

Supporting Information

How to Select Ionic Liquids as Extracting Agent Systematically? Special Case Study for Extractive Denitrification Process

Shurong Gao^{a,b,c,*}, Jiabin Jin^{a,b}, Masroor Abro^c, Ruozhen Song^c, Miao He^d, Xiaochun Chen^{c,*}

^a State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources, North China Electric Power University, Beijing, 102206, China

^b Research Center of Engineering Thermophysics, North China Electric Power University, Beijing, 102206, China

^c Beijing Key Laboratory of Membrane Science and Technology & College of Chemical Engineering, Beijing University of Chemical Technology, Beijing 100029, PR China

^d Office of Laboratory Safety Administration, Beijing University of Technology, Beijing 100124, China

* Corresponding author, Tel./Fax: +86-10-6443-3570, E-mail: gaoshurong@ncepu.edu.cn, chenxc@mail.buct.edu.cn

COSMO-RS Computation

COSMO*therm* allows for simple and efficient processing of large numbers of compounds, i.e., a database of molecular COSMO files; e.g. the COSMO*base* database. COSMO*base* is a database of molecular COSMO files available from COSMOlogic GmbH & Co KG. Currently COSMO*base* consists of over 2000 compounds including a large number of industrial solvents plus a wide variety of common organic compounds. All compounds in COSMO*base* are indexed by their Chemical Abstracts / Registry Number (CAS/RN), by a trivial name and additionally by their sum formula and molecular weight, allowing a simple identification of the compounds. We obtained the anions and cations of different ILs and the molecular structure of typical N-compounds directly from the COSMO*base* database in this manuscript.

Different conformers of one compound have to be given in a so-called “Conformer Block” denoted by square brackets. All molecules found inside the square brackets are assigned as conformers of one compound. If no compound name is given (via the *comp=name* command), the name of the first molecule in the “Conformer Block” will be used. The square brackets have to be in the same lines as the compound/molecule input lines. The conformers will be weighted internally by COSMO*therm* using their COSMO-energy and chemical potential. In addition, a conformer weight prefactor accounting for degeneracy or symmetry of conformer distributions is taken into account: the conformer weight prefactor accounts for cases where it is possible to form a conformer in several different ways. The conformer weight prefactor is determined automatically via comparison of the symmetry properties of all conformer geometries found in a conformer block. Alternatively, it is possible to give a conformer weight factor in the input using the *wc=value* command, overriding the default derived from symmetry considerations. The gas phase energy conformers can be weighted independently using the *wg=value* command. It is possible to use zero as argument of the *wc* or *wg* command, which means that the *cosmo*- or gas phase energy conformer of the compound input line where the *wc* or *wg* command is given, will not be used in the conformer equilibrium.

As an example of an explicit choice of the conformer weight prefactor, the input lines for a mixture of the two compounds glycerol (composed of three conformers, with zero, one and two internal hydrogen bonds) and water are shown in Example 1.

Example 1:

```

ctd=BP_TZVP_C30_1301.ctd # 1st Global command line
wcmn wconf # 2nd Global command line
!! Glycerol conformer equilibrium in Water !! # Comment line
f = h2o.cosmo comp=Water # Water input (no conformers)
f = glycerol0.cosmo comp=Glycerol [ # Glycerol 1st conformer
f = glycerol1.cosmo wc=2 # Glycerol 2nd conformer
f = glycerol2.cosmo ] # Glycerol 3rd conformer

```

The first compound input line of the example contains the file input for the first compound “H₂O”, which is renamed by the *comp=Water* command and will be used to identify this compound in the output file. The following line identifies the start of a conformer block via the ‘*[*’ command. In addition, the *.cosmo* file for the first conformer and the name “Glycerol” is given. The next line contains the second conformer of this compound, which is weighted by a degeneracy factor of two via the *wc=2* command. The next line holds the third conformer and the “end of conformer block” identifier ‘*]*’ .

As an alternative and/or supplement to the input of conformers with the conformer block concept, it is possible to use all COSMO files that are found in the directory as specified by the *f_{dir}* command as conformers with the global or local compound line command *autoc*. If *autoc* is given either in the global command line or in a local compound input line of the input file, *COSMOtherm* will automatically search for conformer *.cosmo*, *.ccf* or *.mcos* files of a given COSMO filename in the *f_{dir}* directory and use these COSMO files as conformers of the given compound. To be able to do this, it is necessary that the COSMO filenames must follow the name convention of conformer COSMO files in *COSMObase*, i.e., for a given “base”-filename *name.cosmo*, the conformer COSMO files have to be named by the “base”-filename with subsequent numbers starting with zero:

```

“
name0.cosmo
name1.cosmo
...
name9.cosmo
”

```

In addition the *autoc* command also recognizes files for the alternative the name convention of conformer COSMO files in COSMObase, which claims that for a given “base”-filename *name.cosmo*, the conformer COSMO files have to be named by the “base”-filename with a “_c” conformer identifier followed by subsequent numbers starting with zero:

```
“  
name_c0.cosmo  
name_c1.cosmo  
...  
name_c9.cosmo  
”
```

By default, the *autoc* command urges COSMOtherm to use all conformers of numbers 0 to 9 plus the given “base”-filename COSMO file if they are found in the given COSMO file directory. The number of conformers to be used automatically with the autoc command can be restricted via the global or local compound line supoption *usec = {i₁ i₂ ...}*. If given, only use the conformer COSMO files with numbers *i₁, i₂, ...* will be used as conformers in the *autoc* option (*i₁, i₂, ...* are integer numbers between 0 and 9).

We obtained the anions and cations of different ILs and the molecular structure of typical N-compounds directly from the COSMObase database in this manuscript. The structures in the COSMObase is available for the following quantum chemical methods and basis sets: Turbomole BP-TZVP, Turbomole BP-SVP-AM1.

One aspect of the easy processing of large datasets is the use of the auxiliary program CT_CREATE that is shipped with COSMOtherm. CT_CREATE is a simple tool that is able to automatically create COSMOtherm input files over a dataset of COSMO files, run these COSMOtherm jobs and scan their output for the required data. For further information, cf. the CT_CREATE user’s manual: Eckert, F., CT_CREATE Users Manual, Version C3.0 Release 13.01, COSMOlogic GmbH & Co. KG (2013).

All COSMOtherm parameter files (CTDATA-files name.ctd) are identified by the quantum chemical calculation method they were parameterized for plus eventually the basis set of the quantum chemical calculation. Additionally, they are denoted by the program version and release number of the COSMOtherm version they are shipped with. Thus any CTDATA-file is identified

METHOD_BASIS_VERSION_RELEASE.ctd. For example, the CTDATA-file for Turbomole COSMO files with BP functional and Ahlrichs-SVP basis set shipped with COSMO*therm* Version C3.0 Release 13.01, is denoted BP_SVP_C30_1301.ctd. The parameterizations based on single point DFT calculations upon semiempirically optimized geometries are denoted by the additional shorthand index _AM1 (e.g. BP_SVP_AM1_C30_1301.ctd).

Parameterizations shipped with COSMO*therm* Version C3.0 Release 13.01:

1. **BP_TZVP_C30_1301.ctd**: Use with quantum chemical COSMO calculations (Turbomole, GAMESS, PQS, Molpro, Columbus, ORCA, or Q-Chem (RI-)DFT with BP-functional and def-TZVP basis set or Gaussian03/09 DGA1-DFT with BP86 functional and Ahlrichs-TZVP basis set) based on fully optimized geometries.
2. **BP_SVP_AM1_C30_1301.ctd**: Use with quantum chemical single point COSMO calculations (Turbomole (RI-)DFT with BP-functional and def-SVP basis set or Gaussian03/09 DGA1-DFT with BP86 functional and Ahlrichs-SVP basis set) based on geometries optimized by MOPAC/AM1/COSMO.
3. **BP_TZVPD_FINE_HB2012_C30_1301.ctd**: New quantum chemical level “**BP-TZVPD-FINE**” parameter set with a novel Hydrogen Bond interaction term (**HB2012** term). This parameter set is delivered in an “as is” state for testing purposes. Use with quantum chemical COSMO calculations based on Turbomole BP86 functional and def2-TZVPD basis set single point calculations with the novel fine grid marching tetrahedron cavity (**FINE**) COSMO, based upon geometries optimized with BP86 functional and TZVP basis set.

DMOL3_PBE_C30_1301.ctd: Use with quantum chemical COSMO calculations (DMOL3 PBE functional,DNP basis set) on fully optimized geometries.

COSMO-RS is based upon the information that is evaluated by quantum chemical methods. The quantum chemical basis of COSMO-RS is COSMO, which supposes that molecules are surrounded by a virtual conductor environment and the interactions take place on segments of this perfect/ideal conductor interface taking into account the electrostatic screening and the back-polarization of the solute molecule. The discrete surface around the solute molecule and each segment are characterized by their geometry and screening charge density (σ). The complete description of the molecule is

achieved by a distribution function designed by σ -profile, i.e., $P_s(\sigma)$, which describes the composition of the surface segment ensemble with respect to the interactions. The σ -profile of the whole system/mixture $P_s(\sigma)$, a sum of the σ -profiles of the components X_i in a solvent S , and the σ -potential $\mu_s(\sigma)$ are calculated by the following equations:

$$E_{MF}(\sigma, \sigma') = a_{eff} \frac{\alpha'}{2} (\sigma + \sigma')^2 \quad (1)$$

$$E_{HB}(\sigma, \sigma') = a_{eff} c_{HB} \min(0; \min(0; \sigma_{donor} + \sigma_{HB}) \times \max(0; \sigma_{acceptor} - \sigma_{HB})) \quad (2)$$

$$E_{vdW}(\sigma, \sigma') = a_{eff} (\tau_{vdW} + \tau'_{vdW}) \quad (3)$$

$$P_s(\sigma) = \sum_{i \in S} x_i P_i(\sigma) \quad (4)$$

$$e(\sigma, \sigma') = (E_{vdW}(\sigma, \sigma') + E_{HB}(\sigma, \sigma') + E_{MF}(\sigma, \sigma')) / a_{eff} \quad (5)$$

$$\mu_s(\sigma) = -\frac{RT}{a_{eff}} \ln \left[\int p_s(\sigma') \exp\left(\frac{a_{eff}}{RT} (\mu_s(\sigma') - e(\sigma, \sigma'))\right) d\sigma' \right] \quad (6)$$

where a_{eff} is the effective contact area between two surface segments; α' is an interaction parameter; c_{HB} and σ_{HB} are adjustable parameters; τ_{vdW} and τ'_{vdW} are element specific adjustable parameters; R is the gas constant; and T is temperature. Equations (1), (2) and (3) define the COSMO-RS energies, i.e., electrostatic misfit energy, hydrogen-bonding energy, van der Waals energy, respectively.

In this work, the solubility of IL in shale oil is calculated by Eq. (7),

$$\log_{10}(\chi_{IL}^{sol}) = [\mu_{IL}^{(P)} - \mu_{IL}^{(Oil)} - \max(0, \Delta G_{fus})] / (RT \ln(10)) \quad (7)$$

where $\mu_{IL}^{(Oil)}$ and $\mu_{IL}^{(P)}$ denote the chemical potential of the IL in shale oil and pure solute, respectively;

ΔG_{fus} stands for Gibbs free energy of fusion.

Based on the chemical potential thermodynamic properties, i.e., the σ -potential $\mu_s(\sigma)$, COSMO-RS can easily simulate the selectivity of IL for N-compounds, which calculated by the following procedures.

The activity coefficient at any concentration can be calculated by the following equation.

$$\ln(\gamma_1) = (\mu_1^{\text{ILs(Solvent)}} - \mu_1^{\text{N(Pure Solute)}}) / RT \quad (8)$$

where subscript 1 represents the nitrogen content, $\mu_1^{\text{ILs(Solvent)}}$ and $\mu_1^{\text{N(Pure Solute)}}$ are chemical potential for the N-compounds such as quinoline, pyridine, and indole in the solvent (ILs) and in the pure solute (N), respectively. To study the activity coefficient of molecular solutes in their infinitely diluted solution with ILs, the infinite dilution activity co-efficient (γ^∞) can be predicted by

$$\gamma_1^\infty = \lim_{x_1 \rightarrow 0} \gamma_1 \quad (9)$$

where the infinite dilution activity co-efficient (γ^∞) describes the interaction between solvent and solute at very low solute concentrations which can estimate the suitability of the given ILs for the separation.

Though the difference in γ^∞ perceived for the selectivity (S^∞) of ILs and N-compounds presents in solvent represented through the inverse of the infinite dilution activity coefficient for nitrogen-containing compounds in ILs, as shown in Eq. (2), concentration of solute N-compounds is extremely low in solvent. With the help of COSMO-RS, the selectivity of IL for N-compounds is calculated by Eq. (10),

$$S_{12,\text{max}} = S_{12}^\infty = \frac{\gamma_2^\infty}{\gamma_1^\infty} \quad (10)$$

Where, γ_1^∞ is the activity coefficient of component; N-compounds (pyridine, quinoline and indole) at infinite dilution in ILs; γ_2^∞ is the activity coefficient of component; N-compounds at infinite dilution in shale oil.

It should be noted that, the information used to input in the COSMO-RS to calculate the solubility of IL in shale oil and the selectivity of IL for N-compounds is the molecular structure of anions and cations of different ILs and the molecular structure of typical N-compounds. Then COSMO-RS will output the shielding charge density of the molecular surface. The chemical potential and molecular interactions of a component are calculated employing statistical thermodynamics by

obtaining σ -profiles with the chemical potential evidence. The thermodynamic properties, for example, the liquid-liquid equilibrium, the solubility, the activity coefficient and the selectivity can be determined for different ILs systems. Through the standard quantum chemical calculations, σ -profiles of user-defined ILs can be conveniently obtained. In a word, COSMO-RS can calculate the chemical potential of each compound as well as the chemical potential of solute and solvent system by means of statistical thermodynamics.

Table S1. Abbreviations and chemical structures of the studied anions

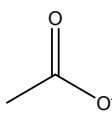
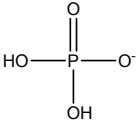
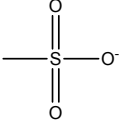
| Number | Description | Abbreviation | Chemical structure |
|--------|----------------------|-----------------------------------|---|
| 1 | bromide | [Br] | Br ⁻ |
| 2 | chloride | [Cl] | Cl ⁻ |
| 3 | acetate | [Ac] |  |
| 4 | dihydrogen phosphate | [H ₂ PO ₄] |  |
| 5 | methane sulfonate | [MeSO ₃] |  |

Table S2. Abbreviations and chemical structures of the studied cations

| | Description | Abbreviation | Chemical structure |
|----|----------------------------------|--------------|--------------------|
| 1 | 1,3-dimethyl-imidazolium | [C1mim] | |
| 2 | 1-ethyl-3-methyl-imidazolium | [C2mim] | |
| 3 | 1-propyl-3-methyl-imidazolium | [C3mim] | |
| 4 | 1-butyl-3-methyl-imidazolium | [C4mim] | |
| 5 | 1-methyl-pyridinium | [C1py] | |
| 6 | 1-ethyl-pyridinium | [C2py] | |
| 7 | 1-butyl-pyridinium | [C4py] | |
| 8 | 1-ethyl-3-methylpyridinium | [C2mpy] | |
| 9 | 1-butyl-3-methyl-pyridinium | [C4mpy] | |
| 10 | 1,1-dimethyl-pyrrolidinium | [C1mpyr] | |
| 11 | 1-ethyl-1-methyl-pyrrolidinium | [C2mpyr] | |
| 12 | 1-butyl-1-methyl-pyrrolidinium | [C4mpyr] | |
| 13 | 1-butyl-2,3-dimethyl-imidazolium | [C4mmim] | |
| 14 | 1-ethyl-2,3-dimethyl-imidazolium | [C2mmim] | |

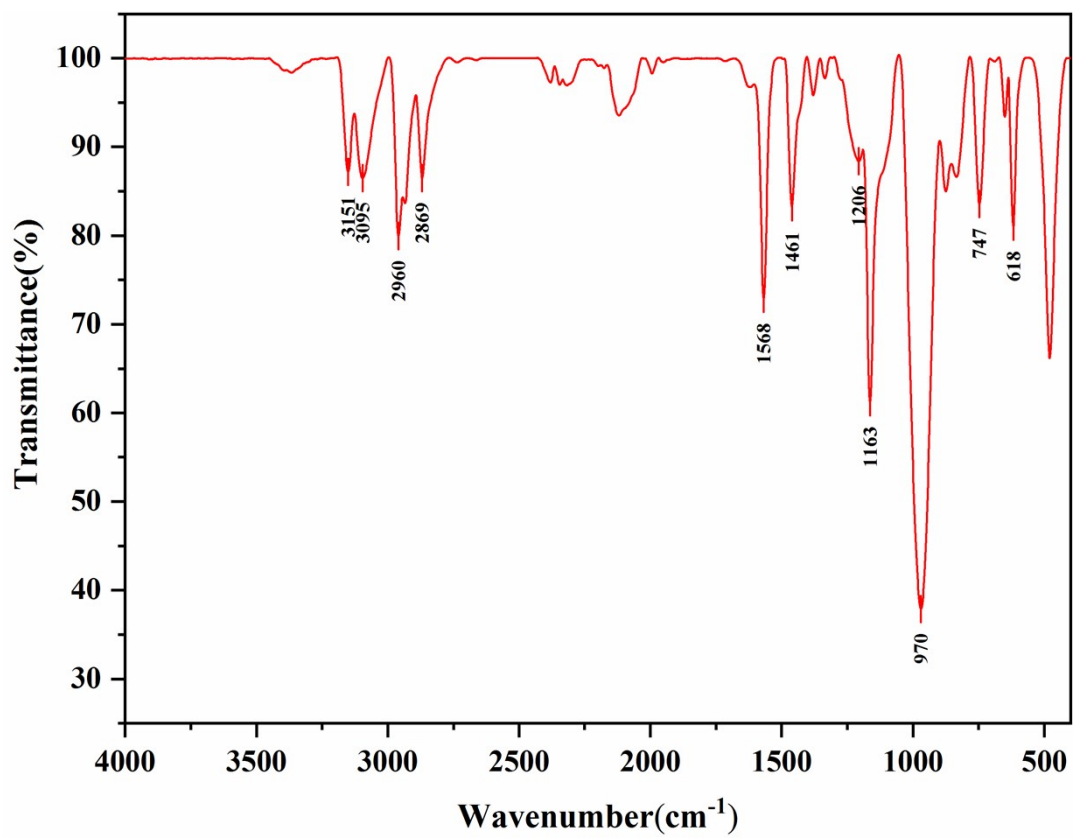


Fig. S1 FT-IR spectrum of [C₄mim]H₂PO₄

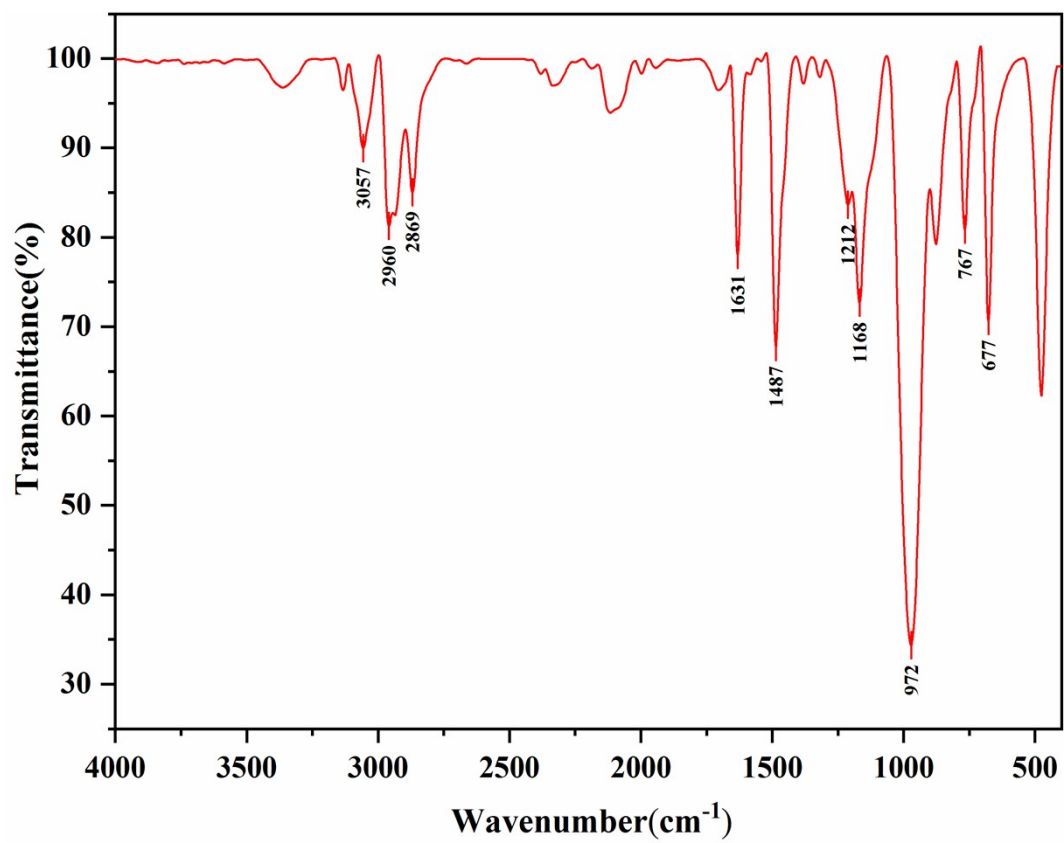


Fig. S2 FT-IR spectrum of [C₄py]H₂PO₄

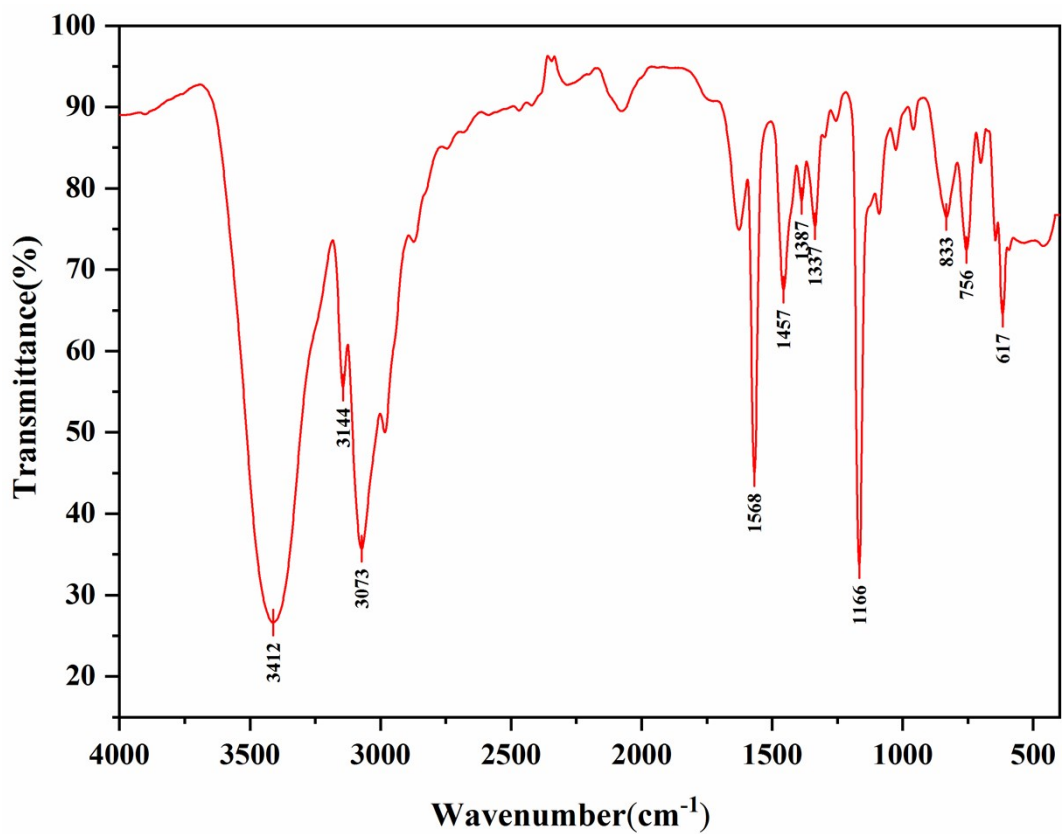


Fig. S3 FT-IR spectrum of [C₂mim]Ac

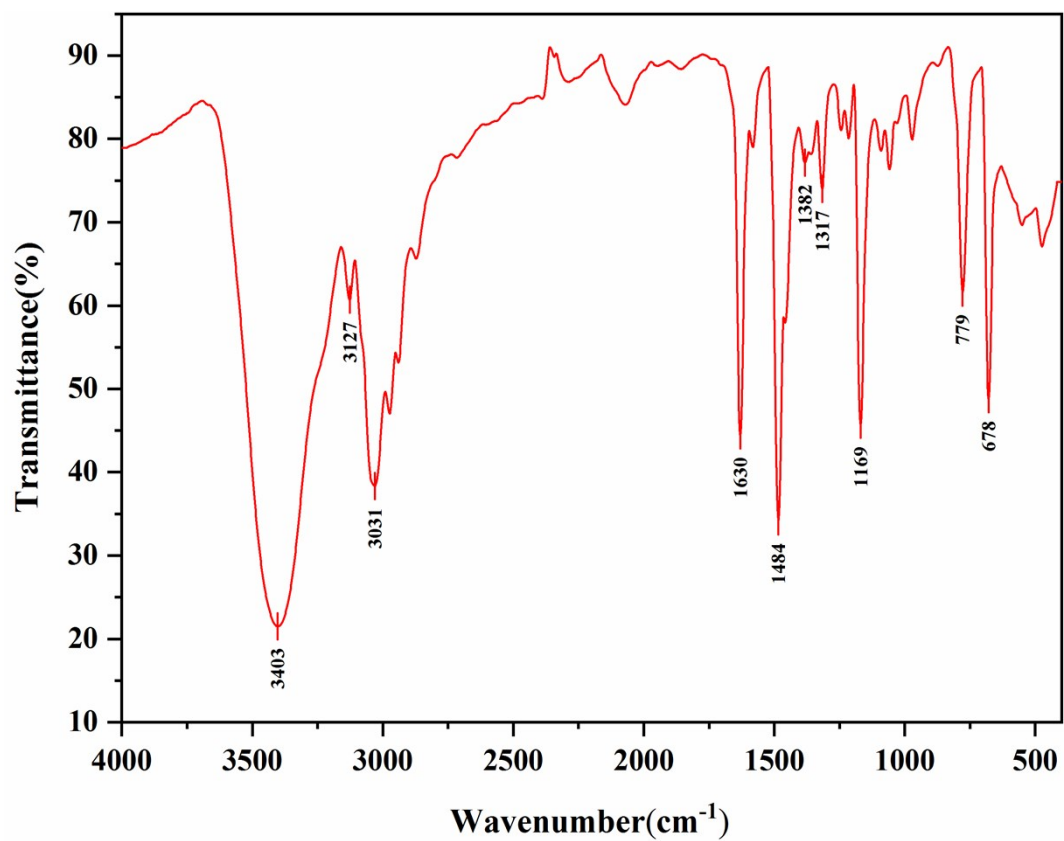


Fig. S4 FT-IR spectrum of [C₂py]Ac