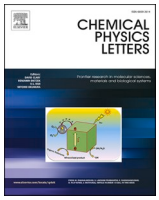
Chemical Physics Letters 860 (2025) 141813

Contents lists available at ScienceDirect

Chemical Physics Letters

journal homepage: www.elsevier.com/locate/cplett



Research paper

Uniconf: An alternative conformer generator with broad applicability

Yury Minenkov *

N. N. Semenov Federal Research Center for Chemical Physics RAS, Kosygina Street 4, 119991 Moscow, Russian Federation

ARTICLE INFO

Keywords:
Conformational search/sampling
Conformer generator
Transition metal compounds
Peptides
Microsolvated clusters

ABSTRACT

The Uniconf program aimed at conformational search/sampling in molecular clusters and isolated molecules with rotatable bonds has been developed. Unlike the other approaches, the spatial structural diversity of generated conformations is prioritized over targeting a semiempirical method or a force field minimum energy structures. The code is tested in generating of the gas-phase conformers for transition metal complexes, biologically active oligopeptides, and sodium ion microsolvated clusters. In one third of cases, either more or equally stable conformers compared to the previously found ones were generated. Energetically more diverse structures than available in the literature were obtained for practically all compounds.

1. Introduction

The recent years have seen many attempts to speed up tedious conformational search through the development of universal conformer generators, i.e. applicable to the systems consisting of any number of separate fragments and bearing various functional groups with/without rotatable bonds. Among a number of specific conformational search routines,[1–5] the literature analysis reinforced with our own expertise [6–11] suggests the two *popular* software solutions that are universal, easy-to-use and computationally inexpensive.

The ABCluster[12,13] builds up the conformers through the application of the artificial bee colony global optimization algorithm to minimize selected potential energy functions. The latter are either internal force fields or quantum chemical approximations called from external software codes. The conformer – rotamer sampling tool called CREST[14] is another powerful computer program to explore the conformational space of isolated molecules and clusters. It is based on the metadynamic trajectories using the GFNn-xTB[15] semiempirical Hamiltonian or GFN-FF[16] force field.

Despite both programs have earned recognition and are actively used for the generation of microsolvated structures, we expect that the computational chemistry community might gain from yet another one universal conformer generator. The more conformer generators are used on particular systems, the more chances all relevant stable structures are identified.

In this work, we introduce our conformer generator named Uniconf. In contrast to other approaches, we prioritize the spatial structural diversity of the conformations over targeting a semiempirical method or a force field minimum energy structures. Anyway, the Gibbs free energy surface of a molecule in a solvent quite often differs from the potential energy surface of an isolated molecule in a vacuum. As a pilot test of the Uniconf program, it has been employed for conformer generation of: a) 20 transition metal compounds from TMCONF40 database;[17] b) 29 model bioactive oligopeptides from PEPCONF database;[18] c) 30 singly charged gas-phase sodium clusters with protic and aprotic (conformationally flexible) solvents.[7] The obtained conformers along with their relative conformational energies are merged into a single database. The database named **MBS-Conf** (metals, biomolecules, solvent clusters) is expected to be promising for the development of computationally efficient methods for the molecular modeling.

2. Theory

2.1. Conformer generation with the Uniconf program

The Uniconf computer program is written in the C programming language[19] and available in the form of the 64 bit Linux binary, see the Data and Software Availability Statement. The code workflow is detailed below.

2.1.1. Starting configuration and connectivity issues

To initiate the conformer generation process, the following essential information characterizing the system has to be provided: a) Cartesian coordinates of all atoms comprising the system; b) discrete atomic charges; c) all rotatable bonds with torsion angle increment (step), its maximum value and allowed variation in degrees.

* Corresponding author.
E-mail address: Yury.Minenkov@chph.ras.ru.

https://doi.org/10.1016/j.cplett.2024.141813
Received 30 July 2024; Received in revised form 2 December 2024; Accepted 5 December 2024
Available online 7 December 2024

Upon start, the program deals with molecular connectivity describing the way atoms are connected together. By default, the connectivity is assigned automatically based on atomic covalent radii.[20] A pair of atoms is considered to be bonded if the corresponding interatomic distance is less than the sum of respective covalent radii plus 0.45 Å. Afterward, connectivity for each atom is additionally checked for the maximum valence not exceeded or the absence of angles between bonds less than 50 degrees. If these tests are not passed, the longest bond for the atom is removed.

2.1.2. On-the-fly conformer generation

In the conformer generation procedure initial structure is used as a reference. Based on the connectivity analysis, the N_{FRG} independent fragments are identified opening a definition of independent variables to set up a particular conformer. For each fragment, the center of mass (COM) position ($\vec{R}_{com}$) as well as three principal axes of rotation are computed.

Then, an arbitrary conformer is formed by:

a) setting a random displacement vector of each individual fragment as a rigid body with respect to the center of mass of the whole molecular cluster ($\vec{R}_{COM}$). To preserve the formation of excessively expanded structures the maximum distance between $\vec{R}_{COM}$ and $\vec{R}_{com}$ is allowed to be twice as large as the largest $\vec{R}_{COM} - \vec{R}_{com}$ distance in the initial configuration. For the sake of simplicity these three variables are treated in a spherical coordinate system. Alternatively, the position of any atom (e.g. metal ion) can be utilized as a center of molecular cluster instead of $\vec{R}_{COM}$;

b) prescribing random rotation angles around the three principal rotational axes of each individual fragment;

c) specifying random torsion values for all rotatable bonds (N_{ROT}).

Thus, every conformer is characterized by $6N_{FRG} + N_{ROT}$ variables and their random values, see Fig. 1 for details. The number of conformers (N_{CONF}) to be generated should be provided for molecular clusters. The whole conformer generation procedure is parallel with respect to N_{CONF}. The OpenMP single-node interface is employed to take advantage of the parallel architecture.

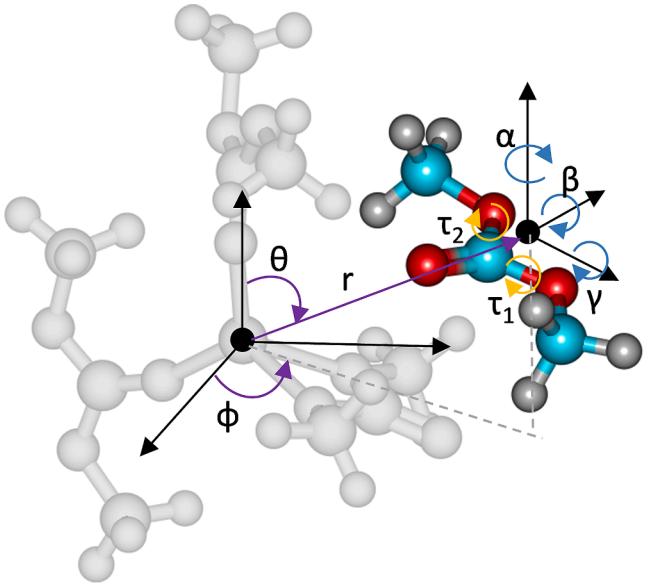


Fig. 1. Illustration of the spatial variables for a separate dimethyl carbonate (DMC) unit within the $Li^+(DMC)_4$ cluster. Translational (**3**) variables (in spherical coordinates) with respect to the whole conformer center of mass are depicted with 1 purple arrow (***r***) and 2 purple curved arrows (φ, θ). Rotational variables (**3**) along three principal axes are sown with blue curved arrows (α, β, γ). Torsion variables (**2**) characterizing rotation along the two C – O bonds are decorated with orange curved arrows (τ_1, τ_2). Atom color coding of the DMC fragment: carbon (turquoise), oxygen (red), hydrogen (gray). The rest of the cluster is shown in gray.

2.1.3. Avoiding molecular clashes through the energy minimization

To get rid of molecular clashes, each generated in Section 2.1.2 conformer is subjected to the energy minimization procedure. The optimization is done via the non-gradient local and global algorithms available via the NLopt library.[21] A set of ($6N_{FRG} + N_{rot}$) generated variables along with the Cartesian coordinates of the initial structure is passed to the potential energy calculation function.

The energy calculation proceeds as follows. First, the Cartesian coordinates of a conformer are generated by the consecutive application of $3N_{FRG}$ translation variables, $3N_{FRG}$ rotational variables and N_{ROT} bond rotational variables to the initial structure. Second, the electrostatic energy (E_{elst}) is calculated for each pair of atoms *ij* and then summed up according to eqn. (1):

$$E_{elst} = \sum_{i=1}^{N_{at}} \sum_{j>i}^{N_{at}} \frac{k_{elst} Z_i Z_j}{r_{ij}} \quad (1)$$

In eqn. (1) k_{elst} is the Coulomb electrostatic constant of 332.0538 kcal (mol Å)$^{-1}$, Z_i and Z_j are the atomic charges in atomic units, and r_{ij} is the interatomic distance in Å.

Third, for each atom pair that does not bear atoms that are either directly bonded or bonded to the same atom, the non-bonded van der Waals energy is calculated and then summed up via eqn. (2):

$$E_{vdW} = \sum_{i=1}^{N_{at}} \sum_{j>i}^{N_{at}} D_{ij} \left\{ -2\left[\frac{x_{ij}}{r_{ij}}\right]^6 + \left[\frac{x_{ij}}{r_{ij}}\right]^{12} \right\} \quad (2)$$

In eqn. (2) *i* and *j* are the atoms that are neither directly bonded nor bonded to the same atom; D_{ij} is the well depth in kcal/mol, x_{ij} is the van der Waals bond length in Å, and r_{ij} is the interatomic distance. The D_{ij} and x_{ij} are obtained as the square roots of multiplication of the individual D_i and D_j, and x_i and x_j atomic values listed in Table 1 in ref.[22] The only exceptions from eqn. (2) are the pairs forming the hydrogen bonds. These are treated according to modified eqn. (2) as proposed in ref,[23] see Section 1.1 in the Supporting Information (SI).

2.1.4. Achieving diversity and reducing conformational space through the clustering

After the initial structure generation detailed in Section 2.1.3, several steps are taken to reduce the conformational space. First, each preliminary optimized structure is subjected to a shortest non-bonded distance test. If a non-bonded distance shorter than the specified threshold (thld) is detected, the structure is eliminated. Second, a further decrease is achieved when all structures are sorted out in ascending order based on energy or other properties, and the first N structures are

Table 1
Relative energies of eight glycine conformers in kcal/mol.

Conformer	ΔE_{rel} (B97-3c)[a]	ΔE_{rel} ref. [37] [b]
I	0.59(p)	0(p)
II	0.01(p)	–
	0.00(n)	1.00(n)
III	2.12(p)	1.70(n)
IV	2.37(n)	1.27(n)
V	3.68(n)	2.61(n)
VI	5.05(p)	4.61(p)
VII	5.74(p)	5.76(p)
VIII	7.00(n)	5.81(n)

[a] B97-3c geometry optimization followed by subsequent harmonic frequencies computation of initially C_S symmetric (planar, p) structures has been performed. A few of these structures became non-planar, n. [b] Reference CCSD(T)/aug-cc-pVTZ//MP2/aug-cc-pVTZ conformational energies were adopted from the study of Kayi et al.[39].

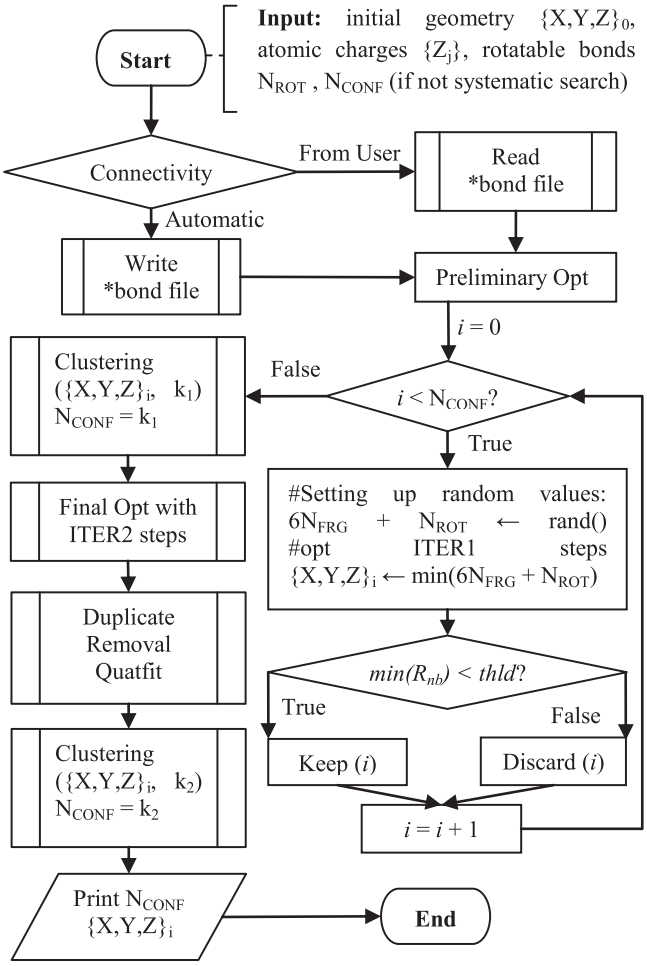
saved for further steps. As an alternative to energy, the average distance from each fragment's center of mass to a certain atom (e.g. Na^+) might be employed for sorting. It will bring more ion-centered compact structures into consideration.

Third, the final diminishing of the conformational space takes place through the grouping of similar structures into clusters. The clustering is organized via the C clustering library[24] and the k-means algorithm. To minimize the memory requirements, for each conformer that survived the previous two steps, the Cartesian coordinates are generated on the fly through the application of the $6N_{FRG}+N_{rot}$ internally stored previously optimized variables to the starting geometry. The descriptors for clustering are then calculated and stored for each conformer. The generated Cartesian coordinates are then cleaned up. As for the descriptors for clustering, we tried: a) sorted conformer principal moments of inertia (I_{XX}, I_{YY}, I_{ZZ}); b) internal UFF-like energy E grouped with (I_{XX}, I_{YY}, I_{ZZ}); c) (I_{XX}, I_{YY}, I_{ZZ}) along with sorted principal moments of inertia of each individual fragment, etc. Finally, either the closest to the centroid or the one bearing the smallest property value (e.g. E) in the cluster structure is saved for further steps in Section 2.1.5.

2.1.5. Final energy minimization and clustering of survived structures

At the final stage, all previously collected conformations are subjected to the re-optimization procedure but with a larger number of steps. All optimized structures are then scanned for possible duplicates as follows. Each structure is compared with another one first based on their energies. If their energies are within the user-defined threshold, their spatial structures are compared with the Quatfit algorithm.[25] If the RMSD difference between the structures is below a certain predefined threshold, one of them is recognized as a duplicate and discarded.

Finally, upon request another clustering is performed, the structures are sorted out and the Cartesian coordinates of the first N structures are obtained. The summary of the program workflow outlined in Section 2.1 is presented in Scheme 1.



Scheme 1. Flowchart (block-scheme) of the Uniconf spatial structure generator.

2.2. Quantum chemical calculations

The reference conformational energies and geometries of the species in the **MBS-Conf** database are based on the RIJCOSX[26] B97-3c calculations using with the def2/J Coulomb density fitting basis set[27] and performed with the ORCA 6.0.0 suite of programs.[28] The atomic charges were calculated with the PBE0[29–31] hybrid GGA DFT functional complemented by D3(BJ)[32,33] empirical dispersion correction and def2-TZVP triple-ζ polarization basis sets[34] within the CHELPG scheme.[35] More details are available in Section 1.2 in the SI.

2.3. MBS-Conf database conformer generation procedure

A few existing conformational energy databases have been employed to test the robustness of our approach and generate the reference structures of **MBS-Conf**. All details on the conformer generation procedure are available in Section 1.3 in the SI.

3. Results and Discussion

3.1. Test case: generation of planar conformers of glycine

Despite accurate determination of glycine conformers is anything but a trivial task,[36–39] we use it as a test case. As our main goal is to show the conformer generation process, we restrict ourselves only with consideration of C_S structures in which all heavy atoms lie in the same plane. By using the discrete angles of 0 and 180 degrees for C – C, C – N, and C – O rotatable bonds and starting from **Ip** C_S symmetry structure, eight glycine conformers have been produced with the Uniconf code as depicted in Fig. 2. It has to be noted, that proper C_S symmetry has been correctly preserved in all cases.

The B97-3c geometry optimization of all structures presented in Fig. 2 resulted in planar (p) conformers **Ip**, **IIp**, **IIIp**, **VIp**, and **VIIp**, and non-planar conformers **IVn, Vn** and **VIIIn**. Calculation of analytical Hessian for all but **IIp** conformer revealed no imaginary frequencies. For conformer **IIp**, an imaginary frequency of 47*i* cm^{-1} has been found. Geometry relaxation along this mode followed by geometry optimization has led to non-planar structure **IIn** which is more stable only by 0.01 kcal/mol. Table 1 compares the B97-3c relative energies of glycine conformers with their CCSD(T)/aug-cc-pVTZ//MP2/aug-cc-pVTZ counterparts adopted as references.[39] According to B97-3c protocol, the most stable conformer is **IIn**, meanwhile the planar configuration **Ip** is ca. 0.6 kcal/mol less stable. This contradicts the reference wave function theory calculations that predict **Ip** to be the most stable conformer. Another remarkable mismatch is the relative stabilities of **VIIp** and **VIIIn** conformers. According to *ab initio* approach, the relative energies of these structures are nearly the same meanwhile B97-3c composite method indicates **VIIp** to be ca. 1.3 kcal/mol more stable. Finally, on average B97-3c predicts higher conformational energies. Despite the detailed comparison of DFT and *ab initio* glycine conformational energies is out of the scope of this study, the differences obtained here and in other works even for fairly small and simple system as glycine are puzzling.

3.2. Conformer generation for transition metal complexes

To further test the performance of our conformer generator, we proceed to transition metal compounds. In total 20 species were selected

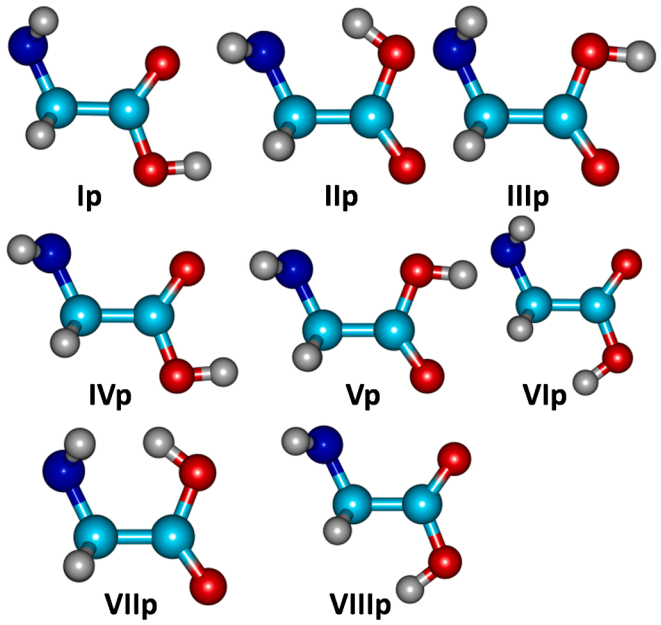


Fig. 2. Spatial structures of eight conformers of glycine in which all the heavy atoms are in the same plane. The nomenclature introduced by Császár is adopted.[36] Atom color coding: carbon (turquoise), oxygen (red), nitrogen (blue), hydrogen (gray).

from TMCONF40 dataset, see Table 2 for details on the number of rotatable bonds, molecular formulas, the number of survived conformers after the duplicate removal procedure,[40–42] and conformational energies in kcal/mol. The structural formulas are available in Fig. 1 of ref. [17].

For 9 out of 20 transition metal complexes, our approach resulted in the more stable conformer than identified in ref.[17] and for another 3 conformers equally stable structures were obtained. The less stable conformers than in ref.[17] were found for 8 complexes. In all cases, the difference between the most stable conformer obtained in the present work and in ref.[17] is within 3.5 kcal/mol, which is reasonable for such challenging systems. A systematic comparison of the most and the least stable structures obtained in this work versus ref.[17], reveals substantially larger conformational energies span for all but 3 structures. All these findings indicate the Uniconf approach as a viable alternative to the CREST[14] software utilized in ref.[17] for conformational sampling, especially when structural as well as energetic diversity is targeted. Our recent work[43] on transition metal compounds reinforces this statement as for all complexes the Uniconf approach resulted in either more or equally stable compared to the XRD conformers.

3.3. Conformer generation for model bioactive peptide molecules

Despite a few other peptide conformational energy databases are available in the literature,[44,45] in this study we focus on the PEPCONF data set. In total, 29 model biologically active oligopeptides were extracted from PEPCONF database for conformer generation with the Uniconf program. Due to immense conformational space formed by the rotation of $N_{ROT} = 15$ to $N_{ROT} = 47$ rotatable bonds, the Uniconf sampling was restricted with 150 000 randomly generated conformers. These conformers were then grouped into 35 clusters with the lowest energy structure of each cluster transferred to DFT optimization; see Section 1.3 in the Supporting Information. Only for 4 out of 29 peptides our approach revealed more stable conformers than in ref.[18] For all other systems our obtained lowest conformational energies are higher, see Table 3 for details. On average, our predicted most stable conformer is 4.3 kcal/mol higher in energy. However, in almost all cases our approach results in substantially larger conformational energy spans ($\Delta E_{MAX} - \Delta E_{MIN}$) than in ref[18] providing energetically diverse structures. On average, ($\Delta E_{MAX} - \Delta E_{MIN}$) difference obtained for the Uniconf generated conformers is 24.7 kcal/mol larger than ($\Delta E_{MAX} - \Delta E_{MIN}$) between the least/most stable conformers from ref.[18].

Less stable yet more energetically diverse oligopeptide structures reported in this work versus their ref[18] counterparts are rooted in the conformer generation procedure. In ref[18] initial gas-phase model structures were produced and then subjected to molecular dynamics conformational search by means of the Amber 16 program[46] designed *specifically* for biomolecular simulations. Further energy minimization of sampled structures with the Amber *ff14sb* force field resulted in quite realistic stable structures and their relative energies. In contrast, in our approach by setting the random values to all rotatable bond dihedrals,

Table 2
CSD codename, molecular formula, number of rotatable bonds and survived conformers, and conformational energies in kcal/mol for all structures of TMCONF40 subset studied in this work.

Molecule [a]	Formula	N_{CONF} [b]	N_{ROT} [c]	ΔE_{MIN} [d]	$\Delta E_{MAX} - \Delta E_{MIN}$ [e]	$\Delta E_{MAX} - \Delta E_{MIN}$ ref. [17] [f]
AXURER	$CdC_6N_4H_{16}S_2Cl_2$	20	8	0	12.0	4.9
AYISEG	$HgC_{21}N_3H_{13}F_{14}$	26	11	–0.9	11.3	9.6
BIDHON	$NbC_{23}NH_{41}O_3Cl_3$	16	9	–0.5	6.8	0.6
CAFKOJ	$IrSiC1_4P_3H_{38}O_{10}$	35	15	–1.6	6.7	3.6
CNETPA	$CrC1_4PN_3H_{12}O_5$	32	7	0.9	9.8	8.0
CODDIJ	$PdC_{32}P_2H_{32}O_6Cl_2$	23	10	–1.6	22.8	4.4
DEFVIT	$CrC1_2P_3H_{27}SeO_{11}$	34	12	3.0	8.5	8.7
DILQAQ	$PtC_{32}P_2N_2H_{36}O_4Cl_2$	32	14	2.3	13.8	16.6
DUGVEH	$RhSi_6C_{24}H_{66}Te_3Cl_3$	34	15	–3.5	8.0	2.1
ESOGEA	$ZnSiC_{29}N_2H_{48}$	27	7	–0.5	8.3	0.3
FODBOP	$IrC_{25}P_3H_{56}$	35	12	2.0	10.3	6.0
FONLID	$HgC_{14}H_{22}O_8$	35	14	–0.3	2.8	0.7
HIGTIA	$HfC_{48}N_4H_{40}$	14	12	0.0	3.8	3.0
LEHJAL	$PdC_{42}N_6H_{46}O_2Cl_2$	31	12	3.0	19.1	11.2
MECYUR	$AgC_{48}P_2N_3H_{41}S_2Br$	26	13	2.8	16.2	2.8
MIYBAZ	$PdC_{30}N_4H_{44}I_2$	35	12	1.2	5.8	3.7
REPFID	$CdC_{20}N_6H_{34}O_2Br_2$	35	14	3.3	5.0	9.4
SIGSUX	$TiBC_{30}N_4H_{32}F_{15}$	35	14	–2.0	7.2	3.3
TITVEX01	$AuGeC_{54}P_3H_{45}Cl_3$	9	13	–0.1	8.7	7.1
YIDHAX	$AuC_{21}NH_{25}O_3F_3$	31	8	0.0	7.1	2.7

[a] CSD code name, see Fig. 1 in ref[17] for chemical structures. [b] Number of conformers survived after the duplicate removal procedure performed for the B97-3c optimized structures. [c] Number of rotatable bonds. [d] Conformer energy relative to the most stable conformer taken from ref[17] and re-optimized according to procedure in Section 2.2. [e] Energetic difference between the most and the least stable conformer found in this work. [f] Energetic difference between the most and the least stable structures from ref[17] and re-calculated according to procedure described in Section 2.2.

Table 3
PEPCONF database codename, molecular formula, total charge, number of rotatable bonds and survived conformers, and conformational energies in kcal/mol for all structures of PEPCONF subset studied in this work.

Molecule [a]	Formula	Charge	N_{CONF} [b]	N_{ROT} [c]	ΔE_{MIN} [d]	$\Delta E_{MAX} - \Delta E_{MIN}$ [e]	$\Delta E_{MAX} - \Delta E_{MIN}$ ref. [18] [f]
GAGAGAGA	$C_{22}N_9H_{37}O_9$	0	35	24	2.3	26.4	10.2
GNNQQNY	$C_{35}N_{13}H_{51}O_{14}$	0	35	36	2.5	37.5	13.3
HHHHHH	$C_{38}N_{19}H_{47}O_7$	0	35	31	–6.3	36.6	8.4
KGD	$C_{14}N_5H_{25}O_6$	0	35	15	–1.4	50.4	5.5
KLVFFAE	$C_{45}N_9H_{67}O_{10}$	0	33	35	9.8	37.1	10.5
NFGAIL	$C_{32}N_8H_{50}O_8$	0	35	26	7.8	21.6	12.2
RGD	$C_{14}N_7H_{25}O_6$	0	34	15	1.8	37.5	4.6
RGDS	$C_{17}N_8H_{30}O_8$	0	35	20	4.8	27.6	6.6
WMNF	$C_{31}N_7H_{39}SO_6$	0	35	21	6.0	16.9	7.3
YGGFL	$C_{30}N_6H_{40}O_7$	0	35	21	6.0	22.0	10.8
YGGFM	$C_{29}N_6H_{38}SO_7$	0	35	23	7.5	20.3	9.4
CNGRC	$C_{20}N_{10}H_{37}S_2O_7$	+1	35	25	11.4	18.4	5.7
CREKA	$C_{25}N_{10}H_{47}SO_8$	+1	35	28	6.8	36.2	5.9
CRGDK	$C_{23}N_{10}H_{43}SO_8$	+1	35	27	5.5	38.8	7.6
IKVAV	$C_{27}N_7H_{52}O_6$	+1	35	23	5.2	45.3	6.5
KLVFF	$C_{37}N_7H_{56}O_6$	+1	35	26	5.4	24.5	6.9
RLGVY	$C_{30}N_9H_{50}O_7$	+1	35	25	8.4	20.3	6.2
RLNVY	$C_{32}N_{10}H_{53}O_8$	+1	35	27	1.4	18.1	2.0
VQIVYK	$C_{38}N_9H_{64}O_9$	+1	35	32	0.9	45.0	6.3
VRN	$C_{17}N_8H_{33}O_5$	+1	35	16	3.5	28.9	6.3
YIGSR	$C_{28}N_9H_{46}O_8$	+1	35	26	6.8	44.4	6.1
FRWWHR	$C_{51}N_{17}H_{67}O_7$	+2	35	34	–1.9	29.4	5.5
KLAK	$C_{23}N_7H_{47}O_5$	+2	35	22	8.8	29.2	13.1
CGKRK	$C_{25}N_{11}H_{52}SO_6$	+3	35	29	8.9	29.1	9.8
KRSR	$C_{23}N_{12}H_{49}O_6$	+3	35	26	1.2	21.9	3.9
RRWWRF	$C_{51}N_{18}H_{73}O_7$	+3	35	36	2.3	27.0	3.6
FEFEFKEK	$C_{56}N_{11}H_{76}O_{15}$	−1	35	47	10.6	64.3	8.6
REDV	$C_{22}N_8H_{37}O_9$	−1	35	23	1.0	47.7	2.6
DGEA	$C_{16}N_5H_{23}O_9$	−2	34	17	–1.1	20.8	2.7

[a] Bioactive oligopeptide code name, see Fig. 1 in ref[18] for spatial structures of individual building blocks. The following amino acid abbreviations are used: Glycine (G), Alanine (A), Leucine (L), Isoleucine (I), Valine (V), Phenylalanine (F), Tryptophan (W), Tyrosine (Y), Threonine (T), Serine (S), Cysteine (C), Methionine (M), Aspartic acid (D), Glutamic acid (E), arginine (R), Histedine (H), Lysine (K), Aspargine (N), Glutamine (Q). [b] Number of conformers survived after the duplicate removal procedure of B97-3c optimized structures. [c] Number of rotatable bonds. [d] Conformer energy relative to the most stable conformer taken from ref[18] and re-optimized according to procedure in Section 2.2. [e] Energetic difference between the most and the least stable conformer found in this work. [f] Energetic difference between the most and the least stable structures from ref[18] and re-calculated according to procedure described in Section 2.2.

we were able to produce diverse structures from completely different parts of the conformational potential energy surface otherwise not accessible in reasonable trajectory times and temperatures. We also expect that more stable structures can also be obtained with our approach if: a) more conformers are initially generated, current N_{CONF} of 150 000 is too small taking into account the huge number of rotatable bonds in these systems; b) any reasonable force field energy minimization of generated structures might help to get rid of the high-energy conformers providing only the most promising structures for subsequent DFT treatment. This is a subject for a future study.

3.4. Conformer generation for microsolvated sodium ion clusters

Finally, the ability of the Uniconf program to generate the spatial structures of molecular clusters was tested. A subset of explicitly solvated sodium ion clusters studied[7] by us previously was taken and specified in Table 4. The following solvent abbreviations are employed with the number of rotatable bonds given in parenthesis: 1,2-DCIE, 1,2-dichloroethane (1); DEA, N,N-diethylacetamide (2); DMA, dimethylacetamide (0); DMF, dimethylformamide (0); DMSO, dimethyl sulfoxide (0); EtOH, ethanol (1); FA, formamide (0); MeOH, methanol (0); NMF, N-methylformamide (2); PC, propylene carbonate (0). The number of survived conformers, and conformational energies in kcal/mol for all structures are listed in Table 4.

For 10 out of 30 clusters, our scheme resulted in more stable structures than in ref,[7] and for another 5 equally stable conformers were located. For all other species, our approach resulted in less energetically stable conformers. Particularly large (> 2 kcal/mol) relative energies of the most stable conformers identified in the present study versus their counterparts from ref.[7] were obtained for $Na^+(EtOH)_7$, $Na^+(H_2O)_8$, $Na^+(MeOH)_8$, and $Na^+(NMF)_6$. Perhaps, more than 35 DFT optimizations and/or more thoughtful choice of molecular properties for clusterization will result in more robust conformational search, but this is a subject for the future work. On a positive side, for all but 5 sodium cluster substantially broader conformational energy spans were obtained. The average (over 30Na^+ clusters) relative energy of our most stable conformer versus ref[7] is only 0.3 kcal/mol, and the average difference between $\Delta E_{MAX} - \Delta E_{MIN}$ span versus ref[7] is 7.7 kcal/mol. In addition, the Uniconf approach has been successfully used to generate the conformers of neutral non-bonded organic carbonate clusters.[47] These results indicate our approach to be a worthy competitor in the field of molecular cluster and explicit solvation environment generation along the ABCluster program[12,13], quantum cluster growth[48] and explicit solvator routine in the ORCA 6 code.

4. Conclusions

The work introduces the Uniconf computer program aimed at the conformational search/sampling in molecular clusters and isolated molecules with rotatable bonds. For initiation, the code requires Cartesian coordinates of the starting structure, atomic charges, and specified rotatable bonds. All geometric parameters of each molecular fragment apart from torsion angles associated with the rotatable bonds are fixed at their starting values. The preliminary set of conformers is constructed by ascribing the random values to the three translational and three rotational variables as well as to the rotatable bond torsions characterizing each molecular fragment. The molecular steric clashes are resolved by optimizing these parameters to minimize the value of the approximate potential energy function combining the universal force field van der Waals and electrostatic Coulomb energy terms. The

Table 4
Sodium ion clusters studied in this work with respective number of rotatable bonds and survived conformers, and conformational energies in kcal/mol.

Cluster	N_{CONF} [a]	N_{ROT} [b]	ΔE_{MIN} [c]	$\Delta E_{MAX} - \Delta E_{MIN}$ [d]	$\Delta E_{MAX} - \Delta E_{MIN}$ ref. [7] [e]
$Na^+(1,2\text{-DCIE})_4$	32	4	0.0	22.2	26.1
$Na^+(1,2\text{-DCIE})_5$	32	5	–1.4	14.9	22.3
$Na^+(1,2\text{-DCIE})_6$	35	6	–1.7	14.8	21.9
$Na^+(DEA)_4$	35	8	0.6	28.7	4.4
$Na^+(DEA)_5$	35	10	0.6	20.0	3.8
$Na^+(DMA)_5$	34	0	–0.9	16.3	0.9
$Na^+(DMA)_6$	35	0	–0.4	8.1	1.2
$Na^+(DMF)_5$	34	0	0.3	13.8	0.6
$Na^+(DMF)_6$	35	0	1.3	9.0	2.4
$Na^+(DMSO)_5$	26	0	–0.1	17.0	0.6
$Na^+(DMSO)_6$	34	0	1.2	18.7	1.5
$Na^+(EtOH)_4$	31	4	0.7	10.2	2.5
$Na^+(EtOH)_5$	33	5	0.9	10.6	2.9
$Na^+(EtOH)_6$	34	6	1.6	13.2	8.9
$Na^+(EtOH)_7$	35	7	5.1	19.5	8.2
$Na^+(EtOH)_8$	34	8	–0.8	15.4	4.9
$Na^+(FA)_6$	34	0	0.2	18.0	20.8
$Na^+(FA)_7$	34	0	1.0	22.5	5.9
$Na^+(FA)_8$	35	0	–2.0	16.8	5.0
$Na^+(H_2O)_6$	17	0	0.0	6.1	5.3
$Na^+(H_2O)_7$	23	0	–0.7	7.6	2.0
$Na^+(H_2O)_8$	32	0	3.1	11.7	7.5
$Na^+(MeOH)_5$	19	0	0.0	5.4	2.3
$Na^+(MeOH)_6$	23	0	0.0	6.2	6.3
$Na^+(MeOH)_7$	35	0	–4.5	15.7	3.3
$Na^+(MeOH)_8$	35	0	2.0	8.2	4.6
$Na^+(NMF)_5$	33	10	0.0	20.0	14.0
$Na^+(NMF)_6$	35	12	3.9	18.8	15.4
$Na^+(PC)_5$	35	0	–1.1	17.1	0.8
$Na^+(PC)_6$	35	0	0.9	16.9	5.3

[a] Number of conformers survived after the duplicate removal procedure of B97-3c optimized structures. [b] Number of rotatable bonds. [c] Conformer energy relative to the most stable conformer taken from ref.[7] and re-optimized according to procedure in Section 2.2. [d] Energetic difference between the most and the least stable conformer found in this work. [e] Energetic difference between the most and the least stable structures from ref.[7] and re-calculated according to procedure described in Section 2.2.

conformational space in then reduced drastically by clustering and taking only the centroids/minima as representative structures. In this work, the Uniconf approach has been tested to generate the conformers of 20 transition metal (TM) complexes, 29 model biologically active oligopeptides, and 30 sodium ion microsolvated clusters. More energetically diverse compared to the literature-available ones were produced for practically all compounds. In half of the cases either more or equally stable conformers were generated for TM compounds and explicitly solvated Na^+ clusters. For oligopeptides, only for 4 species more or equally stable conformers were found requiring more elaborate technique when searching for the most energetically stable structure.

Author contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Funding Sources

The work was financially supported by the Russian Science Foundation (project 24–23-00301).

CRediT authorship contribution statement

Yury Minenkov: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

Prof. Vidar R. Jensen, University of Bergen, Norway, is deeply acknowledged for bringing our attention to the conformational search problem. The author gratefully acknowledges Prof. Luigi Cavallo, King Abdullah University of Science and Technology and Dr. Edrisse Chermak, SABIC, Saudi Arabia, and Dr. Arseniy A. Otlyotov, Dr. Daniil Itkis, Prof. Lada V. Yashina, and Andrey D. Moshchenkov, N. N. Semenov Federal Research Center for Chemical Physics RAS, Moscow, Russia for many insightful discussions. The anonymous reviewers of this work are gratefully acknowledged for their valuable comments and suggestions. The work was financially supported by the Russian Science Foundation (project 24-23-00301).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cplett.2024.141813.

Data availability

The Uniconf program 64-bit Linux binary, user's manual, the MBS-Conf database (MBS-Conf.zip) and a few input files (INPXYZ.zip) are available for download free of charge at https://github.com/QuantumChemistryGroup/uniconf-bin. The ORCA 6.0.0 suite of programs has been used to perform all the density functional theory simulations. The software can be found at https://orcaforum.kofo.mpg.de/index.php.

References

[1] J.A. Rackers, Z. Wang, C. Lu, M.L. Laury, L. Lagardère, M.J. Schnieders, J.-P. Piquemal, P. Ren, J.W. Ponder, Tinker 8: Software Tools for Molecular Design, J. Chem. Theory Comput. 14 (2018) 5273–5289, https://doi.org/10.1021/acs.jctc.8b00529.
[2] M.J. Vainio, M.S. Johnson, Generating conformer ensembles using a multiobjective genetic algorithm, J. Chem. Inf. Model. 47 (2007) 2462–2474, https://doi.org/10.1021/Ci6005646.
[3] N.M. O'Boyle, T. Vandermeersch, C.J. Flynn, A.R. Maguire, G.R. Hutchison, Confab - Systematic generation of diverse low-energy conformers, J. Cheminform. 3 (2011), https://doi.org/10.1186/1758-2946-3-8.
[4] M.A. Miteva, F. Guyon, P. Tufféry, Frog2: Efficient 3D conformation ensemble generator for small compounds, Nucleic Acids Res. 38 (2010) W622–W627, https://doi.org/10.1093/nar/gkq325.
[5] J.P. Ebejer, G.M. Morris, C.M. Deane, Freely Available Conformer Generation Methods: How Good Are They? J. Chem. Inf. Model. 52 (2012) 1146–1158, https://doi.org/10.1021/ci2004658.
[6] D. Itkis, L. Cavallo, L.V. Yashina, Y. Minenkov, Ambiguities in solvation free energies from cluster-continuum quasichemical theory: lithium cation in protic and aprotic solvents, Phys. Chem. Chem. Phys. 23 (2021) 16077–16088, https://doi.org/10.1039/d1cp01454d.
[7] A.A. Otlyotov, Y. Minenkov, Conformational energies of microsolvated Na+ clusters with protic and aprotic solvents from GFNn-xTB methods, J. Comput. Chem. 43 (2022) 1856–1863, https://doi.org/10.1002/JCC.26988.
[8] A.A. Otlyotov, L. Cavallo, Y. Minenkov, Cluster-Continuum Model as a Sanity Check of Sodium Ions' Gibbs Free Energies of Transfer, Inorg. Chem. 61 (2022) 18365–18379, https://doi.org/10.1021/acs.inorgchem.2c02065.
[9] A.A. Otlyotov, A.D. Moshchenkov, L. Cavallo, Y. Minenkov, 16OSTM10: a new open-shell transition metal conformational energy database to challenge contemporary semiempirical and force field methods, Phys. Chem. Chem. Phys. 24 (2022) 17314–17322, https://doi.org/10.1039/d2cp01659a.

[10] Y. Minenkov, D.I. Sharapa, L. Cavallo, Application of Semiempirical Methods to Transition Metal Complexes: Fast Results but Hard-to-Predict Accuracy, J. Chem. Theory Comput. 14 (2018) 3428–3439, https://doi.org/10.1021/acs.jctc.8b00018.
[11] D.I. Sharapa, A. Genaev, L. Cavallo, Y. Minenkov, A Robust and Cost-Efficient Scheme for Accurate Conformational Energies of Organic Molecules, ChemPhysChem 20 (2019) 92–102, https://doi.org/10.1002/cphc.201801063.
[12] J. Zhang, M. Dolg, ABCluster: the artificial bee colony algorithm for cluster global optimization, Phys. Chem. Chem. Phys. 17 (2015) 24173–24181, https://doi.org/10.1039/C5CP04060D.
[13] J. Zhang, M. Dolg, Global optimization of clusters of rigid molecules using the artificial bee colony algorithm, Phys. Chem. Chem. Phys. 18 (2016) 3003–3010, https://doi.org/10.1039/C5CP06313B.
[14] P. Pracht, S. Grimme, C. Bannwarth, F. Bohle, S. Ehlert, G. Feldmann, J. Gorges, M. Müller, T. Neudecker, C. Plett, S. Spicher, P. Steinbach, P.A. Wesołowski, F. Zeller, CREST—A program for the exploration of low-energy molecular chemical space, J. Chem. Phys. 160 (2024), https://doi.org/10.1063/5.0197592.
[15] C. Bannwarth, E. Caldeweyher, S. Ehlert, A. Hansen, P. Pracht, J. Seibert, S. Spicher, S. Grimme, Extended tight-binding quantum chemistry methods, Wiley Interdiscip. Rev. Comput. Mol. Sci. 11 (2021) e1493.
[16] S. Spicher, S. Grimme, Robust Atomistic Modeling of Materials, Organometallic, and Biochemical Systems, Angew, Chemie 132 (2020) 15795–15803, https://doi.org/10.1002/ANGE.202004239.
[17] M. Bursch, A. Hansen, P. Pracht, J.T. Kohn, S. Grimme, Theoretical study on conformational energies of transition metal complexes, Phys. Chem. Chem. Phys. 23 (2021) 287–299, https://doi.org/10.1039/D0CP04696E.
[18] V.K. Prasad, A. Otero-De-la-Roza, G.A. Dilabio, Data descriptor: Pepconf, a diverse data set of peptide conformational energies, Sci. Data 6 (2019) 1–9, https://doi.org/10.1038/sdata.2018.310.
[19] B.W. Kernighan, D.M. Ritchie, The C Programming Language, 2nd ed.,, Prentice Hall Professional Technical Reference, 1988.
[20] B. Cordero, V. Gómez, A.E. Platero-Prats, M. Revés, J. Echeverría, E. Cremades, F. Barragán, S. Alvarez, Covalent radii revisited, J. Chem. Soc. Dalt. Trans. (2008) 2832–2838, https://doi.org/10.1039/b801115j.
[21] S.G. Johnson, The NLopt nonlinear-optimization package, http://ab-initio.mit.edu/nlopt, (2014) http://ab-initio.mit.edu/nlopt. http://ab-initio.mit.edu/nlopt.
[22] A.K. Rappe, C.J. Casewit, K.S. Colwell, W.A. Goddard, W.M. Skiff, UFF, a Full Periodic Table Force Field for Molecular Mechanics and Molecular Dynamics Simulations, J. Am. Chem. Soc. 114 (1992) 10024–10035, https://doi.org/10.1021/ja00051a040.
[23] A.J. Pertsin, A.I. Kitaigorodsky, The Atom-Atom Potential Method, Springer Berlin Heidelberg, Berlin, Heidelberg (1987), https://doi.org/10.1007/978-3-642-82712-9.
[24] M.J.L. De Hoon, S. Imoto, J. Nolan, S. Miyano, Open source clustering software, Bioinformatics 20 (2004) 1453–1454, https://doi.org/10.1093/bioinformatics/bth078.
[25] D.J. Heisterberg The Quatfit program, the CCL archive 1990.
[26] F. Neese, F. Wennmohs, A. Hansen, U. Becker, Efficient, approximate and parallel Hartree-Fock and hybrid DFT calculations. A “chain-of-spheres” algorithm for the Hartree-Fock exchange, Chem. Phys. 356 (2009) 98–109, https://doi.org/10.1016/j.chemphys.2008.10.036.
[27] F. Weigend, Accurate Coulomb-fitting basis sets for H to Rn, Phys. Chem. Chem. Phys. 8 (2006) 1057–1065, https://doi.org/10.1039/b515623h.
[28] F. Neese, Software update: The ORCA program system—Version 5.0, Wiley Interdiscip, Rev. Comput. Mol. Sci. (2022) e1606.
[29] J.P. Perdew, K. Burke, M. Ernzerhof, Generalized Gradient Approximation Made Simple, PhysRevLett. 77 (1996) 1396, https://doi.org/10.1103/PhysRevLett.77.3865.
[30] C. Adamo, V. Barone, Toward reliable adiabatic connection models free from adjustable parameters, Chem. Phys. Lett. 274 (1997) 242–250, https://doi.org/10.1016/s0009-2614(97)00651-9.
[31] C. Adamo, V. Barone, Toward reliable density functional methods without adjustable parameters: The PBE0 model, J. Chem. Phys. 110 (1999) 6158–6170, https://doi.org/10.1063/1.478522.
[32] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu, J. Chem. Phys. 132 (2010) 154104, https://doi.org/10.1063/1.3382344.
[33] S. Grimme, S. Ehrlich, L. Goerigk, Effect of the Damping Function in Dispersion Corrected Density Functional Theory, J. Comput. Chem. 32 (2011) 1456–1465, https://doi.org/10.1002/jcc.21759.
[34] F. Weigend, R. Ahlrichs, Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: design and assessment of accuracy, Phys. Chem. Chem. Phys. 7 (2005) 3297–3305, https://doi.org/10.1039/b508541a.
[35] C.M. Breneman, K.B. Wiberg, Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis, J. Comput. Chem. 11 (1990) 361–373, https://doi.org/10.1002/jcc.540110311.
[36] A.G. Császár, Conformers of Gaseous Glycine, J. Am. Chem. Soc. 114 (1992) 9568–9575, https://doi.org/10.1021/ja00050a041.
[37] H.W. Ke, L. Rao, X. Xu, Y.J. Yan, Theoretical study of glycine conformers, J. Theor. Comput. Chem. 7 (2008) 889–909, https://doi.org/10.1142/S0219633608004192.
[38] V. Barone, M. Biczysko, J. Bloino, C. Puzzarini, Glycine conformers: A never-ending story? Phys. Chem. Chem. Phys. 15 (2013) 1358–1363, https://doi.org/10.1039/c2cp43884d.
[39] H. Kayi, R.I. Kaiser, J.D. Head, A theoretical investigation of the low energy conformers of the isomers glycine and methylcarbamic acid and their role in the interstellar medium, Phys. Chem. Chem. Phys. 13 (2011) 15774–15784, https://doi.org/10.1039/c1cp20656g.
[40] M.W. Walker, L. Shao, R.A. Volz, Estimating 3-D location parameters using dual number quaternions, CVGIP Image Underst. 54 (1991) 358–367, https://doi.org/10.1016/1049-9660(91)90036-O.
[41] W. Kabsch, A solution for the best rotation to relate two sets of vectors, Acta Crystallogr. Sect. A 32 (1976) 922–923, https://doi.org/10.1107/S0567739476001873.
[42] GitHub - charnley/rmsd: Calculate Root-mean-square deviation (RMSD) of two molecules, using rotation, in xyz or pdb format, (n.d.). https://github.com/charnley/rmsd (accessed April 27, 2024).
[43] A.A. Otlyotov, T.P. Rozov, A.D. Moshchenkov, Y. Minenkov, 16TMCONF543: An Automatically Generated Data Set of Conformational Energies of Transition Metal Complexes Relevant to Catalysis, Organometallics (2024), https://doi.org/10.1021/acs.organomet.4c00246.
[44] J. Řezáč, D. Bím, O. Gutten, L. Rulíšek, Toward Accurate Conformational Energies of Smaller Peptides and Medium-Sized Macrocycles: MPCONF196 Benchmark Energy Data Set, J. Chem. Theory Comput. 14 (2018) 1254–1266, https://doi.org/10.1021/acs.jctc.7b01074.
[45] T. Kalvoda, M. Culka, L. Rulíšek, E. Andris, Exhaustive Mapping of the Conformational Space of Natural Dipeptides by the DFT-D3//COSMO-RS Method, J. Phys. Chem. B 126 (2022) 5949–5958, https://doi.org/10.1021/acs.jpcb.2c02861.
[46] R. Salomon-Ferrer, D.A. Case, R.C. Walker, An overview of the Amber biomolecular simulation package, Wiley Interdiscip. Rev. Comput. Mol. Sci. 3 (2013) 198–210, https://doi.org/10.1002/WCMS.1121.
[47] A.S. Ryzhako, A.A. Tuma, A.A. Otlyotov, Y. Minenkov, An influence of electronic structure theory method, thermodynamic and implicit solvation corrections on the organic carbonates conformational and binding energies, J. Comput. Chem. (2024), https://doi.org/10.1002/JCC.27471.
[48] S. Spicher, C. Plett, P. Pracht, A. Hansen, S. Grimme, Automated Molecular Cluster Growing for Explicit Solvation by Efficient Force Field and Tight Binding Methods, J. Chem. Theory Comput. 18 (2022) 3189, https://doi.org/10.1021/acs.jctc.2c00239.