LITERATURE REVIEWS

Transesterification

Biodiesel may be manufactured in batch or continuous processes by transesterification, also called alcoholysis or interesterification. In transesterification, one ester is converted to another. The reaction is catalyzed by either acid or base and involves a reaction with an alcohol. As typically practiced, a basic catalyst, such as sodium or potassium hydroxide, is used to convert the glycerol based triesters (or triacyl glycerides) which make up fats and oils to methanol based monoesters (or methylesters) yielding free glycerol as a byproduct.

A catalyst is usually used to improve the reaction rate and yield. As the reaction is reversible, excess alcohol is used to force the reaction equilibrium to the products side. Alcohols are primary and secondary monohydric aliphatic substances having 1-8 carbon atoms. Among the alcohols that can be used in the transesterification process are methanol, ethanol, propanol, butanol, and amyl alcohol. Methanol and ethanol are used most frequently, especially methanol because of its low price and its physical and chemical advantages – polar and shortest chain alcohol (Fangrui and Hanna, 1999).

Catalysts are classified as alkali, acid or enzymes. Alkali catalyzed transesterification is much faster than acid catalyzed (Freedman *et al.*, 1984). Acid catalyzed transesterification is suitable (Sprules and Price, 1950). The acid could be sulfuric acid, phosphoric acid, hydrochloric acid or organic sulfonic acid. Alkalis are sodium hydroxide, sodium methoxide, potassium hydroxide, potassium methoxide, sodium amide, sodium hydride, potassium amide and potassium hydride (Sprules and Price, 1950). Sodium methoxide is more effective than sodium hydroxide (Freedman *et al.*, 1984) because of the assumption that a small amount of water produced when mixing sodium hydroxide with methanol.

Alcohols are primary and secondary monohydric aliphatic alcohol having 1-8 carbon atoms (Sprules and Price, 1950). The alcohols that can be used in the transesterification process are methanol, ethanol, propanol or butanol. Methanol and ethanol are used most frequently, especially methanol. Because of its low cost and its physical and chemical advantages (polar and shortest chain alcohol). It can quickly react with triglyceride and sodium hydroxide (NaOH) is easily dissolved in it. To complete a transesterification stoichiometrically, 3:1 molar ratio of alcohol to triglyceride is needed (Taguchi, 1989). In addition, the ratio needs to be higher to force the equilibrium to a maximum ester yield. The reaction can be catalyzed by alkalis, acids or enzyme. The alkali includes sodium hydroxide (NaOH), potassium hydroxide (KOH), carbonate, sodium methoxide, and sodium ethoxide. The acid catalysts are sulfuric acid (H_2SO4), hydrochloric acid (HCl). Alkali catalyzed transesterification is much faster than acid catalyzed transesterification and is the most often used commercially (Freedman *et al.*, 1984).

For an alkali catalyzed transesterification, the glyceride and alcohol must be anhydrous (Wright *et al.*, 1994) because water makes the reaction change to saponification that produces soap that lower the yield of esters and make the separation of ester and glycerol and the water washing difficult. Low free fatty acids content in triglyceride is required for alkali catalyzed transesterification. If more water and free fatty acids are in the triglyceride, acid catalyzed transesterification can be appreciated (Keim, 1945). The triglyceride can be purified by saponification and then tranesterified using an alkali catalyst.

After transesterification of triglyceride, the products are a mixture of esters, glycerol, alcohol, catalyst and tri-, di- and monoglyceride. It is not easy to obtain pure ester because there are impurities in ester such as di- and monoglyceride (Ma, 1998). The monoglyceride causes turbidity in the mixture of esters. This problem is very obvious, especially for transesterification of animal fats such as beef tallow. The impurities raise cloud and pour points. The by-product, glycerol, needs to be recovered because of its value added for industrial chemical and it is separated by gravitational settling centrifuging.

Freedman *et al.* (1986) studied the transesterification of kinetics of soybean oil. The S-shaped plot of effects of time and temperature on ester formation for a 30:1 ratio of butanol and soybean oil (SBO), 1% of sulfuric acid (H_2SO_4) and 77-117 °C indicated that the reaction began at slow rate, proceeded at a faster rate and then slowed again as the reaction neared completion. For acid or alkali catalysis, the forward reaction followed pseudo-first-order kinetics for butanol:SBO = 30:1. However, for the alkali catalysis the forward reaction followed second-order-kinetics for butanol:SBO = 6:1. The reaction rate constant for the alkali catalyzed reactions were much higher than those for the acid catalyzed reactions. Rate constant increased with an increasing in the amount of catalyst used.

Wright *et al.* (1944) noted that the starting materials used for alkali catalyzed transesterification of glyceride must meet certain specification. Triglyceride should have an acid value less than 1 (Acid value are milligrams of alkali necessary to neutralize the free fatty acid in 1 gram of oil sample) and all materials should be anhydrous. If the acid value greater than 1, more sodium hydroxide (NaOH) was required to neutralize the free fatty acids. Water also caused soap formation, which consumed the catalyst and reduced catalyst efficiency. The soap caused an increase in viscosity, formation of gels and made the separation of glycerol difficult. Bradshaw and Muely (1994) also concerned about the importance of oils being dry and free of free fatty acids (<0.5%).

Ma *et al.* (1998a) studied the effect of free fatty acids and water on transesterification of beef tallow with methanol. The results showed that the water content of beef tallow should be below 0.06% w/w and free fatty acid content of beef tallow should be below 0.5% w/w in order to get the best conversion. Water content was more critical variable in the transesterification than free fatty acid.

One of the most important variables affecting the yield of ester is the molar ratio of alcohol to triglyceride. The Stoichiometric ratio for transesterification requires three mole of alcohol and one mole of glyceride to obtain three mole of fatty acid ester and one mole of glycerol. The molar ratio is associated with the type of catalyst used. Bradshaw and Meuly (1944) stated that the practical range of molar ratio was from 3.2 to 5.25:1 methanol to vegetable oil. The molar ratio of 4.8:1 was used in some examples, with the yield of 97-98%, depending upon the quality of the oils.

Freedman *et al.* (1984) studied the effect of molar ratio from 1:1 to 6:1 on ester conversion with vegetable oils. Soybean, sunflower, peanut and cotton seed oils behaved similarly and achieved highest conversions (93-98%) at a 6:1 molar ratio of methanol to vegetable oil.

Ma *et al.* (1998a) studied the effect of reaction time on transesterification of beef tallow with methanol. The reaction was very slow during the first minute due to the mixing of and dispersion of methanol into beef tallow. From one to five minutes, the reaction proceeded very fast. The production of beef tallow slowed down and reached the maximum value at about 15 minutes. The di- and monoglyceride increased at the beginning and then decreased. At the end, the amount of monoglyceride was higher than that of diglyceride.

Kusdiana *et al.* (1999) studied the transesterification of vegetable oils with methanol without catalyst. The clearly separation of vegetable oils and methanol was found. So the reaction at supercritical condition was used. The best conditions from the experiment were temperature at 350 °C and the molar ratio of alcohol to vegetable oils was 42:1.

Antolin *et al.* (2001) studied the transesterification of sunflower oil with ethanol and potassium hydroxide (KOH) was used as a catalyst. The results show that the best conditions were three times the stoichiometric quantity of methanol, 0.28% w/w of potassium hydroxide related to the quantity of starting oil and 70°C of reaction temperature. The product was washed for two times, one with slightly acid water by adding phosphoric acid and the other with pure water.

Molecular modeling

Li *et al.* (2002) studied an ab initio UMP2 (full) method has been used to study the reaction mechanism of C_2H_3 radical with O_2 ($a^1\Delta_g$) on the potential energy surface and energies have been calculated at Gaussian-3(G3) level. In this paper, they have optimized the geometry configuration of reactants, products, intermediates and transition states. They have also calculated the activation energies of elementary reactions along the reaction pathways. The results suggest that the reaction mechanism of this reaction involves the formation of three-, four- and five-membered rings and produces different reaction resultants, respectively. According to the calculated activation energies, the CH_2O and CHO are the main products. At the same time, the $C_2H_3 + O_2$ ($a^1\Delta_g$) reaction may form $CH_3 + CO_2$, $CH_2CO_2 + H$, $C_2H_2 + O_2H$, $C_2H_3O + O$ and COHCOH + H.

Shim *et al.* (2003) studied the transition states and the activation energies for the unobserved isomerization reactions between the three possible C2F4S isomers with divalent sulfur, trifluorothioacetyl fluoride 1, tetrafluorothiirane 2, and trifluoroethenesulfenyl fluoride 3, have been determined by ab initio Hartree-Fock (HF), Møller-Plesset second-order perturbation (MP2) calculations and by Gaussian-3 theory. The results show that the unobserved isomerization reactions are feasible. Furthermore, all three isomers should exist as stable species, but the unknown isomer, 3, is considerably less stable than the known isomers, 1 and 2.

So (1999) studied two possible reaction paths of the reaction of CF_3 with NO_2 in their ground electronic states have been studied using the G2 molecular orbital method. CF_3 and NO_2 are predicted to form, exothermically, two intermediates 1 and 2 (-75.8 and -72.1 kcal mol^{-1}). The decomposition of 2 to CF_2O + FON is less favoured thermodynamically with respect to the decomposition of 1 to CF_2O + FNO, and has a higher activation energy (37.5 and 26.9 kcal mol^{-1}). FON is predicted to disappear readily from the reaction products, in line with the experiment that no gas-phase absorption spectrum of FON has been reported.

Kondo *et al.* (2000) studied an ab initio calculations by using Gaussian-2 theory have been carried out for the reactions between halogen atoms and various fuel molecules, i.e. fluorine, chlorine, and bromine atoms vs. hydrogen, methane, ethane, ethylene, acetylene, ammonia, silane, dichlorosilane and phosphine. The activation energy for the reaction between a halogen atom and a fuel molecule seems to indicate whether the reaction between the fuel gas and the corresponding halogen gas occurs spontaneously when they are brought into contact to each other at room temperature.

Langley *et al.* (2005) studied the optimized geometries, relative free energies and related thermodynamic properties, harmonic frequencies, and dipole moments have been calculated at the HF and MP2 levels for ethynyl formate (1a), ethynyl acetate (1b), cyano formate, HCO_2CN (1c), cyano acetate (1d), S-ethynyl thioformate (2a), S-ethynyl thioacetate (2b), S-cyano thioformate (2c), S-cyano thioacetate (2d), N-ethynylformamide (3a), N-ethynylacetamide (3b), N-cyanoformamide (3c), and N-cyanoacetamide (3d) with the GAUSSIAN 98 program. For ethynyl formate, the calculation for 25 ^{o}C at the MP2/6-311++G(df,pd) level predicts that the Z isomer is more stable by 1.23 kcal/mol. For S-ethynyl thioformate, calculations at the MP2/6-311++G(2d,2p) level predict that the E isomer is favored by 0.71 kcal/mol at 25 ^{o}C. The E isomers of N-ethynylformamide and N-ethynylacetamide were found at all levels to be more stable than the Z isomers at 25 ^{o}C. For cyano formate and cyano acetate, calculations at the MP2/6-311++G(df,pd) level predict that the Z isomers are more stable at 25 ^{o}C by 1.50 and 2.72 kcal/mol, respectively. At this level and temperature, the Z isomers of 2c, 2d, 3c, and 3d are predicted to have free energies of 0.46, -0.07, 1.22, and 2.28 kcal/mol, respectively, relative to the E conformations. Z to E free-energy barriers at 25 ^{o}C of 8.63, 10.64, 17.63, 7.39, and 14.03 kcal/mol were calculated for 1a, 2a, 3a, 1c, and 3c at the HF/6-311G(d,p) level, and at the HF/6-311CG(d,p) level, the free-energy barrier for 2c was 7.08 kcal/mol.

Theory

This section will cover the contents of the mechanism of biodiesel production and the general concept of molecular modeling.

Vegetable oils

Vegetable oils have the potential to substitute for petroleum fuel in the near future. Vegetable oil fuels are still not competitive with petroleum fuel now because of their higher course. Anyhow, the increasing of petroleum fuel prices and the uncertainties problem availability make the using of vegetable oils in diesel engine of interest. Vegetables oil and fats are primarily water insoluble, hydrophobic substances that are made of one mole of glycerol and three moles of fatty acids and are commonly referred to as triglyceride (Sonntag, 1979). Fatty acids vary in carbon chain length and in the number of unsaturated bonds (double bonds). Table 1 shows a typical fatty acid composition of common oils. The fatty acid found in vegetable oils is summarized in Table 2.

Table 1 Typical fatty acid composition of common oils

Fatty acid	Soybean	Cotton seed	Palm	Lard	Tallow	Coconut
Lauric	0.1	0.1	0.1	0.1	0.1	46.5
Myristic	0.1	0.7	1.0	1.4	2.8	19.2
Palmitic	10.2	20.1	42.8	23.6	23.3	9.8
Stearic	3.7	2.6	4.5	14.2	19.4	3.0
Oleic	22.8	19.2	40.5	44.2	42.4	6.9
Linoleic	53.7	55.2	10.1	10.7	2.9	2.2
Linolenic	8.6	0.6	0.2	0.4	0.9	0.0

Source: Kincs (1985)

Table 2 Fatty acid composition of vegetable oils

Sample	Fatty acid composition, % by weight						
	16:0	16:1	18:0	18:1	18:2	18:3	Others
Cotton seed	28.7	0	0.9	13.0	57.4	0	0
Poppy seed	12.6	0.1	4.0	22.3	60.2	0.5	0
Rapeseed	3.5	0	0.9	64.1	22.3	8.2	0
Safflower seed	7.3	0	1.9	13.6	77.2	0	0
Sunflower seed	6.4	0.1	2.9	17.7	72.9	0	0
Sesame seed	13.1	0	3.9	52.8	30.2	0	0
Linseed	5.1	0.3	2.5	18.9	18.1	55.1	0
Wheat grain[a]	20.6	1.0	1.1	16.6	56.0	2.9	1.8
Palm	42.6	0.3	4.4	40.5	10.1	0.2	1.1
Corn marrow	11.8	0	2.0	24.8	61.3	0	0.3
Castor[b]	1.1	0	3.1	4.9	1.3	0	89.6
Tallow	23.3	0.1	19.3	42.4	2.9	0.9	2.9
Soybean	13.9	0.3	2.1	23.2	56.2	4.3	0
Bay laurel leaf[c]	25.9	0.3	3.1	10.8	11.3	17.6	31.0
Peanut kernel[d]	11.4	0	2.4	48.3	32.0	0.9	4.0
Hazelnut kernel	4.9	0.2	2.6	83.9	8.5	0.2	0
Walnut kernel	7.2	0.2	1.9	18.5	56.0	16.2	0
Almond kernel	6.5	0.5	1.4	70.7	20.0	0	0.9
Olive kernel	5.0	0.3	1.6	74.7	17.6	0	0.8
Coconut	9.7	0.1	3.0	6.9	2.2	0	65.7

[a] wheat grain oil contains 11.4% of 8:0 and 0.4% of 14:0 fatty acids.

[b] Castor oil contains 89.6% of ricinoloic acid.

[c] Bay laurel oil contains 26.5% of 12:0 and 4.5% of 14:0 fatty acids.

[d] Peanut kernel oil contains about 2.7% of 22:0 and 1.3% of 24:0 fatty acids.

Source: Goering *et al.* (1982)

Table 1 shows the typical fatty acid composition of common oils. A lauric acid has a high composition in coconut oil. A myristic acid has a high composition in coconut oil. A palmitic acid has a high content in palm oil. A stearic has a high composition in tallow. An oleic acid has a high composition in palm oil, lard, and tallow. A linoleic acid has a high composition in soybean oil and cotton seed oil. A linolenic acid has a high composition in soybean oil.

Table 2 shows the fatty acid composition of vegetable oils. A palmitic acid (C16:0) has a high composition in palm oil. A palmitoleic acid (C16:1) has a high composition in almond kernel oil. A stearic acid (C18:0) has a high content in tallow. An oleic (C18:1) has a high composition in rapeseed oil, sesame seed oil, hazelnut seed oil, almond kernel oil, and olive kernel oil. A linoleic acid (18:2) has a high composition in poppy seed oil, safflower seed oil, sunflower seed oil, wheat grain oil, corn marrow oil, soybean, and walnut kernel. A linolenic acid (C18:3) has a high composition in linseed oil.

As the undesirable properties of vegetable oils, so transesterification is needed for improving the properties of vegetable oils and overcome those drawbacks to reach approximately properties and performance of diesel fuel. Transesterification, also called alcoholysis, is the process of using and alcohol (e.g. methanol, ethanol, propanol) in the presence of catalyst such as sodium hydroxide (NaOH), potassium hydroxide (KOH) to break the molecule of the renewable oils chemically into methyl or ethyl ester of renewable oil with the glycerol as a by product. This process has been widely used to reduce the viscosity of triglyceride, enhancing the physical properties of renewable fuels to improve engine performance (Clark *et al.*, 1984).

Biodiesel

Biodiesel, defined as the mono-alkyl esters of fatty acid that is derived from vegetable oils or animal fats, has demonstrated a number of appreciating characteristic include reducing of exhaust emission. It is a substitute for diesel fuel was demonstrated by the inventor of the diesel engine, Rudolph Diesel, around the year 1900, when vegetable oil was proposed as fuel for engines. It is derived from a renewable, domestic resource, thereby relieving reliance on petroleum fuel imports. It is biodegradable and non-toxic. Compared to petroleum-based diesel, biodiesel has a more favorable combustion emission profile, such as low emissions of carbon monoxide, particulate matter and unburned hydrocarbons. Carbon dioxide produced by combustion of biodiesel can be recycled by photosynthesis, thereby minimizing the impact of biodiesel combustion on the greenhouse effect (Körbitz, 1999; Agarwal and Das, 2001). In brief, these merits of biodiesel make it a good alternative to petroleum based fuel and have led to its use in many countries, especially in environmentally sensitive areas (Zhang, 2003).

Transesterification

Transesterification or alcoholysis is the displacement of alcohol from an ester by another in a process similar to hydrolysis, except the alcohol is used instead of water. This process has been widely used to reduce the high viscosity of triglycerides. The transesterification reaction is represented by the general reaction as Fig. 1 (Meher *et al.*, 2004).

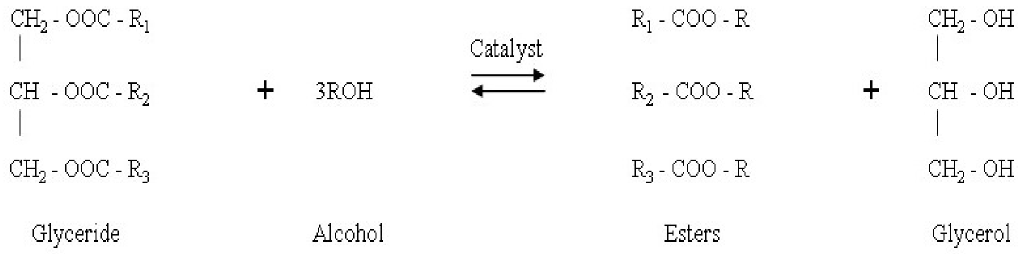


Figure 1 Transesterification of triglyceride with alcohol.
Source: Meher *et al.* (2004)

Transesterification of triglycerides produces fatty acid alkyl esters and glycerol. The glycerol layer settles down at the bottom of the reaction vessel. Diglycerides and monoglycerides are the intermediates in this process. The mechanism of transesterification is described in Fig. 2.

The stepwise reactions are reversible and a little excess of alcohol is used to shift the equilibrium towards the formation of esters. In presence of excess alcohol, the forward reaction is pseudo-first order and the reverse reaction is found to be second order. It was also observed that transesterification is faster when catalyzed by alkali (Meher *et al.*, 2004).

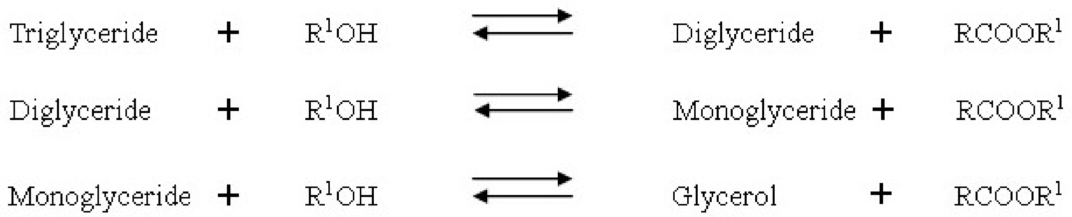


Figure 2 General reaction for transesterification of triglycerides.
Source: Meher *et al.* (2004)

The reaction rate and yield is normally improved by using catalyst. As the reaction is reversible, excess alcohol is used to shift the equilibrium to the products side. Catalysts are classified as alkali, acid or enzymes. Alkali catalyzed transesterification is much faster than acid catalyzed. However a glyceride has a higher free fatty content and more water. Acid catalyzed transesterification is suitable (Sprules and Price, 1950). The acid could be sulfuric acid, phosphoric acid, hydrochloric acid or organic sulfonic acid. Alkalis are sodium hydroxide, sodium methoxide, potassium hydroxide, potassium methoxide, sodium amide, sodium hydride, potassium amide and potassium hydride. Sodium methoxide is more effective than sodium hydroxide (Freedman *et al.*, 1984) because of the assumption that a small amount of water produced when mixing sodium hydroxide with methanol.

For an alkali catalyzed transesterification, the glyceride and alcohol must be anhydrous because water makes the reaction change to saponification. The saponification produces soap that lowers the yield of esters and makes the separation of ester and glycerol and the water washing difficult. Low free fatty acids content in triglyceride is required for alkali catalyzed transesterification. If more water and free fatty acids are in the triglyceride, acid catalyzed transesterification can be appreciated. The triglyceride can be purified by saponification and then transesterified using an alkali catalyst (Wright *et al.*, 1994).

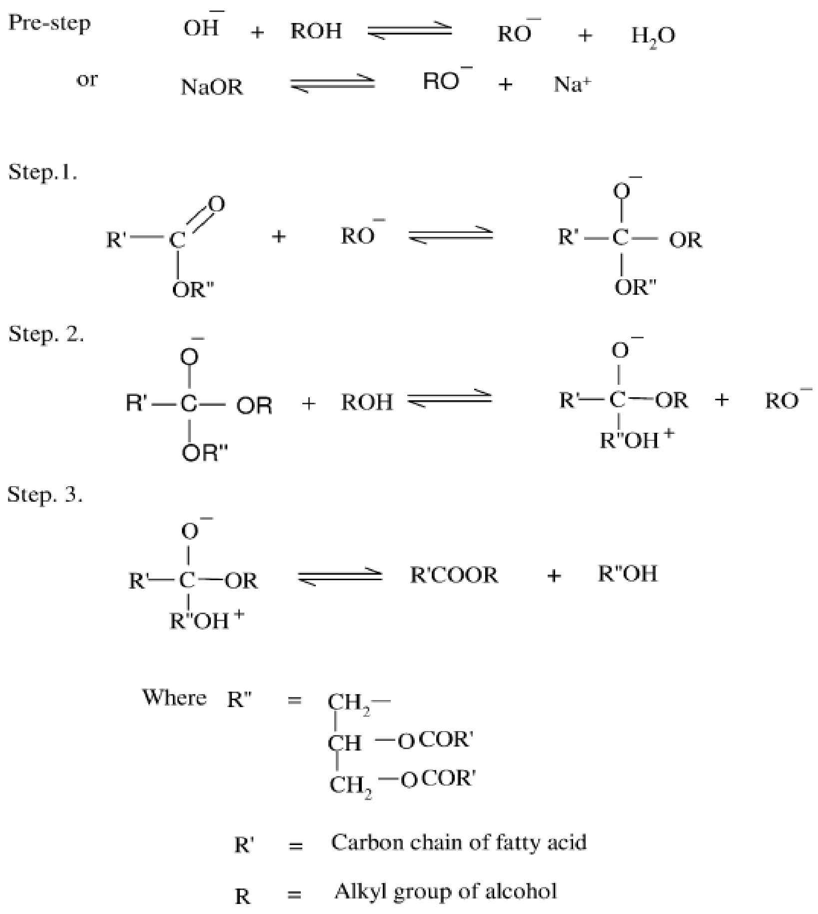


Figure 3 Mechanism of base catalyzed transesterification
Source: Meher *et al.* (2004)

After transesterification of triglyceride, the products are a mixture of esters, glycerol, alcohol, catalyst and tri-, di- and monoglyceride. It is not easy to obtain pure ester because of the impurities in ester such as di- and monoglyceride (Ma, 1998). The monoglyceride causes turbidity in the mixture of esters. This problem is very obvious, especially for transesterification of animal fats such as beef tallow. The impurities raise cloud and pour points. By-product, glycerol, needs to be recovered because of its value added for industrial chemical and it can be separated by gravitational settling centrifuging.

Triolein is a naturally occurring glyceride of oleic acid that is found in fats and oils. Oleic acid is a monounsaturated fatty acid in form of one *cis*- double bond between C_9 and C_{10} as shown in Fig. 4.

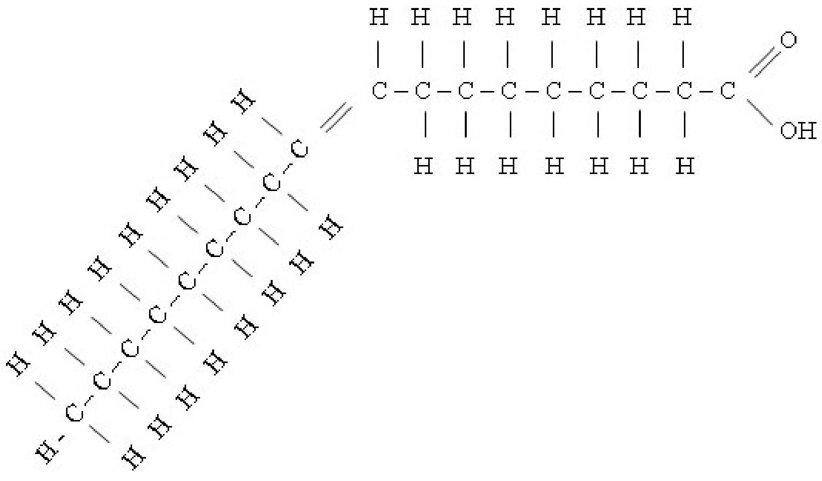
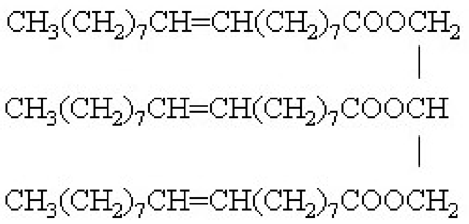


Figure 4 Molecular structure of oleic acid.

Source: http://www.chm.bris.ac.uk/webprojects2003/arbery/back.htm

The triglyceride in natural oil has many kind of fatty acid. Oleic acid is the one of them that has a high content in a various kind of natural oils or natural seeds such as rapeseed oil, palm oil, and sesame seed. This triolein molecule is a critical substance in transesterification. A molecule of triolein consists of carbon 57 atoms, oxygen 6 atoms, and hydrogen 104 atoms as shown in Fig. 5.



$$\begin{array}{r} CH_3(CH_2)_7CH{=}CH(CH_2)_7COOCH_2 \\ | \\ CH_3(CH_2)_7CH{=}CH(CH_2)_7COOCH \\ | \\ CH_3(CH_2)_7CH{=}CH(CH_2)_7COOCH_2 \end{array}$$

Figure 5 Molecular structure of triolein.

Computational models and model chemistry

The computational chemistry simulates chemical structures and reactions numerically, based in full or in part on the fundamental laws of physics. It allows chemists to study chemical phenomena by running calculations on computers rather than by examining reactions and compounds experimentally. Some methods can be used to model not only stable molecules, but also short-lived, unstable intermediates and even transition states. In this way, they can provide information about molecules and reactions which is impossible to obtain through observation. Computational chemistry is therefore both an independent research area and a vital adjunct to experimental studies.

There are two broad areas within computational chemistry devoted to the structure of molecules and their reactivity: molecular mechanics and electronic structure theory. They both perform the same basic types of calculations as following:

- Computing the energy of a particular molecular structure (spatial arrangement of atoms or nuclei and electrons). Properties related to the energy may also be predicted by some methods.

- Performing geometry optimizations, which locate the lowest energy molecular structure close proximity to the specified starting structure. Geometry optimizations depend primarily on the gradient of the energy-the first derivative of the energy with respect to atomic positions.

- Computing the vibrational frequencies of molecules resulting from interatomic motion within the molecule. Frequencies depend on the second derivatives of the energy with respect to atomic to atomic structure, and frequency calculations may also predict other properties which depend on second derivatives. Frequency calculations are not possible or practical for all computational chemistry methods.

Molecular mechanics

Molecular mechanics simulations use the laws of classical physics to predict the structures and properties of molecules. Molecular mechanics methods are available in many computer programs, including MM3, HyperChem, Quanta, Sybyl, and Alchemy. There are many different molecular mechanics methods. Each one is characterized by its particular *force field.* A force field has these components:

- A set of equations defining how the potential energy of a molecule varies with the locations of its component atoms.

- A series of *atom types*, defining the characteristics of an element within a specific chemical context. Atom types prescribe different characteristics and behavior for an element depending upon its environment. For example, a carbon atom in a carbonyl is treated differently than one bonded to three hydrogens. The atom type depends on hybridization, charge and the types of the other atoms to which it is bonded.

- One or more *parameter sets* that fit the equations and atom types to experimental data. Parameter sets define *force constants,* which are values used in the equations to relate atomic characteristics to energy component, and structural data such as bond lengths and angles.

Molecular mechanics calculations don’t explicitly treat the electrons in a molecular system. Instead, they perform computations based upon the interactions among the nuclei. Electronic effects are implicitly included in force fields through parameterization.

This approximation makes molecular mechanics computationally, and allows them to be used for very large systems containing many thousands of atoms. However, it also carries several limitations as well. Among the most important are these:

- Each force field achieves good results only for a limited class of molecules, related to those for which it was parameterized. No force field can be generally used for all molecular systems of interest

- Neglect of electrons means that molecular mechanics methods cannot treat chemical problems where electronic effects predominate. For example, they cannot describe processes which involve bond formation or bond breaking. Molecular properties which depend on subtle electronic details are also not reproducible by molecular mechanics methods.

Electronic structure methods

Electronic structure methods use the laws of quantum mechanics rather than classical physics as the basis for their computations. Quantum mechanics states that the energy and other related properties of a molecule may be obtained by solving the Schrödinger equation:

$$H\Psi = E\Psi$$

For any but the smallest systems, however, exact solutions to the Schrödinger equation are not computationally practical. Electronic structure methods are characterized by their various mathematical approximations to its solution. There are two major classes of electronic structure methods:

- Semi-empirical methods, such as AM1, MINDO/3 and PM3, implemented in programs like MOPAC, AMPAC, HyperChem, and Gaussian, use parameters derives from experimental data to simplify the computation. They solve an approximate form of the Schrödinger equation that depends on having appropriate parameters available for the type of chemical system under investigation. Different semi-empirical methods are largely characterized by their differing parameter sets.

- Ab initio methods, unlike either molecular mechanics or semi-empirical methods, use no experimental parameters in their computations. Instead, their computations are based solely on the laws of quantum mechanics-the first principles referred to in the name ab initio-and on the values of a small number of physical constants such as the speed of light, the masses and charges of electrons and nuclei, and Planck's constant.

The energy and many properties of a stationary state of a molecule can be obtained by solution of the Schrödinger partial differential equation,

$$H\psi = E\psi$$

Here H is Hamiltonian operator, a differential operator representing the total energy. E is the numerical value of the energy of the state. ψ is the wave function, the square of the wave function, ψ^2, is interpreted as a measure of the probability distribution of the particles within the molecule.

The Hamiltonian H, like the energy in classical mechanics, is the sum of kinetic and potential parts,

$$H = T + V$$

The kinetic energy operator T is a sum of differential operators,

$$T = -\frac{h^2}{8\pi^2}\sum_i \frac{1}{m_i}\left(\frac{\partial^2}{\partial x_i^2} + \frac{\partial^2}{\partial y_i^2} + \frac{\partial^2}{\partial z_i^2}\right)$$

The sum is over all particles I (nuclei + electrons) and m_i is the mass of particle i. h is Planck's constant. The potential energy operator is the coulomb interaction,

$$V = \sum_{i<j}\sum \left(\frac{e_i e_j}{r_{ij}} \right)$$

where the sum is over distinct pairs of particles (i,j) with electric charges e_i, e_j separated by a distance r_{ij}. For electrons, e_i =-e, while for nucleus with atomic number Z_i, $e_i = +Z_i e$.

Molecular Orbital Theory

Molecular orbital theory is an approach to molecular quantum mechanics which uses one-electron functions or orbitals to approximate the full wave function. A molecular orbital, $\psi(x, y, z)$, is a function of the Cartesian coordinates x, y, z of a single electron. The spin coordinates, ξ, also has to be included. This coordinate takes on one of two possible values $\left(\pm \frac{1}{2}\right)$, and measures the spin angular momentum component along the z axis in units of $h/2\pi$.

The simplest type of wave function appropriate for the description of an n-electron system would be in the form of a product of spin orbitals,

$$\psi_{product} = \chi_1(1)\chi_2(2)\ldots\chi_n(n)$$

where $\chi_i(i)$ is written for $\chi_i(x_i, y_i, z_i, \xi_i)$, the spin orbital of electron i. The spin orbitals may be arranged in a determinantal wave function.

$$\psi_{\det er \min ant} = \begin{vmatrix} \chi_1(1) & \chi_2(2) & \cdots & \chi_n(1) \\ \chi_1(2) & \chi_2(2) & \cdots & \chi_n(2) \\ \vdots & \vdots & \vdots & \vdots \\ \chi_1(n) & \chi_2(n) & \cdots & \chi_n(n) \end{vmatrix}$$

Here the elements of the first row of the determinant contain assignations of electron 1 to all the spin orbitals $\chi_1, \chi_2, ..., \chi_n$, the second row all possible assignations of electron 2, and so forth.

Some further properties of molecular orbital wave functions are worth noting. It is possible to force the orbitals to be orthogonal to each other, that is,

$$S_{ij} = \int \psi_i^* \psi_j dxdydz$$

This can be accomplished without changing the value of the whole wave function by mixing columns of the determinant. The spin functions, α and β, are orthogonal by integration over spin space (actually summation over the two possible values of ξ):

$$\sum_{\xi} \alpha(\xi)\beta(\xi) = \alpha\left(+\frac{1}{2}\right)\beta\beta\left(+\frac{1}{2}\right) + \alpha\left(-\frac{1}{2}\right)\beta\beta\left(-\frac{1}{2}\right) = 0$$

Molecular orbitals may be normalized, that is

$$S_{ij} = \int \psi_i^* \psi_j dxdydz = 1$$

By multiplication of the individual by a constant, normalization corresponds to the requirement that the probability of finding the electron anywhere in space is unity. Given $S_{ij} = 1$, the determinantal wave function may be normalized by multiplication by a factor of $(n!)^{-1/2}$, that is,

$$\int \cdots \int \Psi * \Psi d\tau_1 \tau_2 \cdots d\tau_n = 1$$

The determinantal wave function may be normalized by multiplication by a factor of $(n!)^{-1/2}$. We can write down a full many-electron molecular orbital wave function for the closed-shell ground state of a molecule with n(even) electrons, doubly occupying n/2 orbitals:

$$\Psi = (n!)^{-\frac{1}{2}} \begin{vmatrix} \Psi_1(1)\alpha(1) & \Psi_1(1)\beta(1) & \Psi_2(1)\alpha(1) & \cdots & \Psi_{n/2}(1)\beta(1) \\ \Psi_1(2)\alpha(2) & \Psi_1(2)\beta(2) & \Psi_2(2)\alpha(2) & \cdots & \Psi_{n/2}(2)\beta(2) \\ \vdots & \vdots & \vdots & & \vdots \\ \Psi_1(n)\alpha(n) & \Psi_1(n)\beta(n) & \Psi_2(n)\alpha(n) & \cdots & \Psi_{n/2}(n)\beta(n) \end{vmatrix}$$

The determinant is referred to as a Slater determinant.

Hatree-Fock Theory

An exact solution to the Schrödinger equation is not possible for any but the most trivial molecular systems. However, a number of simplifying assumptions and procedures do make an approximate solution possible for a large range of molecules. To simplify the treatment further, the next step is to assume that the electrons are non-interacting. This implies that (apart from the constant nuclear-nuclear repulsion term) which can rewrite the total n-electron Hamiltonian as a sum of n one-electron Hamiltonians,

$$H_{el} = \sum_{i=1}^{N} h(i)$$

$$h(i) = \left(-\frac{1}{2}\nabla_i^2 - \sum_{A=1}^{M} \frac{Z_A}{r_{iA}} \right)$$

This is clearly an over implication, since have neglected the electron-electron repulsion term $\frac{1}{r_{ij}}$. $H_{el} = \sum_{i=1}^{N} h(i)$ defines the independent particle model. The one-electron Hamiltonians are termed core-Hamiltonians, since the only interactions included are those between the electrons and the bare nuclei. Including an average interaction term in the {h(i)}, these become effective one-electron Hamiltonians. As a consequence of Equation, the total wave function can be rewritten as a product of n single-particle wave functions,

$$\psi(\vec{r}) = \phi_1(\vec{r}_1)\phi_2(\vec{r}_2)\ldots\phi_n(\vec{r}_n)$$

or, take the electron spin into account,

$$\psi = \chi_1(\vec{x}_1)\chi_2(\vec{x}_2)\ldots\chi_n(\vec{x}_n)$$

The spin orbitals $\{\chi_i(\vec{x}_i)\}$ are the products of the spatial orbitals $\phi_i(\vec{r}_i)$ and the spin functions $(\alpha(\omega))$ and $(\beta(\omega))$; $\vec{x}_i$ denotes both the space and spin coordinates of electron i. The total independent particle spin-orbital wave function is called a Hartree-product. The corresponding eigenvalue is a sum of the single-particle spin-orbital energies,

$$E_{el} = \sum_{i=1}^{N} \varepsilon_i$$

A further requirement on the state wave function is that it must be anti-symmetric with respect to the interchange of coordinate r (both space and spin) of any two electrons,

$$|\psi(\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_n)|^2 = |\psi(\vec{x}_2, \vec{x}_1, \ldots, \vec{x}_n)|^2$$
$$\psi(\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_n) = \pm\psi(\vec{x}_2, \vec{x}_1, \ldots, \vec{x}_n)$$

It is also possible to write Equation in terms of a n x n determinant, a slater determinant, which has the same antisymmetric properties:

$$\psi = (n!)^{-1/2} \begin{vmatrix} \chi_1(\vec{x}_1) & \chi_2(\vec{x}_1) & \cdots & \chi_n(\vec{x}_1) \\ \chi_1(\vec{x}_2) & \chi_2(\vec{x}_2) & & \\ \cdots & & \cdots & \\ \chi_1(\vec{x}_n) & \chi_2(\vec{x}_n) & \cdots & \chi_n(\vec{x}_n) \end{vmatrix}$$

This commonly is written like:

$$|\psi\rangle = (n!)^{-1/2} | \chi_1(\vec{x}_1), \chi_2(\vec{x}_2), \ldots, \chi_n(\vec{x}_n) \rangle$$

It can easily be varied that the slater determinant obeys the Pauli principle, as the determinant then becomes zero. The pre-factor $(n!)^{-1/2}$ is a normalization constant, and the $\{\chi_i\}$ are assumed orthonormal. By antisym-metrizing the Hartree-product in the term of a Slater determinant, that the probability of finding any two electrons at the same point in space (i.e. $\vec{x}_1 = \vec{x}_2$) is zero.

Through the wave functions, the effective potential is generated. This potential allows to refine wave functions, from which a new potential is obtained. The procedure is repeated until a stable, self-consistent solution is reached. Due to the iterative procedure, the initial guess of the wave function, can of course be chosen. However, the better the initial guess is, the easier it is to reach a stable solution to the eigenvalue problems in a relatively short computational time, is provided by the variation principle. This can be stated in the following way. Given any approximate wave function, satisfying the correct boundary conditions, the expectation value of the energy obtained by this wave function never lies below the exact energy of the ground state. Expressed in mathematical terms:

$$E_e = \frac{\langle \psi | H_e | \psi \rangle}{\langle \psi / \psi \rangle} \geq E_{exact}$$

A conceptually appealing model for the (trial) wave function of our molecular system, is to regard it as being constructed from molecular orbitals (MO). This description in analogous to the method used for the atomic orbitals (AO). The MO's, the elements of the wave function determinant, are in turn thought of as being constructed by a Linear Combination of Atomic Orbitals (LCAO),

$$\psi_i^{MO} = \sum_{\mu} c_{\mu i}\phi_{\mu}^{AO}$$

The variational principle leads to following equations describing the molecular orbital expansion coefficients, $c_{\upsilon i}$, derived by Roothaan and by Hall:

$$\sum_{\upsilon=1}^{N}\left(F_{\mu\upsilon} - \varepsilon_i S_{\mu\upsilon}\right)c_{\upsilon i} = 0 \qquad \mu = 1,2,\ldots,N$$

Equation can be rewritten in matrix form:

$$FC = SC\varepsilon$$

With $$F_{\mu\upsilon} = H_{\mu\upsilon}^{core} + \sum_{\lambda\sigma} P_{\lambda\upsilon}\left[(\mu\upsilon/\lambda\sigma) - \frac{1}{2}(\mu\upsilon/\lambda\sigma)\right]$$

$$F_{\mu\upsilon} = H_{\mu\upsilon}^{core} + G_{\mu\upsilon}$$

Where $H_{\mu\upsilon}^{core}$, core-Hamiltonian matrix defined as

$$H_{\mu\upsilon}^{core} = \int dr_1 \varphi_1^*(1)h(1)\varphi_\upsilon(1)$$

The matrix P is the density matrix or change- and bond-order matrix,

$$P_{\mu\upsilon} = 2\sum_{a}^{N/2} C_{\mu\upsilon} C_{\upsilon a}^{*}$$

The matrix S is the overlap matrix, indicating the overlap between orbitals.

$$S_{\mu\upsilon} = \int dr_1 \phi_{\mu}^{*}(1)\phi_{\upsilon}(1)$$

The term $(\mu\upsilon/\lambda\sigma)$ in Equation signified the two-electron repulsion integrals, defined as

$$(\mu\upsilon/\lambda\sigma) = \int dr_1 dr_2 \phi_{\mu}^{*}(1)\phi_{\upsilon}(2) r_{12}^{-1} \phi_{\lambda}^{*}(2)\phi_{\sigma}(2)$$

The (initial) wave function is used to generated an effective orential, which apply this potential in order to refine the coefficient matrix. The modified MO's form the new input in the Roothaan or Pople-Nesbet equations, and a new potential is generated. The iterative procedure is repeated until convergence is reached, i.e. when the changes in energy and/or charge density in two subsequent iterations are below a pre-set threshold value.

Before a more technical description of the SCF-procedure is presented, first need to define a new transformation matrix X, used for orthogonalization of the basis set. This orthogonolization can be either symmetric or canonical. A symmetric orthogonolization implies that X is formed through the relation

$$X = S^{-1/2} = Us^{-1/2}U^{\tau}$$

Where S is the overlap matrix, U is an unitary matrix diagonalizes S, and the diagonal matrix of the eigenvalues of S is given by the relations. In the canonical orthogonalization procedure, X is instead given by

$$X = Us^{-1/2}$$

Consider a new coefficient matrix C' related to the old coefficient matrix C by

$$C' = X^{-1}C, \quad C = XC^{-1}$$

where assumed that X possesses an inverse. Substituting $C = XC^{-1}$ into the Roothaan equations gives

$$FXC' = SXC'\varepsilon$$

Multiplying on the left by X^{τ} gives

$$\left(X^{\tau}FX\right)C' = \left(X^{\tau}SX\right)C'\varepsilon$$

If define a new matrix F^{τ} by

$$F^{\tau} = X^{\tau}FX$$

And use $\sum_{\upsilon=1}^{N}\left(F_{\mu\upsilon} - \varepsilon_i S_{\mu\upsilon}\right)c_{\upsilon i} = 0$, then

$$F'C' = C'\varepsilon$$

Matrix Equation can be solved using standard methods. The matrix element of the Hartree-Fock Hamiltonian operator are dependent of other orbitals through the element , and the Roothaan equations are solved by first assuming an initial set of linear expansion coefficients. The whole process is then reported until the coefficients no longer change within a given tolerance on repeated iteration. The solution is then said to be self-consistent and the method is then referred to as the SCF method.

Semiempirical Theory

Due to their greatly increased requirement for central processing unit time and storage space in the computer memory, ab initio quantum chemical methods are limited in their practical applicability. The approximate quantum chemical methods require significantly less computational resources. Semiempirical methods are based on the Roothaan-Hall .

In ab initio calculations all elements of the Fock matrix are calculated, irrespective of whether the basis functions ϕ_μ, ϕ_ν, ϕ_σ and ϕ_λ are on the same atom, on atoms that are bounded or on atoms that are not formally boned. The semiempirical methods consider the Fock matrix element in three groups: $F_{\mu\mu}$ (the diagonal elements); $F_{\mu\nu}$ (where ϕ_μ and ϕ_ν are on the same atom, and ϕ_σ and ϕ_λ are on different atoms).

The greatest proportion of the time required to perform ab initio Hartree-Fock SCF calculation is invariably calculating and manipulating integrals. The most obvious way to reduce the computational effort is to neglect or approximate some of these integrals. Semiempirical methods achieve this is part by explicit considering into the nuclear core. The overlap matrix, S, is set equal to identity matrix, I. The main implication of this is that the Roothaan-Hall equations are simplified: FC = SCF becomes FC = CE.

The NDDO (neglect of diatomic differential overlap), this theory neglects differential overlap between atomic orbital on different atoms.

The elimination of three- and four-centre integrals greatly reduces the time and storage requirements for an NDDO calculation (which now increase as the square of the number of atoms) relative to that for a full Hartree-Fock treatment.

Three levels of NDDO theory are included in SPARTAN'S SEMI EMPIRICAL module:

MNDO Modified Neglect of Diatomic Overlap
AM1 Austin Model 1
PM 3 MNDO Parametrization Method 3

In all of these formalisms, only the valence electrons are considered. The one-electron terms are given by,

$$H_{\mu\nu}^{core} = U_{\mu\nu} - Z_A \sum_{B \neq A} Z_B \langle \mu\nu / \delta\delta \rangle$$

Here, μ and ν are located on atom A and the summation is over all other atoms. $U_{\nu\nu}$ is related to the binding energy of an electron in atomic orbital ν, and is determined from spectroscopic data. $U_{\mu\nu}$ is set to zero for $\nu \neq \mu$. The second term in Equation represents the attraction on an electron on atom A from the nuclear framework. The two center integral involves only the s function on atom B.ZA is the charge of atom A without its valence electrons.

All one-centre, two-electron integrals $(\upsilon\upsilon / \mu\mu)$ and $(\upsilon\mu / \upsilon\mu)$ are fitted to spectroscopic data. The two-centre, two electron repulsion integrals $(\upsilon\mu / \lambda\sigma)$ are approximated by classical multipole-multipole charge interactions between atoms A and B. The multipole charge separations within an atom are treated as adjustable parameters, i.e. optimized to fit the experimentally derived one-centre integrals.

The $\beta_{\mu\lambda}$ terms appearing in the Fock matrix are the one-electron, two-center core resonance integrals and are approximated as,

$$\beta_{\rho\lambda} = \frac{\beta_\rho + \beta_\lambda}{2} S_{\rho\lambda}$$

Where S is the overlap integral between Slater orbitals and are adjustable parameters optimized using experimental thermo chemical data for simple molecules. Because all of the adjustable parameters are rooted in experimental data, these methods are known as semiempirical. As in ab initio Hartree-Fock calculations, an SCF procedure is used to converge on a density matrix, and the electronic energy.

The three methods differ only in the core-repulsion terms (they also differ in the detailed parametrization). Core repulsion includes nuclear repulsion and non-valence electron-electron repulsion, which are not explicitly considered in the calculation of the electronic energy. In the MNDO model, the core repulsion energy is given by,

$$E^{CR} = \sum_{A \neq B} \sum_{B \neq A} Z_A Z_B \left\langle \delta(A) / \delta(B) \right\rangle \left(e^{-\sigma_a E_{AB}} + e^{-\sigma_B E_{AB}} \right)$$

where is the intermolecular distance and are adjustable parameters fit to give the correct empirical behavior. Details are provided in the original papers. MNDO tends to overestimate core repulsion between two atoms at van der Waals distances. For this reason, the AM1 model was developed.

The AM1 and PM3 are based on MNDO (the name derives from the fact that PM3 is the third parametrization of MNDO, AM1 being considered the second). In AM1 a sum of Gaussians is employed to better represent the core-core repulsion behavior at Van der Waals distances just outside bonding distances. PM3 uses a similar core repulsion function, but differs in the parametrization procedure.