An Investigation on the Effect of Chain Length and Hydroxyl Position on the LogP and Solvent Accessible Surface Area of Alcohols.

Word Count: 2929

Research Question

How does changing the chain length and hydroxyl position of alcohols affect their solvent accessible surface area and logP?

Background Information

I learned in HL Chemistry that a higher chain length of alkanes would correlate to a lower solubility in water as more non-polar groups are being added. This made me wonder whether there was also a correlation between an alcohol's affinity for a polar/non-polar solvent and its chain length or hydroxyl position and how this would affect absorption in the body. After doing some research, I found that alcoholic beverages contain various chained alcohols, with some studies finding that higher chained alcohol are more toxic (Lachenmeier et al., 2008). I also found that the logarithm of the partition coefficient, logP, could be used as a measure of how well a compound passes biological barriers and thus its absorption; the optimal logP for intestinal absorption is 1.35-1.38. (Bhal, 2019) I decided to also measure the solvent accessible surface area so as to be able to understand how intermolecular forces varied with the independent variables, as these affect solubility.

LogP:

The logarithm of the partition coefficient, logP, measures the solubility of a compound in octanol, a non-polar solvent which is known as the organic phase compared with is solubility in water, a polar solvent, which is known as the aqueous phase. (Emery, 2017; Yale 2017)

$$\mathrm{logP} = \log\left(\frac{\text{Concentration of compound in octanol (organic) phase}}{\text{Concentration of compound in water (aqueous) phase}}\right)$$

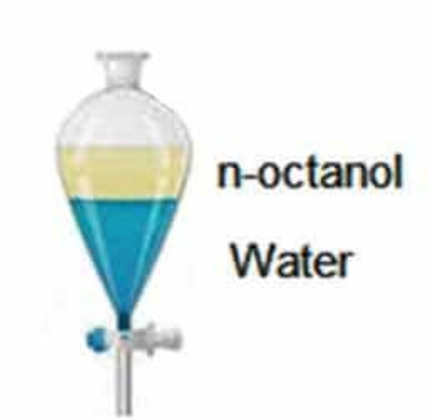

Figure 1: Diagram showing the apparatus used to measure logP.
It shows the polar phase, water, and the non-polar one, octanol. (Emery, 2017)

A logP value of zero indicates that the compound is equally partitioned in both phases, a negative logP value indicates that the compound has a higher affinity for the aqueous phase, while a positive logP value indicates that the compound has a higher affinity for the organic phase. (Bhal, 2019) Since logP is a logarithmic scale, an increase in one unit of logP means that a compound has a 10-fold higher concentration in the non-polar solvent relative to the polar one.

Solvent Accessible Surface Area (SASA):

SASA can be defined as the area formed by the centre of a sphere that rolls over the Van der Waals surface of the molecule (Brick, n.d.). It is a measure of the surface area that can directly interact with a solvent (Bendre, 2018). It is therefore an important factor when measuring the solubility of a molecule.

Methodology

Variables:

Independent Variables:

1. Chain length
    - Refers to the number of carbon atoms in a compound.
    - Compounds with chain lengths from 1-11 are used. Past the chain length of 4-5, water solubility is significantly lost (LibreTexts, n.d.). Lengths of 1-11 are used to develop a relationship. Furthermore, using chain lengths 1-11 allowed for two chain lengths, 10 and 11, that have at least 5 different hydroxyl group positions.
2. Hydroxyl group position
    - A position of 1 means that the hydroxyl group is on the first carbon. Since the compounds are symmetrical, the positions increment only until the middle of the compounds.

Dependent Variable:

1. LogP
    - No unit as it is a ratio.
2. Solvent accessible surface area (SASA)
    - Measured in $Å^2$.

Table 1: Table showing control variables and the methods of controlling them.

Control Variable	Reason for Control	How it is Controlled
Temperature – logP	Since the partition coefficient is essentially an equilibrium constant, temperature will impact it (Raynie, 2023). Also, temperature impacts solubility but its impact may vary in the polar vs non-polar solvents.	The partition coefficient is generally measured at 25 °C (EPA, 1979). It was assumed that the databases adhere to this standard.
Branching of molecules – logP, SASA	Placing carbon molecules as branches versus extensions of the chains affects solubility by impacting a molecule's ability to form London dispersion interactions. (Langdon, n.d.) It will also impact the SASA of the molecule based on its definition earlier.	All carbons were added to the main chain.
Dot density – SASA	Dots are used in the SASA calculation in PyMol. (Pymolwiki, 2021)	The highest dot density of 4 was used for all compounds. A higher dot density leads to more dots and a more accurate result. (Pymolwiki, 2021)
Solvent radius – SASA	Based on the definition of SASA given, the radius of the ball used to represent the solvent would impact the SASA calculation.	The default solvent radius of 1.4 Å was used in both softwares used to calculate SASA. (Chemaxon, n.d.; Pymolwiki, 2021)

Databases:

- PubChem: This is a database provided by the National Institute of Health. The predicted values use the logP of a known reference compound as a starting point, then they use an additive model based on the atoms in the molecule (Cheng, 2007). While using a reference compound provides more reliability, the additive model may have difficulty in differentiating between the position of the hydroxyl functional group since the same atoms are being used.
- ALOGPS: This software was developed by the University of Lausanne. This dataset relies on neural networks that have been trained on 12908 molecules. Although it may be better at differentiating logP for compounds of different hydroxyl positions, it will likely have more uncertainty than an additive model that uses reference compounds.
- ChemSpider ACD/Labs: Their method is not as clearly outlined but it does use thousands of experimental values to base its predictions on (ACD/Labs, n.d.).

- Chemicalize: For logP, this software looks at contributions of different atoms. Thus, it may also neglect the role of hydroxyl position (Chemaxon, n.d.). For SASA, it accounts for the effect of non-polar groups which is vital in this procedure with the changing of carbon chains (Ferrara, 2001)
- PyMol: This is a molecular modelling software developed by Schrödinger. I obtained access to it through a free student license. The .mol files were downloaded from ChemSpider and then inserted into PyMol. It uses a dot-based method to determine the surface area, which may be more precise and help to reveal the surface area of cavities in the molecule (Pymolwiki, 2021).
    - The following command line arguments were used:

      ```
      set dot_solvent, 1
      set dot_density, 4
      get_area {file_name}, load_b=1
      ```

There is no relevant safety, ethical, or environmental concerns with this database IA.

Raw Data

Table 2: Raw data table of logP and SASA for the different alcohols. PC represents PubChem; CS represents ChemSpider; and CA represents Chemicalize.

Compound	No. of carbons	Hydroxyl position	LogP					SASA (Å^2)	
			Exp.	Predicted				Predicted	
			PC ± 0.01	PC ± 0.1	ALOGPS ± 0.01	CS ACD ± 0.01	CA ±0.001	PyMOL ± 0.001	CA ± 0.01
Methanol	1	1	-0.77	-0.5	-1.38	-0.72	-0.519	140.516	197.26
Ethanol	2	1	-0.31	-0.1	-0.40	-0.19	-0.162	164.589	225.44
1-Propanol	3	1	0.25	0.3	0.21	0.34	0.361	188.758	252.31
2-Propanol		2	0.05	0.3	0.04	0.16	0.255	188.355	241.41
1-Butanol	4	1	0.88	0.9	0.84	0.88	0.805	213.017	280.08
2-Butanol		2	0.61	0.6	0.66	0.69	0.777	209.125	269.15
1-Pentanol	5	1	1.51	1.6	1.47	1.41	1.250	237.354	310.51
2-Pentanol		2	1.19	1.2	1.18	1.22	1.222	233.191	300.93
3-Pentanol		3	1.21	1.2	1.22	1.22	1.300	229.863	286.48
1-Hexanol	6	1	2.03	2.0	2.03	1.94	1.694	261.024	331.13
2-Hexanol		2	1.76	1.8	1.75	1.75	1.666	200.648	333.51
3-Hexanol		3	1.65	1.7	1.76	1.75	1.744	219.886	320.39
1-Heptanol	7	1	2.62	2.7	2.53	2.47	2.139	285.371	374.94
2-Heptanol		2		2.3	2.34	2.29	2.111	281.402	355.30
3-Heptanol		3		2.2	2.29	2.29	2.189	237.349	354.45
4-Heptanol		4		2.2	2.26	2.29	2.189	374.387	354.70
1-Octanol	8	1	3.00	3.0	3.21	3.00	2.584	310.054	396.51
2-Octanol		2	2.90	2.9	2.96	2.82	2.556	305.763	398.57
3-Octanol		3		2.8	2.84	2.82	2.634	302.595	378.87
4-Octanol		4		2.7	2.83	2.82	2.634	255.291	388.58
1-Nonanol	9	1	3.77	4.3	3.76	3.53	3.028	333.331	440.30
2-Nonanol		2		3.4	3.56	3.35	3.000	273.905	423.92
3-Nonanol		3		3.3	3.44	3.35	3.078	272.428	420.62
4-Nonanol		4		3.2	3.36	3.35	3.078	272.832	411.32
5-Nonanol		5		3.2	3.33	3.35	3.078	241.882	421.87
1-Decanol	10	1	4.57	4.6	4.24	4.06	3.473	358.814	462.76
2-Decanol		2		4.0	4.13	3.88	3.445	291.368	463.47
3-Decanol		3		3.9	4.11	3.88	3.523	288.379	445.15
4-Decanol		4		3.8	4.01	3.88	3.523	290.448	455.81
5-Decanol		5		3.8	3.97	3.88	3.523	290.448	447.08
1-Undecanol	11	1	4.72	4.6	4.83	4.59	3.917	520.411	505.73
2-Undecanol		2		4.5	4.68	4.41	3.889	377.456	488.19
3-Undecanol		3		4.4	4.66	4.41	3.967	307.508	485.55
4-Undecanol		4		4.3	4.60	4.41	3.967	307.911	479.56
5-Undecanol		5		4.3	4.57	4.41	3.967	307.911	488.83
6-Undecanol		6		4.3	4.55	4.41	3.967	307.911	468.71

*Uncertainties of raw values were taken as the reading error, the last decimal place of the values.

Table 3: Sample calculation for the average logP of methanol.

$\text{Average logP} = \frac{\sum \text{logP}}{\text{n}} = \frac{(-0.5) + (-0.77) + (-1.38) + (-0.72) + (-0.519)}{5} = -0.7778 \approx -0.78$
$\Delta(\text{Average logP}) = \frac{0\ .1 + 0.01 + 0.01 + 0.01 + 0.001}{5} = 0.0262 \approx 0.03$
Therefore, the average logP of methanol is -0.78 ± 0.03.

Table 4: Sample calculation for the average solvent accessible surface area of methanol.

$\text{Average SASA} = \frac{\sum \text{SASA}}{\text{n}} = \left(\frac{140.516\ \text{Å}^2 + 197.26\ \text{Å}^2}{2}\right) = 168.888\ \text{Å}^2$
$\Delta(\text{Average SASA}) = \frac{0.001\ \text{Å}^2 + 0.01\ \text{Å}^2}{2} = 0.0055\ \text{Å}^2 \approx 0.006\ \text{Å}^2$
Therefore, the average SASA of methanol is $168.888\ \text{Å}^2 \pm 0.006\ \text{Å}^2$.

Processed Data

Table 5: Processed data table with average logP and average SASA for the different alcohol compounds.

Compound	No. of carbons	Hydroxyl position	Average logP ± 0.03	Average SASA ± 0.006 Å^2
Methanol	1	1	-0.78	168.888
Ethanol	2	1	-0.23	195.015
1-Propanol	3	1	0.29	220.534
2-Propanol		2	0.16	214.883
1-Butanol	4	1	0.86	246.549
2-Butanol		2	0.67	239.138
1-Pentanol	5	1	1.45	273.932
2-Pentanol		2	1.20	267.061
3-Pentanol		3	1.23	258.172
1-Hexanol	6	1	1.94	296.077
2-Hexanol		2	1.75	267.079
3-Hexanol		3	1.72	270.138
1-Heptanol	7	1	2.49	330.156
2-Heptanol		2	2.26	318.351
3-Heptanol		3	2.24	295.900
4-Heptanol		4	2.23	364.544
1-Octanol	8	1	2.96	353.282
2-Octanol		2	2.83	352.167
3-Octanol		3	2.77	340.733
4-Octanol		4	2.75	321.936
1-Nonanol	9	1	3.68	386.816
2-Nonanol		2	3.33	348.913
3-Nonanol		3	3.29	346.524
4-Nonanol		4	3.25	342.076
5-Nonanol		5	3.24	331.876
1-Decanol	10	1	4.19	410.787
2-Decanol		2	3.86	377.419
3-Decanol		3	3.85	366.765
4-Decanol		4	3.80	373.129
5-Decanol		5	3.79	368.764
1-Undecanol	11	1	4.53	513.071
2-Undecanol		2	4.37	432.823
3-Undecanol		3	4.36	396.529
4-Undecanol		4	4.32	393.736
5-Undecanol		5	4.31	398.371
6-Undecanol		6	4.31	388.311

Analysis

Carbon Chain Length

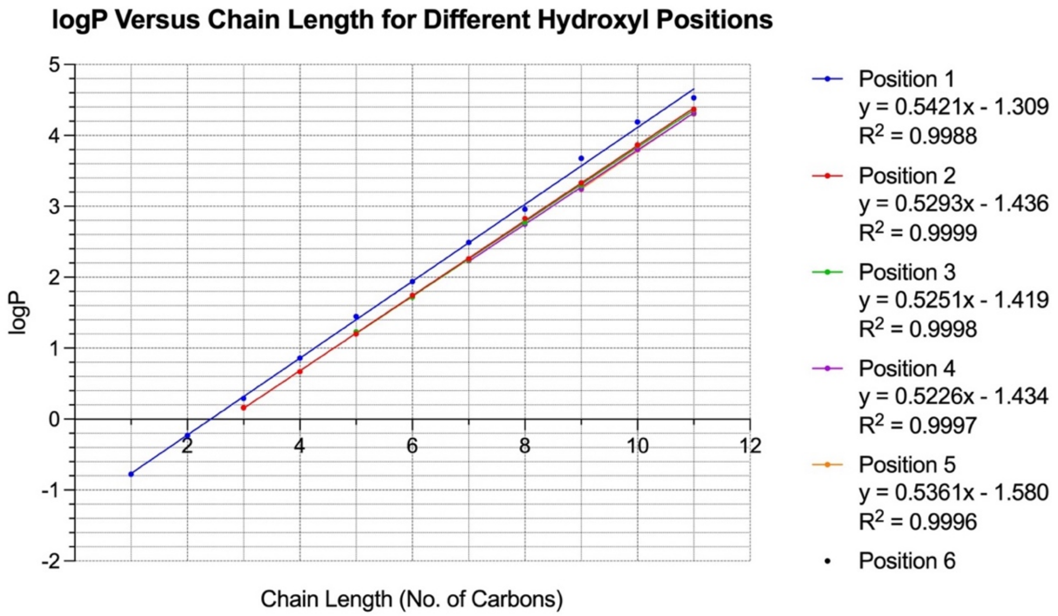

Figure 2: Graph showing the positive, linear correlations between logP and the number of carbons for differing hydroxyl positions. Error bars are too minor to be visualized. Dashed lines are used since the data is discrete. Position 6 does not have a line as it has only one data point.

To see the effect of the carbon chain length, the solvent accessible surface area can be plotted over chain length.

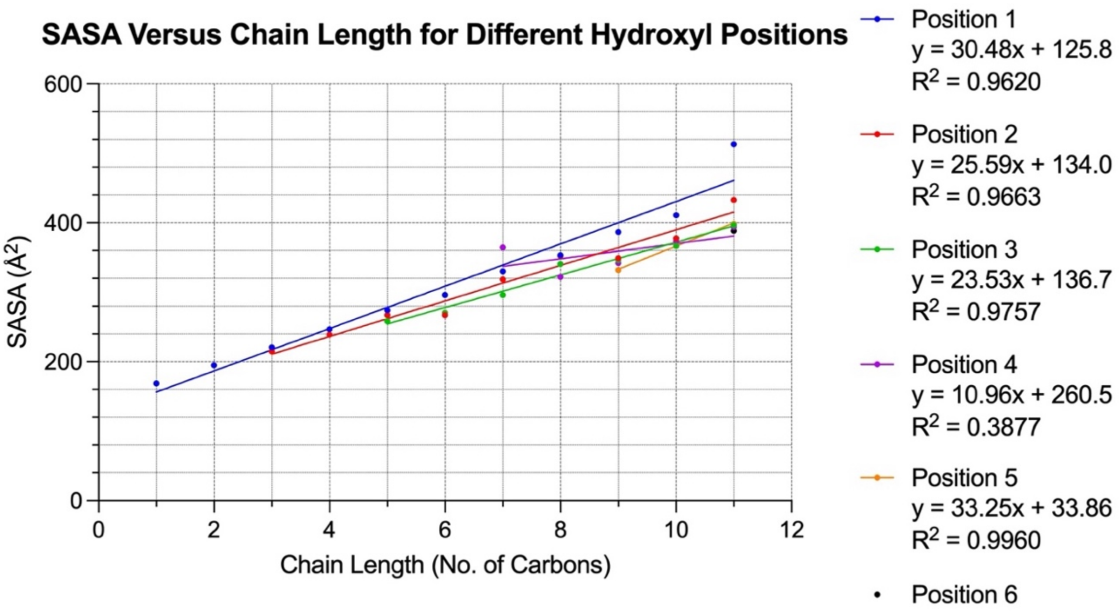

Figure 3: Graph showing the positive, linear correlations between solvent accessible surface area (SASA) and the number of carbons for differing hydroxyl positions. Error bars are too minor to be visualized. Position 6 does not have a line as it has only one data point.

Hydroxyl position

When logP is graphed over hydroxyl position for the different chain lengths, similar curves are obtained for each chain length, however the graphs have differing vertical translations making it difficult to identify the relationship. Therefore, to see this relationship more clearly, the deviation from the 1st position best-fit line in figures 2 and 3 can be used to magnify all the curves and so that they can be placed on the same scale.

Table 6: Table showing the deviation between the best fit line for position 1 and the average values for SASA and logP. The best fit lines come from figures 2 and 3.

Chain length	Hydroxyl position	logP for position 1 (y=0.5421x-1.309 x is the chain length) ± 0.03	Average logP ± 0.03	logP Deviation ± 0.05	SASA for position 1 (y=30.48x+125.8 x is the chain length) ± 0.006 Å^2	Average SASA ± 0.006 Å^2	SASA Deviation ± 0.01 Å^2
7	1	2.49	2.49	0.01	339.160	330.156	-9.00
	2		2.26	-0.23		318.351	-20.81
	3		2.24	-0.24		295.900	-43.26
	4		2.23	-0.25		364.544	25.38
8	1	3.03	2.96	-0.07	369.640	353.282	-16.36
	2		2.83	-0.20		352.167	-17.47
	3		2.77	-0.25		340.733	-28.91
	4		2.75	-0.28		321.936	-47.70
9	1	3.57	3.68	0.11	400.120	386.816	-13.30
	2		3.33	-0.24		348.913	-51.21
	3		3.29	-0.28		346.524	-53.60
	4		3.25	-0.32		342.076	-58.04
	5		3.24	-0.33		331.876	-68.24
10	1	4.11	4.19	0.08	430.600	410.787	-19.81
	2		3.86	-0.25		377.419	-53.18
	3		3.85	-0.26		366.765	-63.84
	4		3.80	-0.31		373.129	-57.47
	5		3.79	-0.32		368.764	-61.84
11	1	4.65	4.53	-0.12	461.080	513.071	51.99
	2		4.37	-0.28		432.823	-28.26
	3		4.36	-0.29		396.529	-64.55
	4		4.32	-0.33		393.736	-67.34
	5		4.31	-0.34		398.371	-62.71
	6		4.31	-0.35		388.311	-72.77

*The uncertainty for the position 1 compounds, as calculated by the best fit line, was taken as the same as that of the average values. Typically, max and min lines should be used however the lines show high correlations and the error bars are very small, so this method was not used.

Table 7: Sample calculation for logP deviation of 1-heptanol.

$\text{logP deviation} = 2.4918 - 2.4857$ $= 0.0061$ ≈ 0.01
$\Delta(\text{logP deviation}) = 0.0262 - 0.0262$ $= 0.0524$ ≈ 0.05
Therefore, the logP deviation of 1-heptanol is 0.01 ± 0.05.

Table 8: Sample calculation for SASA deviation of 1-heptanol.

$\text{SASA deviation} = 330.1555\ \text{Å}^2 - 339.160\ \text{Å}^2$ $= -9.0045\ \text{Å}^2$ $\approx -9.00\ \text{Å}^2$
$\Delta(\text{SASA deviation}) = 0.0055\ \text{Å}^2 + 0.0055\ \text{Å}^2$ $= 0.011\ \text{Å}^2$ $\approx 0.01\ \text{Å}^2$
Therefore, the SASA deviation of 1-heptanol is $-9.00\ \text{Å}^2 \pm 0.01\ \text{Å}^2$

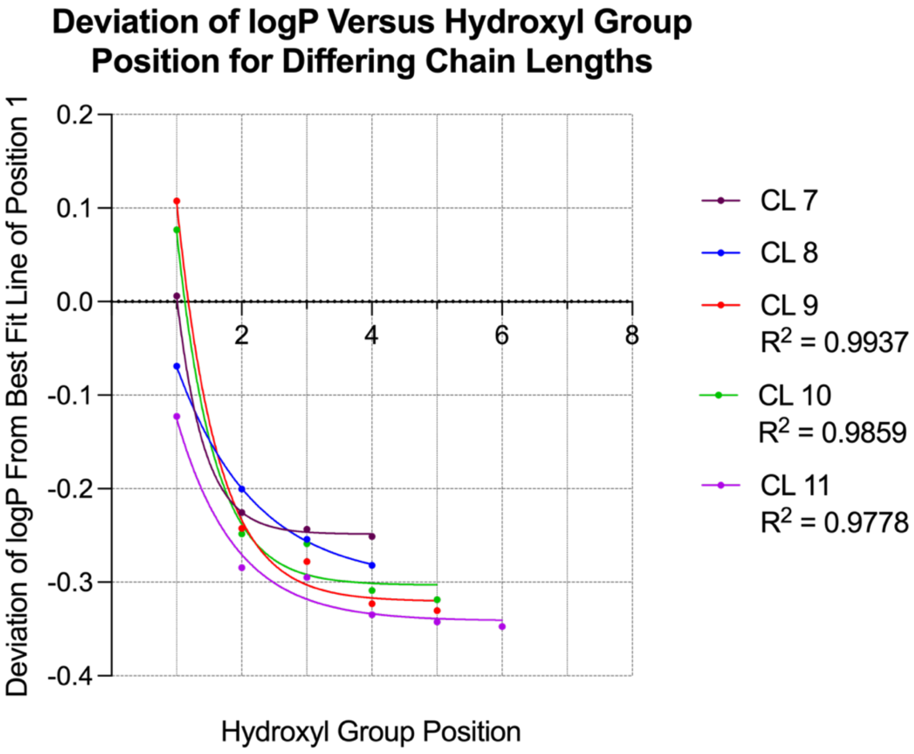

Figure 4: Graph showing the negative decaying correlation between logP deviation and hydroxyl group position for differing chain lengths. Error bars are too minor to be visualized. Sigmoidal functions were used for the best fit lines. R^2 for chain lengths 7 and 8 could not be computed for this type of correlation due to the lack of points.

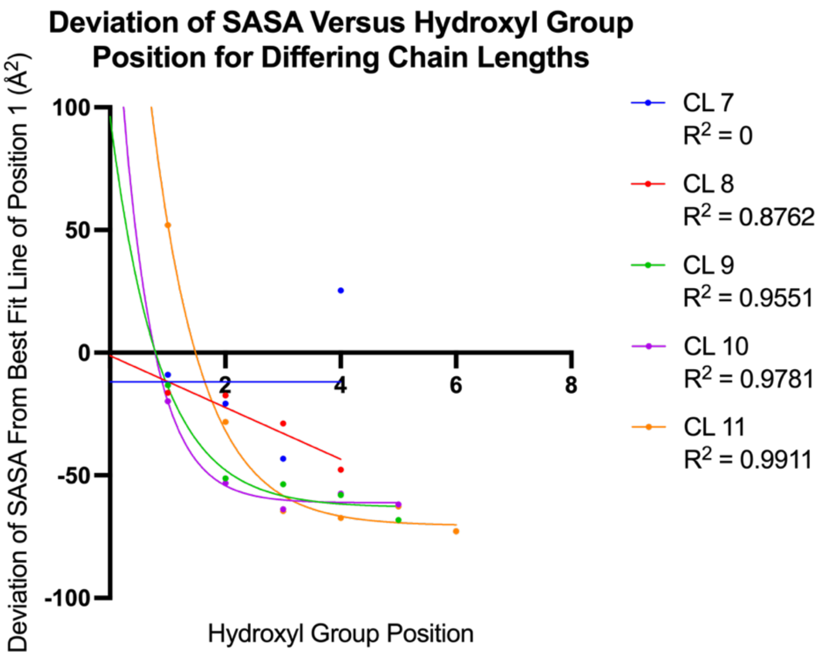

Figure 5: Graph showing the negative decaying correlation between SASA deviation and hydroxyl group position for differing chain lengths. Error bars are too minor to be visualized. Exponential decay functions were used for the best fit lines.

Conclusion

Carbon Chain Length:

The error bars of figure 2 are very small so they cannot be seen. This is a result of the unknown true uncertainty since the values were obtained from databases. That being said, since the R^2 values are very high for all the hydroxyl positions, it provides confidence in the values.

From figure 2, all hydroxyl positions show a very strong, positive linear correlation between logP and carbon chain length; the R^2 values of all the lines are greater than 0.99. Carbon chain length has a very significant impact on logP. Take hydroxyl position 1 for instance, the logP ranges from -0.78 for 1 carbon to 4.53 for 11 carbons. This indicates that increasing the carbon chain length significantly increases the alcohols' affinity for non-polar solvents vs polar ones. Furthermore, all the slopes are greater than 0.5 indicating that for an extra carbon, logP increases by more than 0.5. Since logP is a logarithmic scale, an increase in logP represents a much higher increase in the concentration of the alcohol in octanol divided by that in water. Therefore, the relationships of figure 2 demonstrate that increasing the carbon chain length leads to a substantially higher affinity for non-polar solvents over polar ones.

This can be explained by the polarity of the molecules. Polar molecules are ones where the electric charge is distributed asymmetrically. Water is a polar solvent that has both dipole-dipole forces as well as hydrogen bonds. When a solute molecule is introduced in water, a certain amount of energy is required in order to break the hydrogen-bonded structure. If the solute is polar, new hydrogen bonds or dipole-dipole attractions are formed that release sufficient potential energy to compensate for the energy required to break the initial bonds. If the solute is non-polar, however, then the energy released will not be enough to compensate for the energy required to break the initial bonds. For this reason, polar solutes are soluble in polar solvents while non-polar molecules are soluble in non-polar solvents. (Lower, n.d.)

All alcohols contain a hydroxyl group that is able to form hydrogen bonds with water molecules. However, alcohols with increasing chain length have longer non-polar regions that decrease their affinity for water molecules. (LibreTexts, n.d.)

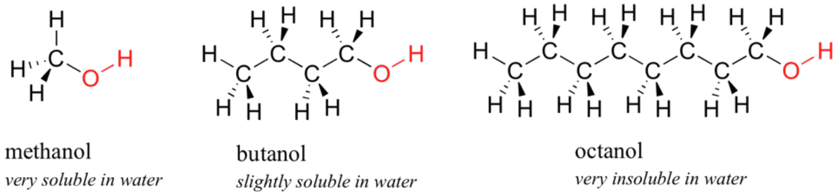

Figure 6: Diagram showing the increasing non-polar area with higher chain length alcohols. (LibreTexts, n.d.)

Therefore, the logP increases with higher chain length of the alcohols since they will have an increasing affinity for octanol, the non-polar solvent, rather than water, the polar one.

Previous research also found a similar relationship between logP and chain length.

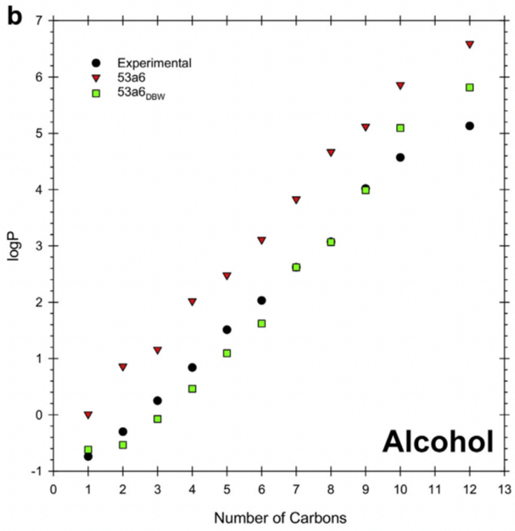

Figure 7: logP versus chain length for alcohol compounds, determined with various methods in a literature source. (Warren et al., 2018)

Figure 7 demonstrates a linear relationship between the logP and the number of carbons for alcohols which confirms the relationships viewed in figure 2.

In figure 3, there are two outliers present for the compounds 4-heptanol, (7, 364.544), and 1-Undecanol, (11, 513.071). The R^2 value for the compounds of the first hydroxyl position is still very high, 0.9620, since there are many other points that follow the linear relationship, but as a result of the outlier for the 4-heptanol, the R^2 value for the fourth position is small, 0.3877.

Figure 3 shows that the SASA has a strong, positive linear correlation to chain length with the exception of hydroxyl position 4, but it does still show a positive, weak correlation. This is of course expected since for every increment in the chain length, one carbon is being added. For every carbon added, the SASA increases from between 23.54-33.25, excluding the hydroxyl position 4 due to the outlier. The increase in SASA as a result of the non-polar carbon chain explains the relationship viewed in Figure 1. Since the non-polar surface area increases, the compounds gain solubility in octanol over water.

Another explanation for the positive linear correlation between logP and chain length from figure 2 is that of steric hindrance. Steric hindrance at an atom is the congestion caused by the physical presence of surrounding atoms. Steric hindrance can prevent reactions with other molecules. As the hydroxyl group is bonded to an increasing long chain, more steric hindrance, or congestion occurs, making it harder to form hydrogen bonds with water molecules. (Gunawardena, n.d.)

The weak trend of hydroxyl position 4 within figure 3 can be attributed to the SASA for 4-Heptanol which seems to be an outlier compared to the other compounds with hydroxyl position 4.

Hydroxyl Position

Figure 2 shows that compounds with the first hydroxyl position have the highest logP value, the highest affinity for the non-polar phase. From the equations of the lines, smaller hydroxyl positions also generally have sleeper slopes and higher y-intercepts. The equation for hydroxyl position 1 is $y = 0.5421x - 1.309$ and for position 5 it is $y = 0.5361x - 1.580$, so the relationship for position 1 shows a steeper slope and a higher y-intercept. A similar relationship is seen with SASA in figure 3. These relationships are better analyzed however when hydroxyl position is placed on the x-axis as in figures 4 and 5.

Figure 4 shows no outleirs and R^2 values all above 0.95. This figure demonstrates that logP has an exponentially decaying relationship with hydroxyl group position. This means that having the hydroxyl position at the end of the chain increases its affinity for the non-polar phase. Moving to the second position leads to a much smaller affinity for the non-polar phase and a larger affinity for the polar phase, but from there, moving the position towards the centre of the molecule has a diminishing effect. For instance, for the alcohols of chain length 7, the logP deviation is 0.01 for hydroxyl position 1, then -0.23 for position 2, but for position 3 it only decreases to -0.24.

Overall, hydroxyl position is a less impactful factor on logP than carbon chain length. The alcohols of chain length 9 have the largest range of values in logP for the hydroxyl positions (figure 4), and still the range in logP deviation is still only 0.44 from hydroxyl position 1 to 5.

The decaying relationship can be explained by the area of the chain accessible to the solvent. Since the carbon bonded to the oxygen is sp^3 hybridized, the oxygen is bonded to the chain with an angle of approximately 109°. Therefore, as the hydroxyl position moves to the centre of the molecule it forces the chain to essentially fold in on itself, leaving less non-polar area accessible to the solvent and making the molecule more like a branched one.

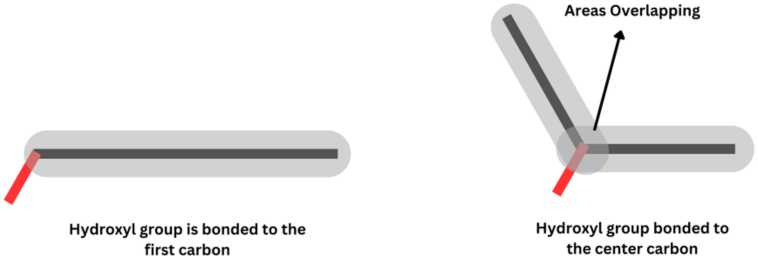

Figure 8: Diagrams created by author to explain the decrease in SASA as the hydroxyl group moves to the centre of the molecule.

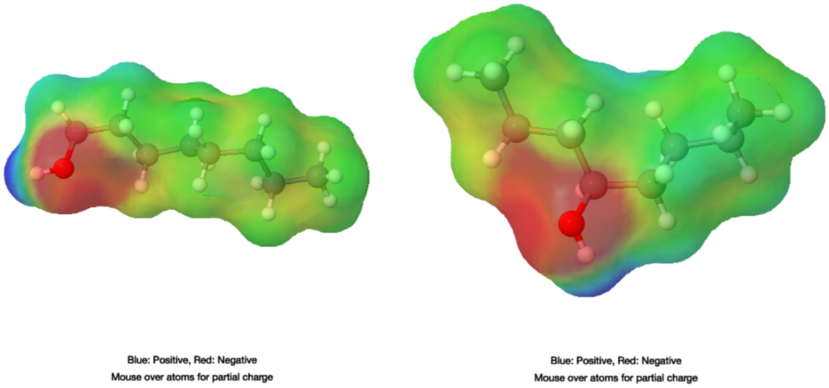

Figure 9: Molecular structures created with MolCalc for 1-octanol (left) and 4-octanol (right). As the hydroxyl group moves to the center of the chain, it bends. (MolCalc, n.d.)

Figures 8 and 9 show that when the hydroxyl position moves to the centre of the molecule, the chain folds. As the molecule loses some non-polar area and becomes more branched, it produces less london dispersion forces and thus loses solubility in the non-polar solvent and gains solubility in the polar one. (Langdon, n.d.) This explains the negative relationship between logP and hydroxyl position. This theory is also supported by the relationship of SASA with hydroxyl group position.

From figure 5, SASA shows the same exponentially decaying relationship with hydroxyl group position, with the exception of chain lengths 7 and 8. However, these chain lengths still show a decreasing trend for SASA with increasing hydroxyl position. This directly supports the idea that the branches converge on one another when the hydroxyl position moves towards the centre of the molecule, with the most significant decrease in SASA occurring when the hydroxyl position moves from the end of the chain to position 2. Then, the changes become less significant.

Figure 5 shows one outlier for 4-heptanol, (4, 25.38), which resulted in an R^2 of 0, but as explained, the other compounds of chain length 7 still show a negative trend.

To summarize the findings found:

1. Chain Length
    a. logP shows a strong linear correlation with increasing chain length. This can be explained by an increasing non-polar surface area (supported by point b) and increasing steric hindrance.
    b. SASA also shows a strong linear correlation with increasing chain length, for all but one hydroxyl position. This explains the linear correlation found between logP and chain length.
2. Hydroxyl Position
    a. logP shows exponentially decaying relationship with hydroxyl position. This means that the most significant decrease occurs when the hydroxyl group leaves the end of the chain. This relationship is explained by the decreasing SASA.
    b. SASA also decreases exponentially with hydroxyl position, for most chain lengths, and for those that do not show an exponential decay, there is still a negative relationship. This is explained by the fact that at higher hydroxyl positions, the molecules branches fold on one another.

Evaluation

Table 9: Systematic and random limitations in the procedure.

Limitation	Impact	Improvement
Systematic Errors		
SASA radius is not fully representative of octanol.	By default, PyMol and Chemicalize calculate the accessible surface area using the ball radius that represents that of a water molecule, 1.4 $Å^2$ (Chemaxon, n.d.). Therefore, the SASA data may not be representative of an octanol solvent molecule which comprises the other phase of logP. Since octanol is a larger molecule than water, the SASA would likely be lower than the values obtained.	Using softwares that allow the solvent radius to be altered, graphs like those in figures 3 and 5 could be plotted using SASA calculations that use the radius of an octanol molecule.
Reliance on predicted data for logP	Four out of the five sources used to gather logP values gave predicted values. The experimental one was also incomplete. Predicted data is prone to systemic error based on misconceptions used in the calculation method. Many of the predicted values failed to differentiate logP for the changing of hydroxyl position with the same carbon chain length. Since logP was shown to decrease with hydroxyl group position, figure 4, it means that true logP values would likely be lower than those obtained from the databases.	It would be better to determine the logP values experimentally. Alternatively, if a theoretical method is used it should not simply be a molecule additive method but one that can differentiate based on the position of the functional group, such as ALOGPS' neural network.
Random Errors		
Lack of SASA sources	SASA is a measurement that must be calculated by softwares. Differences in calculation methods could lead to systematic errors. The two sources showed significant variation in their values with Chemicalize reporting consistently higher values than PyMol.	Since it is a theoretical measurement, one should use more than 2 databases to reduce the effect of random error.
Not enough data points for hydroxyl position	Many of the chain lengths did not have at least 5 hydroxyl positions. As such the plots over hydroxyl position showed some weak R^2 correlations (figure 5) or in some cases R^2 could not be computed (figure 4).	One should use alcohols with greater than 11 carbon bonds which have more hydroxyl positions to develop stronger relationships with hydroxyl position as the independent variable.

References

Ashenhurst, J. (2024, October 23). *Branching, and Its Affect On Melting and Boiling Points*. https://www.masterorganicchemistry.com/. https://www.masterorganicchemistry.com/2010/07/09/branching-melting-boiling-points/

Bendre, A. D., Ramasamy, S., & Suresh, C. G. (2018, July 1). *Analysis of Kunitz inhibitors from plants for comprehensive structural and functional insights*. International Journal of Biological Macromolecules. https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/accessible-surface-area

Bhal, S. K. (2019). *LogP—Making Sense of the Value*. https://www.acdlabs.com/. https://www.acdlabs.com/wp-content/uploads/download/app/physchem/making_sense.pdf

Brick, P. (n.d.). *Areaimol (CCP4: Supported program)*. https://smb.slac.stanford.edu/. https://smb.slac.stanford.edu/facilities/software/ccp4/html/areaimol.html

Gunawardena. (n.d.). *Steric hindrance - Chemistry libretexts*. https://chem.libretexts.org/. https://chem.libretexts.org/Ancillary_Materials/Reference/Organic_Chemistry_Glossary/Steric_Hindrance

Karickhoff, S. W., & Brown, D. S. (1979, April). *Determination of Octanol/Water Distribution Coefficients, Water Solubilities, and Sediment/Water Partition Coefficients for Hydrophobic Organic Pollutants*. https://nepis.epa.gov/. https://nepis.epa.gov/Exe/ZyNET.exe/91013MOV.txt?ZyActionD=ZyDocument&Client=EPA&Index=1976%20Thru%201980&Docs=&Query=&Time=&EndTime=&SearchMethod=1&TocRestrict=n&Toc=&TocEntry=&QField=&QFieldYear=&QFieldMonth=&QFieldDay=&UseQField=&IntQFieldOp=0&ExtQFieldOp=0&XmlQuery=&File=D%3A%5CZYFILES%5CINDEX%20DATA%5C76THRU80%5CTXT%5C00000022%5C91013MOV.txt&User=ANONYMOUS&Password=anonymous&SortMethod=h%7C-&MaximumDocuments=1&FuzzyDegree=0&ImageQuality=r75g8/r75g8/x150y150g16/i425&Display=hpfr&DefSeekPage=&SearchBack=ZyActionL&Back=ZyActionS&BackDesc=Results%20page&MaximumPages=24&ZyEntry=1

Lachenmeier, D. W., Haupt, S., & Schulz, K. (2008, January 16). *Defining maximum levels of higher alcohols in alcoholic beverages and surrogate alcohol products*. https://www.sciencedirect.com/. https://www.sciencedirect.com/science/article/abs/pii/S0273230008000032#:~:text=The%20major%20higher%20alcohols%20found,methyl%2D1%2Dbutanol

Langdon, S. (n.d.). *Introduction to Organic Chemistry*. https://openpress.usask.ca/. https://openpress.usask.ca/intro-organic-chemistry/chapter/2-4/#:~:text=This%20means%20that%20factors%20that,in%20water%20(Table%202.3)

Libretexts. (n.d.-a). *2.12: Intermolecular Forces and solubilities*. https://chem.libretexts.org/. https://chem.libretexts.org/Bookshelves/Organic_Chemistry/Map%3A_Organic_Chemistry_(Wade)_Complete_and_Semesters_I_and_II/Map%3A_Organic_Chemistry_(Wade)/02%

3A_Structure_and_Properties_of_Organic_Molecules/2.12%3A_Intermolecular_Forces_and_Solubilities

Libretexts. (n.d.-b). *13.3: Physical properties of alcohols*. https://chem.libretexts.org/. https://chem.libretexts.org/Bookshelves/Organic_Chemistry/Map%3A_Organic_Chemistry_(Wade)_Complete_and_Semesters_I_and_II/Map%3A_Organic_Chemistry_(Wade)/13%3A_Structure_and_Synthesis_of_Alcohols/13.03%3A_Physical_Properties_of_Alcohols

Lower. (n.d.). *8.2: Thermodynamics of Solutions - Chemistry libretexts*. https://chem.libretexts.org/. https://chem.libretexts.org/Bookshelves/General_Chemistry/Chem1_(Lower)/08:_Solutions/8.02:_Thermodynamics_of_Solutions

Najafi, A. (2017, October 20). *Drug Lipophilicity and Absorption: The Continuous Challenge in Drug Discovery*. https://emerypharma.com/. https://emerypharma.com/blog/drug-lipophilicity-and-absorption-a-continuous-challenge-toward-the-goal-of-drug-discovery/

Raynie, D. E. (2023, March 1). *Practical Understanding of Partition Coefficients*. https://www.chromatographyonline.com/. https://www.chromatographyonline.com/view/practical-understanding-of-partition-coefficients

Warren, D. B., McPhee, E., Birru, W. A., Benameur, H., Chalmers, D. K., & Pouton, C. W. (2018, November 20). Improvement in the predicted partitioning of alcohol and polyethylene oxide groups between water and octanol (logp) in molecular dynamics simulations. https://pubmed.ncbi.nlm.nih.gov/30468829/

Yale University. (2017). *Module 1: Aqueous and lipid solubility*. Center for Green Chemistry & Green Engineering at Yale. https://greenchemistry.yale.edu/modrnu-modules/physicochemical-properties-modules/module-1-aqueous-and-lipid-solubility

Databases:

ChemSpider

Royal Society of Chemistry. (n.d.). ChemSpider. https://www.chemspider.com/

Advanced Chemistry Development, Inc. (ACD/Labs). (n.d.). *ACD/LogP – Predict logP values with industry-leading accuracy.* https://www.acdlabs.com/products/percepta-platform/physchem-suite/logp/

PubChem

National Center for Biotechnology Information. (n.d.). *PubChem* [Database]. https://pubchem.ncbi.nlm.nih.gov/

Cheng, T., Zhao, Y., Li, X., Lin, F., Xu, Y., Zhang, X., Li, Y., Wang, R., & Lai, L. (2007, November 7). *Computation of octanol-water partition coefficients by guiding an additive* *model with knowledge*. Journal of Chemical Information and Modeling. https://pubs.acs.org/doi/10.1021/ci700257y

ALOGPS

Virtual Computational Chemistry Laboratory. (n.d.). ALOGPS: Algorithm for logP and aqueous solubility prediction. https://vcclab.org/lab/alogps/

Chemicalize

Chemaxon Ltd. (n.d.). *Chemicalize* [Online platform]. https://chemicalize.com/

ChemAxon. (n.d.). *logP Plugin.* https://docs.chemaxon.com/display/docs/calculators_logp-plugin.md

Ferrara, P., Apostolakis, J., & Caflisch, A. (2001, November 15). *Evaluation of a fast implicit solvent model for molecular dynamics simulations*. https://onlinelibrary.wiley.com/. http://onlinelibrary.wiley.com/doi/10.1002/prot.10001/full

ChemAxon. (n.d.). *Molecular surface area plugin (3D).* https://docs.chemaxon.com/display/docs/calculators_molecular-surface-area-plugin-3d.md

PyMol

Schrödinger, LLC. (n.d.). *PyMOL* [Molecular visualization software]. https://pymol.org/

Pymolwiki. (n.d.). *Get_area.* https://pymolwiki.org/index.php/Get_area

MolCalc

MolCalc. (n.d.). *Molecular calculations – Solvation.* Retrieved February 25, 2025, from https://molcalc.org/calculations/47e95b8142846c80039586aff0363c0d#/solvation

- https://www.acdlabs.com/products/percepta-platform/physchem-suite/logp/
- https://pubs.acs.org/doi/10.1021/ci700257y
- https://docs.chemaxon.com/display/docs/calculators_logp-plugin.md
- https://pymolwiki.org/index.php/Get_area
- Chemicalize
    - https://onlinelibrary.wiley.com/doi/10.1002/prot.10001
        - (Ferrara, 2001)
- https://docs.chemaxon.com/display/docs/calculators_molecular-surface-area-plugin-3d.md
- https://molcalc.org/calculations/47e95b8142846c80039586aff0363c0d#/solvation

- https://www.sciencedirect.com/science/article/abs/pii/S0273230008000032#:~:text=The%20major%20higher%20alcohols%20found,methyl%2D1%2Dbutanol).
    - (Lachenmeier et al., 2008)
- https://smb.slac.stanford.edu/facilities/software/ccp4/html/areaimol.html
    - (Brick, n.d.)
- https://chem.libretexts.org/Bookshelves/Organic_Chemistry/Map%3A_Organic_Chemistry_(Wade)_Complete_and_Semesters_I_and_II/Map%3A_Organic_Chemistry_(Wade)/13%3A_Structure_and_Synthesis_of_Alcohols/13.03%3A_Physical_Properties_of_Alcohols
- https://www.acdlabs.com/wp-content/uploads/download/app/physchem/making_sense.pdf
    - (Bhal, 2019 - maybe)
- https://emerypharma.com/blog/drug-lipophilicity-and-absorption-a-continuous-challenge-toward-the-goal-of-drug-discovery/
- https://chem.libretexts.org/Bookshelves/General_Chemistry/Chem1_(Lower)/08%3A_Solutions/8.02%3A_Thermodynamics_of_Solutions
    - (Lower, n.d.)

- https://chem.libretexts.org/Bookshelves/Organic_Chemistry/Map%3A_Organic_Chemistry_(Wade)_Complete_and_Semesters_I_and_II/Map%3A_Organic_Chemistry_(Wade)/02%3A_Structure_and_Properties_of_Organic_Molecules/2.12%3A_Intermolecular_Forces_and_Solubilities
  - (LibreTexts, n.d.)
- Improvement in the Predicted Partitioning of Alcohol and Polyethylene Oxide Groups Between Water and Octanol (logP) in Molecular Dynamics Simulations
  - (Warren et al., 2018)
- https://chem.libretexts.org/Ancillary_Materials/Reference/Organic_Chemistry_Glossary/Steric_Hindrance
  - (Gunawardena, n.d.)
- https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/accessible-surface-area
  - (Bendre, 2018)
- https://openpress.usask.ca/intro-organic-chemistry/chapter/2-4/#:~:text=This%20means%20that%20factors%20that,in%20water%20(Table%202.3).
  - (Langdon, n.d.)
- https://www.chromatographyonline.com/view/practical-understanding-of-partition-coefficients
  - (Raynie, 2023)
- https://nepis.epa.gov/Exe/ZyNET.exe/91013MOV.txt?ZyActionD=ZyDocument&Client=EPA&Index=1976%20Thru%201980&Docs=&Query=&Time=&EndTime=&SearchMethod=1&TocRestrict=n&Toc=&TocEntry=&QField=&QFieldYear=&QFieldMonth=&QFieldDay=&UseQField=&IntQFieldOp=0&ExtQFieldOp=0&XmlQuery=&File=D%3A%5CZYFILES%5CINDEX%20DATA%5C76THRU80%5CTXT%5C00000022%5C91013MOV.txt&User=ANONYMOUS&Password=anonymous&SortMethod=h%7C-&MaximumDocuments=1&FuzzyDegree=0&ImageQuality=r75g8/r75g8/x150y150g16/i425&Display=hpfr&DefSeekPage=&SearchBack=ZyActionL&Back=ZyActionS&BackDesc=Results%20page&MaximumPages=24&ZyEntry=1
  - (EPA, 1979)
- https://www.masterorganicchemistry.com/2010/07/09/branching-melting-boiling-points/
  - (Ashenhurst, 2024)

Databaases

- https://www.acdlabs.com/products/percepta-platform/physchem-suite/logp/
- https://pubmed.ncbi.nlm.nih.gov/17985865/
- https://docs.chemaxon.com/display/docs/calculators_logp-plugin.md
- https://pymolwiki.org/index.php/Get_area
- Chemicalize
  - https://onlinelibrary.wiley.com/doi/10.1002/prot.10001
    - (Ferrara, 2001)
- https://docs.chemaxon.com/display/docs/calculators_molecular-surface-area-plugin-3d.md
- https://molcalc.org/calculations/47e95b8142846c80039586aff0363c0d#/solvation

-