Molecular Properties of 3-hydroxy-3-methylglutaryl CoA synthetase Coexpression in Terpenoid Backbone Biosynthesis

Jeff Cromwell, PhD

The Mathematical Learning Space Research Portfolio

1 Abstract

Terpenoid backbone biosynthesis has (1) Glycolysis / Gluconeogenesis, (2) Steroid biosynthesis, (3) Ubiquinone and other terpenoid-quinone biosynthesis, (4) Indole diterpene alkaloid biosynthesis, (5) N-Glycan biosynthesis, (6) Monoterpenoid biosynthesis, (7) Diterpenoid biosynthesis, (8) Carotenoid biosynthesis, (9) Zeatin biosynthesis and Sesquiterpenoid and (10) triterpenoid biosynthesis as a collection of pathways. For the 3-hydroxy-3-methylglutaryl CoA synthetase coexpression network, all of the molecules are stable with only EPRS near the boundary while the binding potential is low for all the molecules. The distinction between hydrophobicity has SC5D, SQLE and HMGCR as positive with the rest negative. Kidera factors and VHSE-scales were computed for the coexpressions based on experimental evidence and cluster analysis performed. Charge differential from positive to negative occurs in all molecules from 5-9 with some variation at pH 7. First degree experimental coexpression relationships for 2.3.3.10 (HMGCS) of CYP51A1 MSMO1 DARS SQLE HMGCR MVD KARS IARS DHCR7 LSS NSDHL IDI1 and FDPS were also examined for Sterol biosynthesis and Disease mutation.

2 Introduction

3-hydroxy-3-methylglutaryl CoA synthetase or 2.3.3.10 (HMGCS) is involved in several pathways such as (1) Valine, leucine and isoleucine degradation, (2) Butanoate metabolism, (3) Terpenoid backbone biosynthesis, (4) Metabolic pathways, (5) Biosynthesis of secondary metabolites and (6) PPAR signaling pathway. [401] 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 1.

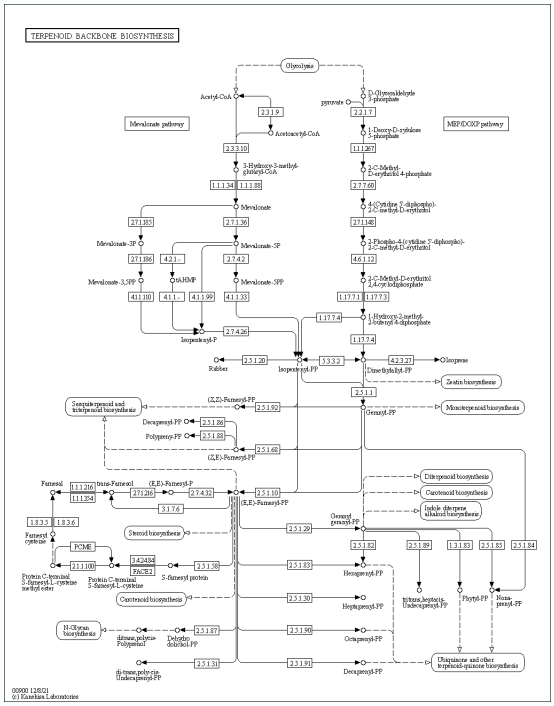

Figure 1: Terpenoid backbone biosynthesis with C5 isoprenoid biosynthesis, mevalonate pathway, C5 isoprenoid biosynthesis, non-mevalonate pathway, C10-C20 isoprenoid biosynthesis, bacteria ,C10-C20 isoprenoid biosynthesis, archaea, C10-C20 isoprenoid biosynthesis, plants , C10-C20 isoprenoid biosynthesis, non-plant eukaryotes, C5 isoprenoid biosynthesis, mevalonate pathway, archaea, Cyclooctatin biosynthesis, dimethylallyl-PP + isopentenyl-PP => cyclooctatin [401]

The chemical reaction for 2.3.3.10 is given by acetyl-CoA + H2O + acetoacetyl-CoA = (S)-3-hydroxy-3-methylglutaryl-CoA + CoA has substrate acetyl-CoA, H2O and acetoacetyl-CoA with product (S)-3-hydroxy-3-methylglutaryl-CoA and CoA. [401]. Table 0 has the ATP and ADP search results from a collection of PMIDs for CID 446925 [601].

ID	PMID	Title
14	26476475	Lycopene protects against atrazine-induced hepatic ionic homeostasis disturbance by modulating ion-transporting ATPases
41	22137263	Lycopene and the LXRÎ± agonist T0901317 synergistically inhibit the proliferation of androgen-independent prostate cancer cells via the PPARÎ³-LXRÎ±-ABCA1 pathway
440	20645919	Reversal of multidrug resitance by natural substances from plants
531	19329757	Differential effects of several phytochemicals and their derivatives on murine keratinocytes in vitro and in vivo: implications for skin cancer prevention
1242	15113052	Modulation of multidrug resistance and apoptosis of cancer cells by selected carotenoids
365	20565070	An efficient and economical MTT assay for determining the antioxidant activity of plant natural product extracts and pure compounds
1325	14534614	Effect of lycopine on the resistance of rat liver microsomes to in vitro induced LPO
1460	11943208	Involvement of NADPH in the cyclization reaction of carotenoid biosynthesis

*

Email address: http://mathlearningspace.weebly.com/ (Jeff Cromwell, PhD)

3 HMGCS Coexpression Network

This transferases class with cyl groups converted into alkyl groups on transfer has interesting coexpression properties. [1]Figure 2 has the coexpression relationships for HMGCS1-Hydroxymethylglutaryl-CoA synthase, cytoplasmic; This enzyme condenses acetyl-CoA with acetoacetyl-CoA to form HMG-CoA a substrate for HMG-CoA reductase based on experimental evidence.

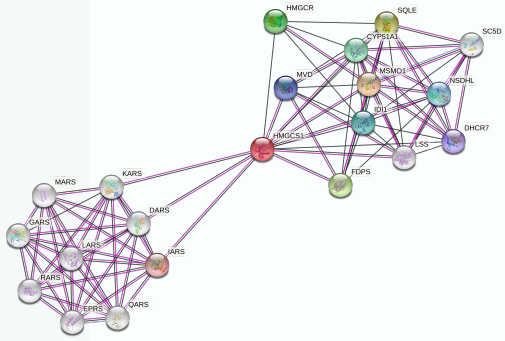

Figure 2: HMGCS Coexpression Network [601]

Table 1 has the associations for HMGCS1 based on Figure 2. [601]

	To	From
6	CYP51A1	HMGCS1
25	DHCR7	HMGCS1
41	FDPS	HMGCS1
56	HMGCR	HMGCS1
57	HMGCS1	CYP51A1
58	HMGCS1	MSMO1
59	HMGCS1	DARS
60	HMGCS1	SQLE
61	HMGCS1	HMGCR
62	HMGCS1	MVD
63	HMGCS1	KARS
64	HMGCS1	IARS
65	HMGCS1	DHCR7
66	HMGCS1	LSS
67	HMGCS1	NSDHL
68	HMGCS1	IDI1
69	HMGCS1	FDPS
74	IARS	HMGCS1
85	IDI1	HMGCS1
94	KARS	HMGCS1
112	LSS	HMGCS1
130	MSMO1	HMGCS1
182	SQLE	HMGCS1

Table 2 has descriptions of the genes from Figure 1. [601]

	Network ID	Description
1	CYP51A1	Lanosterol 14-alpha demethylase; Catalyzes C14-demethylation of lanosterol; it transforms lanosterol into 4,4'-dimethyl cholesta-8,14,24-triene-3-beta-ol; Cytochrome P450 family 51
2	DARS	Aspartate–tRNA ligase, cytoplasmic; Catalyzes the specific attachment of an amino acid to its cognate tRNA in a 2 step reaction: the amino acid (AA) is first activated by ATP to form AA-AMP and then transferred to the acceptor end of the tRNA; Belongs to the class-II aminoacyl-tRNA synthetase family. Type 2 subfamily
3	DHCR7	7-dehydrocholesterol reductase; Production of cholesterol by reduction of C7-C8 double bond of 7-dehydrocholesterol (7-DHC); Belongs to the ERG4/ERG24 family
4	EPRS	Bifunctional glutamate/proline–tRNA ligase; Catalyzes the attachment of the cognate amino acid to the corresponding tRNA in a two-step reaction: the amino acid is first activated by ATP to form a covalent intermediate with AMP and is then transferred to the acceptor end of the cognate tRNA. Component of the GAIT (gamma interferon-activated inhibitor of translation) complex which mediates interferon-gamma-induced transcript- selective translation inhibition in inflammation processes.
5	FDPS	Farnesyl pyrophosphate synthase; Key enzyme in isoprenoid biosynthesis which catalyzes the formation of farnesyl diphosphate (FPP), a precursor for several classes of essential metabolites including sterols, dolichols, carotenoids, and ubiquinones. FPP also serves as substrate for protein farnesylation and geranylgeranylation. Catalyzes the sequential condensation of isopentenyl pyrophosphate with the allylic pyrophosphates, dimethylallyl pyrophosphate, and then with the resultant geranylpyrophosphate to the ultimate product farnesyl pyrophosphate
6	GARS	Glycine–tRNA ligase; Catalyzes the ligation of glycine to the 3'-end of its cognate tRNA. Also produces diadenosine tetraphosphate (Ap4A), a universal pleiotropic signaling molecule needed for cell regulation pathways, by direct condensation of 2 ATPs; Belongs to the class-II aminoacyl-tRNA synthetase family
7	HMGCR	3-hydroxy-3-methylglutaryl-coenzyme A reductase; Transmembrane glycoprotein that is the rate-limiting enzyme in cholesterol biosynthesis as well as in the biosynthesis of nonsterol isoprenoids that are essential for normal cell function including ubiquinone and geranylgeranyl proteins; Belongs to the HMG-CoA reductase family
8	HMGCS1	Hydroxymethylglutaryl-CoA synthase, cytoplasmic; This enzyme condenses acetyl-CoA with acetoacetyl-CoA to form HMG-CoA, which is the substrate for HMG-CoA reductase
9	IARS	Isoleucine–tRNA ligase, cytoplasmic; Catalyzes the specific attachment of an amino acid to its cognate tRNA in a 2 step reaction: the amino acid (AA) is first activated by ATP to form AA-AMP and then transferred to the acceptor end of the tRNA; Aminoacyl tRNA synthetases, Class I
10	IDI1	Isopentenyl-diphosphate Delta-isomerase 1; Catalyzes the 1,3-allylic rearrangement of the homoallylic substrate isopentenyl (IPP) to its highly electrophilic allylic isomer, dimethylallyl diphosphate (DMAPP)
11	KARS	Lysine–tRNA ligase; Catalyzes the specific attachment of an amino acid to its cognate tRNA in a 2 step reaction: the amino acid (AA) is first activated by ATP to form AA-AMP and then transferred to the acceptor end of the tRNA. When secreted, acts as a signaling molecule that induces immune response through the activation of monocyte/macrophages. Catalyzes the synthesis of the signaling molecule diadenosine tetraphosphate (Ap4A), and thereby mediates disruption of the complex between HINT1 and MITF and the concomitant activation of MITF transcriptional activity.
12	LARS	Leucine–tRNA ligase, cytoplasmic; Catalyzes the specific attachment of an amino acid to its cognate tRNA in a two step reaction: the amino acid (AA) is first activated by ATP to form AA-AMP and then transferred to the acceptor end of the tRNA. Exhibits a post-transfer editing activity to hydrolyze mischarged tRNAs; Aminoacyl tRNA synthetases, Class I
13	LSS	Lanosterol synthase; Catalyzes the cyclization of (S)-2,3 oxidosqualene to lanosterol, a reaction that forms the sterol nucleus. Through the production of lanosterol may regulate lens protein aggregation and increase transparency; Belongs to the terpene cyclase/mutase family
14	MARS	Methionine–tRNA ligase, cytoplasmic; Catalyzes the specific attachment of an amino acid to its cognate tRNA in a 2 step reaction: the amino acid (AA) is first activated by ATP to form AA-AMP and then transferred to the acceptor end of the tRNA; Belongs to the class-I aminoacyl-tRNA synthetase family
15	MSMO1	Methylsterol monooxygenase 1; Catalyzes the first step in the removal of the two C-4 methyl groups of 4,4-dimethylzymosterol; Belongs to the sterol desaturase family
16	MVD	Mevalonate diphosphate decarboxylase; Diphosphomevalonate decarboxylase; Performs the first committed step in the biosynthesis of isoprenes
17	NSDHL	Sterol-4-alpha-carboxylate 3-dehydrogenase, decarboxylating; Involved in the sequential removal of two C-4 methyl groups in post-squalene cholesterol biosynthesis; Short chain dehydrogenase/reductase superfamily
18	QARS	Glutamine–tRNA ligase; Glutamine–tRNA ligase. Plays a critical role in brain development; Belongs to the class-I aminoacyl-tRNA synthetase family
19	RARS	Arginine–tRNA ligase, cytoplasmic; Forms part of a macromolecular complex that catalyzes the attachment of specific amino acids to cognate tRNAs during protein synthesis. Modulates the secretion of AIMP1 and may be involved in generation of the inflammatory cytokine EMAP2 from AIMP1; Aminoacyl tRNA synthetases, Class I
20	SC5D	Delta7-sterol 5-desaturase; Lathosterol oxidase; Catalyzes a dehydrogenation to introduce C5-6 double bond into lathosterol; Belongs to the sterol desaturase family
21	SQLE	Squalene monooxygenase; Catalyzes the first oxygenation step in sterol biosynthesis and is suggested to be one of the rate-limiting enzymes in this pathway; Belongs to the squalene monooxygenase family

In the relationship between HMGCS1 and CYP51A1 MSMO1 DARS SQLE HMGCR MVD KARS IARS DHCR7 LSS NSDHL IDI1 and FDPS the following PMIDs shown in Table 3. [601]

ID	PMID	Title	Network IDs
373	PMID:26598836	(2015) Navigating the Shallows and Rapids of Cholesterol Synthesis Downstream of HMGCR.	SQLE, HMGCR, DHCR7
463	PMID:18775413	(2008) CREM modulates the circadian expression of CYP51, HMGCR and cholesterogenesis in the liver.	CYP51A1, HMGCR, FDPS
466	PMID:26632252	(2015) MicroRNA-195 inhibits proliferation, invasion and metastasis in breast cancer cells by targeting FASN, HMGCR, ACACA and CYP27B1.	SQLE, HMGCR, HMGCS1, IDI1
860	PMID:29163687	(2017) Stromal regulation of prostate cancer cell growth by mevalonate pathway enzymes HMGCS1 and HMGCR.	HMGCR, HMGCS1
208	PMID:24711211	(2015) CYP51A1 induced by growth differentiation factor 9 and follicle-stimulating hormone in granulosa cells is a possible predictor for unfertilization.	CYP51A1, SQLE, HMGCS1, LSS
549	PMID:24358204	(2013) Polymorphisms of CYP51A1 from cholesterol synthesis: associations with birth weight and maternal lipid levels and impact on CYP51 protein structure.	CYP51A1, HMGCR, DHCR7
765	PMID:30582412	(2019) RNA-Seq analysis reveals a negative role of MSMO1 with a synergized NSDHL expression during adipogenesis of 3T3-L1.	MSMO1, NSDHL
846	PMID:34759952	(2021) The Clinical Significance and Immunization of MSMO1 in Cervical Squamous Cell Carcinoma Based on Bioinformatics Analysis.	MSMO1 ,IDI1
15	PMID:27871331	(2016) Association of DARS gene polymorphisms with the risk of isolated ventricular septal defects in the Chinese Han population.	RARS, MARS, DARS, QARS, KARS, EPRS, IARS, LARS
717	PMID:30626880	(2019) A chemical biology screen identifies a vulnerability of neuroendocrine cancer cells to SQLE inhibition.	SQLE, HMGCR, LSS
373	PMID:26598836	(2015) Navigating the Shallows and Rapids of Cholesterol Synthesis Downstream of HMGCR.	SQLE, HMGCR, DHCR7
632	PMID:29615062	(2018) KARS-related diseases: progressive leukoencephalopathy with brainstem and spinal cord calcifications as new phenotype and a review of literature.	DARS, KARS, GARS
461	PMID:20804844	(2010) Hair and skin sterols in normal mice and those with deficient dehydrosterol reductase (DHCR7), the enzyme associated with Smith-Lemli-Opitz syndrome.	CYP51A1, HMGCR, DHCR7
555	PMID:32101538	(2020) Metabolic and pathologic profiles of human LSS deficiency recapitulated in mice.	MSMO1, DHCR7, LSS
764	PMID:32877255	(2020) The Polymorphism rs2968 of LSS Gene Confers Susceptibility to Age-Related Cataract.	HMGCR, LSS
27	PMID:26456460	(2015) FR171456 is a specific inhibitor of mammalian NSDHL and yeast Erg26p.	CYP51A1, MSMO1, SQLE, HMGCR, FDPS, NSDHL, LSS
510	PMID:19631568	(2009) Developmental expression pattern of the cholesterogenic enzyme NSDHL and negative selection of NSDHL-deficient cells in the heterozygous Bpa(1H)+ mouse.	HMGCR, DHCR7, NSDHL
765	PMID:30582412	(2019) RNA-Seq analysis reveals a negative role of MSMO1 with a synergized NSDHL expression during adipogenesis of 3T3-L1.	MSMO1, NSDHL

The pathways for this network model are (1) Aminoacyl-tRNA biosynthesisfor RARS, MARS, DARS, QARS, KARS, EPRS, IARS, GARS, LARS (2) Steroid biosynthesis with CYP51A1, MSMO1, SC5D, SQLE, DHCR7, NSDHL, LSS, (3) Metabolic pathways with CYP51A1, MSMO1, MARS, SC5D, SQLE, HMGCR, MVD, QARS, HMGCS1, DHCR7, FDPS, EPRS, NSDHL, IDI1, LSS and (4) Terpenoid backbone biosynthesis with HMGCR, MVD, HMGCS1, FDPS, and IDI1. Table 4 has the gene ontology ids and labels with the genes in the network model. [601]

	ID	Description	Network IDs
1	GO:0004812	aminoacyl-tRNA ligase activity	RARS, MARS, DARS, QARS, KARS, EPRS, IARS, GARS, LARS
2	GO:0003824	Catalytic activity	CYP51A1, RARS, MSMO1, MARS, SC5D, DARS, SQLE, HMGCR, MVD, QARS, HMGCS1, KARS, DHCR7, FDPS, EPRS, NSDHL, IARS, IDI1, GARS, LARS, LSS
3	GO:0036094	Small molecule binding	RARS, MARS, DARS,SQLE, HMGCR, MVD, QARS, HMGCS1, KARS, DHCR7, EPRS, IARS, GARS, LARS
4	GO:0000166	Nucleotide binding	RARS, MARS, DARS, SQLE, HMGCR, MVD, QARS, KARS, DHCR7, EPRS, IARS, GARS, LARS
5	GO:0000049	tRNA binding	RARS, MARS, KARS, IARS
6	GO:0005524	ATP binding	RARS, MARS, DARS, MVD, QARS, KARS, EPRS, IARS, GARS, LARS
7	GO:0043167	Ion binding	CYP51A1, RARS, MSMO1, MARS, SC5D, DARS, SQLE, HMGCR, MVD, QARS, KARS, FDPS, EPRS, IARS, IDI1, GARS, LARS
8	GO:0004819	glutamine-tRNA ligase activity	QARS, LARS
9	GO:0016491	Oxidoreductase activity	CYP51A1, MSMO1, SC5D, SQLE, HMGCR, DHCR7, NSDHL
10	GO:0043168	Anion binding	RARS, MARS, DARS, SQLE, HMGCR, MVD, QARS, KARS, EPRS, IARS, GARS, LARS
11	GO:0016709	Oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, nad(p)h as one donor, and incorporation of one atom of oxygen	CYP51A1, MSMO1, SQLE
12	GO:0016705	Oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	CYP51A1, MSMO1, SC5D, SQLE
13	GO:1901363	Heterocyclic compound binding	CYP51A1, RARS, MARS, DARS, SQLE, HMGCR, MVD, QARS, KARS, DHCR7, FDPS, EPRS, IARS, GARS, LARS
14	GO:0002161	aminoacyl-tRNA editing activity	IARS, LARS
15	GO:0097159	Organic cyclic compound binding	CYP51A1, RARS, MARS, DARS, SQLE, HMGCR, MVD, QARS, KARS, DHCR7, FDPS, EPRS, IARS, GARS, LARS

4 Results

Molecular properties are abundant in dimensional reduction for collections of sequences. Examples such as stability, binding potential aliphatic and hydrophobicity along with the charge at different pH is a few of available molecular properties to examine based on the gene ontology ids in Table 2. The net charge of a protein sequence based on the Henderson-Hasselbalch equation based on pH 5, 7 and 9. The aliphatic index is the relative volume occupied by aliphatic side chains (Alanine, Valine, Isoleucine, and Leucine) and is a positive factor for the increase of thermostability of globular proteins. The potential protein interaction index proposed by Boman (2003) based in the amino acid sequence of a protein and provides an overall estimate of the potential of a peptide to bind to membranes or other proteins as receptors. A protein have high binding potential if the index value is higher than 2.48. This index predicts the stability of a protein based on its amino acid composition, a protein whose instability index is smaller than 40 is predicted as stable, a value above 40 predicts that the protein may be unstable. Hydrophobicity is an important stabilization force in protein folding; this force changes depending on the solvent in which the protein is found. [1001]

Table 5 provides values based on each one of these properties. In the table, all of the molecules are stable with only EPRS near the boundary while for the binding potential is low for all the molecules. The distinction between hydrophobicity has SC5D, SQLE and HMGCR as positive with the rest negative. Charge differential from positive to negative occurs in all molecules from 5-9 with some variation at pH 7.

	Name	Stability Index	Binding Potential	ALiphatic	f.1	CpH5	CpH7	CpH9
1	CYP51A1	8.67	0.2506	18.66	-0.02680	22.565	8.677	-2.05
2	DARS	8.90	0.3962	16.46	-0.08268	13.325	-3.817	-14.56
3	DHCR7	5.58	0.0927	17.91	0.04315	29.879	14.662	-1.96
4	EPRS	38.06	1.7952	78.87	-0.51693	48.504	4.731	-29.75
5	FDPS	8.76	0.2385	15.02	-0.03820	5.898	-3.825	-14.65
6	GARS	15.80	0.5517	27.57	-0.09571	21.038	0.261	-15.91
7	HMGCR	21.15	0.4672	40.17	0.03455	17.557	-2.441	-28.99
8	HMGCS1	6.07	0.3197	17.06	-0.05787	0.582	-12.888	-25.58
9	IARS	28.14	1.1018	65.57	-0.17077	19.168	-13.323	-41.26
10	IDI1	3.87	0.1794	9.65	-0.03474	15.222	4.134	-6.26
11	KARS	11.63	0.4887	21.30	-0.12313	22.117	-2.152	-16.10
12	LARS	27.42	1.0302	52.01	-0.25644	37.667	3.458	-23.43
13	LSS	15.48	0.5027	25.17	-0.10179	21.243	-5.603	-30.38
14	MARS	20.04	0.6086	38.62	-0.10202	12.668	-8.981	-31.34
15	MSMO1	4.00	0.0991	9.14	-0.01161	16.294	1.617	-9.72
16	MVD	6.66	0.2149	13.32	-0.01490	12.622	1.137	-9.35
17	NSDHL	6.37	0.1879	12.65	-0.02109	15.217	4.249	-3.90
18	QARS	12.12	0.5801	29.32	-0.10574	30.850	1.811	-19.36
19	RARS	12.15	0.4764	26.27	-0.07821	15.568	-2.498	-18.80
20	SC5D	3.69	0.0896	10.08	0.00371	21.715	5.486	-2.88
21	SQLE	10.16	0.2533	22.76	0.00653	26.994	10.404	-1.03

In the keyword analysis, Table 6 shows the relationships between HMGCS1 and sterol biosynthesis with first degree relationships of CYP51A1 MSMO1 DARS SQLE HMGCR MVD KARS IARS DHCR7 LSS NSDHL IDI1 and FDPS. [601]

	ID	Description	Network IDs
1	KW-0752	Steroid biosynthesis	CYP51A1, MSMO1, SC5D, HMGCR, MVD, HMGCS1, DHCR7, FDPS, NSDHL, IDI1, LSS
2	KW-0756	Sterol biosynthesis	CYP51A1, MSMO1, SC5D, HMGCR, MVD, HMGCS1, DHCR7, FDPS, NSDHL, IDI1
3	KW-0030	Aminoacyl-tRNA synthetase	RARS, MARS, DARS, QARS, KARS, EPRS, IARS, GARS, LARS
4	KW-0152	Cholesterol biosynthesis	CYP51A1, HMGCR, MVD, HMGCS1, DHCR7, FDPS, NSDHL, IDI1
5	KW-0648	Protein biosynthesis	RARS, MARS, DARS, QARS, KARS, EPRS, IARS, GARS, LARS
6	KW-0225	Disease mutation	RARS, MSMO1, MARS, SC5D, DARS, MVD, QARS, KARS, DHCR7, FDPS, EPRS, NSDHL, IARS, GARS, LARS, LSS
7	KW-0067	ATP-binding	RARS, MARS, DARS, MVD, QARS, KARS, EPRS, IARS, GARS, LARS
8	KW-0007	Acetylation	RARS ,DARS, MVD, QARS, HMGCS1, KARS, FDPS, EPRS, NSDHL, IARS, IDI1, GARS, LARS, LSS
9	KW-0560	Oxidoreductase	CYP51A1, MSMO1, SC5D, SQLE, HMGCR, DHCR7, NSDHL
10	KW-0256	Endoplasmic reticulum	CYP51A1, MSMO1, SC5D, SQLE, HMGCR, DHCR7, NSDHL, LSS
11	KW-0144	Charcot-Marie-Tooth disease	MARS, KARS, GARS
12	KW-0414	Isoprene biosynthesis	FDPS, IDI1
13	KW-0523	Neurodegeneration	MARS, KARS, EPRS, GARS
14	KW-1026	Leukodystrophy	RARS, EPRS

The Kidera Factors were originally derived by applying multivariate analysis to 188 physical properties of the 20 amino acids and using dimension reduction techniques. A 10-dimensional vector of orthogonal factors was then obtained for each amino acid.The first four factors are essentially pure physical properties; the remaining six factors are superpositions of several physical properties, and are labelled for convenience by the name of the most heavily weighted component presented in Table 7. [1001]

	Names	HBF	SCS	ESP	H	DBP	PSV	FEP	OAR	PKC	SH
1	CYP51A1	18	-34	-3	5	-16	-76	21	-6	4	9
2	DARS	-33	-27	2	33	-16	-49	4	-4	10	-3
3	DHCR7	-2	-21	17	-33	-19	-61	9	11	5	11
4	EPRS	-115	-155	-47	178	-46	-319	42	-37	-43	-2
5	FDPS	-24	-15	-4	8	-29	-54	13	-6	1	2
6	GARS	-46	-56	-12	38	-32	-106	18	-8	11	14
7	HMGCR	-63	-118	34	14	-47	-105	14	-17	-13	-7
8	HMGCS1	-4	-55	-3	17	-23	-64	-7	8	-5	6
9	IARS	-67	-79	31	60	-73	-239	27	-30	7	33
10	IDI1	-20	-12	-4	11	-18	-25	10	1	10	-4
11	KARS	-34	-29	-10	40	-18	-71	18	3	8	12
12	LARS	-72	-66	-38	75	-15	-206	45	-5	-18	29
13	LSS	-17	-50	-15	5	-60	-69	27	-2	-9	27
14	MARS	-43	-77	-12	24	-58	-141	49	-49	6	-3
15	MSMO1	-7	8	-1	-12	-4	-31	3	7	17	0
16	MVD	-13	-49	-1	6	-27	-46	-3	-19	-1	9
17	NSDHL	-9	-26	10	7	-4	-43	6	-0	8	-2
18	QARS	-49	-53	-11	32	-43	-107	24	-18	14	40
19	RARS	-51	-33	-1	38	-30	-91	18	12	14	4
20	SC5D	1	2	8	-14	2	-39	-8	8	18	6
21	SQLE	-23	-48	5	1	-16	-98	3	-0	1	17

Table 1: Values multipled by 1000 with 0 decimal places. HBF=Helix/bend preference, SCS=Side-chain size, ESP=Extended structure preference, H=Hydrophobicity, DBP=Double-bend preference,PSV=Partial specific volume, FEP=Flat extended preference, OAR=Occurrence in alpha region, PKC=pK-C, and SH=Surrounding hydrophobicity. [1001]

A cluster analysis was performed based on the Kidera Factors from Table 7 and the cluster dendrogram presented in Figure 3.

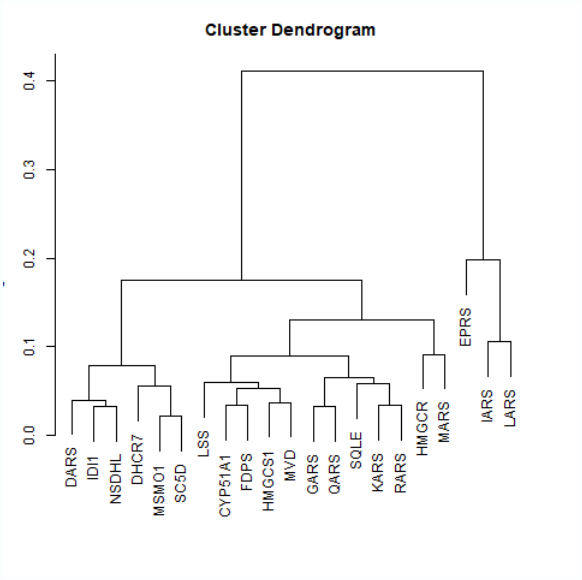

Figure 3: Cluster Analysis of Kidera Factors [1001]

Table 1 has VHSE-scales (principal components score Vectors of Hydrophobic, Steric, and Electronic properties), is derived from principal components analysis (PCA) on independent families of 18 hydrophobic properties, 17 steric properties, and 15 electronic properties, respectively, which are included in total 50 physicochemical variables of 20 coded amino acids. [1001]

	Names	H1	H2	S1	S2	E1	E2	E3	E4
1	CYP51A1	6	-15	-27	-18	-4	-52	33	-1
2	DARS	-21	-11	-19	-23	-18	-30	36	9
3	DHCR7	39	-9	-16	-20	13	-50	5	-17
4	EPRS	-125	-63	-132	-82	-73	-188	225	32
5	FDPS	0	2	-9	-14	-8	-36	22	-6
6	GARS	-22	-26	-44	-33	-25	-64	63	11
7	HMGCR	4	-90	-91	-77	-27	-117	52	-24
8	HMGCS1	-14	-28	-51	-18	-29	-39	33	-11
9	IARS	-15	-19	-50	-98	-55	-149	92	10
10	IDI1	-6	-5	-8	-8	-4	-18	26	-4
11	KARS	-23	-7	-23	-17	-22	-43	59	11
12	LARS	-37	-17	-53	-38	-40	-113	123	7
13	LSS	-3	-19	-46	1	-21	-58	49	-8
14	MARS	-9	-34	-61	-49	-30	-110	61	13
15	MSMO1	18	10	10	-10	3	-16	1	-3
16	MVD	-5	-30	-41	-26	-10	-41	23	13
17	NSDHL	1	-15	-20	-20	-4	-35	21	5
18	QARS	-17	-23	-40	-35	-20	-67	65	19
19	RARS	-17	-8	-21	-30	-23	-47	58	-15
20	SC5D	20	5	3	-14	4	-22	2	5
21	SQLE	11	-28	-39	-30	-5	-65	39	-5

Table 2: Computed average of VHSE-scales of all the amino acids in the corresponding peptide sequence. Each VSHE-scale represent an amino-acid property as follows: VHSE1 and VHSE2: Hydrophobic properties H1 and H2, VHSE3 and VHSE4: Steric properties S1 and S2, and VHSE5 to VHSE8: Electronic properties E1, E2, and E3 and E4. [1001]

For comparison, a cluster analysis was also performed based on the VHSE-scales from Table 8 and the cluster dendrogram presented in Figure 4.

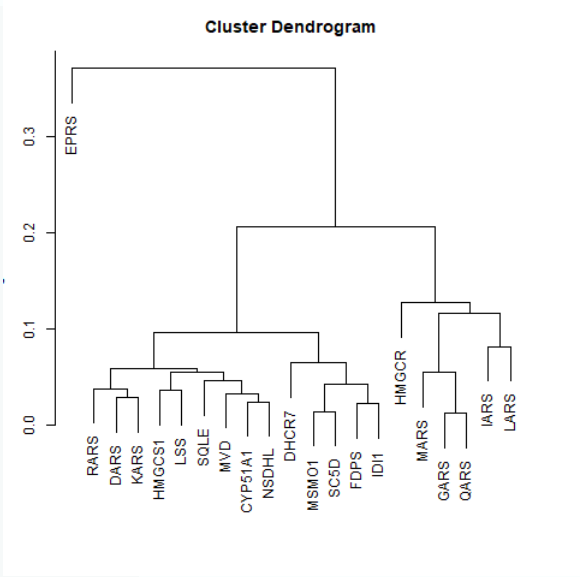

Figure 4: Cluster Analysis of VHSE-scales [1001]

In both examples, EPRS the Bifunctional glutamate/proline–tRNA ligase; is unique as a component of the GAIT (gamma interferon-activated inhibitor of translation) complex which mediates interferon-gamma-induced transcript- selective translation inhibition in inflammation processes.

5 Conclusions

In this brief mathematical note, the human genomic relationships based on interactions were modeled and the sequences examined with selected molecular properties. All of the molecules are stable with only EPRS near the boundary while for the binding potential is low for all the molecules. The distinction between hydrophobicity has SC5D, SQLE and HMGCR as positive with the rest negative. Charge differential from positive to negative occurs in all molecules from 5-9 with some variation at pH 7. Kidera factors were computed and cluster analysis peformed for the experimental coexpression network. First degree experimental coexpression relationships for 2.3.3.10 (HMGCS) of CYP51A1 MSMO1 DARS SQLE HMGCR MVD KARS IARS DHCR7 LSS NSDHL IDI1 and FDPS were also examined for Sterol biosynthesis and Disease mutation.

6 Bibliography

[1] Okamura, Eiji et al. “Unprecedented acetoacetyl-coenzyme A synthesizing enzyme of the thiolase superfamily involved in the mevalonate pathway.” Proceedings of the National Academy of Sciences of the United States of America vol. 107,25 (2010): 11265-70. doi:10.1073/pnas.1000532107.

[2] S. Rackovsky Global characteristics of protein sequences and their implications. PNAS (2010) 107 (19) 8623-8626; https://doi.org/10.1073/pnas.1001299107

[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. “R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.” (2015). 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).