs nearly instantaneously due to Web technologies. Our goal in designing a GUI front end for Wolf$_2$Pack is to provide its accessibility to anyone with Internet access. This also allows us to retain our intellectual property and the complete control over the code itself. It also facilitates the development and growth of a force-field and molecular database that is centralized to one location.

Portal Framework and Components. The portal is built upon the open source Liferay portal server framework.[31] Liferay offers various out-of-the-box services that eases many programming tasks and are maintained by a worldwide community. Such services include user and rights management, database access, and a complete portlet life-cycle management. Through Java portlets[15,16] the framework is adaptable and easily expanded for Wolf$_2$Pack's current and future requirements. The portlet GUIs are implemented using Vaadin,[32] a public domain application framework for Rich Internet Applications. Through the use of existing Liferay and Vaadin technologies, a stable interactive Web site can be created and customized in a relatively short time period.

As seen in Figures 2 and 3, Wolf$_2$Pack's portal is currently composed of the following seven portlets: a) Main Application, b) File Browser, c) Molecular Viewer, d) Sign In, e) Documentation, f) Message Board, and g) Announcement (not shown). Dividing the GUI based on these portlets creates an intuitively logical environment for the programmers and users. Such a modular GUI also simplifies site maintenance, freeing up time for scientific improvements.

Main Application. The scientific workflow[14] is realized in the Main Application portlet, where most of the Wolf$_2$Pack specific Java code resides. Within this portlet, the user can navigate through the following six steps:

dx.doi.org/10.1021/ci300290g | J. Chem. Inf. Model. 2013, 53, 802−808

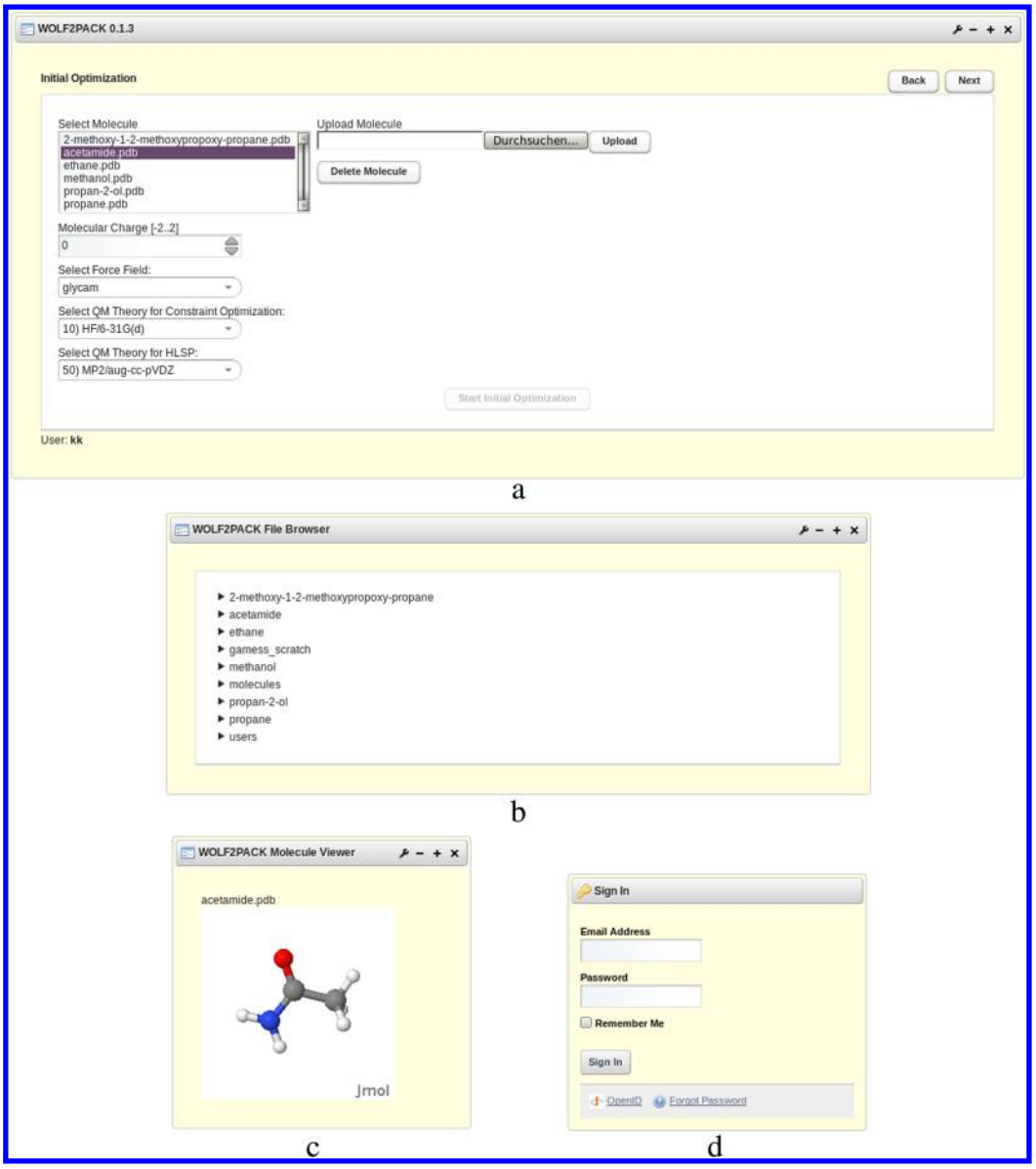

Figure 2. Screenshots of Wolf$_2$Pack's Web portlets: a) Main Application, b) File Browser, c) Molecule Viewer, and d) Sign In.

1. Choose a molecule, a force field, and QM theory levels. Perform an initial optimization.
2. Select internal coordinate and perform QM constraint optimizations.
3. Perform QM high-level-single-point (HLSP) calculations.
4. Perform partial atomic charge calculations via R.E.D.
5. Assign atom types.
6. Perform MM constraint optimization and adjust parameters.

A user can only navigate to the next step after the current one has been executed successfully; the one exception is the QM HLSP (step 3), which the user can bypass.

With the Main Application portlet steps, the user has to make specific decisions that impact the scientific workflow. Foremost, the user can select an existing molecule from the molecular database or upload a new molecule. [Note that the conformation of the molecule will have a direct impact upon the forthcoming potential energy curves. For example, different bond stretching curves between two atoms can be generated from several conformations if the molecule possess several heavy-atom torsion minima.] The user must also select the supported QM theory levels that will be used, select the desired force field, and assign the molecular charge. Each of these choices impacts the subsequent steps.

Documentation. The Documentation wiki portlet serves as an information source for the user. It is a combination of a user guide and tutorials helping newcomers to orient themselves with Wolf$_2$Pack's workflow and parametrization. This wiki can be edited by the development team at any time and any place, enabling them to distribute information quickly to the user community via the Web site.

Molecule Viewer. This portlet incorporates the open source Jmol viewer[33] for chemical structures. Graphical visualization allows the user to see their uploaded or selected molecule, which is helpful for ensuring that it contains the internal coordinate that is desired. Portlet integration of the JmolApplet was done using the Vaadin add-on Applet Integration.[34]

Message Board. The Message Board portlet allows users to interact with each other and the developers. Using a message board instead of other communication options has the

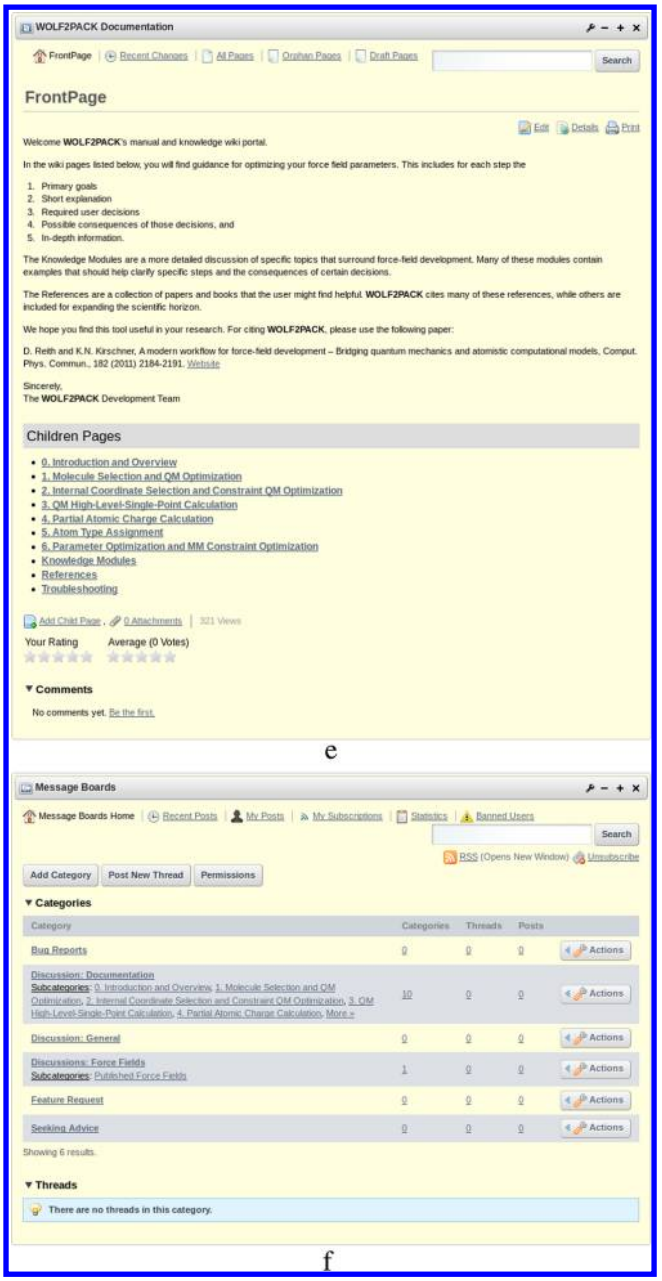

Figure 3. $Wolf_2Pack$'s Web portlets continued: e) Documentation wiki and f) Message Board.

advantage that all discussions are present and searchable in a single online location. Message boards allow for topics to be easily categorized; examples of such categories for community discussion might include program (e.g., GROMACS, AMBER) specific force fields, force-field equations, migrating force fields between programs, and so on. The message board also includes the ability to distribute messages via a mailing list for all users who wish to stay informed of a particular message thread. Users will also be able to directly contact the development team for requesting future functionalities and to report found bugs. Thus, this portlet will represent an additional resource beyond the Documentation wiki for users to gain information about $Wolf_2Pack$, force fields, and their optimization. All registered users can contribute to the message board.

File Browser. The File Browser portlet allows the user to view relevant input and output data generated or accessed in their workspace. For users who are experienced in QM and MM calculations, this portlet allows them access to the unprocessed raw data. Such access may be desired by users for additional external investigations, trouble shooting, or archiving purposes.

Announcements. The Announcements portlet is where important announcement will appear that will be of user interest. Such an announcement could include functionality upgrades, bug fixes, the addition of new force fields, or the addition of new molecules.

Sign In. This portlet allows new users to apply for an account, while registered users can login to gain access to the databases and workflow toolset. User database management is done by Liferay, including resetting forgotten passwords. Only a single login is necessary to access all portlets and workflow steps within $Wolf_2Pack$.

Portal Technology. The implementation of the $Wolf_2Pack$ portal makes use of Liferay,[31] an established enterprise Web platform that is popular for building business solutions. It is available under an open source license and offers ready-to-use components for many tasks. Thus, several desires for the $Wolf_2Pack$ portal could easily be realized by integrating existing out-of-the-box Liferay solutions, such as user management, wiki integration, and news sections.

An additional standard Web technology used in the $Wolf_2Pack$ portal is Vaadin.[32] Vaadin is a user interface framework for rich Web applications and is also available under an open source license. The implementation of $Wolf_2Pack$'s workflow using Vaadin was very effective since the application logic runs entirely on the server side. Thus, the portal was completely implemented in Java, making the tricky and error prone Javascript or PHP programming dispensable.

The portal's GUI components were realized by Vaadin, which offers a great variety of predefined widgets (e.g., buttons, forms, and tables). Additionally, a growing list of free third party add-ons extends the standard functionality of Vaadin, offering helpful tools, such as interportlet communication, applet integration, and 2-dimensional graph visualization.

In total, portal programming with Vaadin using Liferay was quick, efficient, and reduced the chance for errors — resulting in a flexible and reliable Web portal. To protect $Wolf_2Pack$'s intellectual property, the code itself is completely hidden from the user. All active components are executed on the Web server; only GUI components and their local interactions are executed on the user's side in the Web browser.

Access Rights. The portal users can have different access rights to data and tools. In $Wolf_2Pack$ four different roles are realized: a) standard user, b) enhanced user, c) super user, and e) administrator. Standard users gain access to the complete molecular and force-field databases, to most of the implemented portlets, and the ability to perform MM calculations. Enhanced users gain access to all implemented portlets, access to input and output data files, be able to upload new molecules, and gain the ability to perform QM calculations. Super users have the additional rights to access the modular scripts and intermediate files and become members of the development team. Finally, administrators have full access to the entire package for installation purposes.

Shared Workspace. The implementation of a shared workspace is extremely important in the development of $Wolf_2Pack$. As stated in the Introduction, a user will access $Wolf_2Pack$ with the goal of optimizing missing parameters for use with an existing force fields, develop a new specialized force

field, or validate existing force-field parameters. Each of these goals will involve the same scientific workflow, which requires access to a variety of molecules and their internal coordinate data. To save time and resources, molecules uploaded and subsequent QM calculations performed by a user will be accessible to all users through the shared workspace. Thus, a user can browse the molecular database for molecules containing their interested internal coordinate. If a particular molecule is not present, then an enhanced user can upload it as a PDB formatted file,[35] automatically adding it to the database. By this approach, a worldwide community of force-field developers will be established. Therefore, the force-field database will ideally include popular force fields from each MD program that has been integrated into Wolf$_2$Pack (e.g., GAFF, Parm99SB, Glycam06 for AMBER) and community derived force fields, which can be tested and validated by other community members.

As the number of users and the variety of chemical systems grows, so too will the databases. The molecular database contains a collection of molecules with computed QM data. Specifically, the QM data includes a HF/6-31G(d) optimized geometry as well as constraint optimized geometries and relative energies computed at various QM theory levels. Having a large molecular database within a shared workspace will enable users to train and test their parameters over a variety of local chemical environments. The user's time and efforts will be efficiently used toward the parameter optimization directed at their own scientific goals, rather than spending time setting up and performing QM calculations to obtain target potential energy curves.

TREATMENT OF COMPUTE INTENSIVE CALCULATIONS

Of the calculations performed by Wolf$_2$Pack, QM calculations are the most computationally expensive. Currently, Wolf$_2$Pack makes use of GAMESS[36,37] as its QM engine, which requires computer wall times ranging from minutes for small molecules to hours and days for large molecules. Four workflow steps require QM calculations — initial optimization, constraint optimizations, HLSP calculations for energetics, and the generation of molecular electrostatic potentials for partial atomic charge determination. Fortunately, these calculations can make use of parallel computer architectures.

For constraint optimizations, the user is offered two possible execution models for parallelization within Wolf$_2$Pack, which are illustrated in Figure 4. First, several tasks (i.e., jobs) are independently and simultaneously started via a farming approach, with each task being computed on available cores. In this execution model, the portal serves as job supervisor holding a list of computations to be executed. This supervisor distributes the tasks, in an arbitrary order, one-by-one to the cluster nodes who act as slaves. When a slave node has completed a task, its results are written to the shared workspace such that the portal and its users get access to these data. The supervisor then assigns any of the open tasks to the slave node until all tasks have been executed. Farming distributes the work to the available nodes, incorporating different execution speeds of the slaves, such that fast nodes receive more tasks than slow nodes. In total, the compute load gets balanced, and the completion time of the jobs ensemble gets minimized.

The second option involves sequentially submitting jobs, where an input conformation is the output of the previous minimization. [The obvious exception is the first computation, which relies on an initial "guess" conformation as an input.] This execution model will be slower than the first model. However, this model's benefit is that the resulting potential energy curve along the selected internal coordinate will more likely be continuous. For example, in the first execution model a particular input may have a severe steric clash between two atoms, causing significant conformation changes in torsion angles other than the interested internal coordinate. The resulting potential energy curve will be a mixture of conformations and thus will be discontinuous. By using the second execution model, each input is an almost fully relaxed conformation along all internal coordinates except for the investigated one — this leads to less severe steric clashes between atoms and increases the chances of generating a continuous energy curve. Regardless of the execution model employed, after all results are available the supervisor starts the postprocessing, summarizes the single results, and prepares them for the user in form of tables or as graphs.

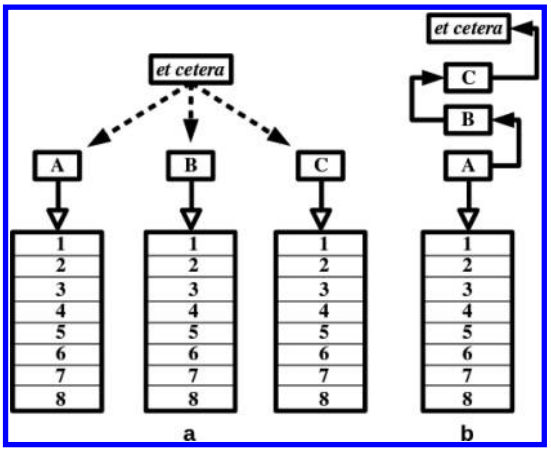

Figure 4. Parallel execution models within Wolf$_2$Pack. Represented are hypothetical compute nodes with 8 cores each and three constraint optimization jobs (A, B, and C). a) Task parallelism executes jobs simultaneously on different compute nodes. b) Sequential job execution, where the optimized geometry of the previous job is used as input into the next job.

FORCE-FIELD OPTIMIZATION

Wolf$_2$Pack does not currently contain automated optimization algorithms; instead, we encourage users to manually adjust bonded parameters (i.e., bonds, angles, and torsions). Our experience has shown that fully automated torsion parameter optimizations do not outperform systematic adjustments done manually. This is partially due to the coupling between bonded parameters (e.g., angle−torsion, torsion−torsion). An automated optimization algorithm has trouble accounting for how a specific parameter adjustment affects the energies arising from coupled parameters (e.g., other torsion angles, Lennard-Jones). Some optimization methods use relative energies from minimizations whose input included unbalanced parameters and/or torsion parameters whose values were set to zero. The resulting optimized parameter may or may not generate forces that couple well with other parameters. However, an iterative process of manually adjusting a parameter and reminimizing the conformations allow a researcher to choose the value that best couples with other parameters.

For example, imagine a molecule has an A−B−C−D−E heavy-atom-type sequence with computed A−B−C−D and C−D−E QM energy curves. The researcher notices that the C−

dx.doi.org/10.1021/ci300290g | J. Chem. Inf. Model. 2013, 53, 802−808

D−E parameter requires optimization. Adjusting the C−D−E angle parameter results in a better angle curve, but the A−B−C−D torsion angle subsequently moves by several degrees. Computing the MM torsion curves reveals that a minimum conformation position has changed by 30° — possibly due to increased steric repulsion arising from a shorter distance between atoms B and E as a result of the modified angle parameter. Iteratively adjusting both the angle and torsion parameters will inform the researcher how coupled these parameter are and subsequently which parameters best reproduce both curves. Furthermore, systematically adjusting the torsion parameters will inform the researcher how sensitive the potential is to its individual terms. With experience the researcher will gain intuition concerning parameter optimization. To support the researcher in quantifying the accuracy of their parameters for a given molecule, $Wolf_2Pack$ displays two outcome tables — one containing the RMSD value for each conformation relative to the QM geometry and one containing the MM and QM relative energies (e.g., Supporting Information Figure 8). As a caveat, a researcher could use the reported energies to perform an automated optimization offline (e.g., using least-squares fit methods,[38] genetic algorithms[18]), record the resulting parameters into their $Wolf_2Pack$'s online force-field file, and check their accuracy.

An important concern for optimizing parameters within a force field is to have them "balanced". Having a balanced force field ensures that the forces generated during an MM minimization (or an MD simulation) are consistent throughout the model. As alluded to above, a force field is considered balanced when the bonded parameters collectively generate reliable potential energy surfaces across internal coordinates and for a variety of molecules. Such a force field is said to have transferable parameters.

However, more often a balanced force field is in reference to its nonbonded parameters (e.g., partial atomic charges, Lennard-Jones) — specifically if their generation was done in a methodological way consistent with how the force field and residues were originally derived. For example, several AMBER force fields (e.g., Parm99SB) and residues (e.g., amino acids) were developed using a two-stage RESP-weighted fitting to a HF/6-31G(d) molecular electrostatic potential.[39] Thus, when using $Wolf_2Pack$ to optimize parameters for an existing force field, we use a workflow that is consistent with how the original parameters were developed (e.g., same QM theory levels, same RESP weighting parameters). In all instances, partial atomic charges are derived prior to MM calculations that use a selected force field. Furthermore, if a researcher uses existing atom types in their parameters, then the Lennard-Jones parameters that are established in that force field are also used.

CURRENT STATE AND EXAMPLES

$Wolf_2Pack$ was designed to be modular, allowing for the integration of different software and force fields. The current version is suitable for investigating AMBER's nonpolarizable force fields (both all atom and united atom). These force fields, and subsequently $Wolf_2Pack$, can model any closed shell organic or biological molecule containing hydrogen, carbon, nitrogen, oxygen, flourine, phosphorus, sulfur, chlorine, and bromine atoms. The following AMBER programs are used in the workflow: Antechamber for initial atom type assignment,[21] tleap for constructing coordinate and topology files, Sander for minimization, and ptraj for computing RMSD.[26] Fixed partial atomic charges are computed using the R.E.D. algorithm.[11] All QM calculations are performed using GAMESS.[36,37] The inputs for these programs are largely generated using Java code and shell scripts. The integration of other MM programs (e.g., GROMACS) is currently being pursued. Pymol is an additional program used to generate files that are needed throughout the workflow.[40]

Two detailed examples of how researchers can make use of $Wolf_2Pack$'s portal can be found in the Supporting Information. The first example demonstrates how a researcher can check the performance of existing force-field parameters in comparison to QM data. Three force fields are examined for how well their parameters perform for the heavy-atom torsion rotation of butadiene. The second example demonstrates how one would optimize missing parameters for an existing force field, using an analogue of a cysteine sulfate residue as input.

CONCLUSIONS AND OUTLOOK

$Wolf_2Pack$ is a tool for computational chemists to optimize and test intramolecular force-field parameters. It can be used by both experts and nonexperts in force-field development. Its foundation concept is to optimize parameters using a user-driven semiautomatic workflow. $Wolf_2Pack$ makes force-field optimization more accessible to researchers by a) providing an existing workflow that encompasses each step in the optimization process, b) automating input file generation and output analysis, c) automating the transferal of data between programs (e.g., output → input), d) reducing the decisions needed at each step, and e) facilitating the iteration of trail parameters. In this paper we have presented the development of $Wolf_2Pack$'s Web-based front end, which provides worldwide access and allows its databases to be centralized in one location. These databases collect force fields and molecules that have QM computed observables (e.g., constraint optimized geometries, relative energies, partial atomic charges). Both databases will naturally grow with $Wolf_2Pack$'s usage. We have also presented our philosophy behind $Wolf_2Pack$'s parameter optimization itself, which is done manually by the user in a systematic and iterative process.

Researchers can freely access much of $Wolf_2Pack$'s capabilities by creating a standard user account at www.wolf2pack.com. Standard users will have access to the force-field and molecular databases, can modify atom types, can modify force-field parameters, and perform MM constraint minimization. This will enable standard users to check, and reoptimize if desired, existing parameters within popular force fields as well as optimize new parameters that may be missing from these force fields. Currently, the database contains ~400 molecules and ~2,100 internal coordinates with computed QM potential curves. These molecules are composed of several chemical functional groups, including saturated and unsaturated hydrocarbons, alcohols, aldehydes, ethers, esters, ketones, amines, amides, carboxylic acids, phosphates, and sulfur containing compounds. Researchers wishing to generate new QM data will have to apply for an enhanced user account by contacting the portal's administrators.

ASSOCIATED CONTENT

Supporting Information

Two detailed examples for how a standard user might use $Wolf_2Pack$. These examples involve AMBER force fields, and thus the functional form of the AMBER force-field equation is also given. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*Phone: +49 (0)2241 14 2052. Fax: +49 (0)2241 14 1328. E-mail: karl.kirschner@scai.fraunhofer.de.

Present Address

[#]Bonn-Rhein-Sieg University of Applied Sciences, Department of Mechanical Engineering (EMT), Grantham-Allee 20, 53757, Sankt Augustin, Germany.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We would like to thank Astrid Maaß, Thorsten Köddermann, and Martin Hofmann-Apitius for helpful discussions and suggestions.

REFERENCES

(1) Galeazzi, R. Molecular Dynamics as a Tool in Rational Drug Design: Current Status and Some Major Applications. *Curr. Comput.-Aided Drug Des.* **2009**, *5*, 225−240.

(2) Morra, G.; Genoni, A.; Neves, M. A. C.; Merz, K. M., Jr.; Colombo, G. Molecular Recognition and Drug-Lead Identification: What Can Molecular Simulations Tell Us? *Curr. Med. Chem.* **2010**, *17*, 25−41.

(3) Durrant, J. D.; McCammon, J. A. Molecular Dynamics Simulations and Drug Discovery. *BMC Biol.* **2011**, *9*, 71.

(4) Bottegoni, G.; Favia, A. D.; Recanatini, M.; Cavalli, A. The Role of Fragment-Based and Computational Methods in Polypharmacology. *Drug Discovery Today* **2012**, *17*, 23−34.

(5) *Simulation Methods for Polymers*; Kotelyanskii, M., Theodorou, D. N., Eds.; Marcel Dekker Inc.: New York, 2004.

(6) Peter, C.; Kremer, K. Multiscale Simulation of Soft Matter Systems. *Faraday Discuss.* **2010**, *144*, 9−24.

(7) Dubbeldam, D.; Snurr, R. Q. Recent Developments in the Molecular Modeling of Diffusion in Nanoporous Materials. *Mol. Simul.* **2007**, *33*, 305−325.

(8) Morales Medina, G.; Martinez Rey, R. Molecular and Multiscale Modeling: Review on the Theories and Applications in Chemical Engineering. *CT&F, Cienc., Tecnol. Futuro* **2009**, *3*, 205−224.

(9) Ungerer, P.; Tavitian, B.; Boutin, A. *Applications of Molecular Simulation in the Oil and Gas Industry*; IFP Publications: Paris, 2005.

(10) Dupradeau, F.-Y.; Cézard, C.; Lelong, R.; Stanislawiak, É.; Pêcher, J.; Delepine, J. C.; Cieplak, P. R.E.DD.B.: A Database for RESP and ESP Atomic Charges, and Force Field Libraries. *Nucleic Acids Res.* **2008**, *36*, D360−D367.

(11) Dupradeau, F.-Y.; Pigache, A.; Zaffran, T.; Savineau, C.; Lelong, R.; Grivel, N.; Lelong, D.; Rosanski, W.; Cieplak, P. The R.E.D. Tools: Advances in RESP and ESP Charge Derivation and Force Field Library Building. *Phys. Chem. Chem. Phys.* **2010**, *12*, 7821−7839.

(12) Hülsmann, M.; Köddermann, T.; Vrabec, J.; Reith, D. GROW: A Gradient-Based Optimization Workflow for the Automated Development of Molecular Models. *Comput. Phys. Commun.* **2010**, *181*, 499−513.

(13) Hülsmann, M.; Vrabec, J.; Maaß, A.; Reith, D. Assessment of Numerical Optimization Algorithms for the Development of Molecular Models. *Comput. Phys. Commun.* **2010**, *181*, 887−905.

(14) Reith, D.; Kirschner, K. N. A Modern Workflow for Force-Field Development − Bridging Quantum Mechanics and Atomistic Computational Models. *Comput. Phys. Commun.* **2011**, *182*, 2184−2191.

(15) *Java*. http://www.java.com (accessed January 23, 2013).

(16) Patil, S. What is a Portlet. 2005. http://oreilly.com/java/archive/what-is-a-portlet.html (accessed June 20, 2012).

(17) Lii, J.-H.; Yuh, Y. H. *TORSFIND*: Unpublished program, distributed with the MM3 program from the group of Norman L. Allinger.

(18) Wang, J.; Kollman, P. Automatic Parameterization of Force Field by Systematic Search and Genetic Algorithms. *J. Comput. Chem.* **2001**, *22*, 1219−1228.

(19) Guvench, O.; MacKerell, A. D., Jr. Automated Conformational Energy Fitting for Force-Field Development. *J. Mol. Model.* **2008**, *14*, 667−679.

(20) Waldher, B.; Kuta, J.; Chen, S.; Henson, N.; Clark, A. E. ForceFit: A Code to Fit Classical Force Fields to Quantum Mechanical Potential Energy Surfaces. *J. Comput. Chem.* **2010**, *31*, 2307−2316.

(21) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graphics Modell.* **2006**, *25*, 247−260.

(22) RESP ESP charge Derive Server. http://q4md-forcefieldtools.org/REDS (accessed December 6, 2012).

(23) *ParamChem*. http://paramchem.org (accessed December 6, 2012).

(24) Vanommeslaeghe, K.; MacKerell, A. D. Automation of the CHARMM General Force Field (CGenFF) I: Bond Perception and Atom Typing. *J. Chem. Inf. Model.* **2012**, *52*, 3144−3154.

(25) Vanommeslaeghe, K.; Raman, E. P.; MacKerell, A. D. Automation of the CHARMM General Force Field (CGenFF) II: Assignment of Bonded Parameters and Partial Atomic Charges. *J. Chem. Inf. Model.* **2012**, *52*, 3155−3168.

(26) Case, D. A. *AMBER 9;* University of California: San Francisco, CA, 2006.

(27) Hess, B.; Kutzner, C.; van der Spool, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.

(28) Gil, Y.; Deelman, E.; Ellisman, M.; Fahringer, T.; Fox, G.; Gannon, D.; Goble, C.; Livny, M.; Moreau, L.; Myers, J. Examining the Challenges of Scientific Workflows. *Computer* **2007**, *40*, 24.

(29) Tiwari, A.; Sekhar, A. K. Workflow Based Framework for Life Science Informatics. *Comput. Biol. Chem.* **2007**, *31*, 305−319.

(30) De Roure, D.; Goble, C. Software Design for Empowering Scientists. *IEEE Software* **2009**, *26*, 88−95.

(31) *Liferay − Enterprise Open Source Portal and Collaboration Software*. http://liferay.com (accessed January 23, 2013).

(32) *Vaadin − Java Framework for Building Modern Web Applications*. http://vaadin.com (accessed January 23, 2013).

(33) *Jmol: an Open-Source Java Viewer for Chemical Structures in 3D*. http://jmol.sourceforge.net (accessed June 20, 2012).

(34) Ekblad, S. AppletIntegration for Vaadin, version 1.2.9. https://vaadin.com/directory#addon/appletintegration (accessed June 20, 2012).

(35) PDB Format. http://www.pdb.org/pdb/static.do?p=file_formats/index.jsp (accessed June 20, 2012).

(36) Schmidt, W. M.; Baldridge, K. K.; Boatz, A. J.; Elbert, T. S.; Gordon, S. M.; Jensen, H. J.; Koseki, S.; Matsunaga, N.; Nguyen, A. K.; Su, S.; Windus, L. T.; Dupuis, M.; Montgomery, A. J. General Atomic and Molecular Electronic Structure System. *J. Comput. Chem.* **1993**, *14*, 1347−1363.

(37) Gordon, M. S.; Schmidt, M. W. In *Theory and Applications of Computational Chemistry: the First Forty Years*; Gordon, M. S., Schmidt, M. W., Dykstra, C. E., Eds.; Elsevier: Amsterdam, Boston, 2005; pp 1167−1189.

(38) Hopfinger, A.; Pearlstein, R. Molecular Mechanics Force-Field Parameterization Procedures. *J. Comput. Chem.* **1984**, *5*, 486−499.

(39) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.

(40) *PyMOL Molecular Graphics System LLC*. http://sourceforge.net/projects/pymol/ (accessed December 6, 2012).