Theses of the doctoral dissertation

András Bence Nacsa

Ab initio conformational analysis of neutral and protonated gas-phase amino acids

Supervisor:

Dr. Gábor Czakó

Associate professor

University of Szeged

Faculty of Science and Informatics

Department of Physical Chemistry and Materials Science

Doctoral School of Chemistry

MTA-SZTE Lendület Computational Reaction Dynamics Research Group

Szeged, 2024

INTRODUCTION

The study of the conformers, i.e., stable minimum geometries, of amino acids is a hot topic for both experimental and theoretical researchers, as these molecules have a variety of stable forms due to their flexible chemical bonds. The individual conformers can be identified by their different vibrational and rotational properties ("fingerprints") using infrared or microwave spectroscopy measurements. To be able to do this, we need to map the conformational space of the molecules with theoretical tools and identify the arrangements corresponding to the minima on their potential energy surfaces (PES). This can be done using algorithms and/or artificial intelligence, or by the "traditional", systematic search, the latter being used in the work related to this dissertation. The idea is to generate a large number of initial structures for geometric optimizations by systematic rotations along the corresponding chemical bonds. For the smallest amino acid, glycine, the eight stable minimum conformers were determined theoretically decades ago, and five of them could be identified experimentally, but for larger amino acids such as cysteine and serine, the conclusion of the literature is not so clear. The theoretically determined geometries, relative energies and vibrational frequencies allow the identification of conformers, for example by astrochemists studying the formation, presence and reactions of molecules in interstellar space that are essential for the origin of life. One of the simplest reactions we can describe is the protonation:

$$B_{(g)} + H^+_{(g)} \rightarrow BH^+_{(g)}, \tag{1.}$$

where B is the analogue neutral base, H^+ is the free proton, while BH^+ is the protonated conjugate acid. To quantify the changes occurring during the reaction, the quantities proton affinity (PA, enthalpy) and gas-phase basicity (GB, Gibbs free energy) were introduced. Absolute values can be determined mostly theoretically, experimentally usually only relative values can be obtained by studying the kinetics, rate and thermodynamics of different protonation reactions. The PA and GB values of the amino acids influence many processes: protein fragmentation, enzyme activity, but also, of course, ion-molecule reactions in space.

After studying the available literature, we set three main objectives for the dissertation. The first is to identify the conformers of neutral and protonated glycine, cysteine and serine amino acids by a more detailed search than ever before, going beyond the Hartree−Fock method. In the case of protonated forms, it is also necessary to investigate possible protonation sites, since so far only the thermodynamically most favourable amino-protonation has been analysed, but the high-accuracy quantum chemical calculations could provide even more precise information

on the differences between the different functional groups. The next step will be to determine the first coupled-cluster level geometries and/or harmonic vibrational frequencies for the same neutral and protonated molecules, and to investigate the impact of different theoretical corrections. The final task is to determine the proton affinity and gas-phase basicity of the three amino acids at the coupled-cluster level, taking into account various corrections for different protonation sites for both global minima and conformer mixtures. Although there are still aspects of conformer search and PA/GB determination that have not been investigated, this does not detract from the value of our work. Our publications, which form the basis of this dissertation, contain many new, previously unexplored results that can be used in a variety of research projects: developing reactive or conformational PESs, validating machine learning-based algorithms, performing anharmonic vibration analysis, etc.

METHODS

To map the conformational space for both neutral and protonated amino acids, the data points were generated by a systematic method of varying the corresponding torsion angles by 60° (and 30° along certain axes for cysteine), starting from the global minima found in the literature for neutral forms, while for the protonated forms, we rotated the most stable structures obtained from the protonation site trials. For the geometry optimizations and the determination of the harmonic vibrational frequencies, we applied the MP2/aug-cc-pVDZ level of theory in the case of the protonated glycine. For the neutral cysteine, we used the MP2 method with the 3-21G, 6-31G, 6-31++G, 6-31G**, 6-31++G** and cc-pVDZ basis sets (later we used HF/3-21G, as well), while for the protonated cysteine and the two serine forms we only employed the 6-31++G** and cc-pVDZ basis sets. We reoptimized the stationary points belonging to the two larger amino acids at the MP2/aug-cc-pVDZ level of theory and determined their harmonic vibrational frequencies. To investigate the protonation sites, we utilized the aforementioned level of theory and performed geometry optimizations and vibrational analysis on different trial structures, which were generated by attaching a proton to different functional groups of the neutral amino acids in different variations. For the neutral glycine we took the CCSD(T)-F12b/aug-cc-pVTZ geometries from one of our previous work and performed single-point energy computations with the aug-cc-pVQZ basis set. In the case of cysteine and serine (both for neutral and protonated forms), due to the large number of conformers, we only investigated the 10 structures with the lowest MP2/aug-cc-pVDZ relative energies. We completed CCSD(T)-F12a/cc-pVDZ-F12 geometry optimizations (and vibrational analysis for the neutral

forms), and computed single-point energies with TZ and QZ basis sets (with CCSD(T)-F12b method as well, for the latter). At the available most accurate structures, we examined the values of the post-(T) corrections, the core correlation corrections and the second-order Douglas–Kroll relativistic corrections. I used MOLPRO (2015 version) for the quantum chemical computations, and MRCC for computing the post-CCSD(T) corrections. The large quantity of data obtained during this research was processed, evaluated and classified by Fortran and Python codes written by me.

RESULTS

All of the following thesis points are based on the three publications on which the dissertation is based, thus I do not indicate them separately.

T1. *For the neutral cysteine 85 and for the neutral serine 95 stable minimum conformer structures and harmonic vibrational frequencies have been determined at the MP2/aug-cc-pVDZ level of theory after the systematic mapping of their conformational space (and some additions).*

Since the conformers of neutral glycine are well established in the literature, including our own work on them, the task of this dissertation was to find the conformers of cysteine and serine. This was done using a systematic method of varying the corresponding torsion angles by 60°, which led to a large number of data points. The trial geometries were optimized using MP2 with different basis sets, for cysteine 6-31++G** and cc-pVDZ proved to be useful, thus for serine only these were used for the search. The structures obtained for cysteine were complemented by a critical analysis of the literature sources, and then the geometries obtained for the two amino acids were further optimized and their harmonic vibrational frequencies were determined at the MP2/aug-cc-pVDZ theoretical level. For both amino acids, we identified several new geometries, including some in the low relative energy region.

T2. *We identified the possible protonation sites of glycine, cysteine and serine amino acids (amino, carbonyl and, in the case of cysteine, the thiol group) and then also searched for the minimum geometries of the resulting forms on the corresponding potential energy surfaces. At the MP2/aug-cc-pVDZ level, we found 3 N- and 8 O-protonated glycine, 21 N-, 64 O-, and 37 S-protonated cysteine, and 15 N- and 46 O-protonated serine minima, and determined their harmonic vibrational frequencies.*

In the case of glycine, protonation of amino, carbonyl and hydroxyl groups was expected. Consistent with the literature, our studies also showed that the protonation of the hydroxyl group cannot lead to a stable form, only the protonation of the two other groups. For the two larger amino acids, chemical considerations suggest that both the thiol and hydroxyl groups of the side chain can be protonated, but only thiol-protonation leads to a stable molecular ion. Although thermodynamically the protonation of the amino group is the most favourable, we also searched for the other forms, which have been neglected in the literature. The method was similar to the one used for neutral amino acids, for protonated glycines the search was performed at the MP2/aug-cc-pVDZ theoretical level, but for protonated cysteine and serine

the orders of magnitude larger number of starting geometries did not allow this, therefore the 6-31++G** and cc-pVDZ basis sets were applied, and the potential minimum geometries were re-optimized (and their harmonic vibrational frequencies were determined) with the aug-cc-pVDZ basis. Also for these forms, we found several new geometries not reported in previous publications.

T3. *For the lowest-energy neutral and protonated glycine, cysteine and serine conformers, we determined the first explicitly-correlated coupled-cluster-level relative energies and/or geometries and/or harmonic vibrational frequencies, and found that the interrupted coupled-cluster series, the correlation of the core electrons and the second-order Douglas–Kroll relativistic correction affect the accuracy of our results on the order of tenths and hundredths of a kcal/mol.*

For neutral glycine, geometries and relative energies at the CCSD(T)-F12b/aug-cc-pVTZ level of theory were available from one of our previous work, thus in addition to the determination of the corrections, only the QZ energies were needed to be computed in this case. For the other two amino acids, the ten conformers with the deepest MP2/aug-cc-pVDZ relative energies were further investigated (both neutral and protonated cases, separately for the protonation sites) using coupled-cluster methods. Geometry optimizations were performed at the CCSD(T)-F12a/cc-pVDZ-F12 level of theory, and harmonic vibrational frequencies were also determined in the neutral cases. Single-point energies were obtained for these structures using TZ and QZ basis sets. The corrections are in the order of tenths and hundredths of kcal/mol, with the exception of a few (neutral/protonated) cysteine conformers, the relativistic correction is below hundredths of kcal/mol. Making various considerations, an estimated margin of error can be given for the computed relative energies/enthalpies.

T4. *We have determined the coupled-cluster level proton affinity and gas-phase basicity values for glycine, cysteine and serine for all possible protonation sites and, in addition to the corrections mentioned in point T3, we have also analysed the effect of considering conformer mixtures instead of global minima. Our values for the amino protonation are in good agreement with the proposals in the literature, mostly based on experimental work, while the protonation of the carbonyl group was found to be thermodynamically much more favourable than previously assumed.*

In the literature, the two quantities characterising protonation have been considered only for the thermodynamically most favourable amino-protonation. Using our most accurate, explicitly-

correlated coupled-cluster-level relative energies, we have determined these values for all functional groups (considering only global minima or conformer mixtures), showing that the energy difference between the protonation of individual sites is much lower than previously estimated (especially for the larger amino acids). The magnitude of the corrections listed in point T3 has also been investigated here and used to estimate the accuracy of our results (along with our approximations made in the determination of vibrational frequencies). Our proposed PA/GB values are 212.4±0.3/204.8±0.6 kcal/mol for glycine, 216.4±0.4/208.2±0.6 kcal/mol for cysteine, and 218.1±0.4/209.9±0.6 kcal/mol for serine.

LIST OF PUBLICATIONS (MTMT ID: 10073207)

Publications covered by the doctoral dissertation ($\sum$IF = 10.145)

A. B. Nacsa & G. Czakó: Benchmark ab initio proton affinity of glycine, *Phys. Chem. Chem. Phys.* **23**, 9663 (2021), IF = 3.945

A. B. Nacsa & G. Czakó: Benchmark ab initio determination of the conformers, proton affinities and gas-phase basicities of cysteine, *J. Phys. Chem. A* **126**, 9667 (2022), IF = 2.9

A. B. Nacsa, M. Kígyósi, G. Czakó: Protonation of serine: Conformers, proton affinites ad gas-phase basicities at the "gold standard" and beyond, *Phys. Chem. Chem. Phys.* **25**, 8891 (2023), IF = 3.3

Publication related to, but not covered by the doctoral dissertation ($\sum$IF = 3.376)

E. M. Orján, **A. B. Nacsa**, G. Czakó: Conformers of dehydrogenated glycine isomers, *J. Comput. Chem,* **41**, 2001 (2020), IF = 3.376

Publications not related to the doctoral dissertation ($\sum$IF = 11.1)

A. B. Nacsa, V. Tajti, G. Czakó: Dynamics of the $Cl^- + CH_3I$ reaction on a high-level ab initio analytical potential energy surface, *J. Chem. Phys.* **158**, 194306 (2023), IF = 4.4

T.Gstir, T. Michaelsen, B. A. Long, **A. B. Nacsa**, A. Ayasli, D. Swaraj, F. Zappa, F. Trummer, S. G. Ard, N. S. Shuman, G. Czakó, A. A. Viggiano, R.Wester: The influence of fluorination on the dynamics of the $F^- + CF_3CH_2I$ reaction, *Phys. Chem. Chem. Phys.* **25**, 18711 (2023), IF = 3.3

A. B. Nacsa, C. Tokaji, G. Czakó: High-level analytical potential-energy-surface-based dynamics of the $OH^- + CH_3CH_2Cl$ S_N2 and E2 reactions in full (24) dimensions, *Faraday Discuss.* DOI: 10.1039/D3FD00161J (2024), IF = 3.4

Sum of IF = 24.621

TALKS

High-level analytical potential-energy-surface-based dynamics of the $OH^- + CH_3CH_2Cl$ S_N2 and E2 reactions in full (24) dimensions
A. B. Nacsa, C. Tokaji, G. Czakó
New directions in molecular scattering Faraday Discussion, Edinburgh, United Kingdom, 2024

A szerin kvantumkémiai konformációs analízise és ab initio protonaffinitása
M. Kígyósi, **A. B. Nacsa**, G. Czakó
XXXVI. National Scientific Students' Associations Conference, Chemistry and Chemical Industry Section, Szeged, Hungary, 2023

Dehidrogénezett és protonált aminosavak ab initio konformeranalízise
A. B. Nacsa, E. M. Orján, G. Czakó
Hungarian Academy of Sciences: Reaction Kinetics and Photochemistry Committee, Mátrafüred, Hungary, 2022

Ab initio conformational analysis of dehydrogenated and protonated amino acids
A. B. Nacsa, E. M. Orján, G. Czakó
Hungarian Academy of Sciences: Material and Molecular Structures Committee, Mátrafüred, Hungary, 2022

A glicin protonaffinitásának nagypontosságú ab initio meghatározása
A. B. Nacsa, G. Czakó
KeMoMo-QSAR Symposium, Szeged, Hungary, 2021

POSTERS

Theoretical insights into the $Cl^- + CH_3I$ and $F^- + CF_3CH_2I$ reactions
A. B. Nacsa, V. Tajti, G. Czakó
Dynamics of Molecular Collisions, Snowbird, UT, USA, 2023

Towards the first-principles dynamics of the $Cl^- + CH_3I$ and $F^- + CF_3CH_2I$ reactions
A. B. Nacsa, V. Tajti, G. Czakó
Molecular Interactions and Dynamics Gordon Research Conference, Easton, MA, USA, 2022