

VALIDATION OF COMPUTATIONAL METHODS APPLIED IN MOLECULAR MODELING OF CAFFEINE WITH EPITHELIAL ANTICANCER ACTIVITY: THEORETICAL STUDY OF GEOMETRIC, THERMOCHEMICAL AND SPECTROMETRIC DATA**Josivan da Silva Costa^{a,b,c,*}, Cleudson Breno Rodrigues dos Santos^{a,b}, Karina da Silva Lopes Costa^b, Ryan da Silva Ramos^b, Carlos Henrique Tomich de Paula da Silva,^d and Williams Jorge da Cruz Macêdo^c**^aUniversidade Federal do Pará, Rua Augusto Corrêa, 1 - Guamá, 66075-110 Belém – PA, Brasil^bDepartamento de Ciências Biológicas, Universidade Federal do Amapá, Rod. Juscelino Kubitschek, Km 02, s/n, Jardim Marco Zero, 68902-280 Macapá – AP, Brasil^cUniversidade Federal Rural da Amazônia, Rua João Pessoa, 121, Campus Capanema Centro, 68700-030 Capanema – PA, Brasil^dEscola de Ciências Farmacêuticas de Ribeirão Preto, Universidade de São Paulo, 14040-903 Ribeirão Preto – SP, Brasil

Recebido em 29/01/2018; aceito em 12/04/2018; publicado na web em 22/05/2018

Models validation in QSAR, pharmacophore, docking, and others, can ensure the accuracy and reliability of future predictions in design and selection of molecules with biological activity. In these study, the caffeine molecule was optimized using Hartree-Fock (HF) and Density Functional Theory (DFT/B3LYP) methods, with seven basis sets. Linear correlation data, errors and RMSD values between theoretical and experimental data allowed us to classify the methods and basis sets for evaluating their correspondence with experimental data (geometric parameters, heat capacity, as well as Infrared, Raman and NMR spectra). The HF method has shown the highest correspondence with the experimental data, occupying the top-five rank of the general classification. The HF/6-31G** method was the best one classified and it can be used to model the biological activity of the caffeine molecule.

Keywords: validation; DFT/B3LYP; HF; basis set; correlation.

INTRODUCTION

Use of computational models for prediction of a certain phenomenon is an important stage of the scientific method, since the quantification of variables and the prediction of how the event can be processed are important information that help in the elucidation of the mechanism (detailed steps) by which such phenomenon occurs. The validation of a model or method used in the description and explanation of a phenomenon depends mainly on information and data obtained experimentally, since the model or the method functions as a theoretical representation of a fact / phenomenon.^{1,2}

An important fact is the role of regulatory agencies in different countries in the production of drugs and food, and in the guidelines for the validation and use of different bioanalytical techniques for this purpose. The authors highlight the similarities and variations of the guidelines for validation of bioanalytical methods issued by the main regulatory agencies in the world (United States Food and Drug Administration - USFDA, European Medical Agency - EMA, Brazilian Agency for National Health Surveillance - ANVISA, and the Ministry of Health, Labour and Welfare - MHLW, Japan) and that the diffusion of these validation regulatory standards assists in the quality of the food and drug production process, resulting in a product with better hygienic-sanitary profiles, stability, resistance to temperature variations and increased expiration date (degradation time) of such product.³

Validation of methods and models has been used in several areas of the knowledge, in special for in silico determinations, and the strategy for such use depends on what the scientist aims to predict, such as example the assessment of the toxicity of nanomaterials using Quantitative Structure-Activity Relationships (QSAR) models, then validated in agreement with the principles of the OECD (Organization for Economic Co-operation and Development)

to ensure clarity, consistency and reliability of model forecasts for regulatory purposes.⁴

Paramagnetic NMR techniques and computational modeling approaches for analyses of molecular structures and receptor-ligand complexes have been used as alternatives to reduce the difficulty of obtaining structural data from biological systems because they are complex, dynamic and with high degree of disorder. However, they point out that 3D modeling, validated from paramagnetic NMR experimental data, is an important tool useful for understanding mechanism, recognition and biomolecular function of complex biological systems.⁵

Other different strategies for validation and use of models can be thus cited:

- 1) Selection of active molecules with potential inhibition of tumors and possible action in the treatment of heart failure with the aid of pharmacophore-based virtual screening, validated from its specificity and sensitivity to active molecules;⁶
- 2) The obtaining new caffeine analogues molecules with epithelial anticancer activity using molecular docking and thus validated by redocking and low RMSD between the experimental crystallographic pose and the calculated one;⁷
- 3) Study with QSAR models, built using PLS and PCR, of new molecules derived from artemisinin, validated from the prediction of the biological activity, followed by comparison with experimental data;⁸
- 4) Validation of the HF/6-31G** method used in the molecular modeling of artemisinin with antimalarial activity, from the correspondence with experimental data of geometric parameters.⁹

We have studied the caffeine molecule and analogues using pharmacophore-based virtual screening and statistical analysis and obtained results indicating the caffeine epithelial anticancer potential.⁷ We will continue here with the search for information that allows the choice of computational method for future studies from the molecular modeling of caffeine and analogues.

*e-mail: josivan.chemistry@gmail.com

In this study, different experimental data of the caffeine molecule were used to evaluate the efficiency of the Hartree-Fock (HF) and Density Functional Theory (DFT/B3LYP) methods, here used with seven different basis sets. The best evaluated method here compared with the experimental data was validated, and it may be indicated to model the caffeine molecule and similar ones to investigate epithelial anticancer activity.

METHODS

Selection of molecules for pharmacophore perception

A set of 19 caffeine analogues with biological activity against epithelial cancer ($ICT_{50} < 0.24$ mM) were selected from the literature.¹⁰ This activity corresponds to Epidermal Growth Factor (EGF) prevention in the malignant transformation of epidermal cells from JB6 lineage mice, C141 (JB6 P+) sensitive of developmental, and ICT_{50} corresponds to inhibition of cell transformation. The most active molecules were selected because the activity is a determining factor in the construction of the pharmacophore.^{11,12} The caffeine was inserted in this set because it is the prototype molecule, i.e. the common basis for the studied activity, whereas xanthine is the structural basis for the set of selected molecules. Selection resulted in a total of 21 molecules.

The crystallographic structure of the caffeine was retrieved from the Cambridge Structural Database (CSD, <http://webcsd.ccdc.cam.ac.uk/>) and the other 20 molecules were drawn using the ChemSketch software.¹³ The HyperChem 7 software¹⁴ was used to pre-optimize the geometry of the molecules, at the MM+ Force Field Molecular Mechanics level.

The pharmacophore was then generated using the PharmaGist¹⁵ web server, adopting the caffeine molecule as pivot (kept frozen). The set of molecules aligned with greater number of ligands and higher score was considered the best candidate for a pharmacophore model.¹⁶

Caffeine molecule optimization using HF and DFT/B3LYP methods

The optimization of the caffeine molecule was performed using the Gaussian 09 software¹⁷ with the HF and DFT/B3LYP methods and the basis sets STO-3G*, 3-21G*, 3-21G**, 6-31G*, 6-31G**, 6-311G* and 6-311G**, totalizing 14 different calculations. The choice of basis sets with polarization function is due to an improved efficiency in the optimization, since the atomic orbitals suffer distortions, or they polarize the molecules formed. Also, possibility of displacement of charges outside the atomic nucleus results in an accumulation of charges in a given region (polarization).¹⁸ Studies suggest that most of calculations using basis sets without polarization function are inadequate choices to reproduce experimental values.^{19,20}

Thus, theoretical data regarding geometric parameters, heat capacity (at constant pressure), IR, Raman and NMR spectra were here calculated for the caffeine molecule.

Evaluation of methods and basis sets

The Root Mean Square Deviation (RMSD) values observed

Table 1. Score and features of the best pharmacophore model

Score	GF	SF	Ar	Hyd	Don	Acc	Neg	Pos
60.865	6	6	2	0	0	3	0	1

GF = General Characteristics; SF = Spatial Characteristics; Ar = Aromatic; Hyd = Hydrophobic; Don = Hydrogen bond donor; Acc = Hydrogen bond acceptor; Neg = Anion; Pos = Cation. Pivot molecule: Caffeine.

between experimental caffeine crystallographic data (bond length, bond angle and torsion angle)²¹ and the theoretical ones here calculated using the different methods and basis set were analyzed. Theoretical errors regarding the experimental values of heat capacity,²² spectroscopic data of IR and Raman,²³ as well as RMSD values regarding to NMR spectroscopic data,²⁴ were also analyzed.

Linear correlation analyses (theoretical \times experimental data, theoretical data \times RMSD) were carried out for comparison of the methods.²⁵ Classifications (A - D) were assigned to the different methods and basis sets, based on the best matches with the experimental data and as well as correlations, for final classification of each method and basis set regarding the different experimental data.

RESULTS AND DISCUSSION

Pharmacophore model

The pharmacophore model candidate with the highest scoring and highest number of aligned molecules (21 molecules) was selected as the best model. Score and features of the pharmacophore model thus chosen are available in Table 1.

In Figure 1 the final pharmacophore can be visualized. PharmaGist indicates a total of 6 features (or general characteristics - GF) equivalent to different pharmacophoric groups; 6 spatial characteristics (SF) are equivalent to conformation and location in the 3D space of pharmacophoric regions.

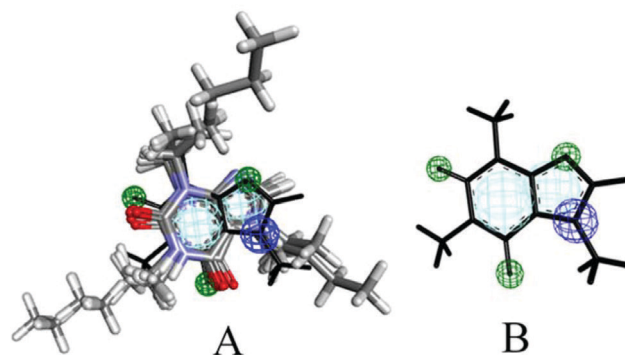


Figure 1. Pharmacophore model. A) Multiple alignment and B) Caffeine molecule

The GF and SF characteristics are respectively represented by two aromatic regions (Ar), three hydrogen bonding acceptor regions (Acc) and one cationic atom (Pos). These regions or pharmacophoric groups could be strongly relevant to biological activity against epithelial cancer in this chemotype here investigated. The 21 molecules then selected to build the pharmacophore model can be visualized in Figure 2.

Quantum chemical calculations

Geometric Parameters

With the spatial arrangement accessed from the pharmacophore model here built for caffeine it was possible to calculate bond

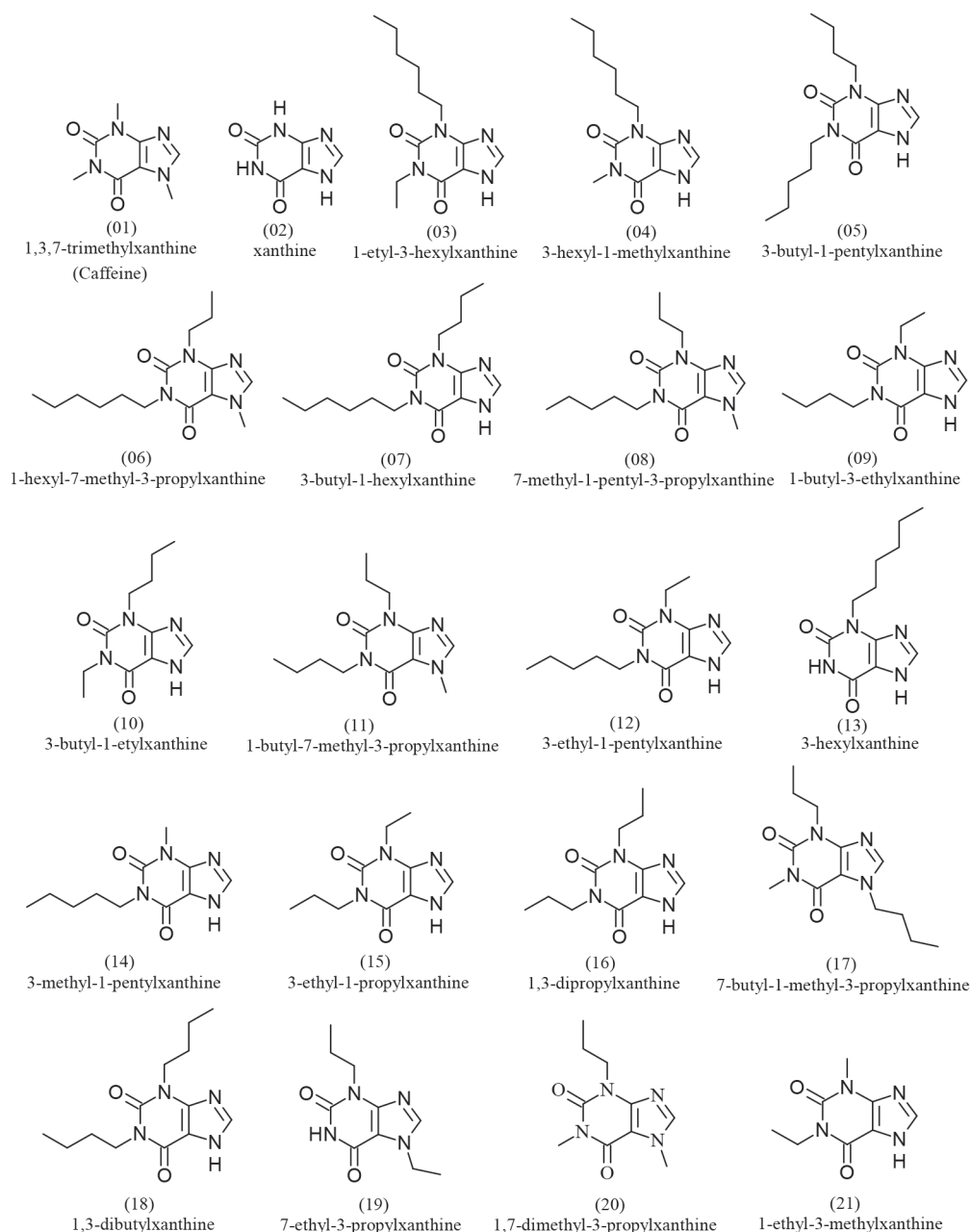


Figure 2. Structures of the molecules used in pharmacophore generation

lengths, bond angles and torsion angles between the atoms of the molecule.

Data on bond lengths, bond angles and torsion angles between the atoms of the caffeine molecule, here calculated using different methods and basis sets, as well as experimental data of crystallography of caffeine are presented in Tables 2 and 3. 2D Structure of caffeine, with its respective numbering is available in Figure 3.

From Tables 2 and 3 it can be inferred that the HF and DFT/B3LYP methods showed good agreement with the experimental data for bond length. The mean of the errors (module values) in relation to the experimental values for HF were: STO-3G* = 0.025; 3-21G* = 0.004; 3-21G** = 0.004; 6-31G* = 0.010; 6-31G** = 0.010; 6-311G* = 0.011; 6-311G** = 0.011. The values for DFT/B3LYP were: STO-3G* = 0.053; 3-21G* = 0.014; 3-21G** = 0.014; 6-31G* = 0.007; 6-31G** = 0.007; 6-311G* = 0.005; 6-311G** = 0.011. The basis set 3-21G*, 3-21G**, 6-31G** e 6-31G** showed the best results for binding length in both HF and DFT/B3LYP, which indicates a better

structural representation of the caffeine molecule in these bases in relation to the parameter bond length.

The geometric parameter of bond angle was better represented in the structures calculated by the DFT/B3LYP method, with the STO-3G base set with the lowest value (in module) for error mean (0.048) and with most other bases presenting values for mean error around 0.3. The HF method presented error mean values for all base sets above 0.4, which can also be considered a good result if one evaluates the magnitude of the error that is relatively low.

For the parameter torsion angle all the bases in both the HF and the DFT/B3LYP methods presented mean error values around 1.2 which attests similar structural representativeness for this parameter in both methods in all basis sets.

The DFT/B3LYP method have indicated lower values of RMSD, compared to the experimental data, considering the basis set 3-21G* (RMSD = 1.973), 3-21G** (RMSD = 1.972) and 6-31G** (RMSD = 1.931). Using the Hartree-Fock method, lower RMSD

Table 2. The theoretical (HF) and experimental geometric parameters of the caffeine structure

Parameters	Hartree-Fock (HF)						EXP. ²¹	
	STO-3G*	3-21G*	3-21G**	6-31G*	6-31G**	6-311G*		6-311G**
Bond length								
N1C2	1.438	1.394	1.394	1.392	1.391	1.392	1.391	1.429
C2N3	1.428	1.375	1.375	1.372	1.372	1.372	1.372	1.355
N3C4	1.407	1.361	1.361	1.369	1.369	1.368	1.368	1.430
C4C5	1.355	1.357	1.357	1.357	1.357	1.356	1.356	1.345
C5C6	1.478	1.424	1.424	1.433	1.433	1.432	1.432	1.440
N1C6	1.441	1.395	1.394	1.395	1.395	1.395	1.395	1.360
C4N9	1.401	1.360	1.360	1.346	1.346	1.346	1.345	1.313
C5N7	1.403	1.387	1.386	1.384	1.383	1.384	1.384	1.403
N7C8	1.377	1.346	1.347	1.329	1.329	1.328	1.328	1.335
N9C8	1.329	1.322	1.323	1.310	1.310	1.310	1.309	1.337
C2O11	1.221	1.215	1.216	1.198	1.199	1.193	1.193	1.200
C6O13	1.223	1.223	1.224	1.202	1.202	1.196	1.196	1.259
Bond angle								
N1C2O11	121.34	121.20	121.17	120.81	120.80	120.84	120.83	122.49
N3C2O11	122.64	121.76	121.78	121.25	121.25	121.30	121.30	123.24
N1C6O13	121.88	121.91	121.89	122.67	122.66	122.65	122.65	120.10
C5C6O13	127.46	126.39	126.39	125.77	125.77	125.80	125.78	125.38
C6C5N7	130.40	131.57	131.50	131.53	131.54	131.59	131.60	133.46
N3C4N9	125.10	126.74	126.78	126.74	126.74	126.67	126.69	126.81
C2N3C4	119.54	119.29	119.28	119.27	119.26	119.26	119.26	121.71
C2N1C6	127.22	126.84	126.83	126.30	126.29	126.34	126.32	127.73
C5C4N9	112.33	110.70	110.69	111.55	111.54	111.49	111.48	111.75
C4N9C8	102.87	104.99	105.04	103.73	103.76	103.79	103.83	105.25
C5N7C8	105.46	105.98	106.01	105.20	105.23	105.16	105.21	104.51
N7C8N9	113.73	112.47	112.36	114.27	114.23	114.29	114.22	112.44
Torsion angle								
C10N1C2O11	-179.99	-180.00	-180.00	-179.99	-179.99	-179.99	-179.99	-179.60
O11C2N3C12	179.96	180.00	180.00	180.00	180.00	180.00	180.00	175.76
C12N3C4N9	-179.97	-179.99	-180.00	-179.99	-179.99	-180.00	-180.00	-179.25
C10N1C6O13	-179.97	-180.00	-180.00	-180.00	-180.00	-180.00	-180.00	-179.88
O13C6C5O7	-179.99	-180.00	-180.00	-179.99	-179.99	-179.99	-179.99	-178.80
C6C5N7C14	-179.99	-180.00	-180.00	-180.00	-179.99	-179.99	-179.99	-176.27
C4N9C8N7	-179.99	-180.00	-180.00	-180.00	-179.99	-179.99	-178.99	-174.76
C5C4N9C8	180.00	179.99	180.00	179.99	179.98	179.98	179.97	176.96
C5N7C8N9	180.00	180.00	180.00	179.99	179.99	179.99	179.99	179.68
C2N3C4C5	-179.98	-180.00	-180.00	-180.00	-180.00	-180.00	-179.99	-174.48
C6N1C2N3	-179.94	-180.00	-180.00	-179.99	-179.99	-179.99	-179.98	-176.96
N3C4C5C6	-180.00	-180.00	-180.00	-180.00	-179.99	-179.99	-179.99	-178.00
RMSD	2.060 (B)	1.973 (A)	1.975 (A)	2.041 (B)	2.038 (B)	2.035 (B)	1.964 (A)	

Classification according to the lowest RMSD: A if RMSD < 2.000; B if RMSD ≥ 2.000; C if RMSD ≥ 3.000 and D if RMSD ≥ 4.000; EXP = experimental.

values are obtained for the basis sets 3-21G* (1.973), 3-21G** (1.975) and 6-311** (1.964). These results and the above discussions on the three structural parameters evaluated (Bond Length, Bond Angle, and Torsion Angle) show that the DFT / B3LYP method of general way presented the best results (lower error values) in relation to the experimental data than HF, which resulted in the lower values of RMSD presented by the DFT / B3LYP method. However, the HF method presented low RMSD values in all base sets and can also be considered good results.

Classification of RMSD values within A and B limits (error values lower or around 2.0) allow us to verify that, for the analysis of the caffeine geometry, all the methods and basis set here investigated agree with the experimental data. DFT/B3LYP and HF methods have been used and reported in vast literature for calculation of

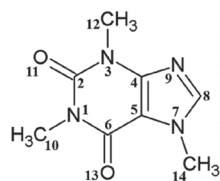
geometric parameters, also indicating excellent agreement with the experimental data, such as observed in this work. Calculations were performed, for example, with 2,2'-Bithiophene,²⁶ silatran molecules²⁷ and 2-amino-4-Methoxy-6-Methyl-Pyrimidine as well.²⁸

Table 4 shows correlations between the theoretical vs experimental values and between theoretical vs RMSD values for geometric parameter of the caffeine structure. Best correlations with the experimental values were observed for the HF method, indicating classification A (Correlation > 0.70) for most of the basis set used, and B classification (0.50 < correlation < 0.70) for most of the basis set, when comparison was performed with the RMSD values. The DFT/B3LYP method showed low correlations between experimental vs RMSD values, with almost all of the basis sets showing

Table 3. The theoretical (DFT/B3LYP) and experimental geometric parameters of the caffeine structure

Parameters	DFT/B3LYP						EXP. ²¹	
	STO-3G*	3-21G*	3-21G**	6-31G*	6-31G**	6-311G*		6-311G**
Bond length								
N1C2	1.468	1.414	1.412	1.408	1.408	1.408	1.391	1.429
C2N3	1.466	1.397	1.400	1.392	1.391	1.391	1.372	1.355
N3C4	1.417	1.369	1.368	1.376	1.376	1.374	1.368	1.430
C4C5	1.394	1.383	1.382	1.382	1.382	1.380	1.356	1.345
C5C6	1.476	1.427	1.425	1.433	1.433	1.432	1.432	1.440
N1C6	1.476	1.419	1.419	1.418	1.418	1.418	1.395	1.360
C4N9	1.431	1.377	1.376	1.360	1.360	1.357	1.345	1.313
C5N7	1.419	1.392	1.390	1.390	1.387	1.387	1.384	1.403
N7C8	1.406	1.373	1.375	1.356	1.356	1.354	1.328	1.335
N9C8	1.374	1.346	1.347	1.330	1.330	1.327	1.309	1.337
C2O11	1.253	1.237	1.239	1.223	1.223	1.216	1.193	1.200
C6O13	1.263	1.250	1.250	1.228	1.229	1.222	1.196	1.259
Bond angle								
N1C2O11	122.99	121.79	121.25	121.41	121.40	121.43	120.83	122.49
N3C2O11	122.65	121.98	122.65	121.43	121.45	121.54	121.30	123.24
N1C6O13	122.39	121.90	121.82	122.63	122.61	122.52	122.65	120.10
C5C6O13	128.30	126.97	127.03	126.35	126.36	126.43	125.78	125.38
C6C5N7	128.47	131.00	131.01	131.00	131.01	131.12	131.60	133.46
N3C4N9	125.87	126.56	126.35	126.84	126.84	126.80	126.69	126.81
C2N3C4	120.66	119.75	119.64	119.74	119.74	119.77	119.26	121.71
C2N1C6	128.45	127.38	127.54	126.86	126.87	126.91	126.32	127.73
C5C4N9	112.95	111.25	111.29	111.84	111.84	111.71	111.48	111.75
C4N9C8	101.74	104.41	104.37	103.59	103.61	103.81	103.83	105.25
C5N7C8	105.88	106.30	106.30	105.64	105.68	105.64	105.21	104.51
N7C8N9	113.94	112.36	112.30	113.82	113.77	113.71	114.22	112.44
Torsion angle								
C10N1C2O11	-180.00	-180.00	-179.99	-180.00	-179.91	-180.00	-179.99	-179.60
O11C2N3C12	180.00	179.99	179.99	180.00	179.96	180.00	180.00	175.76
C12N3C4N9	-179.99	-179.99	-179.96	-180.00	-179.97	-180.00	-180.00	-179.25
C10N1C6O13	-180.00	-180.00	-179.99	-180.00	-179.93	-180.00	-180.00	-179.88
O13C6C5O7	-179.99	-179.99	-179.98	-180.00	-179.69	-180.00	-179.99	-178.80
C6C5N7C14	-179.99	-179.99	-179.99	-180.00	-179.68	-180.00	-179.99	-176.27
C4N9C8N7	-179.99	-179.99	-179.99	-180.00	-179.69	-180.00	-179.99	-174.76
C5C4N9C8	179.99	179.98	179.98	180.00	179.37	180.00	179.97	176.96
C5N7C8N9	180.00	180.00	180.00	180.00	179.86	180.00	179.99	179.68
C2N3C4C5	-179.98	-179.99	-179.99	-180.00	-179.87	-180.00	-179.99	-174.48
C6N1C2N3	-180.00	-180.00	-179.97	-180.00	-179.95	-180.00	-179.98	-176.96
N3C4C5C6	-180.00	-180.00	-180.00	-180.00	-179.91	-180.00	-179.99	-178.00
RMSD	2.228 (B)	1.973 (A)	1.972 (A)	2.021 (B)	1.931 (A)	2.003 (B)	2.031 (B)	-

Classification according to the lowest RMSD: A if RMSD < 2.000; B if RMSD ≥ 2.000; C if RMSD ≥ 3.000 and D if RMSD ≥ 4.000; EXP = experimental.

**Figure 3.** Caffeine molecule with numbered atoms

classification D (correlation < 0.30). Only the 3-21G* basis set shows classification C for correlation with experimental data.

These correlations express the relationship of the theoretical data obtained with the RMSD values and the experimental geometric parameters values. However, for the study performed here, these results are not sufficient to disqualify the evaluated methods. More

experimental data (heat capacity and IR, Raman and NMR spectra) will be compared with theoretical data and discussed in the next sections for better qualification and classification of the methods in evaluation.

Thermochemical parameter - Heat capacity

In Table 5, the different methods and basis set here used for calculation of thermochemical heat capacity as well as the experimental value for caffeine are listed. The Gaussian 09 software¹⁷ returns thermochemical results for the heat capacity, simulating a constant pressure environment, at a temperature of 298.15 Kelvin. The heat capacity is an important parameter for the validation of the computational method to be used in the caffeine molecule, because

Table 4. Correlations obtained between the theoretical vs experimental values and between theoretical vs RMSD values for geometric parameter of the caffeine structure

	HF						
	STO-3G*	3-21G*	3-21G**	6-31G*	6-31G**	6-311G*	6-311G**
Experimental	0.57 (B)	0.58 (B)	0.59 (B)	0.71 (A)	0.70 (A)	0.70 (A)	0.70 (A)
RMSD	0.22 (D)	0.57 (B)	0.56 (B)	0.57 (B)	0.57 (B)	0.57 (B)	0.57 (B)
	DFT/B3LYP						
	STO-3G*	3-21G*	3-21G**	6-31G*	6-31G**	6-311G*	6-311G**
Experimental	0.28 (D)	0.31 (C)	0.29 (D)	0.21 (D)	0.22 (D)	0.23 (D)	0.17 (D)
RMSD	0.08 (D)	0.12 (D)	0.10 (D)	-0.03 (D)	-0.03 (D)	-0.01 (D)	-0.12 (D)

Classification regarding correlation (Corr): A if $0.70 \leq \text{Corr} \leq 1.00$; B if $0.50 < \text{Corr} < 0.70$; C if $0.30 \leq \text{Corr} \leq 0.50$ and D if $\text{Corr} < 0.30$.

Table 5. Theoretical and experimental data for heat capacity of caffeine

Methods	Heat capacity	Error
Experimental ²²	54.409	—
HF/STO-3G*	44.386	10.023 (C)
HF/3-21G*	42.956	11.453 (C)
HF/3-21G**	43.106	11.303 (C)
HF/6-31G*	41.895	12.514 (D)
HF/6-31G**	45.827	8.582 (B)
HF/6-311G*	43.881	10.528 (C)
HF/6-311G**	43.971	10.438 (C)
DFT/B3LYP – STO-3G*	48.154	6.225 (A)
DFT/B3LYP – 3-21G*	46.256	8.153 (B)
DFT/B3LYP – 3-21G**	45.299	9.110 (B)
DFT/B3LYP – 6-31G*	47.064	7.345 (A)
DFT/B3LYP – 6-31G**	47.160	7.249 (A)
DFT/B3LYP – 6-311G*	47.049	7.360 (A)
DFT/B3LYP – 6-311G**	43.972	10.437 (C)

Classifications: A if $\text{Error} < 8$; B if $8 \leq \text{Error} < 10$; C if $10 \leq \text{Error} < 12$; D if $\text{Error} \geq 12$. Values of heat capacity and error recorded in $\text{Cal mol}^{-1} \text{K}^{-1}$.

it can be expressed from contributions of translational, rotational, and vibrational energies, important thermodynamic parameters intrinsically related to the structural geometric parameters.²⁹

The DFT/B3LYP method indicated better results for calculation of the heat capacity, since it resulted in lower error compared to the experimental value. For this method, the basis set STO-3G* (Error = 6.225), 6-31G** (Error = 7.249), 6-31G* (Error = 7.345) and 6-311G* (Error = 7.360) indicated better agreement or adjust to the experimental values. In these analyses, results obtained using the HF method were all close to the worst value obtained for the DFT/B3LYP method, with error values between 10.023 and 12.514. Due to the lower error values obtained using the DFT/B3LYP method, best classifications were then observed (most of the basis set classified as A - error < 8). The HF method indicated classification C ($10 \leq \text{Error} < 12$) with most of the basis set used, whereas classification B (error = 8.582) was only indicated for the 6-31G** basis set.

Theoretically (with calculations carried out using the Gaussian 09 software), the heat capacity can be obtained from a partition function that comprises contributions of four different components: translational motion ($3/2 R$, where R is the gas constant), electronic (null contribution), rotational ($3/2 R$, similar to translational) and vibrational ($3Rf$, where f is a function derived by Einstein and

improved by Debye, considering the number of oscillating atoms at a given temperature, which contribute to the vibrational energy).²⁹

It can be inferred that calculations for the caffeine molecule using the DFT/B3LYP method indicated better results for these contributions, compared to the HF method. The efficiency of the DFT/B3LYP method for calculation of the heat capacity may be related to the integration of the density functional (configured by the integration of several derivatives of the density functional), step that leads to an increase in the accuracy of the calculation performed^{30,31} which does not exist in the HF calculation.

However, the good result obtained using HF/6-31G** (error = 8.582 and classification B) could be due to the presence of the maximum polarization function (** = insertion of polarization functions for the d, p orbitals), which may provide a lowering of the error values for that basis set.^{28,32-34}

Spectroscopic parameters

The spectroscopic parameters (Infrared, Raman and NMR) are excellent methods for identification of substances and play an important role in the structural characterization of organic molecules such as caffeine.

Infrared and Raman Spectra

Results for the IR and Raman spectra can be seen in Figure 4 and Tables 6 and 7. The error values (in brackets) show how the main bands vary according to the experimental frequencies.

In general, the bases 6-31G* and 6-31G** in both HF and DFT/B3LYP methods showed better agreement with the experimental values in IR and Raman spectra. The HF / 6-31G* method in the Raman spectrum showed lower mean error values for frequencies and good agreement with the intensities of the main experimental bands. The HF / 6-31G** method in the same spectrum showed better correspondence with the intensities of the main experimental bands. Also, in the Raman spectrum, both the 6-31G* and the 6-31G** basis set of the DFT/B3LYP method presented low values of mean of error and good correspondence (being similar) to the intensities of the experimental bands (see Figure 4).

From Figure 4 it can be verified that for IR analysis the HF method in the basis set 6-31G* (E), 6-31G** (F), 6-311G* (G) e 6-311G** (H), although they have a slight shift to the left, has a greater similarity both in the intensities of the bands and in the frequencies with the experimental spectra (A). However, for the same method, the basis set STO-3G (B), 3-21G* (C), HF / 3-21G** (D) showed little similarity in both frequencies and intensity of the bands. Similar observations for Raman spectra data can also be seen in Figure 4. These observations confirm the good match of the spectra calculated by the HF method in the basis set 6-31G* (E), 6-31G** (F), 6-311G* (G) e 6-311G**

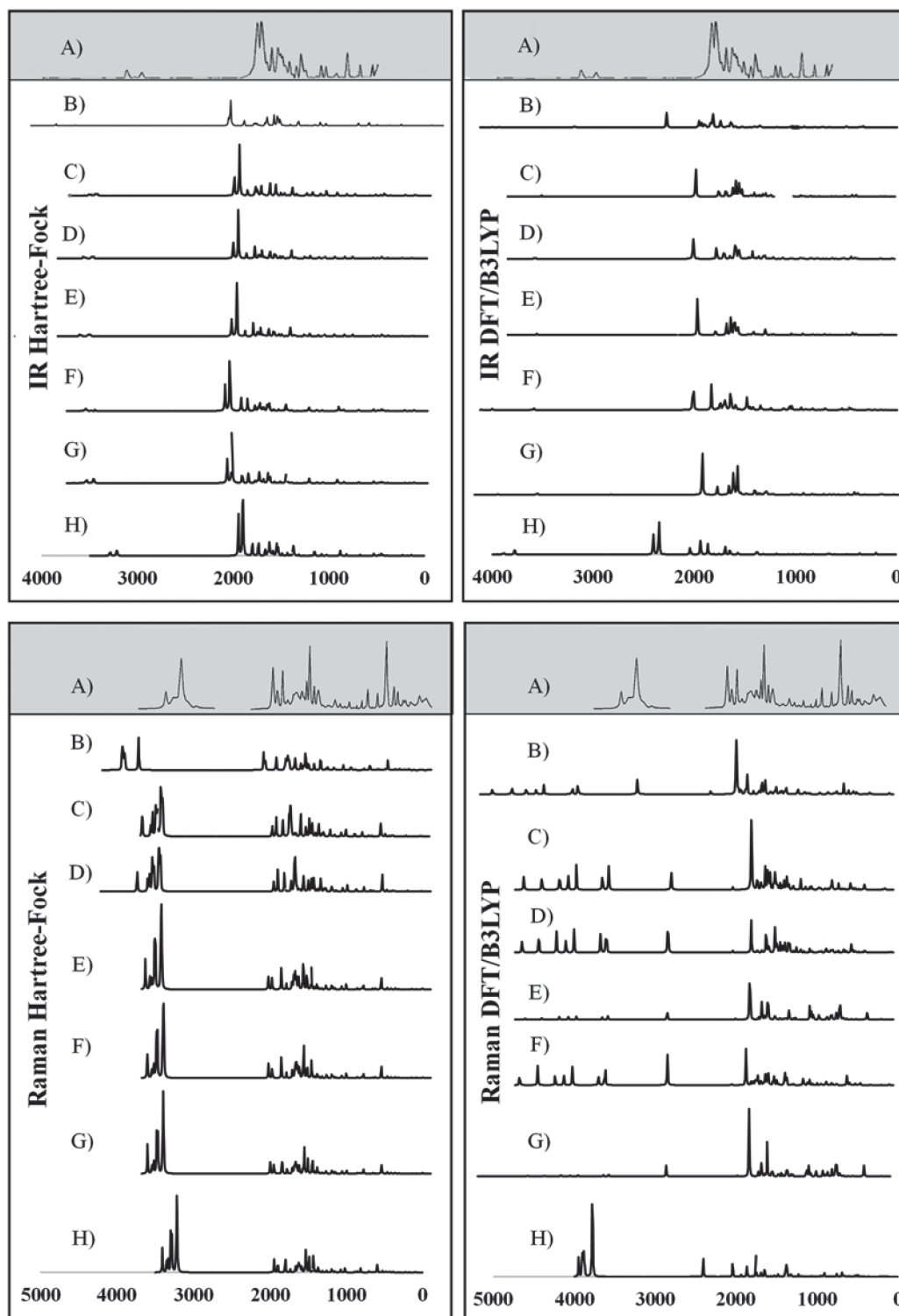


Figure 4. Experimental IR and Raman spectra (A),²³ calculated for caffeine using HF and DFT/B3LYP methods with different basis set: (B) STO-3G*; (C) 3-21G*; (D) 3-21G**; (E) 6-31G*; (F) 6-31G**; (G) 6-311G*; (H) 6-311G**

(H) with the experimental spectra, and less similarity of the spectra in the basis set STO-3G (B), 3-21G* (C), HF/3-21G** (D).

In relation to the IR spectra it can be observed that the experimental data of the C=N bond (1700 VS – Stg), when simultaneously checking the frequency and intensity of the bands, there is a better match with HF/6-31G**, HF/6-311G* e HF/6-311G**. C=O bond (1) (1660 VS – Stg. asym) presented good correspondence with all bases of the HF method except STO-3G*. In C=O bond (2) (1548 S – Stg. sym) all bases in the HF and

DFT / B3LYP methods showed similar intensity of the bands and close to the experimental ones, with emphasis on the DFT / B3LYP method with most of the basis set with low value of error compared to experimental frequency. The bonds C-N (1237 S – Along) and O=C-C (743 S – Deform), in general, had better agreement with the basis set of the HF method, mainly in relation to the intensity of the bands with the frequencies showing a sharp shift to the left.

In the Raman spectra, HF/6-31G* e HF/6-31G** presented better agreement with the experimental data in the bonds C-H (2963 VS

Table 6. Experimental and calculated IR frequencies using the HF and DFT/B3LYP methods, with different basis set

Experimental ²³	1700 VS – Stg. C=N	1660 VS – Stg. asym. C=O (1)	1548 S – Stg. sym. C=O (2)	1237 S – Along. C-N	743 S – Deform. O=C-C	Mean error	Corr. exp.
HF/STO-3G*	2078 S (378)	2057 MS (397)	1708 MS (160)	1601 MS (364)	1583 W (840)	427.8 (D)	0.80 (A)
HF/3-21G*	1899 MS (199)	1849 VS (189)	1555 MS (7)	1498 MS (261)	1335 MS (592)	249.6 (B)	0.89 (A)
HF/3-21G**	1898 MS (198)	1848 VS (188)	1687 MS (139)	1538 MS (301)	1327 MS (584)	282.0 (C)	0.97 (A)
HF/6-31G*	1962 MS (262)	1916 VS (256)	1805 MS (257)	1554 MS (317)	1377 W (634)	345.2 (B)	0.98 (A)
HF/6-31G**	1961 S (261)	1915 VS (255)	1805 MS (257)	1742 MS (505)	1374 MS (631)	381.8 (B)	0.98 (A)
HF/6-311G*	1940 S (240)	1894 VS (234)	1795 MS (247)	1734 MS (497)	1371 MS (628)	369.2 (B)	0.98 (A)
HF/6-311G**	1939 VS (239)	1893 VS (233)	1794 MS (246)	1731 MS (494)	1366 MS (623)	367.0 (B)	0.98 (A)
DFT/B3LYP – STO-3G*	2207 S (507)	1764 W (104)	1690 MS (142)	1587 W (350)	1315 W (572)	335.0 (D)	0.85 (A)
DFT/B3LYP – 3-21G*	1941 VS (241)	1715 MS (55)	1549 MS (1)	1490 MS (253)	1264 W (521)	214.2 (C)	0.90 (A)
DFT/B3LYP – 3-21G**	1941 S (241)	1717 MS (57)	1532 MS (-16)	1507 W (270)	1371 W (628)	236.0 (C)	0.82 (A)
DFT/B3LYP – 6-31G*	1910 VS (210)	1590 MS (-70)	1558 MS (10)	1547 MS (310)	1256 W (513)	194.6 (C)	0.85 (A)
DFT/B3LYP – 6-31G**	1954 S (254)	1780 S (120)	1595 MS (47)	1436 MS (199)	1305 W (562)	236.4 (C)	0.90 (A)
DFT/B3LYP – 6-311G*	1850 VS (150)	1560 S (-100)	1520 S (-28)	1353 W (116)	1246W (503)	128.2 (C)	0.86 (A)
DFT/B3LYP – 6-311G**	2401 S (701)	2346 VS (686)	1934 MS (386)	1861 MS (624)	1690 W (947)	668.8 (D)	0.87 (A)

Asy. = Asymmetric; Sym. = symmetric; Stg = Stretching; Deform. = Deformation; VS = Very Strong; S = Strong; MS = Medium Strong; W = Weak; Corr. exp. = Correlation with the experimental value; Mean error = Mean of error between experimental and theoretical frequencies; Error in relation to the experimental value shown in parentheses, below the frequency values; The classification represented by the upper case letters to the right of the mean of the error was determined taking into account: average of the error and the level of correspondence with intensity of the experimental bands; For the classification of Corr. Exp., A if Corr. exp. \geq 0.70.

Table 7. Experimental and calculated Raman frequencies using the HF and DFT/B3LYP methods with different basis set

Experimental ²³	2963 VS – Stg. C-H	1700 VS – Stg. C=N	1600 S – Stg. C=C	1331 VS – Stg. C-N	556 VS – Deform. C-N-CH ₃	Mean error	Corr. exp.
HF/STO-3G*	3561 S (598)	2077 MS (377)	1930 MS (330)	1581 MS (250)	607 MS (51)	151.0 (C)	0.97 (A)
HF/3-21G*	3223 S (260)	1849 MS (149)	1670 MS (70)	1447 MS (116)	585 MS (29)	326.2 (C)	0.95 (A)
HF/3-21G**	3267 S (304)	1848 MS (148)	1641 S (41)	1538 MS (207)	584 MS (28)	353.4 (C)	0.96 (A)
HF/6-31G*	3225 S (262)	1805 MS (105)	1537 MS (-63)	1442 MS (111)	596 MS (40)	13.4 (B)	0.94 (A)
HF/6-31G**	3226 VS (263)	1805 MS (105)	1534 S (-66)	1440 MS (109)	596 MS (40)	382.6 (B)	0.98 (A)
HF/6-311G*	3231 VS (-1023)	1795 MS (194)	1525 S (195)	1430 MS (403)	593 MS (815)	-6.6 (D)	0.97 (A)
HF/6-311G**	3216 VS (253)	1794 MS (94)	1523 S(-77)	1428 MS (97)	592 MS (36)	953.6 (D)	0.98 (A)
DFT/B3LYP – STO-3G*	3092 MS (129)	1899 VS (199)	1764 S (164)	1551 MS (220)	599 MS (43)	321.2 (C)	0.97 (A)
DFT/B3LYP – 3-21G*	3438 MS (475)	2684 S (984)	1715 VS (115)	1427 MS (96)	517 W (-39)	124.8 (C)	0.95 (A)
DFT/B3LYP – 3-21G**	3537 MS (574)	2725 MS (1025)	1717 S (117)	1429 MS (98)	509 W (-47)	145.6 (C)	0.95 (A)
DFT/B3LYP – 6-31G*	2733 W (-230)	1735 S (35)	1590 MS (-10)	1515 MS (184)	644 MS (88)	91.0(C)	0.99 (A)
DFT/B3LYP – 6-31G**	3480 MS (517)	2731 S (1031)	1780 S (180)	1514 MS (183)	558 MS (2)	90.2 (C)	0.95 (A)
DFT/B3LYP – 6-311G*	2709 W (-254)	1709 VS (9)	1556 MS (-44)	1490 S (159)	653 MS (97)	116.8 (C)	0.99 (A)
DFT/B3LYP – 6-311G**	3878 MS (915)	3778 VS (2078)	2137 MS (537)	1752 MS (421)	1373 MS (817)	80.6 (C)	0.83 (A)

Stg = Stretching; Deform. = Deformation; VS = Very Strong; S = Strong; MS = Medium Strong; W = Weak; Corr. exp. = Correlation with the experimental value; Mean error = Mean of error between experimental and theoretical frequencies; Error in relation to the experimental value shown in parentheses below the frequency values; The classification represented by the upper case letters to the right of the mean of the error was determined taking into account: average of the error and the level of correspondence with intensity of the experimental bands; For the classification of Corr. Exp., A if Corr. exp. \geq 0.70.

– Stg), C=N (1700 VS – Stg.), C=C (1600 S – Stg), C-N (1331 VS – Stg) e C-N-CH₃ (556 VS – Deform), when analyzed simultaneously the frequencies and intensity of the bands. However, most basis set (both in HF and in DFT/B3LYP methods) presented a reasonable agreement with the experimental spectra data (see tables 6, 7 and Figure 4), since, typically the ab initio harmonic vibrational wave numbers are larger than those obtained experimentally.³⁵

Negative values denote shift of the rightmost frequency values compared to experimental spectra, whereas positive values mean that the shift is more to the left. The positive pattern of the error values indicates that, on average, the calculated IR and Raman spectra are shifted more to the left, in relation to the experimental spectra (see figure 4). In most quantum chemical methods this disagreement can be attributed to the anharmonicity and partially to the general tendency of these methods to overestimate the force constants in the

exact equilibrium geometry. Anharmonicity can be large and show important deviations, specifically as in vibrations involving central atoms with isolated pairs and C-H vibrations.

Checking the tables 6, 7 and Figure 4 note the tendency of the minimum basis (STO-3G*, 3-21G* e 3-21G**) lower agreement with the experimental values of frequency and intensity of the bands when compared to the split-valence basis (6-31G*, 6-31G**, 6-311G*, 6-311G**) with better concordances. This trend is in agreement with the one investigated by Stephens and coworkers, which obtained results of frequencies and intensity of the bands closest to the experimental ones with the use of larger basis set.³⁶

Classification assigned to each method was based on the mean error values observed between the frequencies calculated using different methods (and basis sets) and the experimental frequencies values, as well as the correspondence between the intensity of the main bands.

Although none of the methods indicated A (maximum) classification, the HF method showed good agreement with the experimental data for the IR spectra, with most of the basis set with B classification. The HF method also indicated two basis sets (6-31G* and 6-31G**) with B classification for the Raman spectra. The good classification of the HF method is mainly due to the good agreement between the intensities of the theoretical and experimental bands.^{23,37}

Correlations between theoretical and experimental values indicate results, respectively, between 0.80 and 0.99 for the DFT/B3LYP and HF methods, for both IR and Raman spectroscopic data. These strong correlations obtained signalize that frequencies of the major bands would well correspond to the experimental values if the left of the most displacements (represented by the mean error) were disregarded and the spectra overlapped.

NMR spectra

The chemical shift values for hydrogen and carbon atoms of the caffeine molecule can be seen in table 8. In ¹H, The individual analyzed displacements showed better correspondence with the experimental data in HF/3-21G* for H8, DFT/B3LYP – 6-311G** for N1-CH₃, HF/6-31G* and HF/6-31G** for N3-CH₃ and HF/3-21G* for N7-CH₃. Already for ¹³C the best matches were verified in HF/3-21G*, HF/6-311G* and HF/6-311G** for C2, DFT/B3LYP – 3-21G* for C4, DFT/B3LYP – 6-311G** for C5, DFT/B3LYP – 3-21G* and DFT/B3LYP – 3-21G** for C6 and HF/STO-3G* and DFT/B3LYP – 6-31G** for C8. A general analysis of table 8 allows to infer that the DFT/B3LYP method indicates better agreement with the experimental values for chemical displacement data. However, the data obtained for the HF method also present reasonable agreement with the experimental data. Different from IR and Raman spectra (where the minimum bases showed results with less correspondence with the experimental data), in the calculation of chemical shifts in NMR data, all basis sets presented satisfactory results.

The lower RMSD values and the good classifications (majority with B) of this method confirm such consideration. HF method indicates most of the classifications equal to C, and only the basis set STO-3G* and 3-21G** with classification B. Both methods indicate excellent correlation between theoretical and experimental data for all basis set here evaluated, with A classification in all the cases, which it can be attributed to the low values of RMSD, where the theoretical displacements range from 3.18 to 133.57.

Correlation between the theoretical data and RMSD showed low values in both methods. DFT/B3LYP method showed classifications C for values equal or higher than 0.30, whereas the HF method showed a low correlation with of the RMSD values below 0.30, with classification D.

Good results for NMR data calculated with the DFT/B3LYP method have been also obtained, such as reported in works with 4', 7-dihydroxy-8-prenylflavan,³⁸ as well as 1-methylpyrrolidine-2-one-3-Halo-derivatives.³⁹ Here, the HF method used with basis set 6-31G** has also shown good results for NMR data prediction, with B classification.

Selection of the best computational method

The different experimental data (Geometric parameters, Heat capacity, IR spectra, Raman and NMR) to which the theoretical data using HF and DFT/B3LYP methods were compared and allowed the construction of Table 9.

In this table, the classification data obtained using the HF and DFT/B3LYP methods were organized for each basis set analyzed. In general, the HF method indicated better agreement with the experimental data, since it occupies the top-five positions of the seven possible for this method, with a difference of only two points from the former to the fifth. The best-ranked basis set is the 6-31G** (23 points). The best basis set (3-21G**) for the DFT/B3LYP method reached only the sixth position (20 points). The STO-3G* basis set of both methods were among the worst ranked (18 points), followed by the 6-311G* and 6-311G** of the DFT/B3LYP method, the last two ones of the ranking (18 and 15 points, respectively).

Regards to the caffeine molecule, the 6-31G** basis set, in general, showed better results for each method in comparison to both the STO-3G and 3-21G bases with their respective polarization functions. These data are similar to those reported in the literature,^{32,33,40} where the polarization functions (**) in the separate valence basis sets (6-31G and 6-311G) provide more meaningful and accurate results with lower error values, when compared with experimental data.⁴¹⁻⁴³

The best performance of the HF method in most of our comparisons with the experimental data of the caffeine molecule could be due to the accuracy of the DFT/B3LYP method in performing the calculations be influenced for errors sources that are depending on the integration grid and the number of points used in the numerical integration, where

Table 8. NMR data for displacements of ¹H and ¹³C, in ppm

Methods	¹ H (ppm)				¹³ C (ppm)					RMSD	Corr. exp.	Corr. RMSD	
	H8	N1-CH ₃	N3-CH ₃	N7-CH ₃	C2	C4	C5	C6	C8				
Experimental ²⁴	7.89	3.36	3.54	3.96	86.01	81.75	41.15	89.67	76.9	—	—	—	
HF	STO-3G*	6.09	4.76	4.64	4.20	99.93	90.19	89.99	107.43	79.17	19.29 (B)	0.95 (A)	0.32 (C)
	3-21G*	7.11	4.32	4.05	3.88	82.74	110.97	79.45	132.09	90.32	23.22 (C)	0.96 (A)	0.23 (D)
	3-21G**	5.25	3.97	3.98	4.10	72.52	92.57	54.09	121.25	79.91	13.60 (B)	0.97 (A)	0.16 (D)
	6-31G*	6.62	3.69	3.49	3.28	77.95	123.02	83.62	128.3	99.54	26.41 (C)	0.95 (A)	0.25 (D)
	6-31G**	6.73	3.89	3.60	3.41	78.05	123.06	83.69	128.33	100.09	26.49 (C)	0.95 (A)	0.25 (D)
	6-311G*	6.51	3.66	3.44	3.18	82.50	128.05	89.86	133.49	106.47	30.26 (D)	0.95 (A)	0.25 (D)
	6-311G**	6.55	3.78	3.46	3.26	82.47	128.04	89.93	133.57	53.82	29.57 (C)	0.89 (A)	0.39 (C)
DFT/B3LYP	STO-3G*	5.39	4.5	3.96	4.03	96.88	65.46	77.06	97.63	44.69	18.65 (B)	0.89 (A)	0.36 (C)
	3-21G*	4.91	4.01	3.32	3.71	56.44	82.23	71.60	91.24	57.88	16.49 (B)	0.91 (A)	0.33 (C)
	3-21G**	6.61	4.22	3.82	3.68	54.08	86.45	73.36	92.61	57.93	17.50 (B)	0.90 (A)	0.33 (C)
	6-31G*	6.07	3.5	3.06	3.17	52.92	95.80	79.77	95.54	73.93	18.81 (B)	0.90 (A)	0.31 (C)
	6-31G**	6.08	3.65	3.16	3.28	52.77	95.85	79.89	95.55	74.41	18.87 (B)	0.90 (A)	0.30 (C)
	6-311G*	6.44	3.47	3.07	3.13	59.60	102.02	86.07	102.74	79.7	20.33 (C)	0.90 (A)	0.31 (C)
	6-311G**	4.03	3.33	3.26	3.46	53.85	70.30	49.05	94.53	85.16	12.92 (B)	0.95 (A)	0.08 (D)

Corr. exp = Correlation with the experimental value; Corr. RMSD = Correlation with RMSD value; ppm = parts per million. A - D = Classifications.

Table 9. Classification of the methods and basis set, with the HF and DFT/B3LYP methods used

Class.	Theoretical Methods	GP		TP-HC		SP-IR		SP-RN		SP-NMR			Score (points)
		RMSD	Exp. Corr.	Corr. RMSD.	Error	Mean of error	Corr. exp.	Mean of error	Corr. exp.	RMSD	Corr. exp.	Corr. RMSD	
1	HF/6-31G**	B	A	B	B	B	A	B	A	C	A	D	23
2	HF/6-311G**	A	A	B	C	B	A	D	A	C	A	D	22
3	HF/3-21G**	A	B	B	C	C	A	C	A	B	A	D	21
4	HF/3-21G*	A	B	B	C	B	A	C	A	C	A	D	21
5	HF/6-31G*	B	A	B	D	B	A	B	A	C	A	D	21
6	DFT/B3LYP – 3-21G*	A	C	D	B	C	A	C	A	B	A	C	20
7	DFT/B3LYP – 6-31G**	A	D	D	A	C	A	C	A	B	A	C	20
8	HF/6-311G*	B	A	B	C	B	A	D	A	D	A	D	19
9	DFT/B3LYP – 3-21G**	A	D	D	B	C	A	C	A	B	A	C	19
12	DFT/B3LYP – 6-31G*	B	D	D	A	C	A	C	A	B	A	C	19
11	HF/STO-3G*	B	B	D	C	D	A	C	A	B	A	C	18
12	DFT/B3LYP – STO-3G*	B	D	D	A	D	A	C	A	B	A	C	18
13	DFT/B3LYP – 6-311G*	B	D	D	A	C	A	C	A	C	A	C	18
14	DFT/B3LYP – 6-311G**	B	D	D	C	D	A	C	A	B	A	D	15

Class. = Classification; GP = Geometric Parameters; TP – HC = Thermochemical Parameter – Heat capacity; SP – IR = Spectroscopic Parameter - Infrared; SP-RN = Spectroscopic parameter - Raman; SP-NMR = Spectroscopic Parameter - Nuclear Magnetic Resonance; Corr. exp = Correlation with the experimental value; Corr. RMSD = Correlation with the RMSD value. For scoring purposes: A = 3, B = 2, C = 1 and D = 0.

a small integration grid is not recommended for the DFT/B3LYP calculation. Other additional sources of error other than those cited for the DFT/B3LYP method are those similar to the HF method (accuracy of integration and convergence of SCF and CPHF).^{30,31,36,44}

CONCLUSIONS

Pharmacophore generated from caffeine and analogs with biological epithelial anticancer activity indicates pharmacophoric regions that comprised practically the entire extent of the caffeine molecule, which resulted in the choice of bond lengths, bond angles and torsion angles of the atoms coincident with these regions (see tables 2 and 3).

Both HF and DFT/B3LYP methods had good agreement with the experimental data of geometric parameters of caffeine. In general, the HF method revealed better agreements with the experimental data for both IR and Raman spectra. The DFT/B3LYP method indicated better agreement with experimental data of both heat capacity and NMR spectra.

The general classification of the methods and basis set here used indicated that HF method with the basis set 6-31G** has the best agreement with the experimental datasets evaluated (GP, TP-HC, SP-IR, SP-RN e SP-NMR).

The best rating of HF/6-31G** could be attributed to your excellent individual ratings, mainly because it is the only basis set of the HF method to show B classification for TP-HC and the only one among all basis sets to have most of the classifications in A and B categories, and only one classification in C and D (both in SP-NMR). For comparison purposes, the second best classified method and basis set (HF/6-311G**) indicated two C classifications and two D.

The HF / 6-31G** method that indicated better agreement with the experimental data was properly validated, and it can be used to optimize and obtain theoretical data for the caffeine molecule and similar molecules with epithelial anticancer activity.

ACKNOWLEDGEMENTS

We acknowledge support provided by the Postgraduate Program in Biodiversity and Biotechnology of Amazon (PPG-Bionorte),

To the, Paraense Museum Emilio Goeldi (MPEG). Laboratory of Molecular Modeling and Simulation System of Federal Rural University of Amazônia (UFRA-Brazil) and Computational Laboratory of Pharmaceutical Chemistry, School of Pharmaceutical Sciences of Ribeirão Preto, University of São Paulo, Ribeirão Preto (USP/RP) and PROPESP/UFPA.

REFERENCES

1. Yang, X. S.; *Engineering Mathematics with Examples and Applications*, Academic Press: San Diego, 2017.
2. Chicone, C.; *An Invitation to Applied Mathematics: Differential Equations, Modeling and Computation*, Academic Press: San Diego, 2016.
3. Kadian, N.; Raju, K. S.; Rashid, M.; Malik, M. Y.; Taneja, I.; Wahajuddin, M.; *J Pharm. Biomed. Anal.* **2016**, *126*, 83.
4. Puzyn, T.; Jeliaskova, N.; Sarimveis, H.; Marchese, R. R. L.; Lobaskin, V.; Rallo, R.; Richarz, A. N.; Gajewicz, A.; Papadopoulos, M. G.; Hastings, J.; Cronin, M. T. D.; Benfenati, E.; Fernández, A.; *Food. Chem. Toxicol.* **2017**, *17*, 30556.
5. Pilla, K. B.; Gaalswyk, K.; MacCallum, J. L.; *Biochim. Biophys. Acta.* **2017**, *1865*, 1654.
6. Chen, L.; Wang, D.; Cao, Y.; Zhu, Z.; Chai, Y.; *Chin. Herb. Med.* **2016**, *8*, 126.
7. Costa, J. S.; Costa, K. S. L.; Cruz, J. V.; Ramos, R. S.; Silva, L. B.; Barros, B. D. S.; H. T. P. C.; Santos, C. B. R.; Macêdo, W. J. C.; *Curr. Pharm. Des.* **2017**, *23*, 1.
8. Vieira, J. B.; Braga, F. S.; Lobato, C. C.; Santos, C. F.; Costa, J. S.; Bittencourt, J. A.; Brasil, D. S.; Silva, J. O.; Hage-Melim, L. I.; Macêdo, W. J. C.; Carvalho, J. C. T.; Santos, C. B. R.; *Molecules* **2014**, *19*, 10670.
9. Santos, C. B. R.; Vieira, J. B.; Formigosa, A. S.; Costa, E. V. M.; Pinheiro, M. T.; Silva, J. O.; Macêdo, W. J. C.; Carvalho, J. C. T.; *J. Comput. Theor. Nanosci.* **2014**, *11*, 553.
10. Rogozin, E. A.; Lee, K. W.; Kang, N. J.; Yu, H.; Nomura, Miyamoto, M.; K. I.; Conney, A. H.; Bode, A. M.; Dong, Z.; *Carcinogenesis* **2008**, *29*, 1228.
11. Kandakatla, N.; Ramakrishnan, G.; *Adv. Bioinformatics* **2014**, *812148*.

12. Starosyla, S. A.; Volynets, G. P.; Bdzholo, V. G.; Golub, A. G.; Protopopov, M. V.; Yarmoluk, S. M.; *Bioorg. Med. Chem. Lett.* **2014**, *24*, 4418.
13. ACD/Chemsketch Freeware, version 12.00: Advanced Chemistry Development, Inc, Toronto, ON, Canada, **2010**.
14. ChemPlus, Modular Extensions to HyperChem, Release 6.02. Gainesville: Molecular Modeling for Windows Hyper Inc. **2000**.
15. Schneidman-Duhovny, D.; Dror, O.; Inbar, Y.; Nussinov, R.; Wolfson, H. J.; *Nucleic Acids Res.* **2008**, *36*, 223.
16. Schneidman-Duhovny, D.; Dror, O.; Inbar, Y.; Nussinov, R.; Wolfson, H. J.; *J. Comput. Biol.* **2008**, *15*, 737.
17. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A. Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J.; *Gaussian 09, Revision A.1*, Gaussian, Inc., Wallingford CT, **2009**.
18. Santos, C. B. R.; Lobato, C. C.; Vieira, J. B.; Brasil, D. S. B.; Brito, A. U.; Macêdo, W. J. C.; Carvalho, J. C. T.; Pinheiro, J. C.; *Comput. Mol. Biosci.* **2013**, *3*, 66.
19. Leach, A. R.; *Molecular Modelling: Principles and Applications*, Prentice Hall: Englewood Cliffs, 2001.
20. Zuschneid, T.; Fischer, H.; Handel, T.; Albert, K.; Häfelinger, G.; *Z. Naturforsch., B: J. Chem. Sci.* **2004**, *59*, 1156.
21. Sutor, D. J.; *Acta Crystallogr.* **1958**, *11*, 453.
22. Dong, J. X.; Li, Q.; Tan, Z. C.; Zhang, Z. H.; Liu, Y. *J. Chem. Thermodyn.* **2007**, *39*, 108.
23. Gunasekaran, S.; Sankari, G.; Ponnusamy, S.; *Spectrochim. Acta, A* **2005**, *61*, 117.
24. Kan, L. S.; Borer, P. N.; Cheng, D. M.; Ts'o, P. P.; *Biopolymers* **1980**, *19*, 1641.
25. Anderson, J. E.; Bond, A. M.; Jones, R. D.; O'halloran, R. J.; *Electroanal. Chem.* **1981**, *130*, 113.
26. Ortí, E.; Viruela, P. M.; Shnchez-Marín, J. C.; Tomás, F.; *J. Phys. Chem.* **1995**, *99*, 4955.
27. Csonka, G. I.; Hencsei, P.; *Theochem.* **1996**, *362*, 199.
28. Kumar, L.; Chaudhary, J.; *J. Appl. Phys.* **2017**, *9*, 30.
29. Atkins, P.; Paula, J.; *Physical Chemistry*, 10th ed. rev., Oxford University Press: Oxford, 2014.
30. Becke, A. D.; *J. Chem. Phys.* **1992**, *97*, 9173.
31. Perdew, J. P.; Chevary, J. A.; Vosko, S. H.; Jackson, K. A.; Pederson, M. R.; Singh, D. J.; Fiolhais, C.; *Phys. Rev. B* **1992**, *46*, 6671.
32. Hariharan, P. C.; Pople, J. A.; *Theor. Chim. Acta* **1973**, *28*, 213.
33. Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M. S.; *J. Chem. Phys.* **1982**, *7*, 3654.
34. DeTar, D. F.; *J. Phys. Chem. A* **1998**, *102*, 5128.
35. Pavel, I.; Szeghalmi, A.; Moigno, D.; Cînt, S.; Kiefer, W.; *Biopolymers* **2003**, *72*, 25.
36. Santos, C. B. R.; Carvalho, J. C. T.; Macêdo, W. J. C.; *Rational Design of Antimalarial Drugs*, Novas Edições Acadêmicas, **2016**, 204.
37. Muthu, S.; Prasath, M.; Paulraj, E. I.; Balaji, R. A.; *Spectrochim Acta, A* **2014**, *120*, 185.
38. Alcântara, A. F. C.; Teixeira, A. F.; Silva, I. F.; *Quim. Nova* **2004**, *27*, 371.
39. Melo, U. Z.; Silva, R. G. M.; Yamazaki, D. A. S.; Pontes, R. M.; Gauze, G. F.; Rosa, F. A.; Rittner, R.; Basso, E. A.; *J Phys. Chem. A* **2015**, *119*, 2111.
40. Rassolov, V. A.; Ratner, M. A.; Pople, J. A.; Redfern, P. C.; Curtiss, L. A.; *J. Comput. Chem.* **2001**, *22*, 976.
41. McLean, A. D.; Chandler, G. S.; *J. Chem. Phys.* **1980**, *72*, 5639.
42. Levine, I. N. Q.; *Quantum Chemistry*. 4th ed., New York: Prentice-Hall, 1991.
43. Stephens, P. J.; Devlin, F. J.; Ashvar, C. S.; Chabalowski, C. F.; Frisch, M. J.; *Faraday Discuss.* **1994**, *99*, 103.
44. Ricca, A. C.; Bauschlicher Jr., W.; *J. Phys. Chem.* **1995**, *99*, 9003.