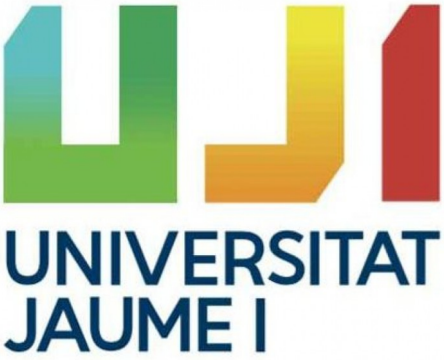
Computational Chemistry Laboratory:

Calculating the Energy Content of Food



Degree final project

Raluca Florina Florea

Supervisor: Sergio Martí Forés

2020/2021

Index

Abstract

The nutritional value of foods has been evaluated to assess whether the theoretical value obtained through computational calculations correlates with the one provided on the food labeling. For this purpose, the combustion reactions for several components of foods have been calculated using different computational methods: AM1, xTB, SCC-DFTB, B3LYP, and PBE/D3(BJ).

1. Introduction

1.1 Food

Food is part of our daily life. This is the reason why studies are carried out on its composition, either in the original or in the additional ingredients that have been incorporated: they are monitored from the time they are harvested/manufactured until they are finally consumed.

Chemistry is an important part in all of these processes since allows to identify the different compounds existing in the food, whether they are naturally present, such as nutrients, carbohydrates, fats, proteins, mineral salts, etc., or those which have been added through different processes to improve a particular aspect the food, such as resistance over time.

Within chemistry, the computational branch has evolved in recent years in such a way that it is increasingly possible to make useful predictions about chemical processes without going through the laboratory. It is also used to solve application-oriented molecular problems for a wide range of systems. Thus, computational chemistry has naturally found its way into applied sciences such as soil science, pharmacy, materials science, food science and industry.

In this study we will take into account the different compositions in kcal/100 g for several kind of food, mainly of carbohydrates, fats, and proteins, obtained from different food labels investigated, to be able to compare them later with the data obtained from calculations made using different computational methods.

1.2 Computational Methods

The two most common models used in molecular modeling are quantum mechanics (QM) and molecular mechanics (MM). These models allow the energy of any arrangement of atoms to be

calculated, and also provide how the energy of the system changes along the atom positions (or gradient).

While both are suitable for the study the thermodynamics of a reaction, only those methods based on QM are appropriate for studying the kinetics, since they account for the electrons and their rearrangements in the bond breaking and forming. In addition, MM depends on an initial parametrization of the chemical characteristics (such as bond energies) which are not needed in the QM based ones. In present study, we will rely on the QM approach.

1.2.1 Self Consistent Field (SCF)

Our starting point is the time-independent Schrödinger equation:

$$\hat{H} = \hat{T} + \hat{V} \qquad\qquad \hat{H}\,\Psi = E\,\Psi \tag{1}$$

where $\hat{T}$ is the operator accounting for the kinetic energy and $\hat{V}$ is the corresponding operator associated with the potential energy. When $\hat{H}$ is applied to the wave function describing the quantum system (Ψ), it provides the stationary energy of the system. In the case of atoms or molecules, the system is made up of nuclei and electrons, which allows each operator to be decomposed into several terms:

$$\hat{H} = \hat{T}_n + \hat{T}_e + \hat{V}_{nn} + \hat{V}_{ee} + \hat{V}_{ne} \tag{2}$$

were the subindex “n” accounts for the nuclei and “e” for the electrons.

The ability to solve the Schrödinger equation analytically disappears as the level of complexity of the problem increases slightly, and therefore several approximations have to be resorted to.

One is the Born-Oppenheimer approximation, which allows to skip the term associated with nuclear kinetics ($\hat{T}_n$), since it is based on the great difference between the masses of the electron and the proton. This approximation also affects the wave function, which now describes only the electrons of the system, and now depends parametrically on the coordinates of the nuclei.

The second difficulty arises from the repulsive interaction among the electrons ($\hat{V}_{ee}$), which prevents the variables for the different electrons to be separated, thus leading to a differential equation for each electron. The common solution to this problem is to treat toe repulsions among the electrons in an approximate way.

In addition, since electrons are fermion-like particles, there is an anti-symmetry constraint on the global wave function concerning the electron pair exchange. The most common way to introduce such behavior is to express the wave function as a determinant-type function from the wave functions for each electron, also taking into account the corresponding electron spin contribution (so-called spin orbitals). This is known as a Slater determinant:

$$\Psi = \frac{1}{\sqrt{N!}} \begin{vmatrix} \psi_1(r_1) & \psi_2(r_1) & \cdots & \psi_N(r_1) \\ \cdots & \cdots & \cdots & \cdots \\ \psi_1(r_N) & \psi_2(r_N) & \cdots & \psi_N(r_N) \end{vmatrix} \tag{3}$$

Regarding the treatment of electronic repulsion, the most appropriate methodology is the so-called Self Consistent Field, initially proposed by Hartree and later modified by Fock. In this approach, the system is reduced to a set of mono-electronic problems, in which the motion of each electron is solved considering that it moves under an average electrostatic field generated by the remaining of the electrons at rest. As a result, an improved wave function is obtained for that electron, which will be used in the subsequent resolution of the rest of the electrons. The overall process ends when the changes in the wave function of the system (Ψ) are less than a given tolerance

In the case of having molecules, consisting of two or more atoms, the procedure is similar: the molecular wave function for a given electronic configuration is expressed as a molecular wave function for a given electronic configuration is expressed as a Slater determinant formed from the spin-orbitals for each of the electrons in the molecule of the molecule. The main difference lies in the fact that the space parts of the spin-orbitals are usually expressed as a linear combination of basis functions, and the energy of the molecule is minimized as a function of the coefficients of this combination, using the variational method (Roothaan-Hall equations).

Simulations carried out at the atomic level with quantum mechanical calculations start with only the atomic numbers of the elements, and correctly represent the bonding between atoms (intra-molecular forces).

1.2.2 Basis sets

Instead of using the wave functions obtained for the hydrogen atom for the spatial part of the spin orbitals, other wave functions are used for reasons of computational efficiency. These are usually Gaussian-type function fits.
In most cases, the valence electrons are the ones which principally take part in the bonding process. In recognition of this fact, it is common to represent valence and core orbitals with different sets of fitted basis function, which leads to the commonly known split-valence basis sets.

The notation for the split-valence basis sets arises from the work of John Pople[1]. Under the generic representation X-YZg, X represents the number of primitive Gaussians comprising each core atomic orbital basis function. The Y and Z indicate that the valence orbitals are composed of two basis functions each, the first one composed of a linear combination of Y primitive Gaussian functions, the other composed of a linear combination of Z primitive Gaussian functions. In this case, the presence of two numbers after the hyphens implies that this basis set is a split-valence double-zeta basis set.

Here is a list of commonly used split-valence basis sets of this type:

- 6-31G
- 6-31G* (polarization functions on heavy atoms)
- 6-31G** (polarization functions on heavy atoms and on the hydrogen ones)
- 6-31+G (diffuse functions on heavy atoms)
- 6-31++G (diffuse functions on heavy atoms and on the hydrogen ones)
- 6-31+G** (combination of diffuse and polarization functions)

For our calculations we have chosen the 6-31G** basis set, also known as 6-31g(d,p).

1 https://en.wikipedia.org/wiki/Basis_set_(chemistry) [changed: 4th July 2021]

1.2.3 Density Functional Theory (DFT)

One of the problems of the Hartre-Fock method is the lack of correlation in the motion of the electrons. Indeed, when solving each electron individually, assuming the rest fixed, the component corresponding to the relative motion between the electrons is lost.

There are several methods, called post Hartee-Fock methods, that allow correcting this error to a greater or lesser extent. Among them are those based on the electron density as a functional.

The expression of the energy for a system in terms of the electron density, under the Born-Oppenheimer approximation, is:

$$E[\rho(r)]=T_{ni}[\rho(r)]+V_{n,e}[\rho(r)]+V_{e,e}[\rho(r)]+V_{n,n}(R)+\Delta T[\rho(r)]+\Delta V_{e,e}[\rho(r)] \quad (4)$$

where, $T_{ni}[\rho(r)]$ accounts for the kinetic energy of a set of non interacting electrons, $\Delta T[\rho(r)]$ represents the correction to the kinetic energy due to the interaction between the electrons and $\Delta V_{e,e}[\rho(r)]$ is the non-classical correction to the electronic repulsion. These last two terms can be grouped together giving rise to the exchange-correlation potential ($V_{XC}[\rho(r)]$), which would contain all those terms that are unknown, and which is approximated by a certain electron density functional:

$$E[\rho(r)]=T_{ni}[\rho(r)]+V_{n,e}[\rho(r)]+V_{e,e}[\rho(r)]+V_{n,n}(R)+V_{XC}[\rho(r)] \quad (5)$$

In general, it is common to express the exchange-correlation potential as the sum of each contribution:

$$V_{XC}[\rho(r)]=V_X[\rho(r)]+V_C[\rho(r)] \quad (6)$$

Among the different expressions for this equation we can find:

- Local density approximation (LDA): in which the ideal situation of a uniform electron gas is considered, and for which both exchange and correlation functionals are known.

- Generalized gradient approximation (GGA): comprising all those functionals that consider both the density and its gradient. Among the different options, the most common are the LYP[2], PBE[3] functionals for correlation, and the B88[4] for exchange. The most recent functionals called meta-GGA belong to this family of functionals, but incorporate higher-order density gradient terms.
- Hybrid functionals: these are functionals normally based on GGA or meta-GGA, but which in turn incorporate a certain amount, determined empirically, of the Hartree-Fock exchange. One of the most widely used is the so-called B3LYP[5], whose expression is:

$$E_{XC}^{B3LYP} = \left(0.20\,E^{HF} + 0.72\,E^{B88,GGA} + 0.08\,E^{S,LDA}\right)_X + \left(0.81\,E^{LYP,GGA} + 0.19\,E^{VWN-3/5,LDA}\right)_C \quad (7)$$

This functional has been the most common choice in the last years, but its being gradually replaced[6] by other more sophisticated functionals like M06 or ωB97M-V.

1.2.4 Semi-empirical Methods

Although semi-empirical methods can be considered to be more "approximate" than their *ab initio* counterparts, they are normally much faster and so can be applied to systems or processes that it would not otherwise be possible to investigate with QM methods.

It would be difficult to give a comprehensive overview of semi-empirical methods because a huge diversity of schemes, derived from different *ab initio* theories, have been developed. Anyways, the most common and accepted approximation relies on the neglect of certain interactions (SCF integrals) or the use of experimentally derived parameters to represent them. One of the most

2 C. Lee, W. Yang, R. G. Parr, Phys. Rev. B, 1988, 37, 785

3 J. P. Perdew, M. Ernzerhof, K. Burke, J. Chem. Phys., 1996, 105, 9982

4 A. D. Becke., Phys. Rev. A, 1988, 38, 3098

5 https://en.wikipedia.org/wiki/Hybrid_functional [changed: 29th June 2021]

6 N. Mardirossiana, M. Head-Gordon, Mol. Phys., 2017, 115, 2315

popular is the Modified Neglect Differential Overlap (MNDO), which is the scaffold for other methods such as AM1[7] or PM3[8].

In addition, is quite common that only the electrons belonging to the valence shell are considered (although all electrons are considered for the calculation of the repulsion term) and a minimal set of basis functions is employed with only one radial function for each value of the angular momentum.

The popular semi-empirical method AM1 is an attempt to improve the MNDO model by reducing the repulsion of atoms at close separation distances. The atomic core-atomic core terms in the MNDO equations were modified through the addition of off-center attractive and repulsive Gaussian functions. The complexity of the parameterization problem increased in AM1 as the number of parameters per atom increased from 7 in MNDO to 13-16 per atom in AM1.

The results of AM1 calculations are sometimes used as the starting points for parameterizations of forcefields in molecular modeling.

On the other hand, two of the most common methods which emerge from the DFT approach are SCC-DFTB[9] and xTB[10].

The SCC-DFTB method is initially obtained from a second order expansion of the electron density with respect to the charge density fluctuation (δρ) from a reference density (ρ_o):

$$E = T_{ni}[\rho(r)] + V_{ne}[\rho(r)] + V_{ee}[\rho(r)] + V_{XC}[\rho(r)] + V_{nn}(R)$$
$$E = \sum_i n_i \left\langle \psi_i \left| -\frac{1}{2}\nabla^2 + V_{ext} + \frac{1}{2}\int \frac{\rho(r')}{|r-r'|} dr' \right| \psi_i \right\rangle + E_{XC}[\rho(r)] + E_{nuc}(R) \quad (8)$$
$$\rho(r) \approx \rho_o(r) + \delta\rho_o(r)$$

7 https://en.wikipedia.org/wiki/Austin_Model_1 [changed: 20th February 2021]

8 https://en.wikipedia.org/wiki/PM3_(chemistry) [changed: 28th April 2020]

9 M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, Th. Frauenheim, S. Suhai, G. Seifert, Physical Review B., 1998, 58, 7260

10 C. Bannwarth, S. Ehlert, S. Grimme, J. Chem. Theory Comput. 2019, 15, 1652

where the value of the reference density is usually expressed as a combination of the atomic densities of the molecule: $\rho_o = \sum_A \rho_A$.

Introducing such an expansion in the energy expression, together with two further approximations for the calculation of the E_{Coul} and E_{Rep} terms, we will obtain the following equation:

$$
\begin{gathered}
E = \sum_i n_i \langle \psi_i | \hat{H}_o | \psi_i \rangle + E_{Coul} + E_{Rep} \\
E_{Rep} \approx \sum_i \sum_j V_{Rep}^{i,j}(R_{i,j}) \qquad E_{Coul} \approx \frac{1}{2} \sum_i \sum_j \gamma_{i,j}(R_{i,j}) \Delta q_i \Delta q_j \\
\gamma_{i,j}(R_{i,j}) \begin{cases} U_i & i = j \\ \dfrac{erf(C_{i,j} R_{i,j})}{R_{i,j}} & i \neq j \end{cases} \\
U_i = IE_i - EA_i \qquad C_{i,j}^2 = \frac{4 \ln 2}{FWHM_i^2 + FWHM_j^2} \qquad FWHM_i \approx \frac{1.329}{U_i} \\
\Delta q_i \approx \int_{Vol_i} \delta\rho(r) dr \qquad \delta\rho(r) = \sum_i \Delta q_i \delta\rho_i(r)
\end{gathered}
\tag{9}
$$

The other method, xTB, has a similar starting point to the previous case, in which the density of the system is again approximated from that of the nuclei forming the molecule (ρ_o) and its fluctuation ($\delta\rho$):

$$
E[\rho] \approx E^{(o)}[\rho_o] + E^{(1)}[\rho_o, \delta\rho] + E^{(2)}\left[\rho_o, (\delta\rho)^2\right] \tag{10}
$$

Within the zeroth order terms, we find:

$$
E^{(o)}[\rho_o] \approx \left(E_{rep}^{(o)} + E_{disp}^{(o)}\right) + \sum_A E_{A,core}^{(o)} + E_{A,valence}^{(o)} \tag{11}
$$

The terms of the summation, which extends to all atoms in the neutral state and with spherical symmetry, correspond to the atomic energies ($E_{A,core}^{(o)}$) whose value is constant and can be taken as a reference (making that term zero); and the energies of the valence orbitals.

For the first order terms, the dependence on δρ allows the atoms to have a non-zero net charge making them no longer neutral.

$$E^{(1)}[\rho_o, \delta\rho] \approx E^{(1)}_{disp} + \sum_A E^{(1)}_{A,valence} \quad (12)$$

For the second-order contributions:

$$E^{(2)}\left[\rho_o, (\delta\rho)^2\right] \approx E^{(2)}_{disp} + E^{(2)}_{ES} + E^{(2)}_{XC} \quad (13)$$

This term is where electrostatic interactions between individual atoms and exchange-correlation contributions (for a single center) are taken into account.
Combining the above equations, we obtain an expression for the total energy of the system that includes different contributions:

- Classical repulsion energy (E_{Rep}), which corresponds to a classical Coulomb-type term, affected by an exponential decay with distance.
- Classical scattering energy (E_{disp}), based on the second-order self-consistent scattering model D4 (which makes use of the electron density).
- Extended Hückel energy (E_{EHT}), which allows describing the formation of covalent bonds between the atoms forming the molecule.
- Electrostatic and exchange-correlation contributions, similar to those shown in equation [9], for which different approximations are provided.

1.2.5 Dispersion Corrections

A general drawback of all common DFT functionals (especially those GGA based ones), including hybrid ones, is that they can not properly describe long-range electron correlations that are responsible for van der Waals (dispersive) forces[11]. Since this kind of interactions play an important role in many chemical systems, a proper correction is needed.

11 S. Grimme, J. Comput. Chem. 2004, 25, 1463

In present work, we will use the DFT-D3[12] kind of dispersive corrections, which accounts for both two- and three-body interaction energies.

The expressions used resemble the standard $1/r^n$ ones (with n = 6, 8, …), but the different coefficients are parametrized for each kind of functional, and they are usually damped and modified for ensuring a smooth behavior.

1.2.6 Free Energy

The value of the free energy is associated with a given chemical equilibrium process and provides us with information about it, such as the degree of spontaneity or the ratio that will be established between reactants and products when equilibrium is reached. The calculations performed provide the internal energy of the system, so it will be necessary to resort to additional methods to calculate the value of the free energy.

Initially, we use Maxwell's relation between Gibbs energy and Helmholtz potential:

$$G = F + P\,V \tag{15}$$

On this equation, we will apply the following considerations:

The energy of the different levels will be expressed with respect to that of the fundamental level, which will become zero. will become zero. Thus, the partition function of the system when only the fundamental level is populated is equal to unity. the fundamental level is equal to unity (it would correspond to the case where the temperature at the boundary is 0 K).

We introduce the relation between Helmholtz potential and system partition function (Q), derived from the statistical definition of entropy:

$$F - F(0) = -k_B T \ln Q \tag{16}$$

Assuming that the system behaves as an ideal gas, the value of the free energy when the temperature limit is 0 K is independent of the variables controlling the measurement and

12 S. Grimme, J. Antony, S. Ehrlich, H. Krieg, J. Chem. Phys. 2010, 132, 154104

employing the Stirling approximation to approximate the value of the logarithm of a factorial number, we will obtain the Gibbs free energy:

$$G^o \approx U^o(0) - RT \ln\left(\frac{q}{N_A}\right) \tag{17}$$

were q is the molecular partition function, which depends on the nuclear degrees of freedom of the molecule (translations, rotations and vibrations) as well as of its accessible electronic states.

This approach is known as Rigid Rotor-Harmonic Oscillator (RRHO).

2. Results

In this study, the computational methods AM1, B3LYP, xTB, SCC-DFTB (or DFTB+) and PBE-D3(BJ) have been used to calculate the different free energies of amino acids, different dipeptides, stearin (to represent fats) and glucose (to represent sugars).

Initially, all the geometrical parameters (nuclei positions) for all the compounds were obtained, either through chemical editing programs (such as Jmol[13]) or from databases of chemical compounds[14]. Then, for each of the computational methods selected, they were optimized until convergence (which means a norm for the gradient vector lesser than 1.4 kJ/mol·Å, depending on the program used), followed by a normal modes analysis (known as "frequencies calculation"). The Gibbs free energy was obtained by applying the equation 17.

Once the Gibbs free energy for the different compounds was obtained, we proceeded to the determination of the corresponding combustion free energy. This process is straight forward, since it only makes use of the free energy for each species, plus the stoichiometric ones for molecular oxygen (O_2), carbon dioxide (CO_2), water (H_2O) and urea (CH_4N_2O).

For example, taking into account that the oxidation products obtained for glucose ($C_6H_{12}O_6$) and stearin ($C_{57}H_{110}O_6$) are simply CO_2 and H_2O, we can write (assuming that all reactions take place in the gas phase):

$$C_6H_{12}O_6 + 6\ O_2 \rightarrow 6\ CO_2 + 6\ H_2O \tag{18}$$

$$C_{57}H_{110}O_6 + \frac{163}{2}\ O_2 \rightarrow 57\ CO_2 + 55\ H_2O \tag{19}$$

13 http://jmol.sourceforge.net [visited: 25th June 2021]

14 https://pubchem.ncbi.nlm.nih.gov [visited: 25th June 2021]

Meanwhile for the phenylalanine-aspartic dipeptide we find:

$$C_{13}H_{16}N_2O_5 + 13\ O_2 \rightarrow (NH_2)_2CO + 12\ CO_2 + 6\ H_2O \qquad (20)$$

A summary of the results obtained for each species with the different methods is shown in Table 1. The results are shown in kcal/mol. The generic products of the oxidations are placed at the bottom of the table and highlighted in gray. For the shake of simplicity, we have deliberately skipped the sulphur based amino acids, cysteine and methionine, from present work (just avoiding the sulphur based metabolites). In addition, carbohydrates will be represented by a plain glucose molecule; meanwhile the glyceryl triestearate (Tristearin) for the different kind of fats, present in both animal and vegetal fats. Finally, being this work based on gas phase calculations, we have made neutral all the acid/base groups present in the amino acids species.

The optimized molecular geometry using B3LYP for Tristearin and the Arg-Asn dipeptide are displayed in figures 1 and 2, respectively.

Table 1: Free energies obtained with different methods for each species (kcal/mol).

				kcal/mol		
Molecule	**g/mol**	**B3LYP**	**PBE-D3(BJ)**	**xTB**	**DFTB+**	**AM1**
Ala	89,09	-203110,4	-202881,1	-13158,6	-10507,1	-54,7
Arg	174,20	-380507,9	-380074,6	-24871,1	-19799,4	28,1
Asn	132,12	-308957,2	-308624,1	-19200,7	-15386,9	-78,8
Asp	133,10	-321432,1	-321089,8	-19607,1	-15760,0	-137,7
Gln	146,14	-333615,5	-333246,2	-21170,5	-16926,4	-70,4
Glu	147,13	-346089,3	-345713,8	-21578,9	-17298,7	-129,8
Gly	75,07	-178455,7	-178259,9	-11188,8	-8967,6	-69,5
His	155,16	-344287,3	-343906,5	-21476,8	-17158,2	29,5
Ile	131,17	-277075,4	-276746,5	-19068,5	-15121,3	-22,7
Leu	131,17	-277072,1	-276743,2	-19066,4	-15119,7	-18,0
Lys	146,19	-311796,4	-311428,9	-21203,7	-16828,1	-12,6
Phe	165,19	-348054,0	-347650,1	-22451,7	-17882,6	20,9
Pro	115,13	-251667,9	-251380,2	-16475,6	-13115,5	-29,5
Ser	105,09	-250302,9	-250030,3	-15704,0	-12584,2	-98,7
Trp	204,23	-430600,8	-430110,9	-27292,2	-21753,9	72,5
Tyr	181,19	-395253,4	-394807,4	-25006,8	-19970,9	-21,6
Val	117,15	-252420,5	-252125,1	-17099,0	-13583,2	-31,3
Gluose	180,16	-431109,5	-430645,1	-27112,8	-21714,7	-198,1
Phe-Asp	280,28	-621530,4	-620842,9	-38886,8	-31084,2	-51,3
Arg-Asp	289,29	-653983,4	-653263,0	-41300,7	-32999,8	-45,3
Gly-Gln	203,20	-464115,0	-463606,0	-29182,8	-23337,8	-75,5
Asn-Gln	260,25	-594622,6	-593977,1	-37196,1	-29756,5	-83,9
Arg-Gln	302,33	-666161,0	-665416,1	-42860,2	-34164,8	24,5
Arg-Asn	288,30	-641515,3	-640804,4	-40897,5	-32630,2	12,4
Gly-Asn	189,17	-439459,1	-438988,8	-27214,3	-21801,4	-84,5
Tristearin	891,48	-1687040,0	-1684857,9	-126291,8	-99217,1	348,1
O2	32,00	-94339,1	-94249,6	-4969,5	-4093,7	-36,5
Urea	60,06	-141337,3	-141182,0	-8825,3	-7070,0	-21,7
H2O	18,02	-47952,2	-47898,5	-3180,6	-2557,2	-57,9
CO2	44,01	-118342,1	-118216,2	-6474,4	-5281,3	-86,2

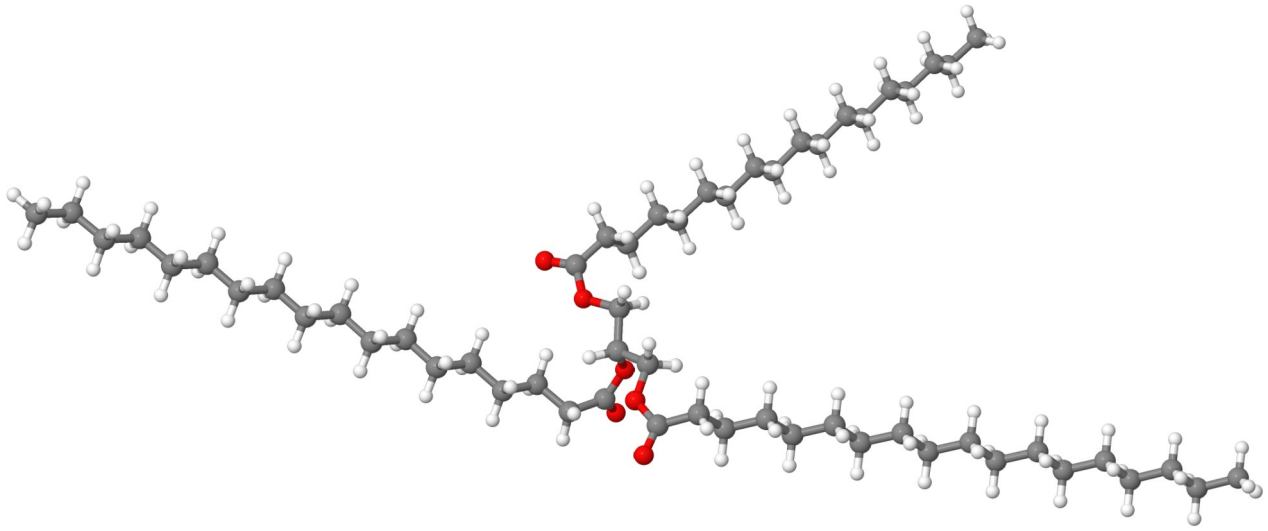


Figure 1: Molecular representation of Tristearin ($C_{57}H_{110}O_6$)

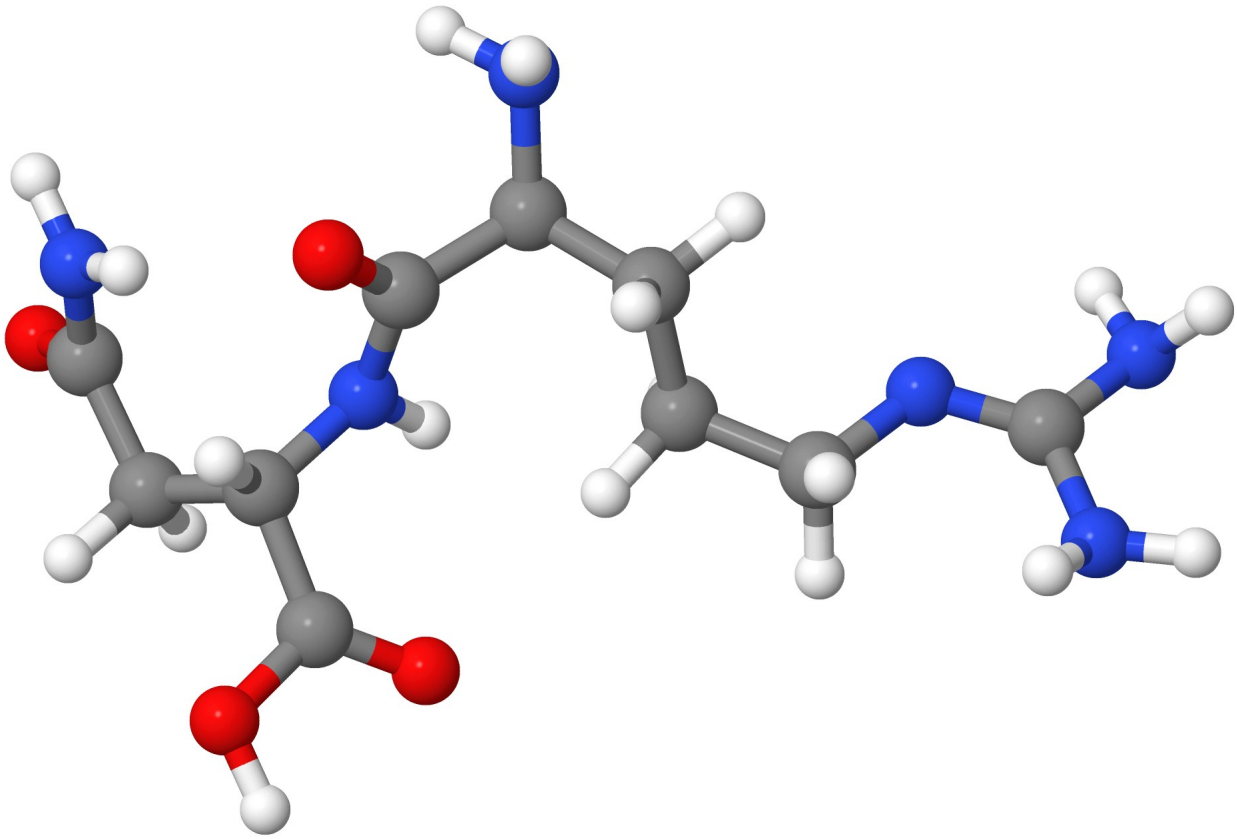


Figure 2: Molecular representation of the Arginine-Asparagine dipeptide (Arg-Asn, $C_{10}H_{19}O_5N_5$)

From the combustion energies of each of the compounds (shown in Table 2), we can proceed to perform the calculations with respect to the selected foods, checking whether the theoretical values obtained correspond to the ones found in their nutritional facts on the food labeling (expressed as energy obtained by the oxidation of 100 g of food).

Table 2: Combustion energies

			kcal/g		
Molecule	**B3LYP**	**PBE-D3(BJ)**	**xTB**	**DFTB+**	**AM1**
Ala	-3,1	-2,8	-5,4	-3,8	-2,3
Arg	-3,0	-2,7	-5,1	-3,6	-2,2
Asn	-2,2	-1,9	-3,8	-2,7	-1,6
Asp	-2,2	-1,9	-3,8	-2,8	-1,6
Gln	-2,9	-2,6	-5,0	-3,5	-2,1
Glu	-2,9	-2,5	-5,0	-3,6	-2,1
Gly	-1,9	-1,7	-3,4	-2,5	-1,4
His	-3,2	-2,8	-5,3	-3,7	-2,3
Ile	-5,1	-4,6	-9,0	-6,3	-3,9
Leu	-5,1	-4,6	-9,0	-6,3	-3,9
Lys	-4,4	-4,0	-7,6	-5,3	-3,2
Phe	-5,5	-5,0	-9,8	-6,7	-4,0
Pro	-4,4	-4,0	-7,6	-5,4	-3,2
Ser	-2,4	-2,1	-4,0	-3,0	-1,7
Trp	-5,2	-4,7	-9,1	-6,3	-3,8
Tyr	-4,9	-4,4	-8,5	-5,9	-3,5
Val	-4,6	-4,1	-8,1	-5,7	-3,5
Gluose	-3,4	-3,0	-5,6	-4,2	-2,5
Phe-Asp	-4,3	-3,9	-7,5	-5,3	-3,1
Arg-Asp	-2,9	-2,5	-4,8	-3,4	-2,1
Gly-Gln	-2,8	-2,5	-4,8	-3,5	-2,0
Asn-Gln	-2,7	-2,4	-4,7	-3,4	-2,0
Arg-Gln	-3,2	-2,8	-5,4	-3,8	-2,3
Arg-Asn	-2,8	-2,5	-4,8	-3,4	-2,1
Gly-Asn	-2,3	-2,0	-4,0	-2,9	-1,7
Tristearin	-8,1	-7,3	-14,2	-9,9	-6,1

For the analysis, 16 foods have been used, as shown in table 3:

Table 3: Nutritional facts

	Kcal/100g
Rosquilletas picos saladitos	449
Jamon serrano	247,8
Pizza Palacios sabor barbacoa	236
Pan	268
Filete de merluza del cabo	82
Rodajas de emperador	111
Helado nocilla de carte d'or	193
Lata atun	172
Cereales rellenos	432
Carne de hamburguesa	221
Carne picada	216
Pate Apis	255
Macarrones	349
Sirope de chocolate	279
Queso fundido	217
Capsula de café con leche	400

Once the combustion energies of each of the different chemical compounds have been obtained and knowing which foods will be used, the grams of each nutrient (proteins, fats and carbohydrates) present in their composition will be obtained from the label of their packaging. These will be shown in the tables below, making a table for each food.

The different results have been obtained as follows: calculations have been made to obtain the amounts of carbohydrates (glucose) and fats (tristerin) of each food. Then we have determined the amino acids, adjusting in each case the remaining quantity between the different options available: amino acids and dipeptides. The main reason for this modus operandi is the lack of a generic species for the protein, and the fact that we don't know which of the amino acids (or dipeptides) is predominant in each kind of food.

Rosquilletas picos saladitos						
	g	**B3LYP/6-31g(d,p)**	**PBE-D3(BJ)**	xTB	**DFTB+**	AM1
Carbohydrates	65,6	-226,4	-196,5	-364,2	-275,1	-162,9
Fats	15,5	-125,1	-113,2	-220,3	-153,5	-95,1
Proteins	10,5	-97,4	-139,7	***	-26,3	-191,3
Total		-448,8	-449,3	***	-454,8	-449,4

Jamon serrano						
	g	**B3LYP/6-31g(d,p)**	**PBE-D3(BJ)**	xTB	**DFTB+**	AM1
Carbohydrates	1,0	-3,5	-3,0	-5,6	-4,2	-2,5
Fats	12,2	-98,5	-81,3	-173,4	-110,9	-173,4
Proteins	33,5	-145,9	-164,2	-112,5	-127,3	-70,0
Total		-247,8	-248,4	-291,4	-242,4	-245,9

Pizza Palacios sabor barbacoa						
	g	**B3LYP/6-31g(d,p)**	**PBE-D3(BJ)**	xTB	**DFTB+**	AM1
Carbohydrates	31,0	-107,0	-93,0	-172,1	-130,2	-77,0
Fats	8,0	-64,6	-58,4	-113,7	-79,2	-49,1
Proteins	9,2	-64,5	-85,6	***	-25,8	-110,0
Total		-236,0	-237,0	***	-235,2	-236,1

Pan						
	g	**B3LYP/6-31g(d,p)**	**PBE-D3(BJ)**	xTB	**DFTB+**	AM1
Carbohydrates	54,0	-186,3	-162,0	-299,8	-226,8	-134,1
Fats	1,3	-10,5	-9,5	-8,0	-12,9	-8,0
Proteins	8,9	-71,1	-97,0	***	-30,3	-125,8
Total		-268,0	-268,5	***	-269,9	-268,0

Filete de merluza del cabo						
	g	**B3LYP/6-31g(d,p)**	**PBE-D3(BJ)**	xTB	**DFTB+**	AM1
Carbohydrates	0,5	-1,7	-1,5	-2,8	-2,1	-1,2
Fats	1,2	-9,7	-8,8	-17,1	-11,9	-7,4
Proteins	18,0	-74,9	-72,0	-68,2	-68,4	-72,9
Total		-86,3	-82,3	-88,0	-82,4	-81,5

Rodajas de emperador						
	g	**B3LYP/6-31g(d,p)**	**PBE-D3(BJ)**	xTB	**DFTB+**	AM1
Carbohydrates	0,0	0,0	0,0	0,0	0,0	0,0
Fats	4,2	-33,9	-30,7	-59,7	-41,6	-25,8
Proteins	18,0	-76,9	-79,2	-60,5	-68,4	-83,7
Total		-110,8	-109,9	-120,1	-110,0	-109,4

Helado nocilla de carte d'or						
	g	**B3LYP/6-31g(d,p)**	**PBE-D3(BJ)**	xTB	**DFTB+**	AM1
Carbohydrates	27,0	-93,2	-81,0	-149,9	-113,4	-67,1
Fats	7,3	-58,9	-53,3	-103,7	-72,3	-44,8
Proteins	3,8	-41,2	-58,2	***	-9,5	-81,1
Total		-193,3	-192,4	***	-195,2	-193,0

Lata atun						
	g	**B3LYP/6-31g(d,p)**	**PBE-D3(BJ)**	xTB	**DFTB+**	AM1
Carbohydrates	0,0	0,0	0,0	0,0	0,0	0,0
Fats	7,5	-60,5	-54,8	-106,6	-74,3	-46,0
Proteins	26,0	-111,1	-117,0	-87,3	-96,2	-126,8
Total		-171,6	-171,8	-193,9	-170,5	-172,9

Cereales rellenos						
	g	**B3LYP/6-31g(d,p)**	**PBE-D3(BJ)**	xTB	**DFTB+**	AM1
Carbohydrates	65,2	-225,0	-195,6	-362,0	-273,8	-162,0
Fats	15,8	-127,5	-115,3	-224,5	-156,4	-97,0
Proteins	5,7	-79,5	-120,3	***	-15,4	-173,2
Total		-432,0	-431,2	***	-445,7	-432,1

Carne de hamburguesa						
	g	**B3LYP/6-31g(d,p)**	**PBE-D3(BJ)**	xTB	**DFTB+**	AM1
Carbohydrates	1,5	-5,2	-4,5	-8,3	-6,3	-3,7
Fats	16,0	-129,1	-116,8	-227,4	-158,4	-98,2
Proteins	17,7	-86,6	-99,1	***	-60,2	-119,3
Total		-220,9	-220,4	***	-224,9	-221,3

Carne picada						
	g	**B3LYP/6-31g(d,p)**	**PBE-D3(BJ)**	xTB	**DFTB+**	AM1
Carbohydrates	1,7	-5,9	-5,1	-9,4	-7,1	-4,2
Fats	15,9	-128,3	-116,1	-225,9	-157,4	-97,6
Proteins	16,5	-81,8	-95,7	***	-56,1	-114,4
Total		-216,0	-216,9	***	-220,7	-216,2

Pate Apis						
	g	**B3LYP/6-31g(d,p)**	**PBE-D3(BJ)**	xTB	**DFTB+**	AM1
Carbohydrates	3,0	-10,4	-9,0	-16,7	-12,6	-7,5
Fats	23,5	-189,6	-171,6	-333,9	-232,7	-144,2
Proteins	8,2	-55,1	-74,3	***	-20,4	-103,0
Total		-255,1	-254,9	***	-265,7	-254,6

Macarrones						
	g	**B3LYP/6-31g(d,p)**	**PBE-D3(BJ)**	xTB	**DFTB+**	AM1
Carbohydrates	70,0	-241,6	-210,0	-388,6	-294,0	-173,9
Fats	1,5	-12,1	-11,0	-21,3	-14,9	-9,2
Proteins	12,0	-95,5	-128,4	***	-40,8	-165,6
Total		-349,2	-349,4	***	-349,7	-348,7

Sirope de chocolate						
	g	**B3LYP/6-31g(d,p)**	**PBE-D3(BJ)**	xTB	**DFTB+**	AM1
Carbohydrates	63,1	-217,7	-189,3	-350,3	-265,0	-156,7
Fats	1,3	-10,5	-9,5	-18,5	-12,9	-8,0
Proteins	2,5	-50,7	80,8	***	-6,3	-114,6
Total		-279,0	-118,0	***	-284,1	-279,3

Queso fundido						
	g	**B3LYP/6-31g(d,p)**	**PBE-D3(BJ)**	xTB	**DFTB+**	AM1
Carbohydrates	5,4	-18,6	-16,2	-30,0	-22,7	-13,4
Fats	17,3	-139,6	-126,3	-245,8	-171,3	-106,2
Proteins	9,8	-58,8	-76,4	***	-24,5	-97,5
Total		-217,0	-218,9	***	-218,5	-217,1

Capsula de café con leche						
	g	**B3LYP/6-31g(d,p)**	**PBE-D3(BJ)**	xTB	**DFTB+**	AM1
Carbohydrates	29,4	-101,5	-88,2	-163,2	-123,5	-73,0
Fats	20,0	-161,4	-146,0	-284,2	-198,0	-122,7
Proteins	21,2	-137,7	-165,4	***	-78,4	-204,4
Total		-400,5	-399,6	***	-399,9	-400,2

Finally, all the results obtained for all the methods have been represented in a single graph (Figure 3), to facilitate the comparison of the theoretical results among the different methods.

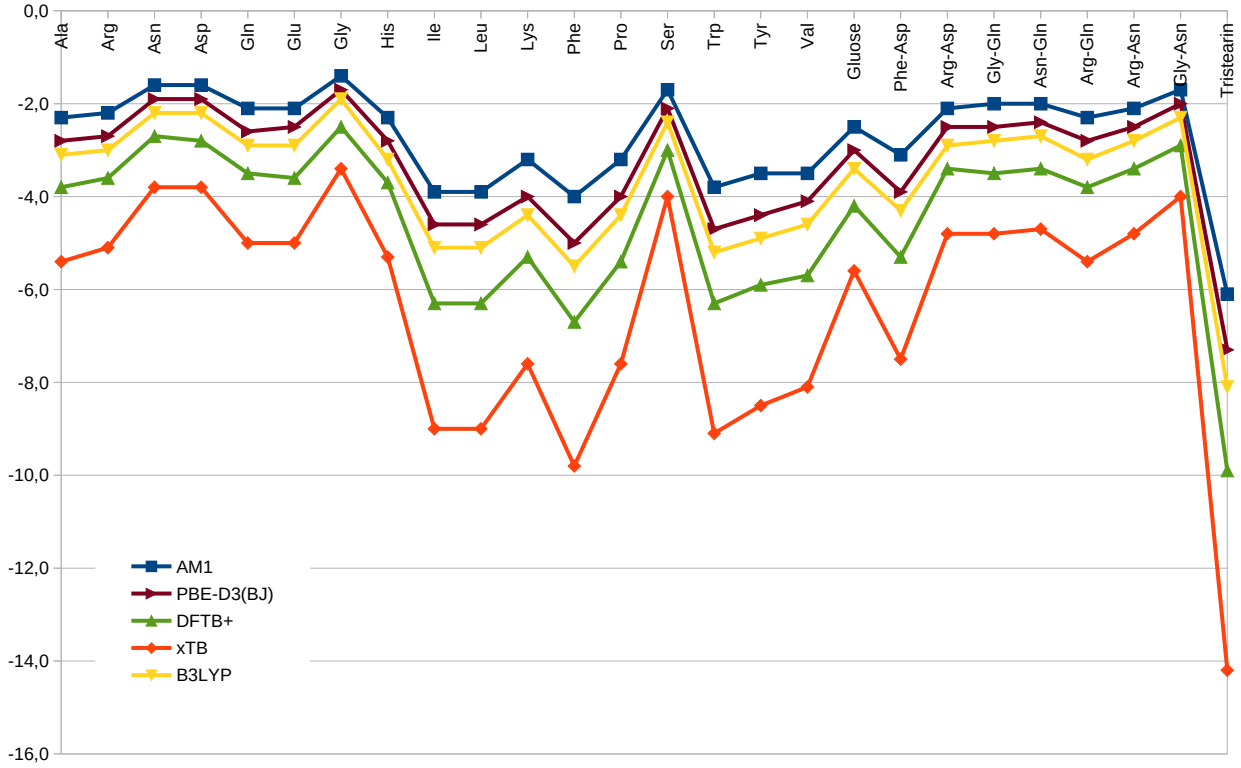


Figure 3: Computational methods comparison

3. Discussion

In the following section, all the energy values previously presented have been arranged by methodology, to perform both a validation against the experimental data, and a comparison about the goodness of each method.

We will initially analyze the data obtained with the hybrid B3LYP functional and 6-31g(d,p) basis set. This method is supposed to provide the most accurate results among all the selected in present work, thus these results will be used as a reference for the rest of methods. As can be seen in Table 4, the energy values are in good agreement with the experimental ones. Anyhow, we have to keep in mind that the protein content has been selected to fit with the remaining experimental value.

Table 4: B3LYP/6-31g(d,p) results

Food	kcal/100g	Carbohydrate+fat	Protein	Total
Rosquilletas picos saladitos	449,0	351,5	97,4	448,8
Jamon serrano	247,8	101,9	145,9	247,8
Pizza Palacios sabor barbacoa	236,0	171,5	64,5	236,0
Pan	268,0	196.83	71,1	268,0
Filete de merluza del cabo	82,0	11,4	74,9	86,4
Rodajas de emperador	111,0	33,9	76,9	110,8
Helado nocilla de carte d'or	193,0	152,1	41,2	193,3
Lata atun	172,0	60,5	111,1	171,6
Cereales rellenos	432,0	352,5	79,5	432,0
Carne de hamburguesa	221,0	134,3	86,6	220,9
Carne picada	216,0	134,2	81,8	216,0
Pate Apis	255,0	200,0	55,1	255,1
Macarrones	349,0	253,7	95,5	349,2
Sirope de chocolate	279,0	228,2	50,7	279,0
Queso fundido	217,0	158,2	50,8	217,0
Capsula de café con leche	400,0	262,9	137,7	400,5

The second method, in terms of accuracy of the results, corresponds to the dispersion corrected GGA PBE-D3(BJ) functional, with the same basis set. As can be seen in Table 5, are also in good agreement with the experimental ones, but it presents more (although small) discrepancies than

the B3LYP one. This can be also checked when comparing the combustion energies for both methods as presented in Figure 3: the PBE values are slightly smaller than those ones obtained with B3LYP. This is a very interesting result, since the PBE calculations are less CPU demanding than the B3LYP ones, and the overload arising from classical dispersion correction is very small. Thus, we can safely choose this functional without apparent loss of accuracy, at least for this kind of calculations.

Table 5: PBE-D3(BJ) results

Food	kcal/100g	Carbohydrate+fat	Protein	Total
Rosquilletas picos saladitos	449,0	309,7	139,7	449,3
Jamon serrano	247,8	84,8	164,2	248,9
Pizza Palacios sabor barbacoa	236,0	151,4	85,6	237,0
Pan	268,0	171,5	97,0	268,5
Filete de merluza del cabo	82,0	10,3	72,0	82,3
Rodajas de emperador	111,0	30,7	79,2	109,9
Helado nocilla de carte d'or	193,0	134,3	58,1	192,4
Lata atun	172,0	54,8	117,0	171,8
Cereales rellenos	432,0	310,9	120,3	431,2
Carne de hamburguesa	221,0	121,3	99,1	220,4
Carne picada	216,0	221,2	95,7	216,9
Pate Apis	255,0	180,6	74,3	254,9
Macarrones	349,0	221,0	128,4	349,4
Sirope de chocolate	279,0	198,8	80,8	279,5
Queso fundido	217,0	142,5	76,4	218,9
Capsula de café con leche	400,0	234,2	165,4	399,6

From now on, the remaining methods fall into the semi-empirical approaches, which makes them extremely fast, but prone to produce inaccurate results.

The next method is the xTB, which is a density functional based one. It is a broad range method, since allows to make calculations with almost all the elements of the periodic table. Unfortunately, quite probably due to this universality, it introduces a large amount of error in the results, at least for the reactions of the chemical species studied in present work.

Table 6: xTB results

Food	kcal/100g	Carbohydrate+fat	Protein	Total
Rosquilletas picos saladitos	449,0	584,5	***	***
Jamon serrano	247,8	178,9	112,5	291,4
Pizza Palacios sabor barbacoa	236,0	285.79	***	***
Pan	268,0	307,8	***	***
Filete de merluza del cabo	82,0	19,8	68,2	88,0
Rodajas de emperador	111,0	59,7	60,5	120,1
Helado nocilla de carte d'or	193,0	253,6	***	***
Lata atun	172,0	106,6	87,3	193,9
Cereales rellenos	432,0	586,5	***	***
Carne de hamburguesa	221,0	235,7	***	***
Carne picada	216,0	235,4	***	***
Pate Apis	255,0	350,6	***	***
Macarrones	349,0	410,0	***	***
Sirope de chocolate	279,0	368,8	***	***
Queso fundido	217,0	275,8	***	***
Capsula de café con leche	400,0	447,4	***	***

In fact, we have not been able to find an appropriate combination of peptides or dipeptides to fit the different results, as can be seen in Table 6 (values with asterisks). This behavior is also observed in the Figure 3: it is clearly the method with the largest deviation for the combustion values (referred to B3LYP) among all the proposed ones. One possible explanation would be, as stated before, its wide range of parametrization along almost all elements of the periodic table, which may be cannot account for the all complexity of each element.

For the second DFT based semi-empirical method, the SCC-DFTB, we have made use of the **mio-1-1** set of parameters, which covers organic compounds built on H, C, N, O, P and S. In contrast to the former case, we have been able to make predictions for all the foods. On the other hand, the final values resemble de experimental ones thanks to the variable protein values: if we compare the "constant" values for the carbohydrates (glucose) and fats (tristerin) we find different values than those obtained with B3LYP. Thus, although the global values are not bad, the protein composition largely differs from the other methods.

Table 7: DFTB+ results

Food	kcal/100g	Carbohydrate+fat	Protein	Total
Rosquilletas picos saladitos	449,0	428,6	26,3	454,8
Jamon serrano	247,8	115,1	127,3	242,3
Pizza Palacios sabor barbacoa	236,0	209,4	25,8	235,2
Pan	268,0	239,7	30,3	269,9
Filete de merluza del cabo	82,0	14,0	68,4	82,4
Rodajas de emperador	111,0	41,6	68,4	110,0
Helado nocilla de carte d'or	193,0	185,7	9,5	195,2
Lata atun	172,0	74,3	96,2	170,5
Cereales rellenos	432,0	430,3	15,4	445,7
Carne de hamburguesa	221,0	164,7	60,2	224,9
Carne picada	216,0	164,6	56,1	220,7
Pate Apis	255,0	245,3	20,4	265,7
Macarrones	349,0	308,9	40,8	349,7
Sirope de chocolate	279,0	277,9	6,3	284,1
Queso fundido	217,0	194,0	24,5	218,5
Capsula de café con leche	400,0	321,5	78,4	399,9

This behavior can be also inferred from Figure 3: the DFTB+ values reproduce the trend of the B3LYP ones, but the values are more negative. They are indeed almost in between the B3LYP and the xTB results. Being the products of the combustions small molecules, one could think about some kind of error compensation which is not present in large molecules (the reactant species). Anyhow, DFTB+ has demonstrated to be a better choice than xTB regarding DFT semi-empircal methods.

The last semi-empirical method used in this work is the AM1. Unlike the previous ones, is not based on the density functional, but on a plain Hartree-Fock approximation. This method is only applicable to molecules constituted by a reduced amount of chemical elements. In its present version it comprises: H, Li, Be, B, C, N, O, F, Na, Mg, Al, Si, P, S, Cl, K, Zn, Br and I. In any case, we must keep in mind that only the valence electrons are considered (just as in the other semi-empirical methods), and that it doesn't provides **d** kind orbitals (it is reduced to s and p kind ones). This can be a large source of error for the elements of the third row, but it is not a problem in our calculations, where we don't include any of them. The different results obtained are shown in Table 8.

Table 8: AM1 results

Food	kcal/100g	Carbohydrate+fat	Protein	Total
Rosquilletas picos saladitos	449,0	258,1	191,3	449,4
Jamon serrano	247,8	175,8	70,1	245,9
Pizza Palacios sabor barbacoa	236,0	126,1	110,0	236,1
Pan	268,0	142,1	125,9	268,0
Filete de merluza del cabo	82,0	8,6	72,9	81,5
Rodajas de emperador	111,0	25,8	83,7	109,4
Helado nocilla de carte d'or	193,0	111,9	81,1	193,0
Lata atun	172,0	46,0	126,8	172,9
Cereales rellenos	432,0	258,9	173,2	432,1
Carne de hamburguesa	221,0	101,9	119,3	221,3
Carne picada	216,0	101,8	114,4	216,2
Pate Apis	255,0	151,7	103,0	254,7
Macarrones	349,0	183,1	165,6	348,7
Sirope de chocolate	279,0	164,7	114,6	279,3
Queso fundido	217,0	119,6	97,5	217,1
Capsula de café con leche	400,0	195,8	204,4	400,2

The global results are quite engaging, but it is again due to the flexibility of the protein composition. If we compare the carbohydrate and fat calculations we face large deviations against B3LYP. Focusing on Figure 3, we can observe a similar behavior to DFTB+, but with larger reaction energies (more positive). Indeed, both semi-empirical methods constitute an upper and a lower boundaries to the B3LYP calculations.

4. Conclusions

We present a computational study comparing four density functional based methods, and a fifth one based on Hartree-Fock. The last three of them are based on semi-empirical approaches, which provides high computational efficiency, but with lacks for the results obtained.

We have applied all the methods to study the combustion reaction of several molecules, intended to represent different chemical compounds present in food. Therefore, we have reduced all the carbohydrates to a glucose molecule, or the various kinds of fats by the glyceryl triestearate. Unfortunately, this goal could not be reached for proteins, where reducing them to a single amino acid or dipeptide has become unreachable. On the other hand, we have opened the door to a fitting procedure which provides flexibility enough to reproduce the experimental data: select those amino acids or dipeptides which meet the remaining energy. So, we can picture it as a method to qualitatively establish the protein composition for a given food.

All in all, the best results are obtained by means of the B3LYP method as expected, since it is the most computationally demanding one of the proposed methods. On the other hand, it is gratifying to note that the B3LYP is closely followed by the dispersion corrected PBE (with the same basis set). This is a very interesting result, given the lower computational resources needed by the latter. Hence, one conclusion of present work is that the PBE-D3(BJ) methodology can be successfully applied to the study of oxidation reactions for organic molecules based on H, C, N and O.

Regarding the use of semi-empirical methods to carry out this kind os studies, the main conclusions can be summarized as follows: i) xTB is not the best choice for this kind of reactions or chemical species. Being a generalist method, seems to over-stabilize the products yielding to a large reaction energies. ii) Both DFTB+ and AM1 reproduce the trend observed in B3LYP or PBE, but with larger and smaller reactions energy, respectively. Thus, both could be used on the base of producing qualitative predictions.

Last but not least: from an educational or formative point of view, this work has offered to me the opportunity to complete several aspects of Physical Chemistry, which have been only briefly considered during the Chemistry degree.

Acknowledgments

I would like to thank Sergio Martí for his patient guidance, encouragement and advice during the development of this project.