Force field refinement for reproducing experimental infrared spectra of ionic liquids

Supplementary Information

András Szabadi,[a,b] Aleksandar Doknic,[c] Jonathan Netsch,[a] Ádám Márk Pálvögyi,[d] Othmar Steinhauser[a] and Christian Schröder[a*]

[a] *Department of Computational Biological Chemistry, Faculty of Chemistry, University of Vienna, Währingerstr. 17, A-1090 Vienna, Austria, email:christian.schroeder@univie.ac.at*
[b] *University of Vienna, Doctoral School in Chemistry (DoSChem), Währingerstr. 42, 1090 Vienna, Austria*
[c] *Research Network Data Science, University of Vienna, Kolingasse 14-16, Vienna, Austria*
[d] *Institute of Applied Synthetic Chemistry, Vienna Technical University, Vienna, Austria*

Contents

1 QM level of theory

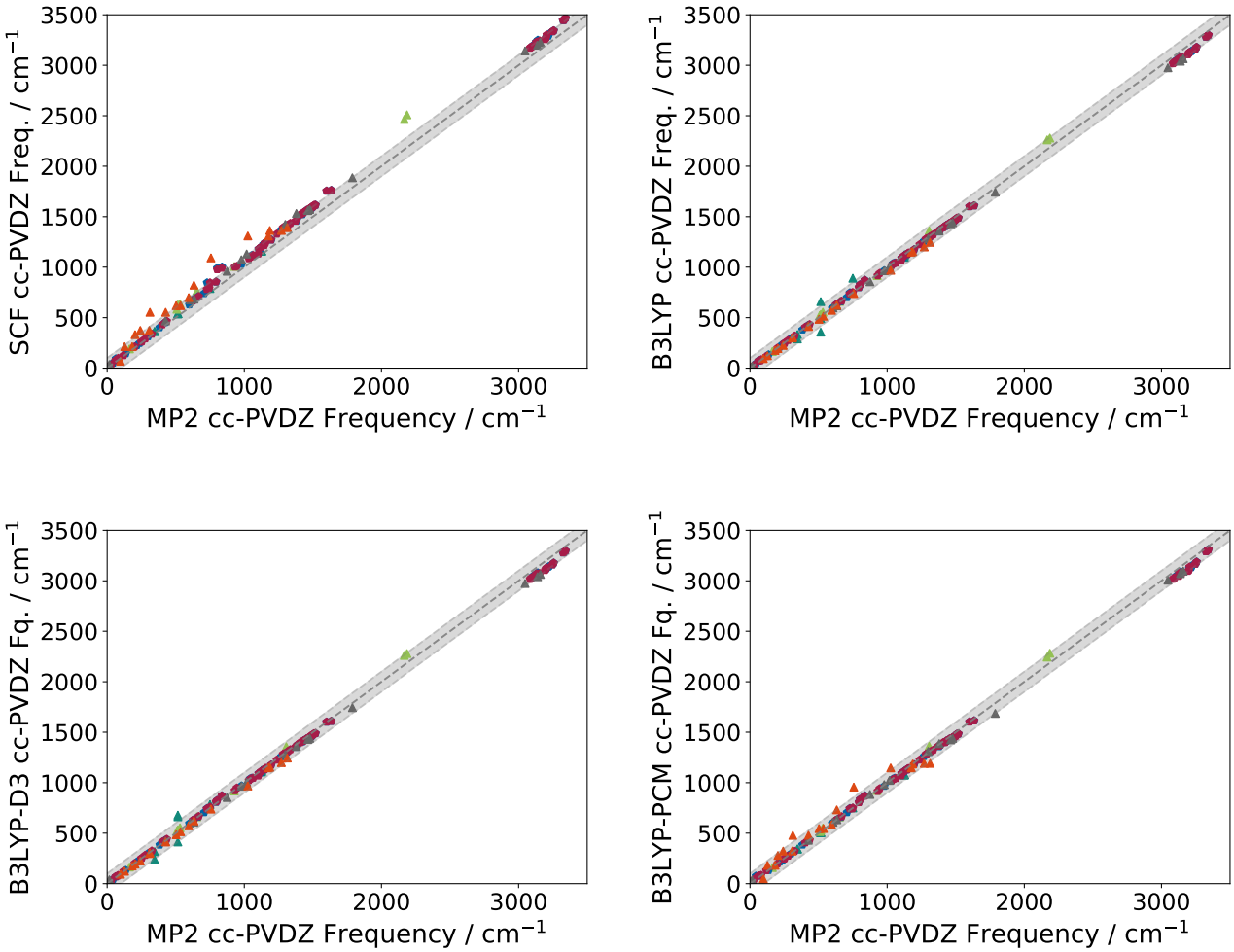
An evaluation of the impact of the level of theory and basis set on the Quantum Mechanical (QM) frequencies was conducted to affirm the integrity of the input data. Using Psi4, QM frequencies across several basis sets

- cc-PVDZ
- cc-PVTZ
- aug-cc-pVTZ

and functionals

- HF
- B3LYP
- B3LYP D3
- B3LYP PCM
- MP2

were calculated and subsequently represented graphically against each other to manifest the extent of the deviation, as demonstrated below. Please note, that the QM gas phase scaling factors have been applied. [1,2]



Even though the Hartree Fock-based method exhibits a small systematic drift when compared to higher levels of theory, all higher levels of theory show insignificant deviations between frequencies. Thus, for the application of FFGenOpt, the MP2 level of theory with a cc-PVDZ basis set can be considered sufficient.

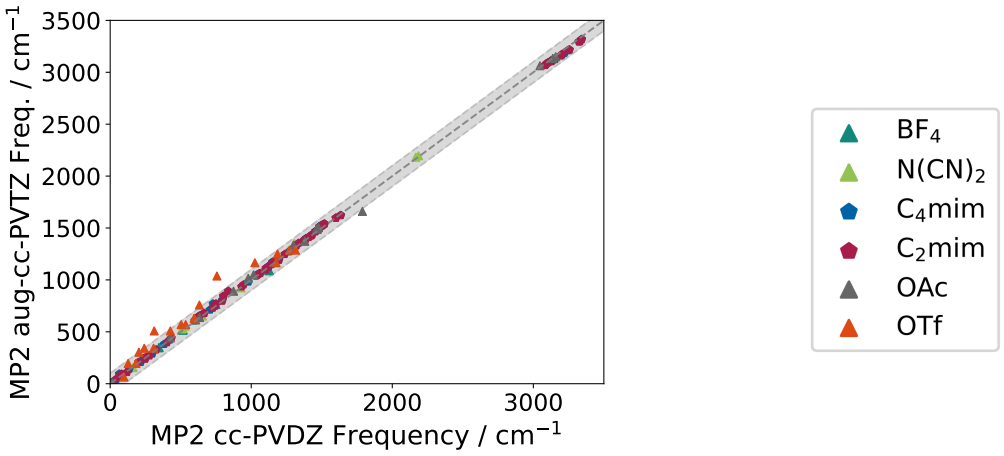


Figure S1: QM frequencies in cm^{-1} plotted against each other.

2 FFGenOpt

FFGenOpt is predicated on the Automated Frequency Matching Method (AFMM),[3] which employs the vibran module in CHARMM or a Hessian calculation in OpenMM to compute normal mode frequencies derived from a specified force field. Initially, modifications were restricted solely to bond and angle force constants. The algorithm is built upon a rudimentary Monte Carlo scheme, offering an effective foundation for the evolution of more advanced methodologies.

The integrity of the generated parameters was assessed via a merit function:

$$Y^2 = \sum_{i=1}^{3N-6} \Omega_i^2 \left(\tilde{\nu}_i^{FF} - \tilde{\nu}_i^{QM} \right)^2 \tag{1}$$

The modes $\tilde{\nu}_i^{FF}$ and $\tilde{\nu}_i^{QM}$ are aligned by projecting the eigenvectors. However, several $\tilde{\nu}_i^{FF}$ may map onto the same $\tilde{\nu}_i^{QM}$ if the QM eigenvector is the most appropriate fit for both $\tilde{\nu}_i^{FF}$. Ω_i^2 serves as a weight parameter that may be utilized to induce a bias towards frequencies of higher relevance.

While the merit function remains a feature within FFGenOpt, mode alignment is now conducted using the Hungarian method,[4] ensuring that each Molecular Mechanics (MM) normal mode corresponds to a distinct QM mode. A key improvement is the substitution of the Monte Carlo method with a genetic algorithm, enhancing optimization by emphasizing ideal frequencies and facilitating simultaneous adjustments of multiple force constants.

2.1 Genetic algorithms

Historically, there have been multiple approaches to what is nowadays known as evolutionary computation. The concept of evolution strategies was explored by Rechenberg[5] since the 1960s when the concept of genetic algorithms was developed by Holland in the 1970s.[6,7] Similar to artificial neural networks, which are inspired by biological neural networks, genetic algorithms are inspired by genetic mutations that occur in nature.[6] Nature successfully uses genetic mutations to solve complex problems.

Supervised machine learning algorithms, including neural networks and support vector machines, present a significant drawback: the necessity for substantial volumes of data to construct an efficacious model, a process colloquially referred to as 'training a model'. Such algorithms are ill-suited for scenarios wherein such data is either unavailable or its generation would be prohibitively costly.

Calculatory methods, though adept at identifying local optima either by equating the gradient of the objective function to zero and solving sets of equations or through local gradient methods such as steepest descent, also harbor noteworthy disadvantages. Primarily, these local methods exclusively explore the vicinity, potentially overlooking superior results situated further from the starting point. Additionally, these methodologies are highly reliant on the existence of a derivative of the objective function. However, objective functions on real-world problems frequently lack a derivative, deviating from the smoothness characteristic of elementary mathematical functions, thereby necessitating the approximation of the gradient for such methodologies.

These beneficial attributes, in conjunction with several other significant advantages over numeric algorithms, have been succinctly encapsulated by Wirsansky:[8]

- Global Optimization: In contrast to many conventional optimization algorithms that limit their search to the vicinity of a starting point and potentially stagnate at local optima, Genetic algorithms operate with a broad array of solution candidates. As such, they can span an extensive search space and exhibit robustness towards local minima, contingent on sufficient population diversity.
- Addressing Problems with Complex Mathematical Representation: GA merely necessitates a fitness function for assessing potential solutions, thereby eliminating dependence on supplementary information about the search space, such as derivatives.
- Tackling Problems Lacking Mathematical Representation: The fitness function is not predicated on a mathematical representation. It is feasible, for instance, to employ subjective user-determined fitness values.
- Noise Resilience: Genetic algorithms may even be implemented for problems wherein the output of the fitness function, with identical input parameters, is non-deterministic. This includes instances of sensor data or user input.
- Parallelism and Distributed Processing Support: Fitness evaluations, mutations, and crossovers among solution candidates can be executed in parallel on distinct members of the population, circumventing many of the parallelizations challenges common to numerous algorithms.
- Appropriateness for Continuous Learning: Genetic algorithms maintain functionality even amidst environmental alterations, and previous populations may be employed to continue the computation.

However, there are also limitations to genetic algorithms which should be considered:

- Requirement for Special Definitions: The task of correctly formulating the problem in terms of chromosome and fitness function can be demanding. The definition of selection, crossover, and mutation operators must also be specifically tailored based on the representation.
- Hyperparameter Tuning: Much like other optimization methodologies, GA necessitate appropriately tuned hyperparameters (such as population size or mutation rate) to discern optimal solutions within an acceptable timeframe.
- Computational Intensity: Owing to their operation on multiple solution candidates as opposed to a singular one, GA can render computations more iterative.
- Premature Convergence: A population characterized by low diversity, that is, elevated similarity among population members, can precipitate convergence to the fittest member, consequently stagnating at a local minimum.
- Absence of Solution Guarantee: There exists no assurance that the global optimum will be located.

It is pertinent to acknowledge that these advantages and constraints are merely a generalized overview. Some optimization algorithms and GA may address the above-mentioned challenges. However, a thorough analysis of the plethora of algorithms exceeds the purview of this thesis. When selecting optimization methodologies, it is recommended to examine domain-specific solutions for optimization problems.

The source code, as well as a tutorial on how to use FFGenOpt, is available free of charge from github(https://github.com/cbc-univie/FFGenOpt).

2.2 Alignment of MM and QM normal modes for the ionic liquid ions

This chapter contains all MM normal mode frequencies and their alignment to the QM modes in the gas phase. A visual representation of these data is given in Fig. 2 of the main article.

2.2.1 Imidazoliums

We focused on the most prominent cation family for ionic liquids, *i.e.* imidazoliums. Although having the most force field parameters and the most normal modes, the optimization process for the imidazoliums 1-ethyl-3-methylimidazolium C_2mim^+ and 1-butyl-3-methylimidazolium C_4mim^+ is straightforward as the normal mode eigenvectors differ significantly from each other and the molecules do not have high symmetry.

	$\tilde{\nu}_{QM}$ [cm^{-1}]	$\tilde{\nu}_{FF}$ [cm^{-1}]		$\tilde{\nu}_{QM}$ [cm^{-1}]	$\tilde{\nu}_{FF}$ [cm^{-1}]		$\tilde{\nu}_{QM}$ [cm^{-1}]	$\tilde{\nu}_{FF}$ [cm^{-1}]
1	64	11	18	1105	1022	35	1268	1719
2	43	48	19	1310	1065	36	1423	1808
3	135	117	20	1140	1122	37	3323	1893
4	237	220	21	1119	1126	38	1432	1973
5	216	226	22	1126	1146	39	1635	2048
6	383	361	23	837	1154	40	1600	2199
7	298	439	24	1041	1172	41	3128	2777
8	630	449	25	1388	1215	42	3140	2823
9	599	559	26	1493	1351	43	3247	2888
10	662	603	27	1492	1363	44	3202	2892
11	803	618	28	1379	1379	45	3256	2896
12	707	671	29	1197	1418	46	1056	3118
13	728	729	30	1455	1525	47	3107	3224
14	422	790	31	1507	1590	48	3217	3341
15	795	813	32	1485	1601	49	3223	3341
16	980	1008	33	1519	1605	50	3343	3427
17	1151	1010	34	1476	1627	51	3331	3444

Table S1: Normal modes of 1-ethyl-3-methylimidazolium C_2mim^+.

	$\tilde{\nu}_{QM}$ [cm^{-1}]	$\tilde{\nu}_{FF}$ [cm^{-1}]		$\tilde{\nu}_{QM}$ [cm^{-1}]	$\tilde{\nu}_{FF}$ [cm^{-1}]		$\tilde{\nu}_{QM}$ [cm^{-1}]	$\tilde{\nu}_{FF}$ [cm^{-1}]
1	30	32	24	1046	970	47	1288	1434
2	75	67	25	931	973	48	1431	1519
3	62	80	26	1033	978	49	1146	1555
4	81	98	27	1155	1045	50	1422	1633
5	119	130	28	1387	1073	51	1634	1743
6	201	186	29	1102	1096	52	1600	1756
7	253	256	30	1065	1118	53	1408	1836
8	326	276	31	1237	1191	54	1475	1848
9	248	317	32	1486	1216	55	3128	3007
10	634	372	33	1195	1232	56	3095	3010
11	278	378	34	1336	1245	57	3090	3047
12	626	445	35	1321	1258	58	3128	3055
13	668	552	36	1519	1262	59	3083	3072
14	403	586	37	1485	1264	60	3154	3111
15	435	628	38	1500	1284	61	3201	3113
16	744	739	39	836	1288	62	3247	3115
17	750	767	40	1122	1293	63	3193	3117
18	802	819	41	1502	1297	64	3256	3117
19	794	840	42	1509	1325	65	3205	3118
20	727	856	43	1490	1342	66	3135	3120
21	1105	891	44	1282	1352	67	3343	3261
22	945	923	45	1455	1388	68	3324	3263
23	1150	944	46	1374	1406	69	3331	3603

Table S2: Normal modes of 1-butyl-3-methylimidazolium C_4mim^+.

2.2.2 Tetrafluoroborate

While the frequencies of BF_4^- are matched well, the experimental spectrum of $[C_4mim]BF_4$ shows a peak in the region between 1000 and $1100\,\mathrm{cm}^{-1}$, where neither the cation nor the anion produce any frequencies in the gas phase.

	$\tilde{\nu}_{QM}$ [cm^{-1}]	$\tilde{\nu}_{FF}$ [cm^{-1}]		$\tilde{\nu}_{QM}$ [cm^{-1}]	$\tilde{\nu}_{FF}$ [cm^{-1}]		$\tilde{\nu}_{QM}$ [cm^{-1}]	$\tilde{\nu}_{FF}$ [cm^{-1}]
1	750	352	4	516	371	7	1127	1130
2	516	370	5	346	514	8	1126	1132
3	516	371	6	346	515	9	1126	1134

Table S3: Normal modes of tetrafluoroborate BF_4^-.

2.2.3 Trifluoromethanesulfonate

	$\tilde{\nu}_{QM}$ [cm^{-1}]	$\tilde{\nu}_{FF}$ [cm^{-1}]		$\tilde{\nu}_{QM}$ [cm^{-1}]	$\tilde{\nu}_{FF}$ [cm^{-1}]		$\tilde{\nu}_{QM}$ [cm^{-1}]	$\tilde{\nu}_{FF}$ [cm^{-1}]
1	97	69	7	308	414	13	1026	832
2	183	178	8	312	414	14	1271	1044
3	204	179	9	595	479	15	1313	1044
4	633	255	10	503	482	16	1186	1304
5	427	256	11	535	483	17	757	1357
6	242	275	12	127	644	18	1177	1358

Table S4: Normal modes of trifluoromethanesulfonate OTf^-.

	$\tilde{\nu}_{QM}$ [cm^{-1}]	$\tilde{\nu}_{FF}$ [cm^{-1}]		$\tilde{\nu}_{QM}$ [cm^{-1}]	$\tilde{\nu}_{FF}$ [cm^{-1}]		$\tilde{\nu}_{QM}$ [cm^{-1}]	$\tilde{\nu}_{FF}$ [cm^{-1}]
1	69	66	7	553	373	13	1092	962
2	213	219	8	554	373	14	1364	1175
3	213	219	9	620	415	15	1364	1175
4	332	266	10	620	415	16	1391	1239
5	374	319	11	699	461	17	1309	1298
6	374	319	12	822	612	18	1309	1298

Table S5: Normal modes of trifluoromethanesulfonate OTf^- of the original force field.[9]

2.2.4 Acetate

	$\tilde{\nu}_{QM}$ [cm^{-1}]	$\tilde{\nu}_{FF}$ [cm^{-1}]		$\tilde{\nu}_{QM}$ [cm^{-1}]	$\tilde{\nu}_{FF}$ [cm^{-1}]		$\tilde{\nu}_{QM}$ [cm^{-1}]	$\tilde{\nu}_{FF}$ [cm^{-1}]
1	16	202	6	426	898	11	979	1452
2	1788	466	7	1477	1112	12	1380	1593
3	875	563	8	1464	1201	13	3047	3092
4	634	767	9	1017	1375	14	3138	3160
5	606	823	10	1299	1408	15	3159	3167

Table S6: Normal modes of acetate OAc^-.

	$\tilde{\nu}_{QM}$ [cm^{-1}]	$\tilde{\nu}_{FF}$ [cm^{-1}]		$\tilde{\nu}_{QM}$ [cm^{-1}]	$\tilde{\nu}_{FF}$ [cm^{-1}]		$\tilde{\nu}_{QM}$ [cm^{-1}]	$\tilde{\nu}_{FF}$ [cm^{-1}]
1	467	414	6	1073	1031	11	1561	1521
2	36	591	7	1130	1035	12	1886	1612
3	683	626	8	1532	1303	13	3142	2908
4	677	687	9	1424	1365	14	3199	2969
5	959	901	10	1572	1520	15	3227	2970

Table S7: Normal modes of acetate OAc^- of the original force field.[10]

2.2.5 Dicyanamide

The QM vibrational frequency of the C≡N bond is difficult to reproduce, both regarding the gas phase normal modes and in the bulk IR spectra. The quantum mechanical frequencies around $2500\,\mathrm{cm}^{-1}$ are significantly higher than expected, while experimental values are in the range typically associated with triple bonds (around 2100 to $2200\,\mathrm{cm}^{-1}$).

The low frequencies are challenging to reproduce in both acetate and dicyanamide, leading to some of the largest discrepancies in the entire investigated set of ions. In all cases, extra care has to be taken to account for degenerate eigenmodes, arising due to symmetry, as well as to perform an RMS calculation and align the coordinates to the reference prior to matching the eigenvectors.

	$\tilde{\nu}_{QM}$ [cm^{-1}]	$\tilde{\nu}_{FF}$ [cm^{-1}]		$\tilde{\nu}_{QM}$ [cm^{-1}]	$\tilde{\nu}_{FF}$ [cm^{-1}]		$\tilde{\nu}_{QM}$ [cm^{-1}]	$\tilde{\nu}_{FF}$ [cm^{-1}]
1	532	28	4	508	716	7	1306	1394
2	516	64	5	654	769	8	2167	2435
3	163	284	6	928	1148	9	2186	2485

Table S8: Normal modes of dicyanamide N(CN)$_2^-$.

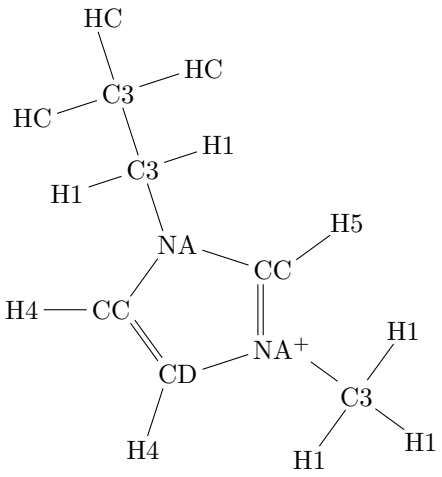
	$\tilde{\nu}_{QM}$ [cm^{-1}]	$\tilde{\nu}_{FF}$ [cm^{-1}]		$\tilde{\nu}_{QM}$ [cm^{-1}]	$\tilde{\nu}_{FF}$ [cm^{-1}]		$\tilde{\nu}_{QM}$ [cm^{-1}]	$\tilde{\nu}_{FF}$ [cm^{-1}]
1	627	61	4	191	634	7	1410	1253
2	637	115	5	584	652	8	2509	2065
3	1007	235	6	765	1006	9	2464	2136

Table S9: Normal modes of dicyanamide N(CN)$_2^-$ of the original force field.[9]

2.3 Optimized force constants of FFGenOpt

All displayed force constants here are not modified by the scaling factor for the liquid phase.

2.3.1 C_2mim^+



Atom type 1	Atom type 2	k_1 [kcal mol^{-1} Å^{-2}]	k_2 [kcal mol^{-1} Å^{-2}]
C3	H1	335.9	318.91
C3	NA	334.7	258.67
CC	NA	438.8	622.3
CD	NA	438.8	518.59
CC	H5	356.0	459.42
CC	CD	504.0	414.37
CC	H4	350.1	150.0
CD	H4	350.1	453.83
C3	C3	303.1	654.61
C3	HC	337.3	427.82

Table S10: C_2mim^+ bond parameters. k_1 taken from Ref. 9, k_2 from FFGenOpt

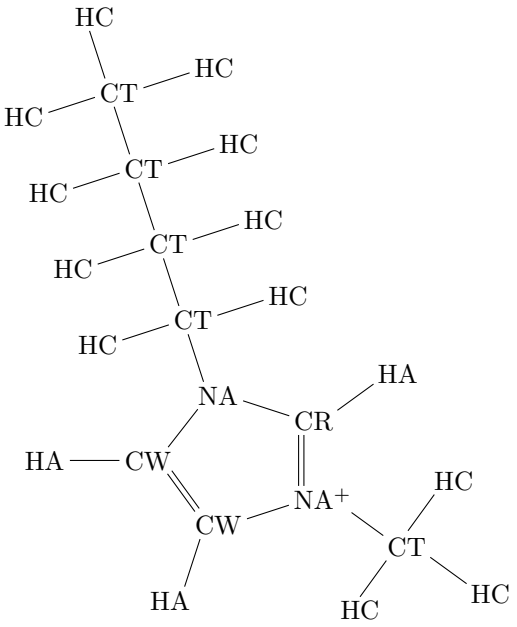
Atom type 1	Atom type 2	Atom type 3	k_1 [kcal mol^{-1} rad^{-2}]	k_2 [kcal mol^{-1} rad^{-2}]
C3	NA	CC	62.56	188.47
H1	C3	H1	39.18	47.5
H1	C3	NA	49.9	45.28
NA	CC	NA	73.65	191.59
NA	CC	H5	49.76	59.48
NA	CD	CC	72.91	200.0
NA	CD	H4	50.22	21.96
NA	C3	C3	65.73	77.82
CC	NA	CD	63.88	211.83
CC	NA	CC	68.94	140.09
NA	CC	CD	72.91	161.11
NA	CC	H4	50.22	210.8
CC	CD	H4	47.19	127.63
CD	NA	C3	62.56	138.34
CD	CC	H4	47.19	185.0
C3	C3	HC	46.37	65.82
H1	C3	C3	46.36	20.0
HC	C3	HC	39.43	28.06

Table S11: C_2mim^+ angle parameters. k_1 taken from Ref. 9, k_2 from FFGenOpt

Atom type 1	Atom type 2	Atom type 3	Atom type 4	n	k_1 [kcal mol^{-1}]	k_2 [kcal mol^{-1}]
X	C3	NA	X	2	0.00	0.00
X	CC	NA	X	2	0.85	0.85
X	CD	NA	X	2	0.85	0.85
X	CC	CD	X	2	2.00	2.00
X	C3	C3	X	3	0.08	0.08
CC	NA	C3	C3	1	0.01	0.01
CC	NA	C3	C3	2	0.12	0.12
CC	NA	C3	C3	3	0.16	0.16
CC	NA	C3	C3	4	0.02	0.02
CC	NA	C	C3	6	0.01	0.01

Table S12: C_2mim^+ dihedral parameters, X denotes a wildcard. k_1 taken from Ref. 9, k_2 from FFGenOpt

2.3.2 C_4mim^+



Atom type 1	Atom type 2	k_1 [kcal mol^{-1} Å^{-2}]	k_2 [kcal mol^{-1} Å^{-2}]
CR	HA	367.0	507.6
CR	NA	476.68	304.13
CT	HC	340.0	372.53
CT	CT	267.82	510.01
CW	HA	367.0	413.75
CW	CW	519.65	408.67
CW	NA	426.72	453.98
HN	NA	474.0	979.10
NA	CT	336.78	219.79

Table S13: C_4mim^+ bond parameters. k_1 taken from Ref. 11, k_2 from FFGenOpt

Atom type 1	Atom type 2	Atom type 3	k_1 [kcal mol^{-1} rad^{-2}]	k_2 [kcal mol^{-1} rad^{-2}]
CA	CA	CA	70.0	140.08
CA	CA	HA	35.0	62.22
CA	CA	NA	70.0	80.73
CA	NA	CA	70.0	24.65
CA	NA	CT	70.0	92.22
CR	NA	CW	69.95	136.73
CT	CT	CT	58.31	171.72
CT	CT	HC	37.47	45.41
CT	CT	NA	58.31	71.63
CT	NA	CW	69.95	89.09
CT	NA	CR	69.95	124.34
CW	CW	HA	34.98	70.56
CW	CW	NA	69.95	149.34
HA	CA	NA	35.0	49.55
HA	CR	NA	34.98	52.05
HA	CW	NA	34.98	28.88
HC	CT	NA	37.47	40.07
HC	CT	HC	33.0	28.78
NA	CR	NA	69.95	81.48

Table S14: C_4mim^+ angle parameters. k_1 taken from Ref. 11, k_2 from FFGenOpt

Atom type 1	Atom type 2	Atom type 3	Atom type 4	n	k_1 [kcal mol^{-1}]	k_2 [kcal mol^{-1}]
CA	NA	CT	CT	2	15.72	12.72
CA	NA	CT	CT	4	26.27	26.27
CA	NA	CT	HC	2	17.43	17.43
CR	NA	CT	CT	1	20.94	20.94
CR	NA	CT	HC	3	0.45	0.45
CT	CT	CT	CT	1	37.76	37.76
CT	CT	CT	CT	2	19.56	19.56
CT	CT	CT	CT	3	10.34	10.34
CT	CT	CT	HC	3	0.14	0.14
CW	NA	CT	HC	3	1.78	1.78
CW	NA	CT	CT	1	28.08	28.08
CW	NA	CT	CT	2	3.49	3.49
CW	NA	CT	CT	3	10.47	10.47
HC	CT	CT	HC	3	0.02	0.02
NA	CT	CT	CT	1	23.78	23.78
NA	CT	CT	CT	2	8.29	8.29
NA	CT	CT	CT	3	10.25	10.25
NA	CT	CT	HC	1	18.99	18.99
X	CA	CA	X	2	28.84	28.84
X	CW	CW	X	2	2.13	2.13
X	NA	CA	X	2	14.44	14.44
X	NA	CR	X	2	3.23	3.23
X	NA	CW	X	2	3.29	3.29

Table S15: C_4mim^+ dihedral parameters, X denotes a wildcard. k_1 taken from Ref. 11, k_2 from FFGenOpt

2.3.3 Tetrafluoroborate

	B - F [kcal mol^{-1} Å^{-2}]	F - B - F [kcal mol^{-1} rad^{-2}]
k_1	386.33	80.00
k_2	100.0	141.67

Table S16: BF_4^- force field parameters. k_1 taken from Ref. 11, k_2 from FFGenOpt

2.3.4 Trifluoromethanesulfonate

Atom type 1	Atom type 2	k_1 [kcal mol^{-1} Å^{-2}]	k_2 [kcal mol^{-1} Å^{-2}]
C3	F	363.8	397.35
C3	S6	254.0	296.58
O	S6	541.1	406.24

Table S17: OTf^- bond parameters. k_1 taken from Ref. 9, k_2 from FFGenOpt

Atom type 1	Atom type 2	Atom type 3	k_1 [kcal mol^{-1} rad^{-2}]	k_2 [kcal mol^{-1} rad^{-2}]
C3	S6	O	41.66	20.0
F	C3	F	71.26	26.81
F	C3	S6	81.22	158.37
O	S6	O	46.66	75.97

Table S18: OTf^- angle parameters. k_1 taken from Ref. 9, k_2 from FFGenOpt

Atom type 1	Atom type 2	Atom type 3	Atom type 4	n	k_1 [kcal mol^{-1}]	k_2 [kcal mol^{-1}]
F	C3	S6	O	3	0.07	0.07

Table S19: OTf^- dihedral parameters. k_1 taken from Ref. 9, k_2 from FFGenOpt

2.3.5 Acetate

Atom type 1	Atom type 2	k_1 [kcal mol^{-1} Å^{-2}]	k_2 [kcal mol^{-1} Å^{-2}]
OD2C2A	CD2O2A	527.00	50.0
CD2O2A	CD33A	110.00	439.47
CD33A	HDA3A	322.00	365.67

Table S20: OAc^- bond parameters. k_1 taken from Ref. 10, k_2 from FFGenOpt

Atom type 1	Atom type 2	Atom type 3	k_1 [kcal mol^{-1} rad^{-2}]	k_2 [kcal mol^{-1} rad^{-2}]
OD2C2A	CD2O2A	OD2C2A	83.70	54.65
OD2C2A	CD2O2A	CD33A	23.80	178.52
CD2O2A	CD33A	HDA3A	27.70	57.47
HDA3A	CD33A	HDA3A	35.50	22.0

Table S21: OAC^- angle parameters. k_1 taken from Ref. 10, k_2 from FFGenOpt

Atom type 1	Atom type 2	Atom type 3	Atom type 4	n	k_1 [kcal mol^{-1}]	k_2 [kcal mol^{-1}]
OD2C2A	CD2O2A	CD33A	HDA3A	6	0.21	0.21

Table S22: OAc^- dihedral parameters. k_1 taken from Ref. 10, k_2 from FFGenOpt

2.3.6 Dicyanamide

Atom type 1	Atom type 2	k_1 [kcal mol^{-1} Å^{-2}]	k_2 [kcal mol^{-1} Å^{-2}]
CG	NE	509.50	584.06
CG	N1	994.70	1429.27

Table S23: $N(CN)_2^-$ bond parameters. k_1 taken from Ref. 11, k_2 from FFGenOpt

Atom type 1	Atom type 2	Atom type 3	k_1 [kcal mol^{-1} rad^{-2}]	k_2 [kcal mol^{-1} rad^{-2}]
CG	NE	CG	66.00	117.11
NE	CG	N1	65.04	80.90

Table S24: $N(CN)_2^-$ angle parameters. k_1 taken from Ref. 11, k_2 from FFGenOpt

Atom type 1	Atom type 2	Atom type 3	Atom type 4	n	k_1 [kcal mol^{-1}]	k_2 [kcal mol^{-1}]
CG	NE	CG	N1	2	0.00	0.00

Table S25: $N(CN)_2^-$ dihedral parameters. k_1 taken from Ref. 11, k_2 from FFGenOpt

2.4 IR Spectra using scaled force constants

The IR spectra have been evaluated utilizing not only the force constants given in Tables S10 to S25, but also from trajectories with force constants for bond and angle terms subjected to scaling by 0.95, 0.90, 0.80, and 0.60. It should be noted, however, that in this particular methodology, the dihedral terms were not subjected to scaling.

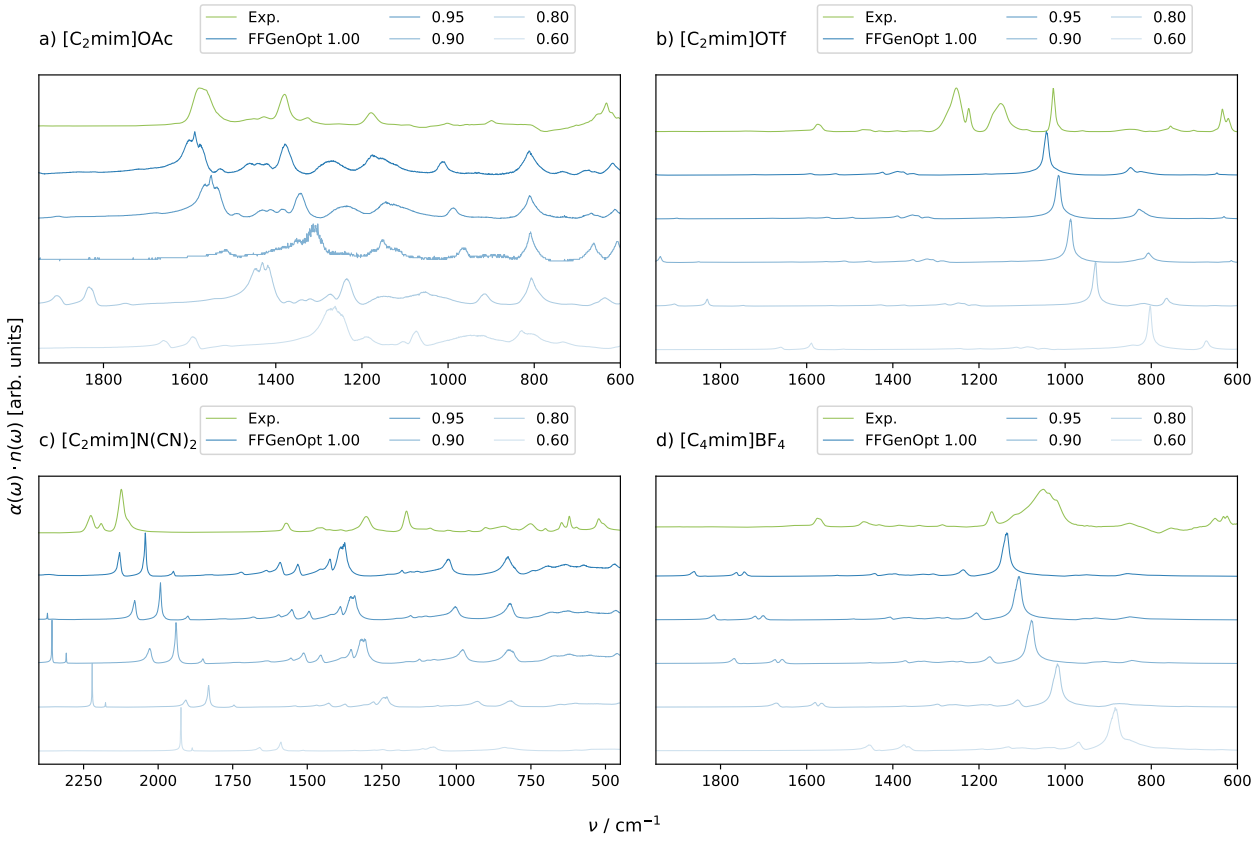


Figure S2: IR spectra calculated from polarizable MD trajectories using scaled force constants for bond and angle terms

The resulting IR spectra, as depicted in Fig. S2, demonstrate a consistent trend of progressive red-shift in correlation with decreasing scaling factors. Upon examination of $[C_2mim]OAc$ and $[C_2mim]N(CN)_2$, it becomes evident that the shapes of the resulting IR spectra undergo significant alterations upon the transition from a scaling factor of 0.80 to 0.60. This essentially establishes a boundary concerning the extent of the scaling magnitude that can be implemented.

3 Technical issues of calculating IR spectra

3.1 Numerical time-derivative versus velocities

The time derivative of the molecular dipole moment can be obtained by numerical derivation of the transient coordinates

$$\dot{\vec{\mu}}_i(t) = \sum_\beta q_{i\beta} \frac{d\vec{r}_{i\beta}(t)}{dt} \tag{2}$$

or by using the atomic velocities

$$\dot{\vec{\mu}}_i(t) = \sum_\beta q_{i\beta}\ \vec{v}_{i\beta}(t). \tag{3}$$

This also applies to the collective dipole moment $\dot{\vec{M}}_D(t) = \sum_i \dot{\vec{\mu}}_i(t)$. However, as visible in Fig. S4 both ways of calculating the derivative yield similar IR spectra. Since writing the velocity trajectories in addition to the coordinate trajectories need extra disk space and computational time, the numerical derivative method is fine.

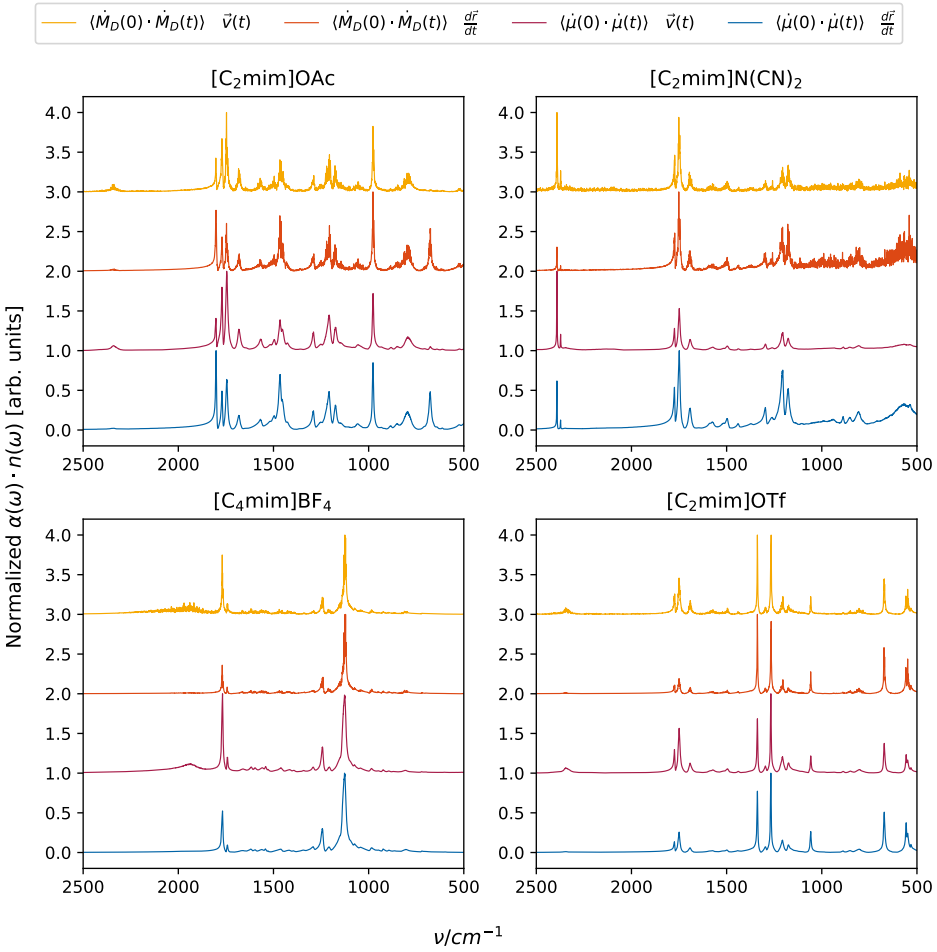


Figure S3: IR spectra calculated from polarizable MD trajectories.

3.2 The influence of SHAKE

In classical MD simulations time steps of 1 or 2 fs are used after applying the SHAKE algorithm to keep bond fixed involving hydrogens. However, the time interval between two frames in our polarizable MD simulations has to be reduced to 0.5 fs due to the Drude oscillators. As visible in Fig. S4 there is no visible difference

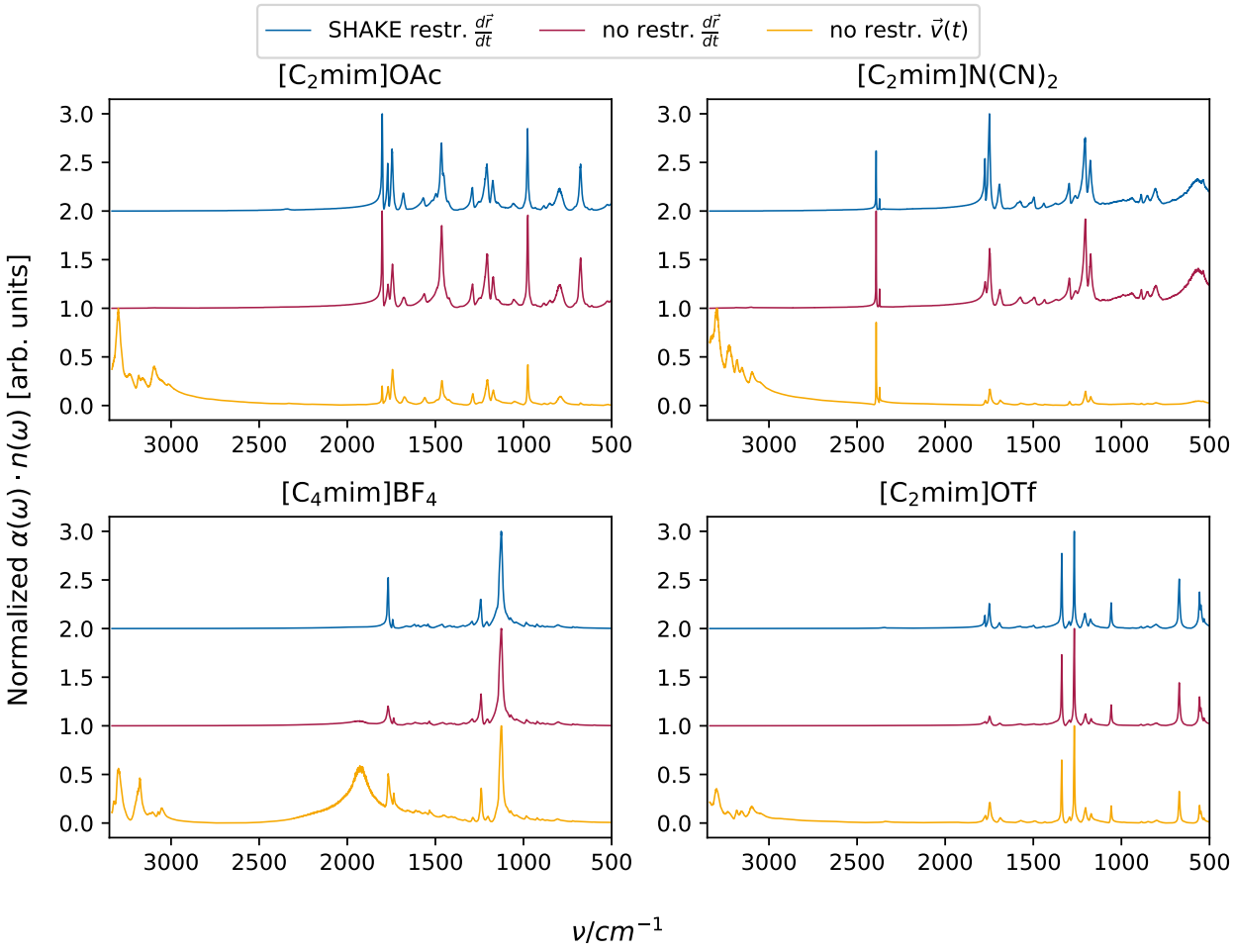


Figure S4: IR spectra calculated using the FFGenOpt force fields with and without hydrogen - heavy atom bonds frozen (SHAKE restraints)

between the IR spectra obtained from the numerical derivative of the molecular dipole moment with (blue) and without (red) SHAKE. Of course, vibrations beyond $3000\,\mathrm{cm}^{-1}$ are not detectable in MD simulations using SHAKE. These vibrations are also not visible in the red curves in the simulations without SHAKE and using the numerical derivative of the molecular dipole moments. This may be due to the fact that not all frames have been saved to disk but every tenth frame. This is common procedure for coordinate trajectories. If one is interested in the vibrations beyond $3000\,\mathrm{cm}^{-1}$, one should use the velocities written at each time step (yellow curve) to compute the corresponding IR spectra.

3.3 Machine learning potentials

Unfortunately, the agreement between IR spetra from numerical derivation and velocities is no longer given in case of machine learning potentials as visible in Fig. S5. Here, the computing of the IR spectra from the

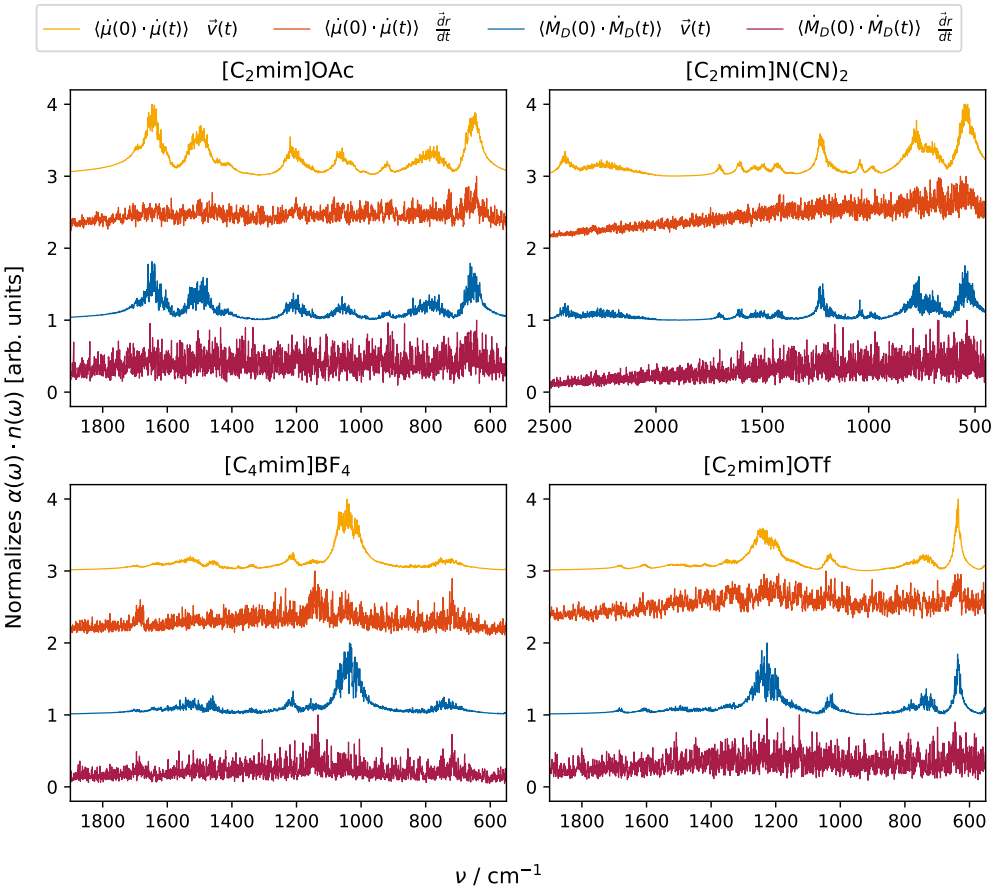


Figure S5: IR spectra calculated using machine learning potentials

numerical derivative of the molecular dipole moments fails (orange curve for $\dot{\vec{\mu}}$ and red curve for $\dot{\vec{M}}_D$). One must use the velocities which may be due to the fact that the machine learning potentials have no functional form.

3.4 Mixed MD/ML approaches

In principle, two methods exist for the mixing of MD and ML approaches:

First, using one approach for the cations and the other for the anions. We applied this in the case of $[C_4mim]BF_4$ (see Fig. S6) as boron is not available in ANI-2x.[12] Consequently, the cation was propagated by an ML potential and the forces for BF_4^- as well as the interactions between cations and anions were computed by a polarizable force field. Interestingly, the peak at $1040\,\mathrm{cm}^{-1}$ (yellow and green curve in Fig. S6) concerns the anion. Despite being modelled by a polarizable force field in the ML + FF trajectories, the agreement is significantly better than that obtained from the pure polarizable trajectories (blue curve). This emphasizes again the importance of the local environment of the molecules for computing IR spectra.

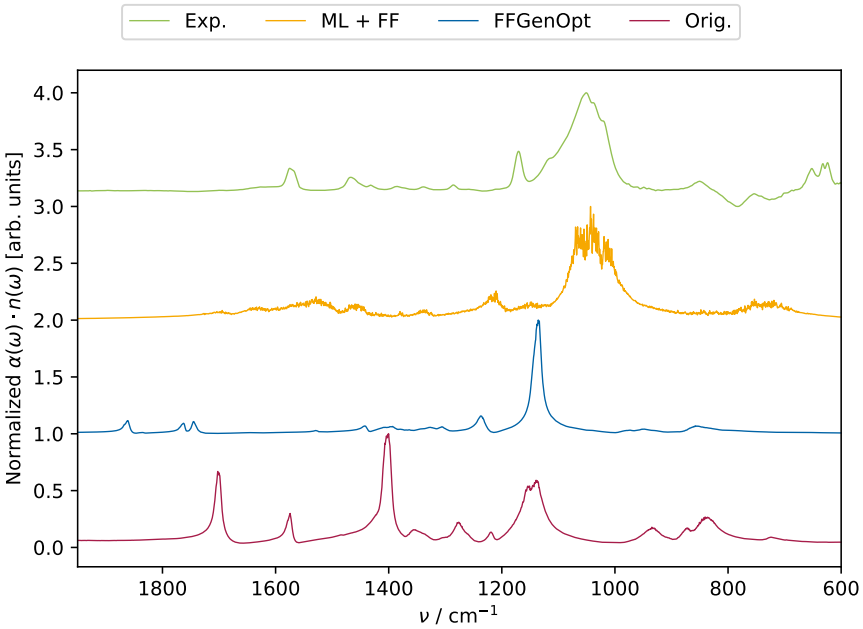


Figure S6: Comparison of IR spectra calculated by polarizable MD (red and blue lines) and a mixed setup using the ML potential for the cation and the MD potential for the anion (yellow)

The second approach to mix MD and ML is a dual topology approach. Here, the molecule is described by two potentials at the very same time, one from ML and one from polarizable MD. The parameter λ denotes the weighting between these two potentials. In our case, we used λ=0.5 for $[C_2mim]N(CN)_2$ and yielded reasonable agreement with the experimental spectrum (see Fig. 5c in the main article).

However, all mixed approaches suffer from the disadvantages of the ML approaches: The additional computational time and the restriction to the velocities.

References

[1] J. P. Merrick, D. Moran and L. Radom, *J. Phys. Chem. A*, 2007, **111**, 11683–11700.

[2] `https://cccbdb.nist.gov/vsfx.asp`.

[3] A. C. Vaiana, Z. Cournia, I. B. Costecsu and J. C. Smith, *Comput. Phys. Commun.*, 2005, **167**, 34–42.

[4] H. W. Kuhn, *Naval research logistics quartely*, 1955, **2**, 83–97.

[5] I. Rechenberg, *Simulationsmehtoden in der Medizin und Biologie*, Springer, Heidelberg, 1975.

[6] J. H. Holland, *Adaption in Natural and Artificial Systems*, University of Michigan Press, Ann Arbor, MI, second edition edn, 1992.

[7] M. Affenzeller, S. Winkler, S. Wagner and A. Beham, *Genetic Algorithms and Genetic Programming: Modern Concepts and Practical Applications*, Chapman and Hall, 2009.

[8] E. Wirsansky, *Hands-On Genetic Algorithms with Python: Applying genetic algorithms to solve real-world deep learning and artificial intelligence problems*, Packt Publishing, 2020.

[9] A. Szabadi, P. Honegger, F. Schöfbeck, M. Sappl, E. Heid, O. Steinhauser and C. Schröder, *Phys. Chem. Chem. Phys.*, 2022, **24**, 15776–15790.

[10] P. Chatterjee, E. Heid, C. Schröder and A. D. MacKerell, *Biophys. J.*, 2019, **116**, 142a.

[11] A. Szabadi, R. Elfgen, R. Macchieraldo, F. L. Kearns, H. Lee Woodcock, B. Kirchner and C. Schröder, *J. Mol. Liq.*, 2021, **337**, 116521.

[12] C. Devereux, J. S. Smith, K. K. Huddleston, K. Barros, R. Zubatyuk, O. Isayev and A. E. Roitberg, *J. Chem. Theory Comput.*, 2020, **16**, 4192–4202.