

基于色散校正的密度泛函理论的食品添加剂太赫兹光谱研究

李天莹, 李春, 章龙, 马卿效, 蒋玲*

南京林业大学信息科学技术学院, 江苏南京 210037

摘要 食品添加剂违规使用或滥用会对身体健康产生严重危害, 相比于传统检测方法, 太赫兹光谱技术具有检测周期短、准确率高等优势, 被广泛应用于食品添加剂检测领域, 但对其缺乏深入的理论分析。主要致力于根据晶体结构特征改进传统计算方法, 提高食品添加剂模拟光谱预测精度并评估不同模型的计算性能。选取了三种富含氢键的食品添加剂, 即苯甲酸、山梨酸和木糖醇, 实验采用太赫兹时域光谱仪, 理论计算采用 CRYSTAL14 软件包计算周期性晶体结构。鉴于富含氢键的体系中伦敦色散作用力不可忽略, 在传统的 Becke-3-Lee-Yang-Par(B3LYP)泛函和 Perdew-Burke-Ernzerhof(PBE)泛函的基础上添加了伦敦色散校正系数, 构建了 B3LYP-D 和 PBE-D 模型。计算得到吸收峰位平均绝对误差(AAE)分别为苯甲酸 0.073(B3LYP-D)、0.096(PBE-D), 山梨酸 0.039(B3LYP-D)、0.047(PBE-D), 木糖醇 0.023(B3LYP-D)、0.087(PBE-D), 相比于色散校正前的模型, AAE 降低了 0.03~0.1。对比了 B3LYP-D 和 PBE-D 模型的计算时长, B3LYP-D 泛函耗时均为 PBE-D 泛函的二倍以上。结果表明, 经色散校正后的模型适用于富含氢键的系统, 能够提高模拟光谱预测精度, 其中 B3LYP-D 模型预测精度更高但较为耗时, PBE-D 泛函预测精度略低于 B3LYP-D 模型, 但计算速度更快。振动来源分析表明, 三种食品添加剂的振动模式均以晶格整体的平移和旋转为主, 分子内的振动贡献较小。研究提出的色散校正模型对于其他类似体系的理论研究具有重要的参考价值。

关键词 太赫兹光谱; 食品添加剂; 周期性晶体结构; 氢键系统; 色散校正

中图分类号: O433.5 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2021)01-0100-05

引言

食品添加剂的违规使用或滥用会对身体健康产生严重的危害, 传统的高效液相色谱法(HPLC)等检测方法存在耗时长、操作难等缺点, 太赫兹光谱技术具有方便快捷、准确率高等优势, 被广泛应用于食品添加剂检测领域。目前, 食品添加剂的太赫兹光谱研究多侧重于实验测量及结合化学计量学方法的定性定量分析^[1-3], 缺乏深入的理论研究。

梁成森等^[4]利用太赫兹时域光谱仪测试了甜味剂 D-木糖及其加氢后得到的木糖醇, 理论计算采用 B3LYP 泛函, 与实验结果吻合较好且能够准确预测木糖醇在低频处相对于 D-木糖的红移现象。Ge 等^[5]利用 GAUSSIAN09 软件结合 B3LYP/6-311++G(d, p)算法对苯甲酸等六种抗氧化剂的单分子进行计算, 部分吸收峰与实验结果吻合较好, 但无法预测分子间作用力或晶格整体振动模式。以往的研究多基于

B3LYP 泛函, 没有根据研究体系的结构特点对比不同泛函的计算性能^[6]。近年来, 许多研究表明广义梯度近似(GGA)泛函在太赫兹光谱计算中表现出良好的计算性能^[7-8], 尤其是对于富含氢键的体系。King 等^[9]利用 GGA 泛函中的 PBE 泛函对腺嘌呤和胸腺嘧啶的共晶体进行理论模拟, 由于共晶体中存在弱色散作用力, 在 PBE 泛函的基础上添加了色散校正系数, 结果表明, 色散修正后的 PBE 泛函对于得到收敛的晶体结构和高质量的振动频谱有显著影响。然而, 对于不同计算方法添加色散修正后的计算精度、计算时长, 缺乏深入的对比分析。

本文采用太赫兹时域光谱仪测试了三种富含氢键的食品添加剂, 即苯甲酸、山梨酸和木糖醇, 理论计算采用 CRYSTAL14 软件包构建周期性晶体结构。考虑到描述分子间相互作用的色散作用力对氢键系统至关重要, 在传统的 B3LYP 和 PBE 泛函基础上添加了色散校正系数, 建立了 B3LYP-D 和 PBE-D 模型, 对比分析了色散校正前后四种模

收稿日期: 2019-12-20, 修订日期: 2020-04-11

基金项目: 国家自然科学基金项目(31200541), 江苏省自然科学基金项目(BK20161526)资助

作者简介: 李天莹, 1994 年生, 南京林业大学信息科学技术学院硕士研究生 e-mail: m18362985762@163.com

* 通讯作者 e-mail: jiangling@njfu.edu.cn

型的预测精度。从结构变化、计算精度、计算时长等多个角度评估了 B3LYP-D 和 PBE-D 模型的计算性能。

1 实验部分

1.1 仪器

采用日本 Advantest 公司 TAS7500SP 型太赫兹时域光谱仪 (THz-TDS), 仪器包含透射、反射、衰减全反射三个模块, 本文采用透射模块。系统分辨率 7.6 GHz, 动态范围高于 60 dB, 理想状态下可测得 0.1~4 THz 范围内的高质量光谱。仪器自带样品腔, 测量时通过充入氮气减少空气中水汽的影响, 所有测试在室温下进行。

1.2 样品

所用样品 (纯度 $\geq 99\%$) 均购于 Sigma-Aldrich 公司, 使用前未经进一步纯化, 三种食品添加剂的结构信息如表 1 所示。为了将吸收峰强度降低到合适的水平, 将苯甲酸、山梨酸和木糖醇分别与聚乙烯以 1:10 的质量均匀混合, 混合后的粉末放置于直径为 13 mm 的真空模具中压制成厚度约为 1 mm 的薄片。同时将 100 mg 的聚乙烯粉末按照相同的方法压制成相同规格的薄片, 作为参考背景。

表 1 三种食品添加剂的结构信息

Table 1 The structures of three food additives

食品添加剂	分子式	分子量	晶体空间组	晶体类型
苯甲酸	C ₇ H ₆ O ₂	122.12	P2 ₁ /c	单斜晶体
山梨酸	C ₆ H ₈ O ₂	112.13	C2/c	单斜晶体
木糖醇	C ₅ H ₁₂ O ₅	152.15	P2 ₁ 2 ₁ 2 ₁	正交晶体

2 理论计算

2.1 模型的建立

采用北京并行科技股份有限公司提供的 HPC 系统 (基于中国国家网格的超算云平台) 进行计算, 利用 ParaCloud (并行云桌面) 平台运行 CRYSTAL14 软件^[12]。

理论计算初始结构来源于剑桥晶体数据库 (CCDC), 在室温 (283~303 K) 下获得。CRYSTAL14 软件依据平移对称性构建周期性晶胞构型, 相比于单分子、多聚体及单晶胞构

型, 更能还原物质的实际形态。在传统的 B3LYP、PBE 泛函基础上添加了 GRIMME 伦敦色散校正系数, 建立了 B3LYP-D 和 PBE-D 模型, 基组函数均采用 6-311G(d, p)。

2.2 模型评价方法

从结构优化、计算精度、计算时长等多角度分析不同计算模型的性能。其中结构优化部分采用均方根偏差 (root-mean-square deviation, RMSD) 来评估计算模型, 计算精度部分采用平均绝对误差 (average absolute error, AAE) 作为评价标准, 见式 (1)

$$\text{RMSD} = \sqrt{\frac{\sum_{i=1}^n (p_{\text{cal},i} - p_{\text{exp},i})^2}{n}} \quad (1)$$

其中, $p_{\text{cal},i}$ 和 $p_{\text{exp},i}$ 分别代表几何结构参数的计算值和对应的实验值, RMSD 的值越小, 证明匹配效果越好, 得到的几何构型越稳定。

$$\text{AAE} = \frac{\sum_{j=1}^n |p_{\text{cal},j} - p_{\text{exp},j}|}{n} \quad (2)$$

其中, $p_{\text{cal},j}$ 和 $p_{\text{exp},j}$ 分别代表吸收峰位置 (THz) 的计算值和对应的实验值, AAE 越小, 证明计算模型的模拟精度越高, 计算结果越可靠。

3 结果与讨论

3.1 收敛标准

CRYSTAL14 计算过程包含两个重要的收敛标准: 自洽场收敛阈值 (TOLDEE) 和双电子积分截断标准 (TOLINTEG), 对计算时长影响较大。表 2 为不同收敛标准下基于 B3LYP 算法的苯甲酸模拟结果。当 TOLDEE 为默认值 10^{-6} , TOLINTEG 为默认值 (6, 6, 6, 6, 12) 时, 计算结果出现虚频 (-0.25 THz)。当 TOLDEE = 10^{-9} , TOLINTEG = (8, 8, 8, 8, 16) 时, 吸收峰位的平均绝对误差 (AAE) 为 0.103, 可以满足基本的计算需求。进一步提高收敛标准至 TOLDEE = 10^{-12} 、TOLINTEG = (10, 10, 10, 10, 20), 模拟精度没有明显提高, 且计算时间为原来的二倍以上。综合考虑, 后续计算收敛标准设置为 TOLDEE = 10^{-9} 、TOLINTEG = (8, 8, 8, 8, 16)。

表 2 不同收敛标准下苯甲酸的模拟结果

Table 2 Simulation results of benzoic acid under different convergence criteria

TOLDEE	TOLINTEG	Exp. /THz	B3LYP/THz	AAE	Time/s
10^{-6}	6, 6, 6, 6, 12		-0.25, 0.72, 1.21, 1.72	—	5 281 634.730
10^{-9}	8, 8, 8, 8, 16	0.64, 0.88, 1.08, 1.88, 2.37	0.78, 1.23, 1.81, 2.28	0.103	11 382 555.395
10^{-12}	10, 10, 10, 10, 20		0.79, 1.24, 1.87, 2.27	0.090	28 518 529.085

3.2 计算结果

3.2.1 结构优化

稳定的几何结构是频率计算的前提, 优化后的结构与实验数据越接近, 证明得到的结构越稳定。为便于举例说明,

图 1 给出了苯甲酸的晶胞结构、氢键二聚体及原子标号, 其中虚线部分表示分子间氢键。苯甲酸晶体为单斜晶体, 空间组为 P2₁/c, 单晶胞包含四个分子。CRYSTAL14 程序通过识别空间组号自动生成对称算子, 结合分子坐标、晶格参数

等信息还原晶体的实际结构。氢键是指氢原子介于两个电负性较大的原子之间,形成的一种 X—H...Y 型分子间或分子内相互作用。在苯甲酸晶体中,存在大量的分子间氢键,即位于二聚体羧基之间的 O—H...O 型氢键,与色散作用力密切相关,考虑色散作用有利于得到收敛的结构以及高质量的 THz 光谱,因此在传统泛函的基础上添加了色散校正系数^[11]。

表 3 给出了苯甲酸基于 B3LYP-D 和 PBE-D 泛函优化后的结构与实验值的对比,包含键长、键角及晶胞体积变化。采用同样的计算方法可以得到木糖醇优化后键长值 RMSD 为 0.031 9(B3LYP-D)和 0.005 5(PBE-D),键角值 RMSD 为 0.696(B3LYP-D)和 0.749(PBE-D),体积变化为 -6.53%(B3LYP-D)和 -8.80%(PBE-D)。结果表明,两种经色散校正后的泛函均能得到较为稳定的结构,其中 PBE-D 泛函计算得到的结构体积更小。

需要注意的是,CRYSTAL14 支持两种结构优化方式:完全优化方式更为精确,能够同时优化晶格参数和原子坐标,但对于初始结构和基态相距较远的体系,会造成长时间无法收敛且浪费大量机时的情况;固定晶格参数法只优化原子坐标,精度略低但能够保证结构快速收敛。根据三种食品添加剂各自的体系特点,苯甲酸和木糖醇采用完全优化方式,而山梨酸采取固定晶格参数法,(CCDC 中山梨酸初始体

积为 1 191.175 Å³,优化后参考基态体积为 595.588 Å³,体积预计减小 50%)。山梨酸优化后键长值 RMSD 为 0.0111(B3LYP-D)和 0.0147(PBE-D),键角值 RMSD 为 0.860(B3LYP-D)和 0.749(PBE-D),与实验结果较为吻合,能够满足计算需求。

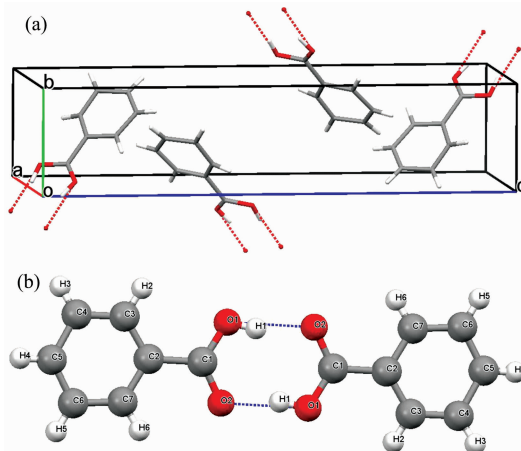


图 1 苯甲酸的晶胞结构(a)及其氢键二聚体(b)
Fig. 1 Molecular packing in unit cell (a) and hydrogen bonding dimer of benzoic acid (b)

表 3 苯甲酸优化后结构参数与实验值对比

Table 3 Optimized structural parameters of benzoic acid compared with experimental values

Bond length/Å				Bond angle/deg			unit cell parameters/Å and volumes/Å ³				
	Exp.	B3LYP-D	PBE-D		Exp.	B3LYP-D	PBE-D	Exp.	B3LYP-D	PBE-D	
C1-O1	1.275	1.317	1.319	O1-C1-O2	123.22	123.11	123.59	a	5.077	5.427	5.348
C1-O2	1.263	1.231	1.245	O1-C1-C2	120.27	122.35	121.60	b	4.893	4.998	4.924
C1-C2	1.484	1.482	1.483	O2-C1-C2	117.67	114.54	114.81	c	21.100	20.826	20.808
C2-C3	1.390	1.397	1.403	C1-C2-C3	118.79	118.61	118.60	V	515.217	557.863	539.808
C3-C4	1.387	1.392	1.393	C2-C3-C4	120.45	120.05	120.06	ΔV		+8.28%	+4.77%
C4-C5	1.390	1.394	1.400	C3-C4-C5	119.90	120.03	120.07				
C5-C6	1.392	1.387	1.400	C4-C5-C6	119.75	120.10	120.03				
C6-C7	1.401	1.396	1.398	C5-C6-C7	119.68	120.03	120.08				
C7-C1	1.392	1.401	1.404	C6-C7-C2	119.85	119.79	119.78				
				C7-C2-C1	118.33	120.39	120.43				
RMSD		0.018 3	0.017 6	RMSD		1.372	1.223				

3.2.2 频率计算

频率计算可以得到研究对象在 THz 波段的吸收峰,从而进行深入的振动来源分析,本文采用 MOLDRAW 软件分析吸收峰的振动来源。采用上述收敛标准 TOLDEE=10⁻⁹、TOLINTEG=(8, 8, 8, 8, 16),计算结果均无虚频。

理论计算得到的吸收峰位与实验结果的匹配程度能够最直观的反映计算模型的精度,表 4 对比了三种食品添加剂基于四种模型的计算结果及对应的振动模式。结果表明,色散校正模型能够提高吸收峰预测精度,其中 B3LYP-D 泛函得到的结果更优。振动来源分析表明,三种食品添加剂的吸收峰振动来源以晶格整体的平移或旋转为主,分子内振动模式贡献较小。需要说明的是,振动模式往往是耦合的,表 4 列

出的是吸收峰主要的振动模式,而省略了振动强度较弱的模式,例如山梨酸在 1.79 THz 处的主要振动模式为晶格沿 a 轴旋转,但该频段处伴随着—CH₃ 强度较弱的扭转。

图 2 给出了基于不同模拟方法的吸收峰位平均绝对误差(AAE),相比于色散校正前的模型,色散校正后的模型 AAE 降低了 0.03~0.1。结果表明,提出的色散校正模型能够提高模拟光谱预测精度,其中, B3LYP-D 模型精度更高。然而,理论计算仍存在误差,一方面源于振动频率计算基于绝对零度,而实验数据则在室温下获得;另一方面,对于高估或低估振动强度的情况,可能是由于优化前后分子结构的细微改变,同时实验测试中食品添加剂与聚乙烯混合比例的不同也会对吸收峰强度造成影响。

表 4 吸收峰位及对应的振动模式

Table 4 The position of peaks (THz) and vibration modes

食品添加剂	Exp.	B3LYP	B3LYP-D	PBE	PBE-D	Description
苯甲酸	0.64	—	—	—	—	—
	0.88	0.78	0.88	0.98	0.92	晶格沿 a 轴平移
	1.08	1.23	1.33	1.34	1.36	晶格沿 a 轴旋转
	1.88	1.81	1.91	1.95	1.93	苯环扭转
	2.37	2.28	2.36	2.43	2.39	a-c 面内平移
	AAE	0.103	0.073	0.123	0.096	
山梨酸	0.95	—	—	—	—	—
	1.20	1.09	1.16	1.32	1.18	晶格沿 a 轴旋转
	1.43	1.39	1.39	—	—	晶格沿 b 轴旋转
	1.53	1.60	1.48	1.69	1.51	晶格沿 c 轴平移
	1.79	1.71	1.71	1.89	1.75	晶格沿 a 轴旋转
	1.99	1.85	1.96	2.07	2.02	a-b 面内平移
	2.12	2.07	2.14	2.21	2.19	—CH ₃ 扭转
	2.32	2.24	2.29	2.42	2.25	晶格沿 c 轴平移
	2.56	2.48	2.58	2.64	2.64	b-c 面内平移
	AAE	0.081	0.039	0.104	0.047	
木糖醇	1.62	1.52	1.60	—	—	晶格沿 a 轴旋转
	1.85	1.69	1.87	1.91	1.92	CH ₂ —OH 面外摇摆
	2.45	—	—	2.16	2.31	a-c 面内平移
	2.64	2.53	2.61	2.73	2.69	晶格沿 a 轴平移
	AAE	0.123	0.023	0.147	0.087	

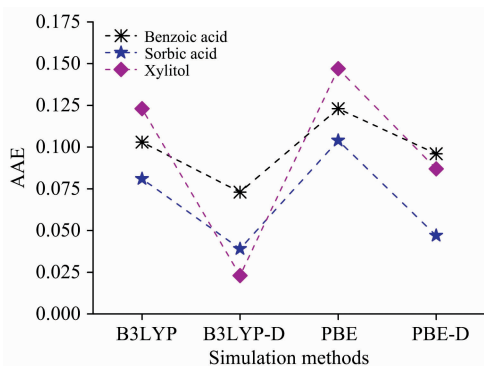


图 2 基于不同模拟方法的吸收峰位平均绝对误差

Fig. 2 Average absolute error of absorption peaks based on different simulation methods

3.2.3 计算时长

采用的服务器 CPU 型号为 Intel Xeon E5-2692 (2.20 GHz), 单节点可完成 24 核运算, 所有计算均基于 4 节点 96 核计算, 基于两种色散校正后的泛函的 CPU 总耗时如表 5 所示。三种食品添加剂基于 B3LYP-D 泛函的耗时分别为 PBE-D 泛函的 3.56 倍、2.89 倍、2.07 倍, PBE-D 泛函所用的计算机时明显更少, 对于较大的研究体系, 可以节省大量的计算时间。因此, 在体系较大、精度要求不高的情况下, 可以优先考虑 PBE-D 泛函。

表 5 两种色散校正后泛函的计算时长

Table 5 The CPU time consumed by two dispersion-correction functionals (s)

	B3LYP-D	PBE-D
Benzoic acid	28 518 529.085	8 010 021.645
Sorbic acid	35 165 303.664	12 177 328.971
Xylitol	34 759 784.620	16 796 994.306

4 结论

利用 THz-TDS 系统测试了苯甲酸、山梨酸、木糖醇的吸收光谱, 理论计算采用 CRYSTAL14 软件构建了基于周期性边界条件的晶体结构, 讨论了两个重要的收敛标准。根据三种食品添加剂富含氢键的结构特点, 在传统的 B3LYP 和 PBE 泛函的基础上添加了色散校正系数, 构建了 B3LYP-D 和 PBE-D 模型。结果表明, 色散校正模型能够提高吸收峰指认精度。B3LYP-D 模型的精度更高, 但较为耗时, PBE-D 模型的精度略低于 B3LYP-D 模型, 但可以节省计算机时, 需要根据具体的体系情况灵活运用。振动来源分析表明, 三种食品添加剂的振动来源均以晶格整体的平移和旋转为主, 是以往的单分子或多聚体模型无法预测到的。本文提出的 CRYSTAL14 部分参数设置以及适用于氢键系统的固态色散校正模型, 对于其他类似体系的理论研究具有重要的参考价值。

References

- [1] Haddad J E, Miollis F D, Sleiman J B, et al. *Analytical Chemistry*, 2014, 86: 4927.
- [2] Sun X D, Zhu K, Liu J B, et al. *Journal of Infrared, Millimeter, and Terahertz Waves*, 2019, 40: 466.
- [3] Du C M, Zhang X, Zhang Z Y. *Vibrational Spectroscopy*, 2019, 100: 64.
- [4] LIANG Cheng-sen, ZHAO Guo-zhong(梁成森, 赵国忠). *Spectroscopy and Spectral Analysis(光谱学与光谱分析)*, 2011, 31(2): 323.
- [5] Ge H Y, Jiang Y Y, Lian F Y, et al. *Optik*, 2016, 127: 4954.
- [6] Su T F, Huang R, Su Y Q, et al. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2015, 139: 456.
- [7] Yan H, Fan W H, Zheng Z P. *Optics Communications*, 2012, 285: 1593.
- [8] Zhang H, Zhang Z H, Zhao X Y, et al. *Chinese Physics B*, 2015, 24(7): 073301.
- [9] King M D, Ouellette W, Korter T M. *J. Phys. Chem. A*, 2011, 115(34): 9467.
- [10] Dovesi R, Orlando R, Erba A, et al. *Int. J. Quantum Chem.*, 2014, 114(19): 1287.
- [11] Grimme S, Antony J, Ehrlich S, et al. *J. Chem. Phys.*, 2010, 132: 154104.

Investigation on Terahertz Spectroscopy of Food Additives Based on Dispersion-Correction Functional Theory

LI Tian-ying, LI Chun, ZHANG Long, MA Qing-xiao, JIANG Ling*

College of Information Science and Technology, Nanjing Forestry University, Nanjing 210037, China

Abstract Illegal use or abuse of food additives can damage human health. Terahertz (THz) spectroscopy was widely used for food additives detection with fast and accurate etc. advantages, but lacked in-depth theoretical analysis. This paper explores to improve traditional methods to produce higher quality simulation spectra, and evaluate different models. THz spectra of sorbic acid, benzoic acid and xylitol were obtained by terahertz time-domain spectroscopy. Simulations were performed using CRYSTAL14 software to calculate the periodic crystals. Considering that dispersion force cannot be ignored in hydrogen-bonded systems, the dispersion correction term was used to augment the traditional functionals to construct the B3LYP-D and PBE-D models. The average absolute error (AAE) of benzoic acid absorption peak positions was 0.073 for B3LYP-D model, 0.096 for PBE-D model. For sorbic acid, AAE was 0.039 and 0.047, xylitol 0.023 and 0.087. The values of AAE were decreased by 0.03~0.1 compared with primitive functionals. The computation time of B3LYP-D model was more than twice of PBE-D model. Results showed that dispersion-corrected models can produce higher quality simulation spectra for hydrogen-bonded systems. The B3LYP-D model holds higher accuracy but is more time-consuming; the PBE-D model provides comparable accuracy to B3LYP-D model with much higher simulation speed. The spectral features were as signed as primarily lattice translations and rotations with lesser intramolecular torsions. The dispersion-corrected model proposed in this paper has important reference value for the theoretical research of other similar systems.

Keywords Terahertz spectra; Food additive; Periodic crystal structure; Hydrogen-bonded system; Dispersion-correction

(Received Dec. 20, 2019; accepted Apr. 11, 2020)

* Corresponding author