Distribution Identification of Lycopene Similarities with Atomic Polarizabilities and Complexity in Carotenoid Biosynthesis

Jeff Cromwell, PhD

The Mathematical Learning Space Research Portfolio

1 Abstract

Lycopene is a linear, unsaturated hydrocarbon carotenoid that has a red pigment found in fruits such as (a) tomatoes, (b) pink grapefruit, (c) apricots, (d) red oranges, (e) watermelon, (f) rosehips, and (g) guava. Lycopene all-trans-Lycopene with formula C40H56 is involved in several reactions in cartenoid biosynthesis as well as biosynthesis of (a) plant secondary metabolites, (b) iosynthesis of terpenoids and steroids, (c) biosynthesis of plant hormones, (d) metabolic pathways, (e) biosynthesis of secondary metabolites and (f) biosynthesis of cofactors. For three molecular properties of complexity, BPOL and APOL, Pearson distributions of type 6 and 5 were obtained with the expected skewness with the Tanimoto similarity for the compounds. Given the design of the geometric structure of Lycopene, the design of compounds in the reaction relationship can provide insight in the development of additional agents for clinical applications.

2 Introduction

Lycopene is used in clinical trials as an Anticarcinogenic Agents that can reduce the frequency or rate of spontaneous or induced tumors independently of the mechanism involved. It can also be used in Radiation-Protective Agents and Drugs used for ionizing radiation protection and used in radiation therapy. In addition, it can be used as Anti-Inflammatory Agents that can reduce and suppress inflammation and antioxidants that inhibit oxidation reactions. Tissue locations include (1) Adipose Tissue, (2) Adrenal Gland, (3) Epidermis, (4) Fibroblasts, (5) Intestine, (6) Kidney, (7) Leukocyte, (8) Liver, (9) Pancreas, (10) Platelet, (11) Prostate and (12) Spleen. [800]

In Terpenoids, the carotenoid Lycopene with chemical arrangement shown in Figure 1

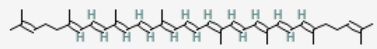

Figure 1: Lycopene [800]

Lycopene is a linear, unsaturated hydrocarbon carotenoid and compounds in photosynthetic organisms (plants, algae, and some types of fungus) that are chemically characterized by a large polyene chain containing 35-40 carbon atoms. These carotenoid polyene chains are terminated by two 6-carbon rings. [800]

Lycopene have antioxidant properties that reduce aging and many degenerative diseases in an animal diet. Having a chemical formula of C40H56, lycopene is a tetraterpene assembled from eight isoprene units of carbon and hydrogen and a non-essential human nutrient, a non-provitamin A carotenoid pigment and because it does not have a terminal beta ionone ring it does not mediate vitamin A activity. However , it is a potent antioxidant molecule for reactive oxygen species (ROS) singlet oxygen. [800]

In Terpenoid backbone biosynthesis, Terpenoids(isoprenoids) are natural products consisting of isoprene (C5) units with two biosynthetic pathways, (1) mevalonate pathway and (2) non-mevalonate pathway or the MEP/DOXP pathway. The terpenoid components are isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP). The action of prenyltransferases creates higher-order components such as (a) geranyl diphosphate (GPP), (b) farsenyl diphosphate (FPP), and (c) geranylgeranyl diphosphate (GGPP) as precursors of (1) monoterpenoids (C10), (2) sesquiterpenoids (C15), and (3) diterpenoids (C20), respectively.[401]

Condensation of these building components have precursors of (a) sterols (C30) and (b) carotenoids (C40). While the MEP/DOXP pathway is absent in higher animals and fungi, green plants have MEP/DOXP and mevalonate pathways co-exist in separate cellular compartments. The MEP/DOXP pathway has the formation of (1) essential oil monoterpenes and (2) linalyl acetate, (3) sesquiterpenes, (4) diterpenes, (5) carotenoids and (6) phytol. The mevalonate pathway in cytosol generates (1) triterpenes, (2) sterols, and (3) sesquiterpenes. [401]

Here the Terpenoid backbone biosynthesis for HMGCS is shown in Figure 2.

*

Email address: http://mathlearningspace.weebly.com/ @

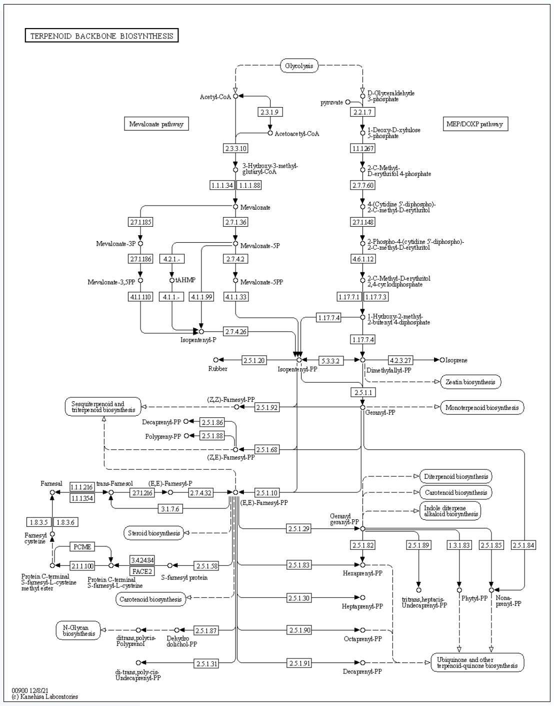

Figure 2: Terpenoid backbone biosynthesis with C5 isoprenoid biosynthesis, mevalonate pathway, C5 isoprenoid biosynthesis, non-mevalonate pathway, C10-C20 isoprenoid biosynthesis, bacteria ,C10-C20 isoprenoid biosynthesis, archaea, C10-C20 isoprenoid biosynthesis, plants , C10-C20 isoprenoid biosynthesis, non-plant eukaryotes, C5 isoprenoid biosynthesis, mevalonate pathway, archaea, Cyclooctatin biosynthesis, dimethylallyl-PP + isopentenyl-PP => cyclooctatin [401]

Figure 3 has the carotenoid biosynthesis pathway

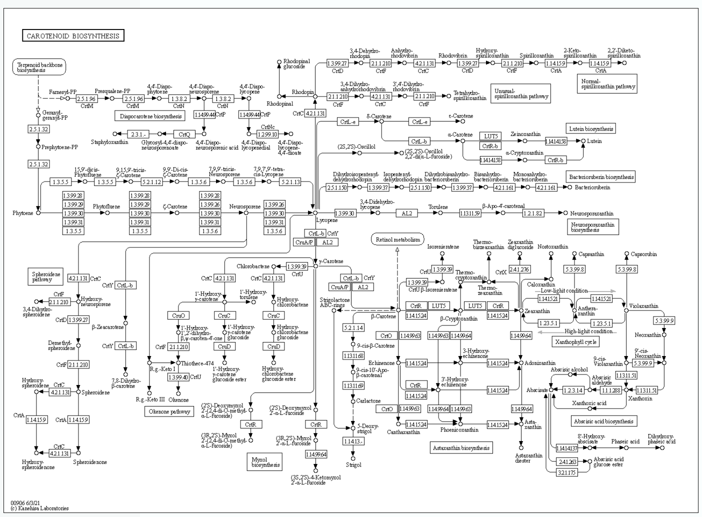

Figure 3: carotenoid biosynthesis pathway [401]

Several reactions with Lycopene are part of the following enzymes 1.3.5.6, 1.3.99.26, 1.3.99.30, 1.3.99.31 2.5.1.149, 2.5.1.150, 4.2.1.131, 5.2.1.13, 5.5.1.18 and 5.5.1.19 are included in the list. [401]

1. beta-Carotene <=> Lycopene
2. all-trans-zeta-Carotene + 2 Oxygen + 2 Reduced acceptor <=> Lycopene + 4 H2O + 2 Acceptor
3. all-trans-Neurosporene + Reduced acceptor + Oxygen <=> Lycopene + Acceptor + 2 H2O
4. Lycopene <=> gamma-Carotene
5. Lycopene <=> delta-Carotene
6. Lycopene + Reduced acceptor + Oxygen <=> 3,4-Dehydrolycopene + Acceptor + H2O
7. Lycopene + H2O <=> 1-Hydroxy-1,2-dihydrolycopene
8. Lycopene <=> (2S,2'S)-Oscillol
9. Lycopene <=> R.g.-Keto I
10. all-trans-zeta-Carotene + 2 Acceptor <=> Lycopene + 2 Reduced acceptor
11. 15-cis-Phytoene + 4 Acceptor <=> Lycopene + 4 Reduced acceptor
12. Dimethylallyl diphosphate + Lycopene + Acceptor + H2O <=> Nonaflavuxanthin + Reduced acceptor + Diphosphate
13. Dimethylallyl diphosphate + Lycopene + H2O <=> Dihydroisopentenyldehydrorhodopin + Diphosphate

Specifically, the Table 1 has the compound and reaction for the beta-Carotene biosynthesis module.

ID	Name	Reaction
C00353	Geranylgeranyl diphosphate	C00353 -> C03427
C03427	Prephytoene diphosphate	C03427 -> C05421
C05421	15-cis-Phytoene	C05421 -> C19765
C19765	15,9'-dicis-Phytofluene	C19765 -> C19764
C19764	9,15,9'-tricis-zeta-Carotene	C19764 -> C15857
C15857	9,9'-dicis-zeta-Carotene	C15857 -> C19759
C19759	7,9,9'-tricis-Neurosporene	C19759 -> C15858
C15858	7,9,7',9'-tetracis-Lycopene	C15858 -> C05432
C05432	Lycopene	C05432 -> C05435
C05435	gamma-Carotene	C05435 -> C02094
C02094	beta-Carotene	C00353 -> C03427

Table 1: Enyzme sequence 2.5.1.32, 1.3.5.5, 5.2.1.12, 1.3.5.6, 5.2.1.13, lycopene, CRI-b, Beta-Carotene [401]

The Gamma-carotene is a cyclic carotene obtained by the cyclisation of lycopene and has a role as a plant metabolite and a fungal metabolite. It is a cyclic carotene and a carotenoid beta-end group. Figure 4 has the first step in the sequence with Geranylgeranyl diphosphate from Table 2.

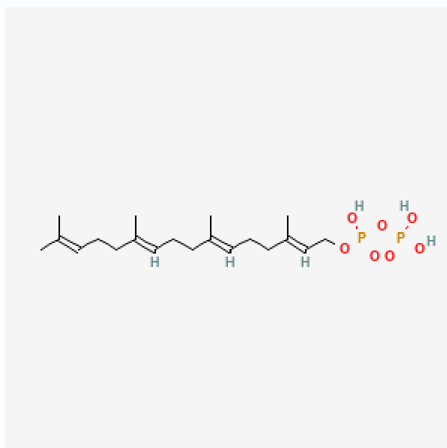

Figure 4: Geranylgeranyl diphosphate [800]

Figure 5 has the chemical structure with its eye of a needle represented by the attached ring for (a) and the double ring with (b). [800]

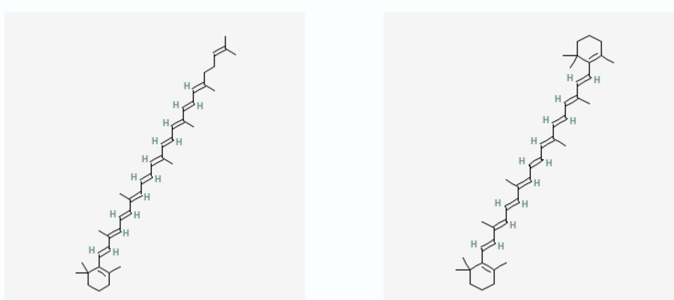

Figure 5: (a) gamma-Carotene and (b) beta-Carotene [800]

Beta-Carotene has potential antineoplastic, chemopreventive activities and antioxidant properties and naturally-occurring retinol (vitamin A) obtained from fruits and vegetables. Beta carotene inhibits (a) free-radical damage to DNA and induces (b) cell differentiation and apoptosis of some tumor cell types-early stages of tumorigenesis, and enhances immune system activity by stimulating the release of (a) natural killer cells, (b) lymphocytes, and (c) monocytes. [800]

Based on the collection of PMIDs for CID 446925 Table 2 has the PMIDs and Titles from an abstract search with the keyword LYMPH. [601]

ID	PMID	Title
67	22550343	Prediagnostic circulating carotenoid levels and the risk of non-Hodgkin lymphoma: the Multiethnic Cohort
147	22530142	Remedial Prospective of Hippophae rhamnoides Linn. (Sea Buckthorn)
443	21190502	Peripheral blood lymphocytes: a model for monitoring physiological adaptation to high altitude
790	17913757	Age-related increases in DNA repair and antioxidant protection: a comparison of the Boyd Orr Cohort of elderly subjects with a younger population sample
1127	15705230	Daily intake of a formulated tomato drink affects carotenoid plasma and lymphocyte concentrations and improves cellular antioxidant protection
1139	16438805	A half-marathon and a marathon run induce oxidative DNA damage, reduce antioxidant capacity to protect DNA against damage and modify immune function in hobby runners
1151	15308586	Effects of supplementing a low-carotenoid diet with a tomato extract for 2 weeks on endogenous levels of DNA single strand breaks and immune functions in healthy non-smokers and smokers
1176	15054415	Lycopene and vitamin C concentrations increase in plasma and lymphocytes after tomato intake. Effects on cellular antioxidant protection
1211	15461275	[Experimental studies of lycopene in inhibiting tumor growth in S180-bearing mice]
1258	14748938	DNA damage and susceptibility to oxidative damage in lymphocytes: effects of carotenoids in vitro and in vivo
1295	14511231	Retinoid- and carotenoid-enriched diets influence the ontogenesis of the immune system in mice
1378	14695926	Statins lower plasma and lymphocyte ubiquinol/ubiquinone without affecting other antioxidants and PUFA
1422	12067573	DNA damage and antioxidants; fluctuations through the year in a central European population group
1431	12180130	Carotenoids induce apoptosis in the T-lymphoblast cell line Jurkat E6.1
1449	12020262	Lymphocyte oxidative DNA damage and plasma antioxidants in Alzheimer disease
1464	11884238	beta-Carotene reduces bleomycin-induced genetic damage in human lymphocytes
1526	11477460	Low-dose supplementation with lycopene or beta-carotene does not enhance cell-mediated immunity in healthy free-living elderly humans

Based on the collection of PMIDs for CID 446925 Table 3 has the PMIDs and Titles from an abstract search with the keyword RADIATION. [601]

ID	PMID	Title
21	22549412	Alleviation of chilling injury in postharvest tomato fruit by preconditioning with ultraviolet irradiation
40	22692080	Decylglucoside-based microemulsions for cutaneous localization of lycopene and ascorbic acid
55	22743816	Nutrition support and therapy in patients with head and neck squamous cell carcinomas
275	21294875	The use of some nanoemulsions based on aqueous propolis and lycopene extract in the skin's protective mechanisms against UVA radiation
285	20853276	The effect of environmental conditions on nutritional quality of cherry tomato fruits: evaluation of two experimental Mediterranean greenhouses
286	20151398	Determination of the influence of IR radiation on the antioxidative network of the human skin
315	20637178	Significant correlations of dermal total carotenoids and dermal lycopene with their respective plasma levels in healthy adults
364	22254062	The role of phytonutrients in skin health
374	20041432	Lycopene protects the structure of the small intestine against gamma-radiation-induced oxidative stress
419	19591120	A simple and rapid method to assess lycopene in multiple layers of skin samples
427	19681037	Resonance Raman spectroscopy as an effective tool for the determination of antioxidative stability of cosmetic formulations
432	20090407	Radical production by infrared A irradiation in human tissue
514	19450652	Lycopene: An antioxidant and radioprotector against gamma-radiation-induced cellular damages in cultured human lymphocytes
610	20155616	Lycopene in the prevention of gastrointestinal toxicity of radiotherapy
627	18959418	Influence of electron-beam irradiation on bioactive compounds in grapefruits (Citrus paradisi Macf.)
645	18696411	Phytochemicals as protectors against ultraviolet radiation: versatility of effects and mechanisms
819	17368731	Evaluation of the antioxidant effects of carotenoids from Deinococcus radiodurans through targeted mutagenesis, chemiluminescence, and DNA damage analyses
838	17189673	Lycopene as a natural protector against gamma-radiation induced DNA damage, lipid peroxidation and antioxidant status in primary culture of isolated rat hepatocytes in vitro
888	17992006	Culture characteristics of carotenoid-producing filamentous fungus T-1, and carotenoid production
986	16736578	[Construction and functional analysis of the crtI gene disruptant in Deinococcus radiodurans]
1017	16689237	[Combined effects of enhanced UV-B radiation and doubled CO2 on tomato growth and its fruit quality]
1080	16018824	Prostate cancer: chemoprevention update 2005
1086	15884827	Bioactive compounds of grapefruit (Citrus paradisi Cv. Rio Red) respond differently to postharvest irradiation, storage, and freeze drying
1255	14678532	Antioxidant activity of topically applied lycopene
1570	11265592	Role of lycopene in recovery of radiation induced injury to mammalian cellular organelles

Based on the collection of PMIDs for CID 446925 Table 3 has the PMIDs and Titles from an abstract search with the keyword GASTRIC CANCER. [601]

ID	PMID	Title
299	21686188	Lycopene enhances antioxidant enzyme activities and immunity function in N-methyl-N'-nitro-N-nitrosoguanidine-enduced gastric cancer rats
581	18669867	Dietary intake of selected micronutrients and gastric cancer risk: an Italian case-control study
693	18339681	Plasma levels of carotenoids, retinol and tocopherol and the risk of gastric cancer in Japan: a nested case-control study
694	18454810	Nutrition and gastric cancer risk: an update
1064	16172214	Fruits, vegetables, and antioxidants and risk of gastric cancer among male smokers
1069	16030430	Combination of S-allylcysteine and lycopene induces apoptosis by modulating Bcl-2, Bax, Bim and caspases during experimental gastric carcinogenesis
1137	16379561	Combination chemoprevention of experimental gastric carcinogenesis by s-allylcysteine and lycopene: modulatory effects on glutathione redox cycle antioxidants
1170	15533906	Prediagnostic levels of serum micronutrients in relation to risk of gastric cancer in Shanghai, China

3 Results

Molecular descriptors provide a dimensional reduction summary measure that can be used for comparisons in drug discovery. In the categories of (1) hybrid (2) constitutional, (3) topological (4) electronic and (5) geometrical, descriptors were selected to determine the probability distribution from the Tanimoto similarity of N=2573 compounds with the methods of maximum likelihood. Here the electronic properties of BPol and APOL are used along with complexity. Consider the constiutional descriptors for N in Table 1. [1001]

	XLogP	MW	LipinskiFailures	nRotB	MLogP	nAtomLAC	nAtomP
n	2573.00	2573.00	2573.00	2573.00	2573.00	2573.00	2573.00
mean	9.43	286.24	1.07	7.14	3.75	11.55	8.46
sd	6.00	179.90	0.49	5.91	1.43	7.37	5.25
median	8.34	240.28	1.00	5.00	3.44	9.00	6.00
trimmed	8.58	256.07	1.07	6.07	3.51	10.21	7.64
mad	3.14	74.15	0.00	1.48	0.65	2.97	2.97
min	3.94	124.13	0.00	0.00	2.45	0.00	2.00
max	109.15	3381.05	2.00	65.00	28.85	90.00	50.00
range	105.21	3256.93	2.00	65.00	26.40	90.00	48.00
skew	7.01	5.71	0.15	3.31	5.77	2.16	1.83
kurtosis	87.36	62.27	0.97	17.46	63.92	7.70	5.87
se	0.12	3.55	0.01	0.12	0.03	0.15	0.10

Table 2: XLogP = XLogPDescriptor, MW = WeightDescriptor, LipinskiFailures = RuleOfFiveDescriptor, nRotB = RotatableBondsCountDescriptor, MLogP = MannholdLogPDescriptor, nAtomLAC = LongestAliphaticChainDescriptor and nAtomP = LargestPiSystemDescriptor [101]

Figure 1 has the scatterplot matrix for XLogPDescriptor-Prediction of logP based on the atom-type method. RotatableBondsCountDescriptor-The number of rotatable bonds (RBN) is the number of bonds which allow free rotation around themselves. These are defined as any single bond, not in a ring, bound to a nonterminal heavy atom. The number of rings (nCIC) counts the rings in a molecule. LargestChainDescriptor-Calculate the count of atoms of the largest chain and ALOGPDescriptor-(Ghose-Crippen LogKow) and the Ghose-Crippen molar refractivity with the histogram on the diagonal.

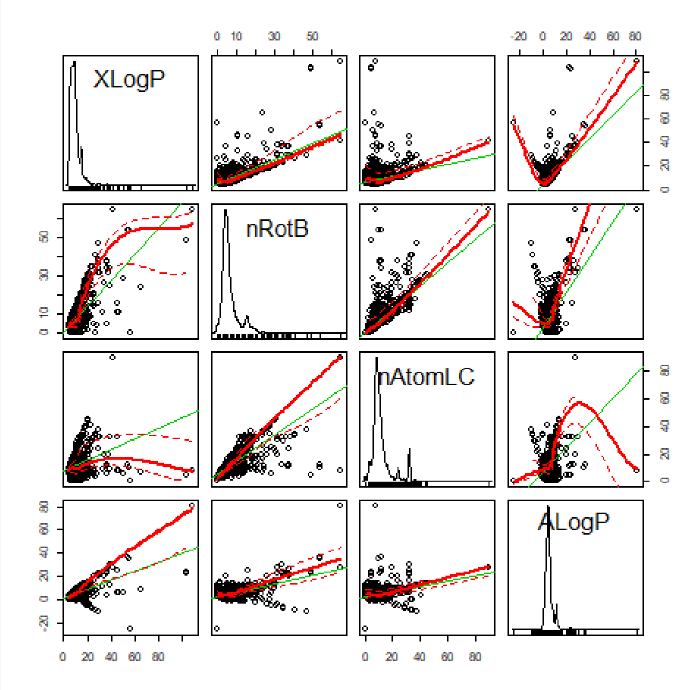

Figure 6: XLogPDescriptor RotatableBondsCountDescriptor LargestChain-Descriptor ALOGPDescriptor [1000]

Figure 6 has of the maximum likelihood estimates of the complexity measure for Tanimoto similarity with CID 446925 and Pearson distribution type=6 with a=6.67 and b=3.30 and location=43.370525888452 and scale=112.14. Type 6 are (location-scale transformations of) Beta prime distributions, and Beta prime distributions are scaled F-distributions. The probability density function with parameters a, b, scale=s and location=λ is given by

$$f(x) = \frac{\Gamma(a+b)}{(|s|\Gamma(a)\Gamma(b))}((x-\lambda)/s)^{(a-1)}(1+(x-\lambda)/s)^{(-a-b)} \tag{1}$$

for $a > 0,\ b > 0,\ s <> 0,\ (x-\lambda)/s > 0.$

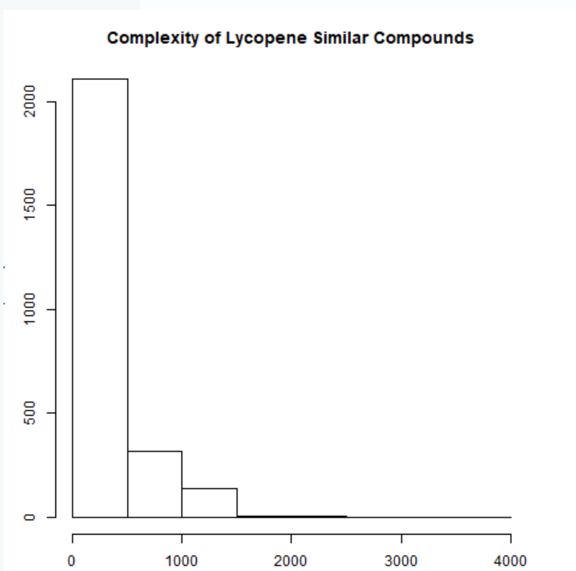

Figure 7: Histogram of Complexity Similarity with Min. 1st Qu. Median Mean 3rd Qu. Max of 62.4 192.0 285.0 365.6 409.0 3960.0 [800]

Figure 7 has Sum of the absolute value of the difference between atomic polarizabilities of all bonded atoms in the molecule (including implicit hydrogens) with polarizabilities. This descriptor assumes 2-centered bonds. The maximum likelihood estimate is distribution is type=5 with shape=3.151 and location=9.611 and scale=60.56. The Type 5 is Inverse Gamma distributions. The probability density function with parameters shape=a, scale=s and location=λ is given by [100]

$$f(x) = |s|^{a}/\Gamma(a)|x-\lambda|^{(-a-1)}e^{-(s/(x-\lambda))} \tag{2}$$

for $s <> 0,\ a > 0$ and $s/(x-\lambda) > 0.$

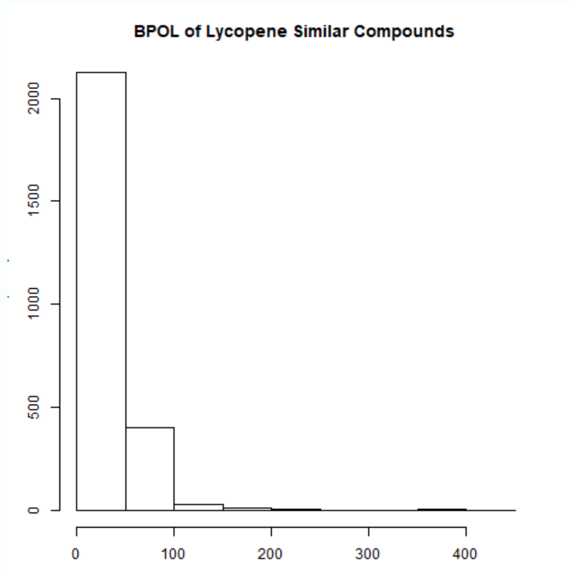

Figure 8: Histogram of BPOL Similarity Min. 1st Qu. Median Mean 3rd Qu. Max. 15.30 24.05 32.80 37.61 41.54 426.35 [800]

Figure 8 Sum of the atomic polarizabilities (including implicit hydrogens). Here the ML estimate is type=5 with shape=3.47 and location=15.09 and scale=110.15.

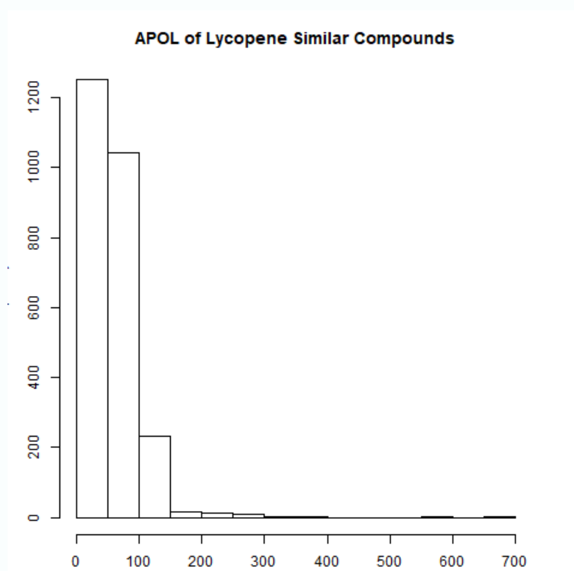

Figure 9: Histogram of APOL Similarity with Min. 1st Qu. Median Mean 3rd Qu. Max. 26.51 41.50 50.83 59.79 63.58 698.29 [800]

Figure 9 has the relationship between each of the three measures with the scatterplot matrix with boxplots on the diagonal for each molecular summary measure.

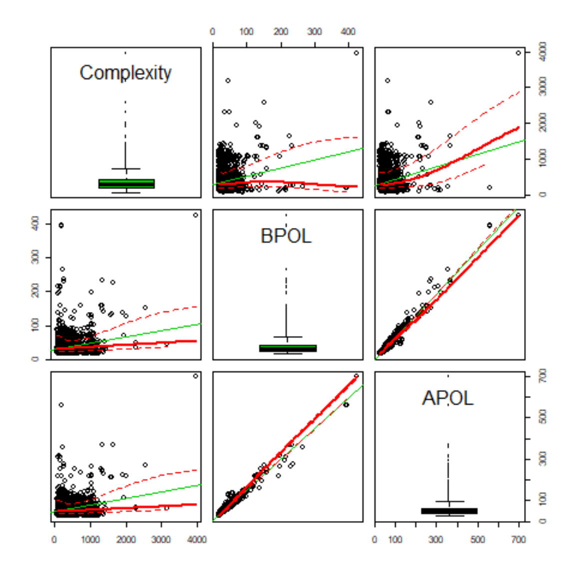

Figure 10: Scatterplot Matrix with Boxplots [800]

The relationship between BPOL and APOL is expected to be linear with a curvilinear relationship with complexity. This reflected in the difference between the Type 5 and 6 distributions.

4 Conclusions

In this mathematical note, a quick review of Terpenoid backbone biosynthesis and carotenoid biosynthesis was conducted with their enzyme reactions with Lycopene. The Tanimoto similarity compounds provides a useful data set to examine the molecular features for possible machine learning application in the design of Anticarcinogenic , Radiation-Protective, Anti-Inflammatory, and antioxidant Agents in various tissue locations. For three molecular properties of complexity, BPOL and APOL, Pearson distributions of type 6 and 5 were obtained with the expected skewness with the Tanimoto similarity for the compounds. Given the design of the geometric structure of Lycopene, the design of compounds in the reaction relationship can provide insight in the development of additional agents for clinical applications.

5 Bibliography

[100] Martin Becker and Stefan Klößner (2017). PearsonDS: Pearson Distribution System. R package version 1.1. https://CRAN.R-project.org/package=PearsonDS

[101] Guha, R. (2007). 'Chemical Informatics Functionality in R'. Journal of Statistical Software 6(18)

[400] Kanehisa, Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K.; KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res. 45, D353-D361 (2017).

[401] Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M.; KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. 44, D457-D462 (2016).

[402] Kanehisa, M. and Goto, S.; KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 28, 27-30 (2000).

[403] Petri, V., Jayaraman, P., Tutaj, M., Hayman, G. T., Smith, J. R., De Pons, J., . . . Jacob, H. J. (2014). The pathway ontology – updates and applications. Journal of Biomedical Semantics, 5, 7. http://doi.org/10.1186/2041-1480-5-7

[601] Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, Jensen LJ, von Mering C. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. 2019 Jan; 47:D607-613.PubMed

[602] Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P, Jensen LJ, von Mering C. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Res. 2017 Jan; 45:D362-68.PubMed

[603] Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, Kuhn M, Bork P, Jensen LJ, von Mering C. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. 2015 Jan; 43:D447-52.PubMed

[604] Franceschini A, Lin J, von Mering C, Jensen LJ. SVD-phy: improved prediction of protein functional associations through singular value decomposition of phylogenetic profiles. Bioinformatics. 2015 Nov; btv696.PubMed

[605] Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, Lin J, Minguez P, Bork P, von Mering C, Jensen LJ. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res. 2013 Jan; 41:D808-15.PubMed

[606] Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P, Doerks T, Stark M, Muller J, Bork P, Jensen LJ, von Mering C. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. Nucleic Acids Res. 2011 Jan; 39:D561-8.PubMed

[607] Jensen LJ, Kuhn M, Stark M, Chaffron S, Creevey C, Muller J, Doerks T, Julien P, Roth A, Simonovic M, Bork P, von Mering C. STRING 8–a global view on proteins and their functional interactions in 630 organisms. Nucleic Acids Res. 2009 Jan; 37:D412-6.PubMed

[608] von Mering C, Jensen LJ, Kuhn M, Chaffron S, Doerks T, Krueger B, Snel B, Bork P. STRING 7–recent developments in the integration and prediction of protein interactions. Nucleic Acids Res. 2007 Jan; 35:D358-62.PubMed

[609] von Mering C, Jensen LJ, Snel B, Hooper SD, Krupp M, Foglierini M, Jouffre N, Huynen MA, Bork P. STRING: known and predicted protein-protein associations, integrated and transferred across organisms. Nucleic Acids Res. 2005 Jan; 33:D433-7.PubMed

[610] von Mering C, Huynen M, Jaeggi D, Schmidt S, Bork P, Snel B. STRING: a database of predicted functional associations between proteins. Nucleic Acids Res. 2003 Jan; 31:258-61.PubMed

[611] Snel B, Lehmann G, Bork P, Huynen MA. STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. Nucleic Acids Res. 2000 Sep 15;28(18):3442-4.PubMed

[700] National Cancer Institute Annual Plan and Budget Proposal FY 2019 https://www.cancer.gov/about-nci/budget/plan

[800] Kim, Sunghwan et al. "PubChem in 2021: new data content and improved web interfaces." Nucleic Acids Res. vol. 49,D1 (2021): D1388-D1395. doi:10.1093/nar/gkaa971

[1000] R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

[1001] Osorio, D., Rondon-Villarreal, P. and Torres, R. Peptides: A package for data mining of antimicrobial peptides. The R Journal. 7(1), 4-14 (2015).