PCCP

Large-scale virtual high-throughput screening for the identification of new battery electrolyte solvents: computing infrastructure and collective properties

Tamara Husch,[a] Nusret Duygu Yilmazer,[a] Andrea Balducci[b] and Martin Korth*[a]

A volunteer computing approach is presented for the purpose of screening a large number of molecular structures with respect to their suitability as new battery electrolyte solvents. Collective properties like melting, boiling and flash points are evaluated using COSMOtherm and quantitative structure–property relationship (QSPR) based methods, while electronic structure theory methods are used for the computation of electrochemical stability window estimators. Two application examples are presented: first, the results of a previous large-scale screening test (*PCCP*, 2014, **16**, 7919) are re-evaluated with respect to the mentioned collective properties. As a second application example, all reasonable nitrile solvents up to 12 heavy atoms are generated and used to illustrate a suitable filter protocol for picking Pareto-optimal candidates.

1. Introduction

The current battery technology cannot meet the demands arising from the electrification of the automobiles, which is of essential importance to meet the world's rising energy demand with renewable energy technologies.[1] Materials science has contributed substantially to the process of developing better, safer and greener batteries, but especially electrolyte systems are still far from being perfect.[2–4] Computational screening can contribute to help with this problem, but comparably little work has been done in this area so far,[5] as most theoretical studies focus exclusively on electrode materials.[6,7] Standard electrolyte formulations consist of a mixture of cyclic and linear carbonates, most often ethylene carbonate (EC) and dimethyl carbonate (DMC), with lithium salts like hexafluorophosphate (LiPF$_6$) and several additives.[8] When searching for alternative materials, properties which have to be taken into account include electrochemical stability windows, melting, boiling and flash points, dielectric constant, viscosity, ionic and electronic conductivity, toxicity and price.[8] Especially the correct prediction of the electrochemical stability of the whole electrolyte systems is a very complex problem, because of the interactions of the electrolyte components with each other (*e.g.*, reduced solvent molecules can abstract hydrogen atoms from other species) and with the electrodes – usually a passivating film, the so-called

solid-electrolyte-interphase (SEI) is formed from decomposed electrolyte species during the first charging cycles.[9] The formation of stable SEI films is of essential importance for the battery performance, but is hard to characterize experimentally and cannot be predicted yet using computational models.[5] Here we do not take electrolyte reactivity into account when screening for new materials, but instead focus on the computing infrastructure necessary for really large-scale screenings, and on the approximate description of important collective properties (melting, boiling and flash points, viscosity, ion solubility *etc.*), as opposed to the 'non-collective' properties of (single-molecule/non-reactive) electrochemical stability, which we have investigated in more detail in a previous screening study.[10] More details on lithium ion battery science and technology can be found in several reviews published over the last few years, for instance, by Goodenough,[9,11–13] Aurbach,[4,14] Scrosati,[2,15] Winter,[16,17] and Tarascon.[3] Excellent reviews on electrolyte materials were published by Xu;[8,18] SEI formation and its properties were reviewed by Novak,[19] and again by Xu.[20,21] Reviews on computational studies in this field were published by Balbuena,[22] Curtiss,[23] Leung[24,25] and Korth.[5] The most important facts to be considered in the work presented here are the following: to improve the LIB technology substantially, the chemical potentials of the anode and the cathode have to be pushed farther apart, *i.e.* advanced electrode materials are needed. As soon as one goes in this direction, the electrolyte is likely to become a bottleneck, as it has to remain functioning under the new conditions. The development of advanced electrolyte systems is thus also a very important field for improving the LIB technology. Theory can contribute to this to help

[a] *Institute for Theoretical Chemistry, Ulm University, Albert-Einstein-Allee 11, 89069 Ulm, Germany. E-mail: martin.korth@uni-ulm.de*
[b] *MEET, University of Münster, Corrensstrasse 46, 48149 Münster, Germany*

understand current systems, but also to suggest new materials. While there is a good number of theoretical studies of the first type, comparably little is published on the virtual (pre-)screening of new electrolyte materials.[5] One reason is that approaches which look at single properties only are not well suited to treat the multidimensional problem of optimizing electrolyte systems. We therefore present an approach which goes beyond the current state-of-the-art systems by including estimates for collective properties and grid-type computing ressources.

2. Volunteer computing

Computational screening offers the possibility to filter a large number of compounds for subsequent experimental work, but the 'chemical space' of possibly suitable small organic molecules is known to be vast.[26] Sufficiently large computing resources are thus of vital importance for systematic screening studies. The largest part of the world's computing power is assumed to be distributed over almost a billion personal computers. These provide a maximum computing capacity of 8 to 21 PetaFLOPS,[27] which is in a similar range to today's supercomputers. 'Volunteer computing' (VC) strives to make these resources available for scientific purposes. In contrast to supercomputers, the computing power cannot be bought, but has to be earned, which also makes it a cheaper alternative to supercomputers. Everybody who owns an internet-connected personal computer can donate computer time. Projects with a larger public appeal therefore attract more volunteers. To encourage contribution to one's project, time has to be spent on promoting the project and communicating with the volunteers. To give them the opportunity to participate, the application should be adapted to a wide range of computer types. The volunteers remain effectively anonymous and are therefore not accountable for projects. They have to trust the project to treat the provided access to their computers with appropriate care. The results, which may be wrong due to malfunctions or intentional obstruction, may be validated by performing each job on several computers. The appropriation of middleware increased the appeal for researchers to set up VC projects. Therefore, the effort of the scientists as well as the required computational knowledge is significantly reduced. One of the most popular providers of middleware systems is the Berkeley Open Infrastructure for Network Computing (BOINC) project. Over the last decade the BOINC platform has established itself as a standard tool for realizing VC projects.[28] The BOINC platform has several advantages, which allow a comparably easy setup for VC projects – the backend server is based on standard web-server components and BOINC provides work-scheduling, data handling and accounting features, as well as a 'core-client' software which needs to be installed on the volunteer's computer. Scientists can thus focus on adapting their computer programs to work within the BOINC framework and administer their project. A project in the language of BOINC is a data unit that uses BOINC for distributing its jobs. Each project is independent, has its own web site and incorporates applications. An application includes

the intended programs as well as a set of workunits and results. A workunit is one computation that is going to be performed, also known as a job. Each result is associated with a workunit and it describes the instance of a computation. Each application can be compiled for different platforms. The application program can in principle be written in any language, but one has to keep the details in mind. One may circumvent altering the source code by using the provided BOINC wrapper. To adapt the program directly, only some minor source code modifications have to be incorporated. To account for the special requirements the existing program has to be interfaced with BOINC. Interfacing the software with BOINC is done *via* implementing message passing interface (MPI)-like calls, which account for the communication between the scientific application and the core client, which in turn organizes the communication with the project server(s). To illustrate how BOINC works, the life-cycle of one job is traced: first a work generator creates a job and its input-files. BOINC then creates one or more instances of the job. The core client requests work *via* a scheduler request from the server, when it has free capacities. The scheduler scans the database for available jobs. The client gets the binary and input files of an application, starts the application, sends the results back to the project server and reports the job as completed. A validator checks the output files and potentially compares matching outputs of the same job. After full completion the file deleter deletes the input and output files. Volunteers have complete control over how much work is done at what times, and can look up their results on the project web pages. Furthermore, they collect the so-called credit points proportional to the work their computers did and are ranked in top lists according to their overall credit value. More details on the BOINC platform can be found elsewhere;[29] an overview of the existing projects is given on the BOINC web pages.[28] In 2005, Korth and Grimme released the first VC project in chemistry, Quantum Monte Carlo at home (QMC@home);[30] more recently, Aspuru-Guzik and co-workers presented the *Harvard Clean Energy Project*.[31] We present here the cleanmobility.now project,[32] which is a re-release of the QMC@HOME project, now with a focus on the search for new electrolyte materials. The results are based on a modified version of ORCA,[33] but to verify the outcome, computations were also performed on local computing resources at this stage. The distribution of other software packages within our project is in preparation. With our VC project, we would like to help in finding safer and greener battery materials. This confronts us with several scientific challenges, amongst others, the estimation of collective properties, addressed in the next section.

3. Methods for estimating collective properties

We aim at an integrated computational approach for the large-scale screening of molecular battery materials. As a first step, we evaluated computational methods for the prediction of (single-molecule/non-reactive) electrochemical stability window rankings.[10] The so-called 'electrochemical stability window' (ESW) of a compound can be computed from its oxidation and reduction

This journal is © the Owner Societies 2015

Phys. Chem. Chem. Phys., 2015, **17**, 3394–3401 | **3395**

potentials (though it needs to be shifted by the computed potential of the reference electrode to match the experimentally measured value):

$$V_{ox} = -\frac{\Delta G_{ox}}{nF} \quad V_{red} = -\frac{\Delta G_{red}}{nF}$$

One thus needs the Gibbs free energies of oxidation and reduction:

$$\Delta G_{ox} = \Delta G(X) - \Delta G(X^+) \quad \Delta G_{red} = \Delta G(X^-) - \Delta G(X)$$

Individual free energies are usually taken from density functional theory (DFT) computations by taking zero-point and thermal enthalpic, entropic, as well as (implicit) solvation effects into account:

$$\Delta G = \Delta H - T\Delta S = \Delta E + \Delta E_{ZVPE} + \Delta H_T - T\Delta S + \Delta G_{solvation}$$

As an estimate of the oxidation and reduction potentials one can look at the electronic energy differences (electron affinity (EA) and ioniziation potential (IP))

$$\Delta G_{ox} \approx IP = \Delta E_{ox} = E(X) - E(X^+)$$
$$\Delta G_{red} \approx EA = \Delta E_{red} = E(X^-) - E(X)$$

which in turn can be estimated from the lowest unoccupied and highest occupied molecular orbital (LUMO/HOMO) energies:

$$IP \approx -E_{HOMO} \quad EA \approx -E_{LUMO}$$

In our previous study we have evaluated several computational approaches and approximations for their impact on ranking compounds with respect to their EWS. We suggested a combination of semiempirical quantum mechanical (SQM) and wave function theory (WFT) methods for an efficient two-step screening procedure. All screening results presented below do include the electrochemical stability as a factor, partly based on SQM and WFT data as previously suggested (the database benchmark) and partly based on SQM data only (the nitrile set), as we have found that SQM estimates are usually good enough for ranking compounds within our extended screening procedure outlined below. In the following, we turn our attention to the approximate treatment of collective properties with lower level methods, as no higher-level methods are available for the fast prediction of these properties. At this point we still do not take solid-electrolyte-interface (SEI) formation into account, but schemes for using estimators for complex properties including SEI formation are in preparation. The results presented here are thus based on simplified model systems and approximate computational methods and should be taken with appropriate care (as simpler problems were shown to require much more advanced methods sometimes).[34] Our main focus is the definition of a screening strategy, not the benchmarking of lower level methods against each other, though we are able to present some data for the comparison of COSMOtherm with 'pure' QSPR type models. We furthermore do not consider ionic liquids here, for which some details of the current screening setup are not optimal, though our scheme can easily be adjusted

to work for ionic liquids also. Several collective properties are relevant for improving electrolyte systems. Here we investigate the possibilities of the COSMOtherm model[35] for predicting boiling and flash points, viscosities (as estimators of ion conductivity), solubilities and free energies of solvation for several ionic species (as an estimator of solubility again) and of a pure quantitative structure property relationship (QSPR) model of Lang for computing melting points.[36]

COSMOtherm predictions are based on empirical models which make use of data from electronic structure theory calculations to allow for the description of hitherto experimentally unknown species also (unlike standard chemical engineering models, which usually require some compound-specific, experimentally determined parameters). For COSMOtherm we compare the performance of density functional theory (DFT) based estimates with semi-empirical (SQM) ones with respect to the ranking of candidate compounds. Semi-empirical PM$_6$-DH$^{+\,37-40}$ calculations were done using MOPAC2012,[41] making use of the COSMO[35] solvation model to generate the input for COSMOtherm. BP86[42,43] DFT calculations have been performed using TURBOMOLE 6.4,[44,45] D2 dispersion corrections,[46] the RI approximation for two-electron integrals,[47,48] and again COSMO to generate the input for COSMOtherm. BP86 DFT calculations (again using D2 and RI) and local pair natural orbital (LPNO) coupled electron pair approximation (CEPA1)[49] (CEPA in the following) calculations were done using a modified version of ORCA 2.8.[50] TZVP, TZVPP and QZVP AO basis sets[51] were employed for TURBOMOLE and ORCA calculations.

More information about the COSMOtherm model can be found, for instance, in a recent review by Klamt,[35] but some details with direct relevance for the following need to be mentioned: in COSMOtherm, the liquid viscosity of a pure compound at room temperature is computed using a QSPR-type model:

$$\ln(\eta_i) = c_A A_i + C_{M^2} M_i^2 + c_{N_{ring}} N_i^{ring} + c_{TS} TS_i + c_0$$

It is based on the surface area A_i of the compound, the second σ-moment M_i, the number of ring atoms N_{ring} and the pure entropy time temperature TS_i, as well as on five parameters, which were derived from a set of 175 neutral organic compounds.

For boiling points at a given pressure, COSMOtherm varies the temperature of the system until the difference between the predicted vapor pressure and the given pressure is below 10^{-4} mbar. The vapor pressure itself is computed via the chemical potential of compound i in system S from the integration of the σ-potential over the surface of the compound

$$\mu_i^S = \mu_i^{C,S} + \int p_i(\sigma)\mu_s(\sigma)d\sigma$$

(using $\mu_i^{C,S}$ as a combinatorial contribution) and an estimate of the pure compound's chemical potential in the gas phase

$$\mu_i^{gas} = E_{gas}^i - E_{COSMO}^i - \omega_{ring}N_{ring}^i + \eta_{gas}$$

(using E as quantum chemical total energies, a ring correction term and two parameters ω and η) according to:

$$p_i^S/1bar = \exp[(\mu_i^{gas} - \mu_i^S)/RT]$$

3396 | Phys. Chem. Chem. Phys., 2015, 17, 3394–3401

This journal is © the Owner Societies 2015

Flash points are computed from the temperature dependent variation of the vapor pressure until the flash point pressure (FPP) is found,[52] which in turn is computed from the molecular surface area A according to:

$$\ln(FPP) = 22.7 - 3 \times \ln(A)$$

The prediction of melting points was not possible using COSMOtherm when we initially finished our study, though this feature has recently been added. We do not present COSMOtherm melting point predictions here, but instead use a QSPR model of A. Lang.

The model of Lang uses readily available molecular descriptors (with the number of hydrogen-bond donors and polar surface area as most important ones here) for the purely empirical estimation of melting points. Melting points are especially hard to predict as rather minor differences between molecular structures can result in large melting point differences due to packing effects. More details on QSPR methods and the available software packages can be found in recent reviews.[53,54]

To get an idea of how well COSMOtherm performs in comparison to 'pure' QSPR models we did some additional QSPR calculations using the T.E.S.T. software package.[55] Although benchmarking such methods is not the focus of our work, this seemed interesting to us, as QSPR models were, for instance, used to estimate viscosities for the purpose of developing new ionic liquids.[56] For our QSPR predictions, we relied on the consensus model (the average over all implemented models) implemented in the T.E.S.T. software package. The details of all included approaches can be found in the T.E.S.T. user guide. In these methods, the properties investigated here are predicted using overall 797 molecular descriptors and relying on experimental data sets for several thousand compounds.

Table 1 shows the predicted and measured[8] results obtained for typical electrolyte solvents. Perusing this table, one finds that mean average deviations (MADs, about 0.2 cP and 18/23/23 degrees of viscosities and melting/flash/boiling points) are in the order of about 10 to 15 percent of the relevant property windows (here, 0.33 to 2.53 cP and −137 to 26/−17 to 160/41 to 270 degrees). The correct ranking of compounds can be investigated by looking at correlation coefficients, such as Pearson's R values for linear correlation and Kendall's τ values for non-linear (rank) correlation. Both correlation measures are very high, especially for the COSMOtherm-based estimates (with R values of 0.95 to 0.98 and τ values of 0.73–0.78), which implies that the ranking of compounds with respect to these properties is even better than the prediction of the actual values. This is a very promising result for integrated computational and experimental screening procedures, in which the computational part acts only as a filter for subsequent experimental high-throughput work.

Table 2 shows a comparison of the performance of the consensus QSPR method implemented in T.E.S.T. with COSMOtherm. Mean absolute deviations (MADs) are higher for the consensus model especially in the case of viscosities. R and τ values are lower for the consensus model, especially the R values of flash and boiling points. To be fair it should be mentioned that the consensus melting point prediction model performed much worse than the one of Lang which we use for our screenings, with an MAD of 37 K (30 K on the fit set) for the former, opposed to 18 K for the latter. This clearly illustrates that better QSPR models compared to the ones implemented in T.E.S.T. are available, which have a strong focus on toxicity prediction, not investigated here. Other available QSPR software packages unfortunately do not supply models for all properties of interest and none seems to be suitable for our high-throughput infrastructure.[54]

Table 1 Calculated estimates (this work) and experimental values (from Xu[8]) of collective properties of common electrolyte solvents

	Viscosity [cP]			Melting point [°C]			Flash point [°C]			Boiling point [°C]		
	Calculated	Measured	Deviation	Calculated	Measured	Deviation	Calculated	Measured	Deviation	Calculated	Measured	Deviation
1,3-DL	0.74	0.59	0.15	−75.7	−95.0	19.3	−24.0	1	−25.0	96.8	78	18.8
2-Me-1,3-DL	0.87	0.54	0.33	−40.9	—	—	−6.5	—	—	123.9	—	—
2-Me-THF	0.62	0.47	0.15	−97.2	−137.0	39.8	−16.6	−11	−5.6	112.1	80	32.1
4-Me-1,3-DL	0.86	0.60	0.26	−40.9	−125.0	84.1	−8.1	−2	−6.1	121.2	85	36.2
BL	1.10	1.73	−0.63	−36.7	−43.5	6.8	63.1	97	−33.9	237.9	204	33.9
DEC	0.76	0.75	0.01	−37.0	−74.3	37.3	−1.5	31	−32.5	124.8	126	−1.2
DEE	0.90	—	—	−69.7	−74.0	4.3	12.1	20	−7.9	144.5	121	23.5
DMC	0.61	0.59	0.02	−15.0	4.6	−19.6	−30.7	18	−48.7	78.9	91	−12.1
DME	0.55	0.46	0.09	−59.6	−58.0	−1.6	−31.5	0	−31.5	80.7	84	−3.3
DMM	0.40	0.33	0.07	−98.4	−105.0	6.6	−61.8	−17	−44.8	36.6	41	−4.4
EA	0.50	0.45	0.05	−83.8	−84.0	0.2	−30.0	−3	−27.0	84.2	77	7.2
EB	0.75	0.71	0.04	−86.9	−93.0	6.1	4.3	19	−14.7	134.0	120	14.0
EC	1.81	1.90	−0.09	22.6	36.4	−13.8	97.5	160	−62.5	284.3	248	36.3
EMC	0.69	0.65	0.04	−38.0	−53.0	15	−15.3	—	—	103.2	110	−6.8
MB	0.65	0.60	0.05	−83.6	−84.0	0.4	−12.2	11	−23.2	109.1	102	7.1
NMO	1.60	2.50	−0.90	40.1	15.0	25.1	111.7	110	1.7	320.0	270	50.0
PC	1.79	2.53	−0.74	−15.2	−48.8	33.6	102.8	132	−29.2	299.0	242	57.0
THF	0.50	0.46	0.04	−100.1	−109.0	8.9	−37.4	−17	−20.4	81.9	66	15.9
VL	1.41	2.00	−0.59	−17.3	−31.0	13.7	85.4	81	4.4	278.4	208	70.4
MAD			0.22			17.69			22.86			22.64
R			0.95			0.87			0.95			0.98
τ			0.78			0.75			0.72			0.76

This journal is © the Owner Societies 2015

Phys. Chem. Chem. Phys., 2015, **17**, 3394–3401 | **3397**

Table 2 Comparison of the performance of the consensus QSPR method implemented in T.E.S.T. with COSMOtherm: mean absolute deviation (MAD), Pearson's R and Kendall's τ values for the correlation between properties computed for the systems in Table 1

Property	QSPR			COSMOtherm		
	MAD	R	τ	MAD	R	τ
Viscosity	1.15	0.83	0.68	0.22	0.95	0.78
Flash point	26.82	0.77	0.63	22.86	0.95	0.72
Boiling point	37.08	0.63	0.65	22.64	0.98	0.76

Table 3 Pearson's R and Kendall's τ values for the correlation between computed properties of the database set, only values with $R > 0.5$ are given

	R	τ
IP/HOMO	−0.84	−0.67
EA/LUMO	−0.57	−0.29
Melting/flash point	0.65	0.49
Melting/boiling point	0.63	0.48
Boiling/flash point	0.99	0.92
$\Delta G_{solv}(Li^+)/\Delta G_{solv}(Mg^{2+})$	0.97	0.83
$\Delta G_{solv}(Li^+)/\Delta G_{solv}(Al^{3+})$	0.98	0.87
$\Delta G_{solv}(Mg^{2+})/\Delta G_{solv}(Al^{3+})$	1.00	0.96

4. Example applications

(A) Database benchmark re-evaluation

In a recent study, we screened 100 000 molecules from public databases for their redox stability.[10] Structures were automatically retrieved in the SMILES format and converted using OpenBabel[57] into force field optimized input structures for DFT calculations. The highest occupied molecular orbital/lowest unoccupied molecular orbital (HOMO/LUMO) gaps, dipole moments and elemental composition were used as filters for identifying 83 (out of 100 000) candidate compounds, which were used for a systematic benchmarking of quantum chemical methods. When investigating the 'hits' of this screening study in more detail, later on, many turned out to have unfavorable collective properties, like high melting points, which nicely illustrate the need for a multi-level approach like it is presented here. As a first example application we thus re-evaluate the results of this earlier study using our improved approach. Using the previous CEPA ionization potential (IP) and electron affinity (EA) values as estimators of electrochemical stability, and after computing viscosities, melting/flash/boiling points, $Li^+/Mg^{2+}/Al^{3+}/LiPF_6$-solubilities and free energies of solvation for all compounds, we applied the following filtering scheme to identify the most promising candidates. The COSMOtherm model is not well suited to describe the properties of small, highly charged ions and thus these results are likely only meaningful for the ranking of rather similar compounds. Furthermore, computed solubilities are just indicated as high for all compounds with reasonable solubility using COSMOtherm, so that we turned to free energies of solvation for the ions as a rough estimator of ion solubility, again to be used only to rank rather similar compounds (which is actually not the case in this example application, but in the next one, see section B). Free energies of solvation are highly correlated for the small ions, so that using one value (we take the one for Li^+) is sufficient for ranking purposes. Compounds with an IP below, an EA above, and free energies of solvation (which are negative) above the average were discarded, as well as compounds with melting/flash/boiling points above 273 K/below 323 K/below 373 K. Calculations at all levels were successful for 8772 candidates out of the subset of about 10 000 small organic molecules from the whole database of 100 000 structures. We take problems with any of the calculations as an indicator of the complicated electronic nature of the compound and thus discard it. Filtering left us with 72 structures and we then restricted our list to

53 Pareto-optimal ones, *i.e.* the candidates which are not equal to or beaten by another candidate with respect to all properties, as non-Pareto optimal candidates would offer no advantages over the remaining stock. To account for the inaccuracy of our approximate models we binned the computed values in 5 percent intervals before checking for Pareto-optimal cases. Further results obtained for these compounds are presented elsewhere, here we concentrate on the evaluation of the COSMOtherm model (but all data are made publicly available on our project web page[32]).

Table 3 shows the correlation between the different properties computed. Not unexpectedly, one finds a very good correlation between IP and HOMO values and much lower values for the correlation between EA and LUMO values. Melting points are correlated with flash and boiling points, which in turn are almost perfectly correlated with each other. Free energies of solvation for different small cations are also highly correlated, but not correlated to the corresponding values of the large anionic PF_6^- ions. These findings will be discussed in more detail below, together with the corresponding data for the second application case.

(B) Nitrile solvents

Abu-Lebdeh and Davidson,[58,59] Isken *et al.*[60] as well as Balducci and co-workers[61] recently proposed adiponitrile (ADN) as a new electrolyte solvent (for different types of applications), which begs the question of whether there are other nitrile solvents that might offer advantages over the currently used ones. To investigate this, we used the Molgen algorithm[62] to construct all 'reasonable' (poly-)nitrile solvents up to 12 heavy atoms. For 'reasonable' structures we hereby assume no C/C double- or triple-bonds apart from those in aromatic systems and no rings other than 5- to 7-membered ones, as compounds with such structural elements would very likely be rather reactive and unstable. The outcome is converted again using OpenBabel into force field optimized structures as starting points for BP86/TZVP and PM_6-DH^+ optimizations as inputs for COSMOtherm. This setup gave 4947 structures, calculations at all levels were successful for 4897 candidates, and the above filtering scheme left us with 20 structures, out of which 17 are Pareto-optimal. Most interestingly, adiponitrile, the compound suggested by several groups, was on our final list and was thus successfully picked out of almost 5000 possible candidates (as well as several other small

Table 4 Pearson's R and Kendall's τ values for the correlation between computed properties of the nitrile set, only values with $R > 0.5$ are given

	DFT		SQM	
	R	τ	R	τ
Viscosity/flash point	0.58	0.83	0.59	0.81
Viscosity/boiling point	0.52	0.78	0.54	0.78
Melting/boiling point	0.55	0.41	0.51	0.36
Flash/boiling point	0.99	0.92	0.99	0.92
Viscosity/$\Delta G_{solv}(PF_6^-)$	−0.47	−0.49	−0.50	−0.44
Flash point/$\Delta G_{solv}(PF_6^-)$	−0.74	−0.44	−0.70	−0.41
Boiling point/$\Delta G_{solv}(PF_6^-)$	−0.70	−0.42	−0.66	−0.39
$\Delta G_{solv}(Li^+)/\Delta G_{solv}(Mg^{2+})$	0.93	0.76	0.95	0.79
$\Delta G_{solv}(Li^+)/\Delta G_{solv}(Al^{3+})$	0.95	0.81	0.97	0.83
$\Delta G_{solv}(Mg^{2+})/\Delta G_{solv}(Al^{3+})$	1.00	0.94	1.00	0.95

di-nitriles previously suggested). Compounds supposedly better than adiponitrile are now investigated experimentally in this group, so that we can again focus on the evaluation of the computational models here (but all data are made publicly available on our project web page[32]).

Table 4 again shows the correlation between different properties, now for DFT as well as SQM based estimates. First of all, values of SQM are very similar to those of DFT, implying that it is possible to obtain DFT-level ranking results with the much faster SQM method, see also the discussion of Table 5 below. For this set, viscosities are highly correlated with both flash and boiling points, which are in turn again perfectly correlated with each other. Also free energies of solvation for different small cations are again highly correlated, but still not correlated to the corresponding values of the large anionic PF_6^- ions. Viscosities, and flash and boiling points are inversely correlated with free energies of solvation for PF_6^-. This implies that for a given compound class high thermal stability and good ion solubility often go hand in hand, but usually come at the price of higher viscosities, *i.e.* very likely lower ion conductivities. The results obtained for the much more diverse database set presented above, on the other hand, did not show a high correlation between viscosities and boiling and flash points. This indicates that different compound classes show different relationships between viscosities and thermal stability.

Table 5 Comparison of the performance of SQM and DFT as the starting point for COSMOtherm: Pearson's R and Kendall's τ values, as well as deviation measures (mean deviation MD, mean absolute deviation MAD, root mean square deviation RMSD and error span MIMA) for the nitrile set, showing the correlation and deviation between property estimates based on SQM calculations and the corresponding property estimates based on DFT calculations

	R	τ	MD	MAD	RMSD	MIMA
HOMO	0.96	0.86	3.70	3.70	3.71	4.36
LUMO	0.93	0.60	−1.72	1.72	1.76	3.33
Viscosity	0.95	0.89	0.40	0.50	1.76	79.34
Boiling point	0.95	0.82	25.35	26.69	31.53	548.91
Flash point	0.95	0.83	14.07	14.85	17.81	322.57
$\Delta G_{solv}(Li^+)$	0.74	0.56	3.64	3.66	3.72	24.61
$\Delta G_{solv}(Mg^{2+})$	0.72	0.50	9.15	9.19	9.31	64.21
$\Delta G_{solv}(Al^{3+})$	0.73	0.50	13.25	13.30	13.49	92.31
$\Delta G_{solv}(PF_6^-)$	0.97	0.84	−0.30	0.34	0.41	6.33

Table 6 Comparison of the performance of the consensus QSPR method implemented in T.E.S.T. with COSMOtherm: mean absolute deviation (MAD), Pearson's R and Kendall's τ values for the nitrile set, showing the correlation and deviation between property estimates from T.E.S.T. and the corresponding property estimates from COSMOtherm

Property	MAD	R	τ
Viscosity	2.51	0.37	0.41
Flash point	15.26	0.76	0.62
Boiling point	72.56	0.71	0.59

The best way of addressing the challenge of balancing thermal stability with ion conductivity thus seems to be a diversity oriented approach, which goes beyond the usual compound classes (carbonates, nitriles, *etc.*).

The used COSMOtherm models are parametrized to work on top of B86/TZVP DFT calculations, but they are also possible to work on top of SQM computations, which are about 2 to 3 magnitudes faster. It is thus of high interest to investigate the effect of using SQM instead of DFT information on the ranking results in more detail. Table 5 shows correlation and error measures – mean deviations (MDs), mean absolute deviations (MADs), root mean square deviations (RMSDs) and error spans (MIMAs) all in kcal mol^{-1} – for the comparison of properties computed at the DFT level with those computed at the SQM level. Perusing this table, first of all one finds very high correlation values for all computed properties. MD and MAD values of similar magnitude indicate that systematic shifts are found for all properties, which are mostly within the accuracy found for the COSMOtherm approach in comparison to experimental values (Table 1). High error span (MIMA) values nevertheless suggest to re-screen preselections of compounds from SQM level computations at the DFT level again to exclude outliers, or to directly use a two-level approach as a consistency check. Correlation measures close to the ones found for comparison of COSMOtherm with experiment (Table 1) allow one to draw the conclusion that the theoretically less appealing SQM computations can also be very valuable for large-scale screening approaches based on the COSMOtherm model.

Finally, Table 6 shows a comparison of the consensus QSPR method implemented in T.E.S.T. with COSMOtherm for the nitrile set. Mean absolute deviation (MAD), Pearson R and Kendall's τ values between 'pure' QSPR and COSMOtherm values are given. Perusing this data one finds that the consensus QSPR model gives substantially different results than the COSMOtherm approach. In light of our evaluation of the consensus model for the systems in Table 1 (see above) and the unavailability of accurate QSPR alternatives that are suitable for our high-throughput approach, COSMOtherm seems to be a better choice for our task.

5. Conclusions

We have presented a volunteer computing approach for screening molecular electrolyte components, evaluated lower-level methods for computing collective properties and described a

protocol for analyzing the results in combination with higher-level estimators for electrochemical stability windows. A comparison with experimental references showed the high value of COSMOtherm and QSPR models for estimating collective properties of electrolyte components and especially for ranking compounds with respect to these properties. Furthermore, much faster available SQM-based COSMOtherm estimates are likely almost as valuable as DFT-based ones for this purpose. Two application examples illustrate the opportunities of our integrated multi-level approach. Comparing the first study on a very diverse set of compounds with the second one on nitriles, we find that a diversity-oriented approach offers more opportunities for balancing thermal stability with ion conductivity. From the systematic study on all reasonable nitrile solvents of up to 12 heavy atoms adiponitrile is found as one of the 17 Pareto-optimal candidates, in accordance with recent suggestions from experimental work (as well as several other small di-nitriles previously investigated).

Acknowledgements

References

1 F. T. Wagner, B. Lakshmanan and M. F. Mathias, *J. Phys. Chem. Lett.*, 2010, **1**, 2204.
2 B. Scrosati, J. Hassoun and Y.-K. Sun, *Energy Environ. Sci.*, 2011, **4**, 3287.
3 J.-M. Tarascon, *Philos. Trans. R. Soc., A*, 2010, **368**, 3227.
4 R. Marom, S. F. Amalraj, N. Leifer, D. Jacob and D. Aurbach, *J. Mater. Chem.*, 2011, **21**, 9938.
5 M. Korth, Computational Studies of Solid Electrolyte Interphase Formation, in *Specialist Periodical Reports: Chemical Modeling: Applications and Theory*, ed. M. Springborg and J.-O. Joswig, Royal Society of Chemistry, London, UK, 2014.
6 G. Hautier, A. Jain, S. P. Ong, B. Kang, C. Moore, R. Doe and G. Ceder, *Chem. Mater.*, 2011, **23**, 3495.
7 G. Hautier, A. Jain, H. Chen, C. Moore, S. P. Ong and G. Ceder, *J. Mater. Chem.*, 2011, **21**, 17147.
8 K. Xu, *Chem. Rev.*, 2004, **104**, 4303.
9 J. B. Goodenough and Y. Kim, *Chem. Mater.*, 2010, **22**, 587.
10 M. Korth, *Phys. Chem. Chem. Phys.*, 2014, **16**, 7919.
11 J. B. Goodenough and K.-S. Park, *J. Am. Chem. Soc.*, 2013, **135**, 1167.
12 J. B. Goodenough, *Acc. Chem. Res.*, 2013, **46**, 1053.
13 J. B. Goodenough, *Energy Environ. Sci.*, 2014, **7**, 14.
14 V. Etacheri, R. Marom, R. Elazari, G. Salitra and D. Aurbach, *Energy Environ. Sci.*, 2011, **4**, 3243.
15 B. Scrosati and J. Garche, *J. Power Sources*, 2010, **195**, 2419.
16 M. Winter and R. J. Brodd, *Chem. Rev.*, 2004, **104**, 4245.
17 R. Wagner, N. Preschitschek, S. Passerini, J. Leker and M. Winter, *J. Appl. Electrochem.*, 2013, **43**, 481.
18 K. Xu and A. von Cresce, *J. Mater. Chem.*, 2011, **21**, 9849.
19 P. Verma, P. Maire and P. Novak, *Electrochim. Acta*, 2010, **55**, 6332.
20 K. Xu, *Energies*, 2010, **3**, 135.
21 K. Xu and A. von W. Cresce, *J. Mater. Res.*, 2012, **27**, 2327.
22 *Lithium-Ion Batteries: Solid-Electrolyte Interphase*, ed. Y. Wang and P. B. Balbuena, Imperial College Press, London, 2004.
23 G. Ferguson and L. A. Curtiss, Atomic-Level Modeling of Organic Electrolytes in Lithium-Ion Batteries, *Applications of Molecular Modeling to Challenges in Clean Energy*, American Chemical Society, Washington D C., 2013, ch. 13, p. 127.
24 K. Leung, *Chem. Phys. Lett.*, 2013, **568–169**, 1.
25 K. Leung, *J. Phys. Chem. C*, 2013, **117**, 1539.
26 J.-L. Reymond, L. Ruddigkeit, L. Blum and R. van Deursen, *WIREs Comput. Mol. Sci.*, 2012, **2**, 717.
27 BOINC stats, http://boincstats.com/en/stats/-1/project/detail/overview, accessed Nov. 1, 2014.
28 BOINC, boinc.berkeley.edu, accessed Jul. 15, 2014.
29 D. Anderson, *Proc. 5th IEEE/ACM Int. Workshop Grid Comp.*, 2004, Proc. Grid '04, 4.
30 M. Korth and S. Grimme, *J. Phys. Chem.*, 2008, **112**, 2104.
31 J. Hachmann, *et al.*, *J. Phys. Chem. Lett.*, 2011, **2**, 2241.
32 cleanmobility.now. www.qmcathome.org/clean_mobility_now.html, accessed Jul. 15, 2014.
33 F. Neese, *WIREs Comput. Mol. Sci.*, 2012, **2**, 73.
34 M. Korth, S. Grimme and M. D. Towler, *J. Phys. Chem. A*, 2011, **115**, 11734.
35 A. Klamt, *WIREs Comput. Mol. Sci.*, 2011, **1**, 699.
36 A. S. I. D. Lang, MeltingPointModel010, http://onchallenge.wikispaces.com/MeltingPointModel010, accessed Jul. 15, 2014.
37 M. Korth, M. Pitonak, J. Rezac and P. Hobza, *J. Chem. Theory Comput.*, 2010, **6**, 344.
38 M. Korth, *J. Chem. Theory Comput.*, 2010, **6**, 3808.
39 M. Korth, *ChemPhysChem*, 2011, **12**, 3131.
40 J. C. Kromann, A. Christensen, C. Steinmann, M. Korth and J. H. Jensen, *PeerJ Preprints*, 2014, http://dx.doi.org/10.7287/peerj.preprints.353v1.
41 OPENMOPAC, www.openmopac.net, accessed Jul. 15, 2014.
42 A. D. Becke, *Phys. Rev. A: At., Mol., Opt. Phys.*, 1988, **38**, 3098.
43 J. P. Perdew, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1986, **33**, 8822.
44 R. Ahlrichs, M. Bär, M. Häser, H. Horn and C. Kölmel, *Chem. Phys. Lett.*, 1989, **162**, 165.
45 TURBOMOLE V6.4 2012, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007, available from http://www.turbomole.com.
46 S. Grimme, *J. Comput. Chem.*, 2006, **27**, 1787.
47 K. Eichhorn, O. Treutler, H. Öhm, M. Häser and R. Ahlrichs, *Chem. Phys. Lett.*, 1995, **242**, 652.
48 K. Eichhorn, F. Weigend, O. Treutler and R. Ahlrichs, *Theor. Chem. Acc.*, 1997, **97**, 119.
49 F. Neese, A. Hansen, F. Wennmohs and S. Grimme, *Acc. Chem. Res.*, 2009, **42**, 641.
50 F. Neese, *ORCA – an ab initio*, Density Functional and Semi-empirical program package, Version 2.9, University of Bonn, 2012.

51 A. Schäfer, C. Huber and R. Ahlrichs, *J. Chem. Phys.*, 1994, **100**, 5829.

52 COSMOlogic GmbH & Co. KG, COSMOthermX UserGuide, Version C30 1401 and A. Klamt, to be published.

53 A. R. Katritzky, *et al.*, *Chem. Rev.*, 2010, **110**, 5714.

54 J. C. Dearden, P. Rotureau and G. Fayet, *SAR QSAR Environ. Res.*, 2013, **24**, 279.

55 *T.E.S.T. 4.1*, U.S. Environmental Protection Agency, 2012, www.epa.gov/nrmrl/std/qsar/qsar.html, accessed Nov. 11, 2014.

56 I. Billard, G. Marcou, A. Quadi and A. Varnek, *J. Phys. Chem. B*, 2011, **115**, 93.

57 N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch and G. R. Hutchison, *J. Cheminf.*, 2011, **3**, 33.

58 Y. Abu-Lebdeh and I. Davidson, *J. Electrochem. Soc.*, 2009, **156**, A60.

59 Y. Abu-Lebdeh and I. Davidson, *J. Power Sources*, 2009, **189**, 576.

60 P. Isken, C. Dippel, R. Schmitz, R. W. Schmitz, M. Kunze, S. Passerini, M. Winter and A. Lex-Balducci, *Electrochim. Acta*, 2011, **56**, 7530.

61 A. Brandt, P. Isken, A. Lex-Balducci and A. Balducci, *J. Power Sources*, 2012, **204**, 213.

62 A. Kerber, *et al.*, *MATCH*, 1998, **37**, 205.

This journal is © the Owner Societies 2015

Phys. Chem. Chem. Phys., 2015, **17**, 3394–3401 | 3401