

universität  
**ulm**

Laboratory Module for Master Programs  
**Biophysics Lab Course**

Experiment:  
**Bioinformatics**

Performed on  
Group 8

**Haiyang Zhang** and **Nicolae Turcan**

[haiyang.zhang@uni-ulm.de](mailto:haiyang.zhang@uni-ulm.de)

[nicolae.turcan@uni-ulm.de](mailto:nicolae.turcan@uni-ulm.de)

Supervisor: Camilla Förster

Haiyang Zhang

Turcan Nicolae

We hereby confirm that we have elaborated the present work independently and have detailed knowledge of the entire contents.

# Contents

<b>1 Project 1</b>	<b>4</b>
1.1 results . . . . .	4
1.1.1 1st BLAST search . . . . .	4
1.1.2 2nd BLAST search . . . . .	5
1.1.3 3rd BLAST search . . . . .	7
1.1.4 4th BLAST search . . . . .	8
1.1.5 5th BLAST search . . . . .	9
1.2 discussions . . . . .	10
<b>2 Project 2</b>	<b>13</b>
2.1 results . . . . .	13
2.1.1 enzyme report from <i>Uniprot</i> . . . . .	13
2.1.2 Protein analysis using Rhea and String . . . . .	14
2.1.3 enzyme report from <i>ENZYME</i> . . . . .	16
2.1.4 KEGG pathways of 2-Methyl Citrate Synthase . . . . .	17
2.1.5 Gene defect analysis using OMIM database . . . . .	20
2.2 discussions . . . . .	21
<b>3 Project 3</b>	<b>24</b>
3.1 results . . . . .	24
3.1.1 Space filling vs Cartoons . . . . .	24
3.1.2 Enzymatic Pocket . . . . .	25
3.1.3 Protein Description . . . . .	25
3.2 discussions . . . . .	29
<b>4 Project 4</b>	<b>30</b>
4.1 results . . . . .	30
4.1.1 The searching result from BRENDA database . . . . .	30
4.2 discussions . . . . .	32
<b>5 Project 5</b>	<b>34</b>
5.1 results . . . . .	34
5.1.1 homologs of 2-Methyl citrate synthases . . . . .	34
5.1.2 Comparing the homologs using ClustalOmega(figure 5.1 and 5.2) . . . . .	34
5.1.3 The 3D image showing the highly conserved regions(figure 5.3) . . . . .	34
5.2 discussions . . . . .	34

**Bibliography**

**38**

# 1 Project 1

In the following section we will display the results of the first project of the bioinformatics laboratory. The objective of the first project was to identify an unknown protein from short peptide sequences given in Table 1.1 . The project consisted in doing 5 BLAST searches , which allows you to search a database for similar sequences. It works by identifying regions of similarity between our query and sequences in the database, assigning scores based on similarity and statistical significance. The results provide a list of sequences from the database that closely match our query [[Alt+90]].

## 1.1 results

<b>Peptide Sequence 1</b>	TVGWIAHWSEMHSNDGMK
<b>Peptide Sequence 2</b>	AMGIPSSMFTVIFAMAR
<b>Reverse Peptide Sequence 2</b>	RAMAFIVTFMSSPIGMA

Table 1.1: Peptide sequences

### 1.1.1 1st BLAST search

The first BLAST search was performed on Peptide Sequence 1 displayed in Table 1.1. The results of this BLAST search are displayed in three parts:

A Graphic summary( Figure 1.1) displaying a distribution of the best entries in comparison with the query sequence, the aminoacid position is signed as well to enable better comparison, the color green signifies a similarity of over 90 percent.

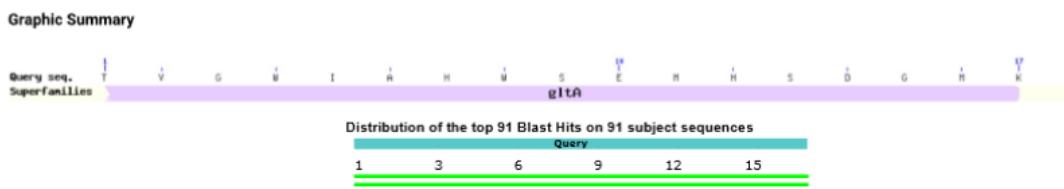


Figure 1.1: Graphic summary of Top 3 results in 1st BLAST search

The Description table in Figure 1.2 submits the sequence of origin and the organism name from which it the sequence was isolated, following that we have multiple parameters that can help us better choose which sequence of origin is more suitable and in particular an E value. The E-value, or Expectation value, refers to the number of random matches one can expect to

## 1.1 results

---

find by chance when searching a database with a particular query, a lower E-value indicates a more significant match.

Descriptions									
Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession	Feedback
hypothetical protein [Escherichia coli]	Escherichia coli	63.0	63.0	100%	1e-10	100.00%	39	MDM862014	
hypothetical protein [Escherichia coli]	Escherichia coli	63.0	63.0	100%	1e-10	100.00%	39	MDF0764298.1	
Citrate synthase [Escherichia coli DORA_A_5_14_21]	Escherichia coli DORA_A_5_14_21	63.0	63.0	100%	2e-10	100.00%	42	ETJ26103.1	

Figure 1.2: Top 3 results of 1st BLAST search

And finally , in Figure 1.3 we can see a detailed set of pairwise alignments bewteen the query and the best candidates, in this case the 3rd candidate provided the most significant result with information on the function of the protein in which it was found.

MAG: Citrate synthase, partial [Escherichia coli DORA\_A\_5\_14\_21]

Sequence ID: ETJ26103.1 Length: 42 Number of Matches: 1

Range 1: 4 to 20

Score	Expect	Identities	Positives	Gaps	Frame
63.0 bits(141)	2e-10()	17/17(100%)	17/17(100%)	0/17(0%)	
Query 1	TVGWIAMNSEMHSMDGNK	17			
Sbjct 4	TVGWIAMNSEMHSMDGNK	20			

Figure 1.3: 3rd Alignment of 1st BLAST search

The Following BLAST search results will be displayed in the same order and without further comment if not necessary.

### 1.1.2 2nd BLAST search

#### Graphic Summary

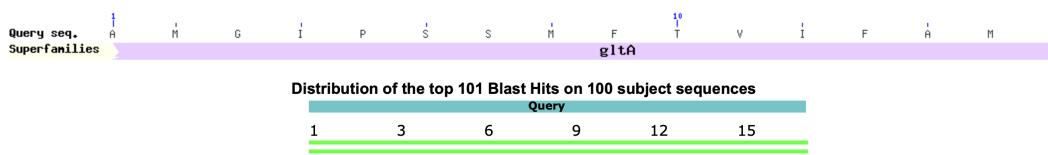


Figure 1.4: Graphic summary of Top 3 results in 2nd BLAST search

## 1.1 results

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
citrate synthase (EC 4.1.3.7) [Escherichia coli]	Escherichia coli	59.6	59.6	100%	1e-09	100.00%	29	AAA23902.1
citrate/2-methylcitrate synthase [Escherichia coli]	Escherichia coli	59.6	59.6	100%	6e-09	100.00%	50	WP_284344613.1
citrate (Si)-synthase [Salmonella enterica]	Salmonella enterica	59.6	59.6	100%	7e-09	100.00%	56	EAM6650575.1

Figure 1.5: Top 3 results of 2nd BLAST search

**Alignments**

Alignment view **Pairwise**  CDS feature

citrate synthase (EC 4.1.3.7), partial [Escherichia coli]  
Sequence ID: **AAA23902.1** Length: 29 Number of Matches: 1  
Range 1: 13 to 29

Score	Expect	Identities	Positives	Gaps	Frame
59.6 bits(133)	1e-09()	17/17(100%)	17/17(100%)	0/17(0%)	
Query 1	AMGIPSSMFTVIFAMAR	17			
Sbjct 13	AMGIPSSMFTVIFAMAR	29			

citrate/2-methylcitrate synthase, partial [Escherichia coli]  
Sequence ID: **WP\_284344613.1** Length: 50 Number of Matches: 1

[https://blast.ncbi.nlm.nih.gov/Blast.cgi#sort\\_mark](https://blast.ncbi.nlm.nih.gov/Blast.cgi#sort_mark)

22/01/2024, 13:45 NCBI Blast:Protein Sequence

Range 1: 7 to 23

Score	Expect	Identities	Positives	Gaps	Frame
59.6 bits(133)	6e-09()	17/17(100%)	17/17(100%)	0/17(0%)	
Query 1	AMGIPSSMFTVIFAMAR	17			
Sbjct 7	AMGIPSSMFTVIFAMAR	23			

citrate (Si)-synthase [Salmonella enterica]  
Sequence ID: **EAM6650575.1** Length: 56 Number of Matches: 1  
Range 1: 1 to 17

Score	Expect	Identities	Positives	Gaps	Frame
59.6 bits(133)	7e-09()	17/17(100%)	17/17(100%)	0/17(0%)	
Query 1	AMGIPSSMFTVIFAMAR	17			
Sbjct 1	AMGIPSSMFTVIFAMAR	17			

Figure 1.6: Top 3 Alignments of 2nd BLAST search

## 1.1 results

---

### 1.1.3 3rd BLAST search

hypothetical protein, partial [Escherichia coli]

Sequence ID: **WP\_244581302.1** Length: 42 Number of Matches: 1  
Range 1: 4 to 29

Score	Expect	Method	Identities	Positives	Gaps	Frame
43.5 bits(101)	4e-04()	Compositional matrix adjust.	19/27(70%)	22/27(81%)	1/27(3%)	
Query 1	TVGWIAHWSEMHSNDGMKAMGIPSSMFT	27				
Sbjct 4	TVGWIAHWSEMHSNDGMK + P ++T IARPRQLYT	29				

citrate (Si)-synthase [Escherichia coli]

Sequence ID: **NAI03858.1** Length: 43 Number of Matches: 1  
Range 1: 5 to 30

Score	Expect	Method	Identities	Positives	Gaps	Frame
43.5 bits(101)	4e-04()	Compositional matrix adjust.	19/27(70%)	22/27(81%)	1/27(3%)	
Query 1	TVGWIAHWSEMHSNDGMKAMGIPSSMFT	27				
Sbjct 5	TVGWIAHWSEMHSNDGMK + P ++T IARPRQLYT	30				

citrate/2-methylcitrate synthase, partial [Escherichia coli]

Sequence ID: **WP\_274536608.1** Length: 44 Number of Matches: 1  
Range 1: 6 to 31

Score	Expect	Method	Identities	Positives	Gaps	Frame
43.5 bits(101)	4e-04()	Compositional matrix adjust.	19/27(70%)	22/27(81%)	1/27(3%)	
Query 1	TVGWIAHWSEMHSNDGMKAMGIPSSMFT	27				
Sbjct 6	TVGWIAHWSEMHSNDGMK + P ++T IARPRQLYT	31				

Figure 1.7: Top 3 alignments of 3rd BLAST search

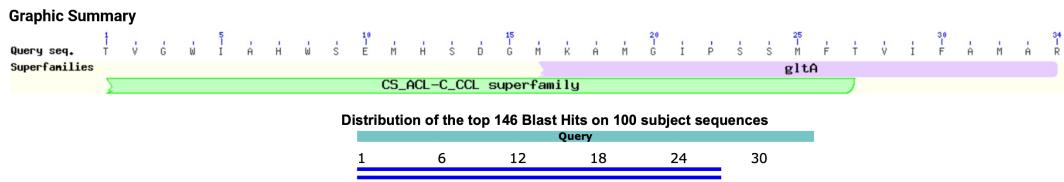


Figure 1.8: Graphic summary of Top 3 results in 3rd BLAST search

## 1.1 results

Descriptions								
Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
hypothetical protein [Escherichia coli]	Escherichia coli	43.5	43.5	79%	4e-04	70.37%	42	WP_244581302.1
citrate (Si)-synthase [Escherichia coli]	Escherichia coli	43.5	43.5	79%	4e-04	70.37%	43	NAI03858.1
citrate/2-methylcitrate synthase [Escherichia coli]	Escherichia coli	43.5	43.5	79%	4e-04	70.37%	44	WP_274536608.1

Figure 1.9: Top 3 results of 3rd BLAST search

### 1.1.4 4th BLAST search

citrate/2-methylcitrate synthase [Klebsiella michiganensis]  
Sequence ID: **MDU4389148.1** Length: 142 Number of Matches: 1  
Range 1: 87 to 120

Score	Expect	Method	Identities	Positives	Gaps	Frame
76.6 bits(187)	3e-16()	Composition-based stats.	33/34(97%)	34/34(100%)	0/34(0%)	
Query 1	AMGIPSSMFTVIFAMARTVGWIAHWSEMHSNDGMK	34				
Sbjct 87	AMGIPSSMFTVIFAMARTVGWIAHW+EMHSNDGMK	120				

citrate (Si)-synthase [Klebsiella oxytoca]  
Sequence ID: **HEC2128142.1** Length: 121 Number of Matches: 1

<https://blast.ncbi.nlm.nih.gov/Blast.cgi>

22/01/2024, 13:44

NCBI Blast:Protein Sequence

Range 1: 66 to 99

Score	Expect	Method	Identities	Positives	Gaps	Frame
75.9 bits(185)	4e-16()	Composition-based stats.	33/34(97%)	34/34(100%)	0/34(0%)	
Query 1	AMGIPSSMFTVIFAMARTVGWIAHWSEMHSNDGMK	34				
Sbjct 66	AMGIPSSMFTVIFAMARTVGWIAHW+EMHSNDGMK	99				

citrate/2-methylcitrate synthase, partial [Klebsiella aerogenes]  
Sequence ID: **WP\_323129656.1** Length: 159 Number of Matches: 1  
Range 1: 104 to 137

Score	Expect	Method	Identities	Positives	Gaps	Frame
76.6 bits(187)	5e-16()	Composition-based stats.	33/34(97%)	34/34(100%)	0/34(0%)	
Query 1	AMGIPSSMFTVIFAMARTVGWIAHWSEMHSNDGMK	34				
Sbjct 104	AMGIPSSMFTVIFAMARTVGWIAHW+EMHSNDGMK	137				

Figure 1.10: Top 3 alignments of 4th BLAST search

## 1.1 results

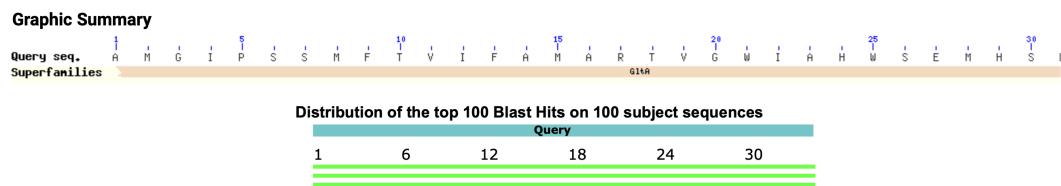


Figure 1.11: Graphic summary of Top 3 results in 4th BLAST search

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
citrate/2-methylcitrate synthase [Klebsiella michiganensis]	Klebsiella michiganensis	76.6	76.6	100%	3e-16	97.06%	142	<a href="#">MDU4389148.1</a>
TPA: citrate (S)-synthase [Klebsiella oxytoca]	Klebsiella oxytoca	75.9	75.9	100%	4e-16	97.06%	121	<a href="#">HEC2128142.1</a>
citrate/2-methylcitrate synthase [Klebsiella aerogenes]	Klebsiella aerogenes	76.6	76.6	100%	5e-16	97.06%	159	<a href="#">WP_323129656.1</a>

Figure 1.12: Top 3 results of 4th BLAST search

### 1.1.5 5th BLAST search

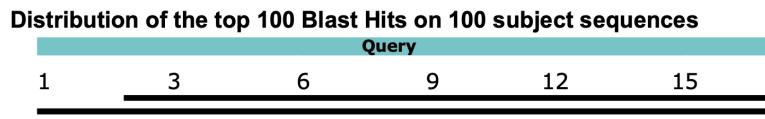


Figure 1.13: Graphic summary of Top 3 results in 5th BLAST search

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
threonine-serine exporter family protein [Shewanella sp. NIFS-20-20]	Shewanella sp. NIFS-20-20	36.7	36.7	88%	1.4	66.67%	253	<a href="#">MBV7315808.1</a>
peptidylprolyl isomerase [Bacteroidales bacterium]	Bacteroidales bacterium	34.6	34.6	100%	8.1	58.82%	455	<a href="#">MDO4496460.1</a>
hypothetical protein ADK66_22520 [Micromonospora sp. NRRL_B-16802]	Micromonospora sp. NRRL_B-16802	33.7	33.7	76%	16	76.92%	167	<a href="#">KOX06643.1</a>

Figure 1.14: Top 3 results of 5th BLAST search

## 1.2 discussions

---

threonine-serine exporter family protein [Shewanella sp. NIFS-20-20]

Sequence ID: **MBV7315808.1** Length: 253 Number of Matches: 1

Range 1: 139 to 153

Score	Expect	Identities	Positives	Gaps	Frame
36.7 bits(79)	1.4()	10/15(67%)	12/15(80%)	0/15(0%)	
Query 3	MAFIVTFMSSPIGMA	17			
Sbjct 139	MAF+ TF+ S IGMA				
	MAFVMTFLASSIGMA	153			

peptidylprolyl isomerase [Bacteroidales bacterium]

Sequence ID: **MDO4496460.1** Length: 455 Number of Matches: 1

Range 1: 7 to 23

Score	Expect	Identities	Positives	Gaps	Frame
-------	--------	------------	-----------	------	-------

last:Protein Sequence

34.6 bits(74) 8.1() 10/17(59%) 12/17(70%) 0/17(0%)

Query 1 RAMAFIVTFMSSPIGMA 17  
RAMAF+ T M +GMA  
Sbjct 7 RAMAFVATLMTAVVGMA 23

hypothetical protein ADK66\_22520 [Micromonospora sp. NRRL B-16802]

Sequence ID: **KOX06643.1** Length: 167 Number of Matches: 1

Range 1: 82 to 94

Score	Expect	Identities	Positives	Gaps	Frame
33.7 bits(72)	16()	10/13(77%)	11/13(84%)	0/13(0%)	
Query 5	FIVTFMSSPIGMA	17			
Sbjct 82	FIV F+SS IGMA				
	FIV AFLSSHIGMA	94			

Figure 1.15: Top 3 alignments of 5th BLAST search

## 1.2 discussions

a)Q: What is the name of your unknown protein

A: The name of our protein is : 2-Methyl Citrate Synthase.

b) Q: Why are there multiple hits for what appears to be the same enzyme?

A: There are multiple hits because the same enzyme exists in different organisms, and since most of the results were from bacteria, we can expect to find this protein not only in organisms which performed vertical gene transfer and differentiation in different species, but also in evolutionarily unrelated species because of horizontal gene transfer( the transfer of plasmids or dna integration)

c) Q: Can you positively identify the organism from which this enzyme derived? If so, what organism is it from? Explain how you used the BLAST output to make this determination. If you were not able to unambiguously make the determination of the organism, why not? Explain using your BLAST output.

A: No, we couldn't identify a single organism. There are many results which have the same sequence, but from different origin organism, this could be attributed to the fact that this protein is part of the energetic metabolism, and proteins related to energetic pathways are extremely conserved across organism, and are also extremely ancient in evolutionary origin, in fact the same enzyme (with minimal variation )can be found in much more complex organisms such as H.Sapiens.

d) Q: What is its E.C. number?

A: the E.C. number of our enzyme is 2.3.3.5.

e) Q: When you entered your peptide in backwards, did you get a significant match to anything in the database? What does this teach you about the importance of directionality and orientation in biological macromolecules?

A: Expecting to find results from this search is unreasonable because it is not the same as performing a reverse search on DNA sequences( which are complimentary and antiparallel, meaning their 5' to 3' direction is opposite). Each aminoacid corresponds to a specific codon ( a combination of three consecutive basepairs ), so searching for the reverse of a peptide sequence is the same as searching for a DNA sequence in which every codon sequence is the same but the codon position is inverted, and as expected the probability of such a construct arising from pure chance, decreases factorially ( e.g  $1/N!$  ) with the amount of codons in the sequence. And even if we consider this probability to be reasonable over all the time and dna replication events, such a sequence wouldn't have any incentive to be conserved over the course of evolution because it most probably doesn't perform any useful function.

f) Q: Compare the results from the two searches that used both peptide fragments (Searches 3 and 4). Did you get identical results? This comparison calls into play gaps and gap penalties. Explain your results from Searches 3 and 4 in relation to gaps and gap penalties.

A: The 3rd search and 4th search don't display the same results. The reason is that the gap is considered to be an unknown gap ( series of aminoacids) in a contiguous sequence, so the 3rd search represented a sequence in which a translocation event at the DNA level occurred,

creating a dysfunctional protein that wasn't replicated further because it resulted as a hinder to the fitness of the host organism. So, even if the above mentioned argument on the probability of such a sequence occurring doesn't hold anymore, we can see how protein functionality , which is a direct result of the sequence can be fundamentally changed.

We must however consider that translocation events are between the most proficient in creating new and functional proteins, or that allow for more structurally complex proteins ( e.g. the MHC Complex ) to evolve from very few sequences that encode for specific smaller protein structures( e.g alpha-helices).

g) Q: Find the peptide fragments you used to identify your unknown protein in the full protein sequence. Highlight them on the print out you are submitting with your problem set. Are the sequences contiguous to one another and does it matter? Please explain.

A: In the figure 1.16 we can see the highlighted peptide fragments in the full sequence. As we can see they are continuous to each other, but the second sequence has much worse identity. It matters because having contiguous sequences and knowing they are so allows us to perform BLAST searches with much greater specificity?

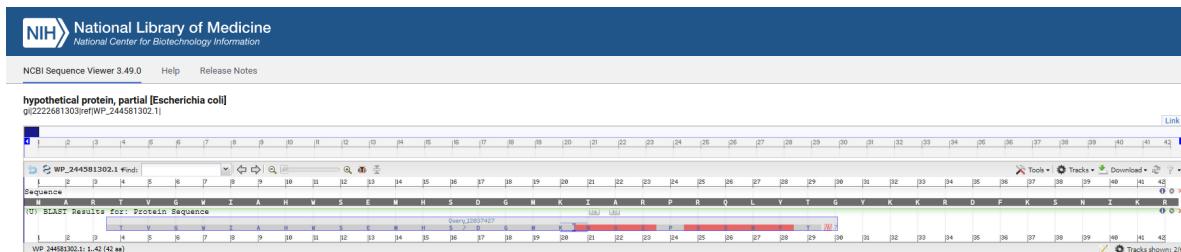


Figure 1.16: Sequences aligned

h) Q: Think of an example of how you might use a BLAST search for something other than simply identifying a protein for which you have partial sequence information.

A: BLAST is often used to reconstruct genomes or genes from shotgun sequences. However the same principle could be used to find a phone number in a phone book in order of likelihood if you start from a washed-away phone number on a napkin, or a teared/mistyped one. One could also use BLAST for Linguistic Studies, like finding which Languages use a certain word or variants of it.

# 2 Project 2

This Project consists of finding various information on our protein of interest (2-methyl Citrate Synthase) from databases such as KEGG [[Kan+23]] and ExPASy [[Gas+03]].

## 2.1 results

### 2.1.1 enzyme report from *Uniprot*

Uniprot is a database which provides general informations of proteins. Further introduction of uniprot is provided at 2.2.a. Some searching results of 2-Methyl citrate synthase(basic informations, names and taxonomy) obtained from uniprot is shown in figure 2.1 and 2.2.

P31660 · PRPC_ECOLI	
Protein <sup>i</sup>	Amino acids
2-methylcitrate synthase	389 ( <a href="#">go to sequence</a> )
Gene <sup>i</sup>	Protein existence <sup>i</sup>
prpC	Evidence at protein level
Status <sup>i</sup>	Annotation score <sup>i</sup>
UniProtKB reviewed (Swiss-Prot)	5/5
Organism <sup>i</sup>	
Escherichia coli (strain K12)	

Figure 2.1: Uniprot code and basic informations of 2-Methyl citrate synthase

### Names & Taxonomy<sup>i</sup>

#### Protein names<sup>i</sup>

Recommended name

2-methylcitrate synthase 1 Publication

EC number

EC:2.3.3.5 (UniProtKB | ENZYME | Rhea ) 3 Publications

Short names

2-MCS 1 Publication ; MCS 1 Publication

Alternative names

(2S,3S)-2-methylcitrate synthase 1 Publication

Citrate synthase 1 Publication (EC:2.3.3.16 2 Publications ) . EC:2.3.3.16 (UniProtKB | ENZYME | Rhea ) 2 Publications

#### Gene names<sup>i</sup>

Name

prpC 1 Publication

Synonyms

yahS,yzzD

Ordered locus names

b0333, JW0324

#### Organism names

Organism<sup>i</sup>

Escherichia coli (strain K12)

Taxonomic identifier<sup>i</sup>

83333 NCBI

Taxonomic lineage<sup>i</sup>

cellular organisms > Bacteria (eubacteria) > Pseudomonadota > Gammaproteobacteria > Enterobacterales > Enterobacteriaceae > Escherichia > Escherichia coli

#### Accessions

Primary accession

P31660

Secondary accessions

P77217

Q2MC91

#### Proteomes<sup>i</sup>

Identifier

UP000000318

Component<sup>i</sup>

Chromosome

Figure 2.2: Names and taxonomy of 2-Methyl citrate synthase

### 2.1.2 Protein analysis using Rhea and String

Except uniprot, there are some other databases, like Rhea[[Ban+22]] and STRING [[Szk+21]], shows some specific properties of protein and enzymes. We obtained the chemical reaction catalysed by our enzyme via Rhea, and the interaction network of our enzyme from String.

## 2.1 results

---

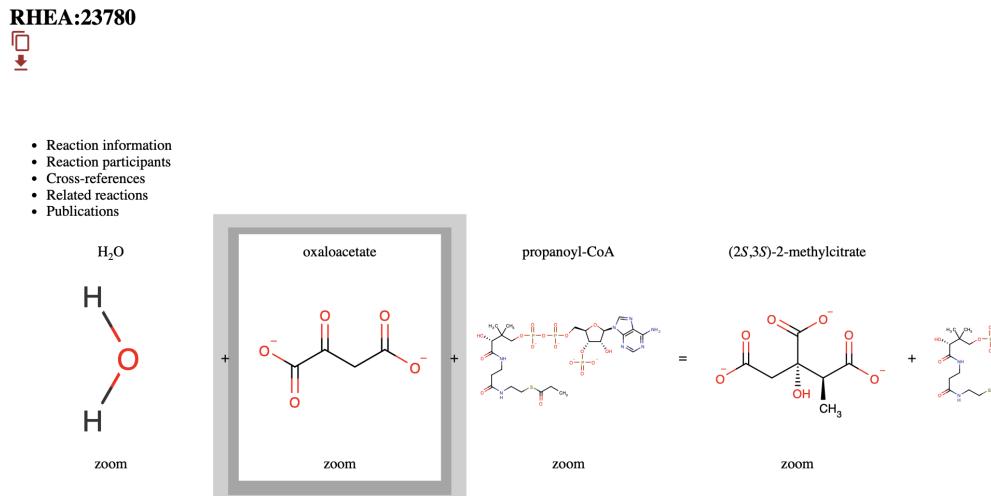


Figure 2.3: The chemical reaction catalysed by 2-Methyl citrate synthase, acquired from Rhea

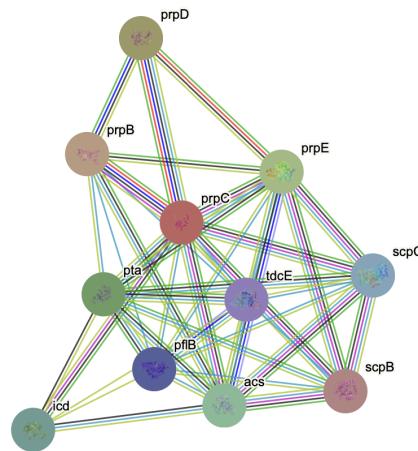


Figure 2.4: The interaction networks of 2-Methyl citrate synthase, acquired from String

## 2.1 results

**Nodes:**

Network nodes represent proteins <i>splice isoforms or post-translational modifications are collapsed, i.e. each node represents all the proteins produced by a single, protein-coding gene locus.</i>	Node Color  <i>colored nodes: query proteins and first shell of interactors</i>  <i>white nodes: second shell of interactors</i>	Node Content  <i>empty nodes: proteins of unknown 3D structure</i>  <i>filled nodes: a 3D structure is known or predicted</i>
---	--	---

**Edges:**

Edges represent protein-protein associations <i>associations are meant to be specific and meaningful, i.e. proteins jointly contribute to a shared function; this does not necessarily mean they are physically binding to each other.</i>	Known Interactions  <i>from curated databases experimentally determined</i>	Predicted Interactions  <i>gene neighborhood</i>  <i>gene fusions</i>  <i>gene co-occurrence</i>	Others  <i>textmining</i>  <i>co-expression</i>  <i>protein homology</i>
---	---	--	--

**Your Input:**

<i>2-methylcitrate synthase; Involved in the catabolism of short chain fatty acids (SCFA) via the tricarboxylic acid (TCA)(acetyl degradation route) and via the 2-methylcitrate cycle I (propionate degradation route). Catalyzes the Claisen condensation of propionyl-CoA and oxaloacetate (OAA) to yield 2-methylcitrate (2-MC) and CoA. Also catalyzes the condensation of oxaloacetate with acetyl-CoA to yield citrate but with a lower specificity. (389 aa)</i>	<i>Neighborhood</i>  <i>Gene Fusion</i>  <i>Cocurrence</i>  <i>Coexpression</i>  <i>Experiments</i>  <i>Databases</i>  <i>Textmining [Homology]</i>  <i>Score</i>
--	---

**Predicted Functional Partners:**

	<i>Neighborhood</i>	<i>Gene Fusion</i>	<i>Cocurrence</i>	<i>Coexpression</i>	<i>Experiments</i>	<i>Databases</i>	<i>Textmining [Homology]</i>	<i>Score</i>
prpB	2-methylisocitrate lyase; Involved in the catabolism of short chain fatty acids (SCFA) via the 2-methylcitrate cycle I (propionate de...							0.999
prpD	2-methylcitrate dehydratase; Involved in the catabolism of short chain fatty acids (SCFA) via the tricarboxylic acid (TCA)(acetyl de...							0.999
prpE	propionate-CoA ligase; Catalyzes the synthesis of propionyl-CoA from propionate and CoA. Also converts acetate to acetyl-CoA b...							0.996
pta	Phosphate acetyltransferase; Involved in acetate metabolism. Catalyzes the reversible interconversion of acetyl-CoA and acetyl p...							0.977
acs	acetyl-CoA synthetase; Catalyzes the conversion of acetate into acetyl-CoA (AcCoA), an essential intermediate at the junction of a...							0.969
icd	Isocitrate dehydrogenase, specific for NADP+; Protein involved in tricarboxylic acid cycle and anaerobic respiration. Belongs to th...							0.959
scpC	propionyl-CoA:succinate CoA transferase; Catalyzes the transfer of coenzyme A from propionyl-CoA to succinate. Could be part o...							0.956
pfb	Formate acetyltransferase 1; Protein involved in anaerobic respiration and cellular amino acid catabolic process.							0.950
tdcE	Pyruvate formate-lyase 4/2-ketobutyrate formate-lyase; Catalyzes the cleavage of 2-ketobutyrate to propionyl-CoA and formate. It ...							0.948
scpB	methylmalonyl-CoA decarboxylase, biotin-independent; Catalyzes the decarboxylation of (R)-methylmalonyl-CoA to propionyl-CoA...							0.938

**Your Current Organism:**

Escherichia coli K12  NCBI taxonomy Id: 511145 Other names: <i>E. coli</i> str. K-12 substr. MG1655, <i>Escherichia coli</i> MG1655, <i>Escherichia coli</i> str. K-12 substr. MG1655, <i>Escherichia coli</i> str. MG1655, <i>Escherichia coli</i> strain MG1655
--

Figure 2.5: Explanations on the interaction network, acquired from String

### 2.1.3 enzyme report from ENZYME

ENZYME database could be used to search for the enzyme type of the protein. In ENZYME database, we found the enzyme type of our protein. 2-Methyl cirtate synthase belongs to EC 2.3.3.5.

## 2.1 results

22/01/2024, 14:32

ENZYME - 2.3.3.5 2-methylcitrate synthase

**Expasy** 

ENZYME (<https://p-enzyme.expasy.org>)

Home (/index.html)  
Search (/enzyme-search-ec.html)  
Downloads (<https://ftp.expasy.org/databases/enzyme/>)

Browse (/cgi-bin/enzyme/enzyme-search-cl?1)  
Documentation (/enzuser.txt)  
Contact (/contact)

**ENZYME entry: EC 2.3.3.5**

Accepted Name

**2-methylcitrate synthase**

Alternative Name(s)

**2-methylcitrate oxaloacetate-lyase**

MCS

**methylcitrate synthase**

**methylcitrate synthetase**

Reaction catalysed

$\text{H}_2\text{O} + \text{oxaloacetate} + \text{propanoyl-CoA} \rightleftharpoons (2S,3S)\text{-2-methylcitrate} + \text{CoA} + \text{H}(+)$

Comment(s)

- The enzyme acts on acetyl-CoA, propanoyl-CoA, butanoyl-CoA and pentanoyl-CoA.
- The relative rate of condensation of acetyl-CoA and oxaloacetate is 140% of that of propanoyl-CoA and oxaloacetate, but the enzyme is distinct
- Oxaloacetate cannot be replaced by glyoxylate, pyruvate or 2-oxoglutarate.
- Formerly EC 4.1.3.31.

Figure 2.6: The EC number and catalysed reaction of our protein

### 2.1.4 KEGG pathways of 2-Methyl Citrate Synthase

In KEGG database [[Kan+23]], we obtained the (propanoate)metabolic pathways of EC 2.3.3.5. (figure 2.8 and 2.9). And then we looked at the metabolic pathways in homo sapines. As is shown in the figure 2.10 and 2.11, EC 2.3.3.5 does not involve in the reaction of homo sapines.

## 2.1 results

---

KEGG ENZYME: 2.3.3.5

<https://www.kegg.jp/entry/2.3.3.5>

**ENZYME: 2.3.3.5**

Help

<b>Entry</b>	EC 2.3.3.5      Enzyme
<b>Name</b>	2-methylcitrate synthase; 2-methylcitrate oxaloacetate-lyase; MCS; methylcitrate synthase; methylcitrate synthetase
<b>Class</b>	Transferases; Acyltransferases; Acyl groups converted into alkyl groups on transfer [BRITIE hierarchy]
<b>Sysname</b>	propanoyl-CoA:oxaloacetate C-propanoyltransferase (thioester-hydrolysing, 1-carboxyethyl-forming)
<b>Reaction(IUBMB)</b>	propanoyl-CoA + H <sub>2</sub> O + oxaloacetate = (2S,3S)-2-hydroxybutane-1,2,3-tricarboxylate + CoA [RN:R00931]
<b>Reaction(KEGG)</b>	R00931 Reaction
<b>Substrate</b>	propanoyl-CoA [CPD:C00100]; H <sub>2</sub> O [CPD:C00001]; oxaloacetate [CPD:C00036]
<b>Product</b>	(2S,3S)-2-hydroxybutane-1,2,3-tricarboxylate [CPD:C02225]; CoA [CPD:C00010]
<b>Comment</b>	The enzyme acts on acetyl-CoA, propanoyl-CoA, butanoyl-CoA and pentanoyl-CoA. The relative rate of condensation of acetyl-CoA and oxaloacetate is 140% of that of propanoyl-CoA and oxaloacetate, but the enzyme is distinct from EC 2.3.3.1, citrate (Si)-synthase. Oxaloacetate cannot be replaced by glyoxylate, pyruvate or 2-oxoglutarate.
<b>History</b>	EC 2.3.3.5 created 1978 as EC 4.1.3.31, transferred 2002 to EC 2.3.3.5, modified 2015
<b>Pathway</b>	ec00640 Propanoate metabolism ec01100 Metabolic pathways
<b>Orthology</b>	K01659 2-methylcitrate synthase
<b>Genes</b>	NCOL: 116245028 DDI: DDB_G0287281 TGO: TGME49_263130 TET: TTHERM_00537060 PTM: GSPATT00014981001 GSPATT00031721001 ECO: b0333(prpC) ECJ: JW0324(prpC) ECOK: ECMDS42_0255(prpC) ECOC: C3026_01630 C3026_24800 ECE: Z0428(prpC) » show all Taxonomy
<b>Reference</b>	1
<b>Authors</b>	Uchiyama, H. and Tabuchi, T.
<b>Title</b>	Properties of methylcitrate synthase from <i>Candida lipolytica</i> .
<b>Journal</b>	<i>Agric Biol Chem</i> 40:1411-1418 (1976)
<b>Reference</b>	2 [PMID:9325432]
<b>Authors</b>	Textor S, Wendisch VF, De Graaf AA, Muller U, Linder MI, Linder D, Buckel W.
<b>Title</b>	Propionate oxidation in <i>Escherichia coli</i> : evidence for operation of a methylcitrate cycle in bacteria.

1 of 2

22/01/2024, 14:48

Figure 2.7: The introduction on EC 2.3.3.5 generated from KEGG database

## 2.1 results

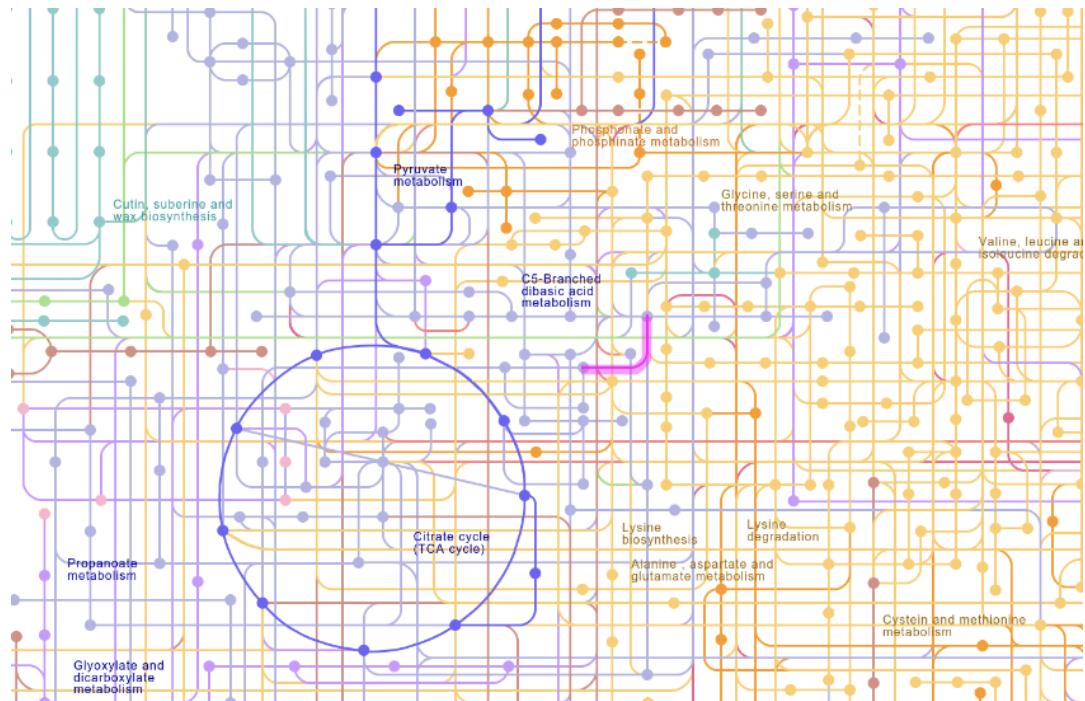


Figure 2.8: Part of the metabolic pathways, the pink sparkling pathway is related to EC 2.3.3.5

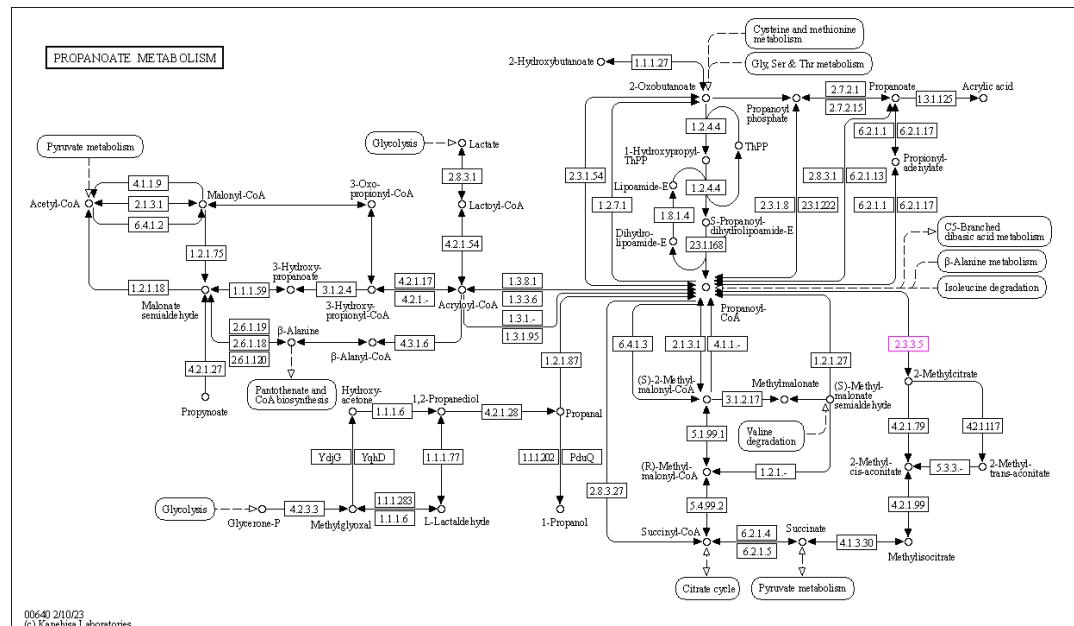


Figure 2.9: The propanoate metabolic pathways, the red part is related to EC 2.3.3.5

## 2.1 results

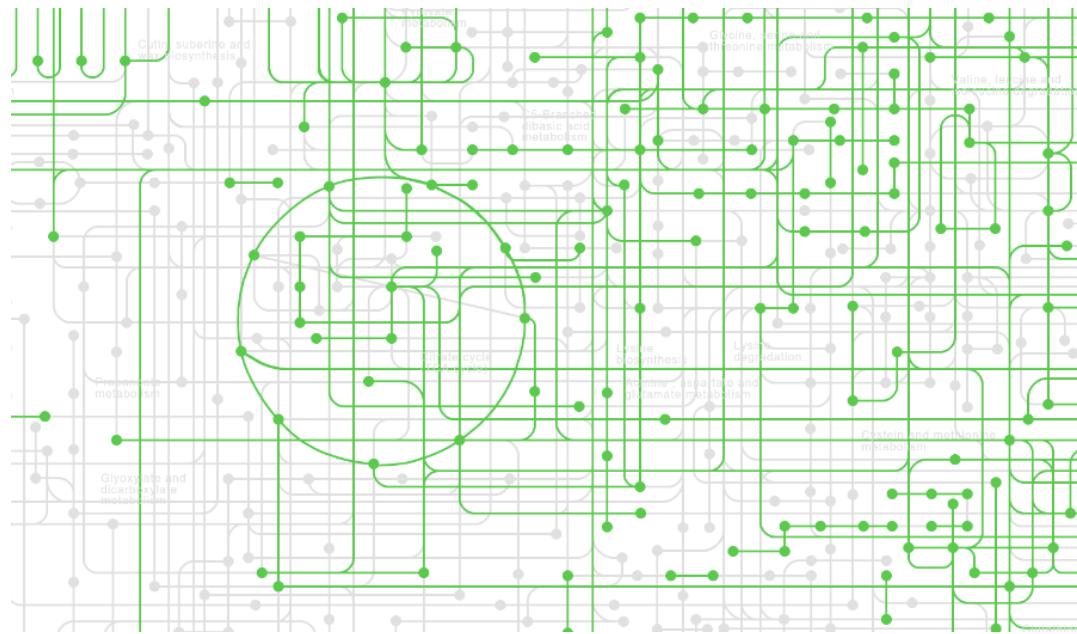


Figure 2.10: Part of the metabolic pathways in homo sapines. The green pathways are the pathways exist in homo sapines.

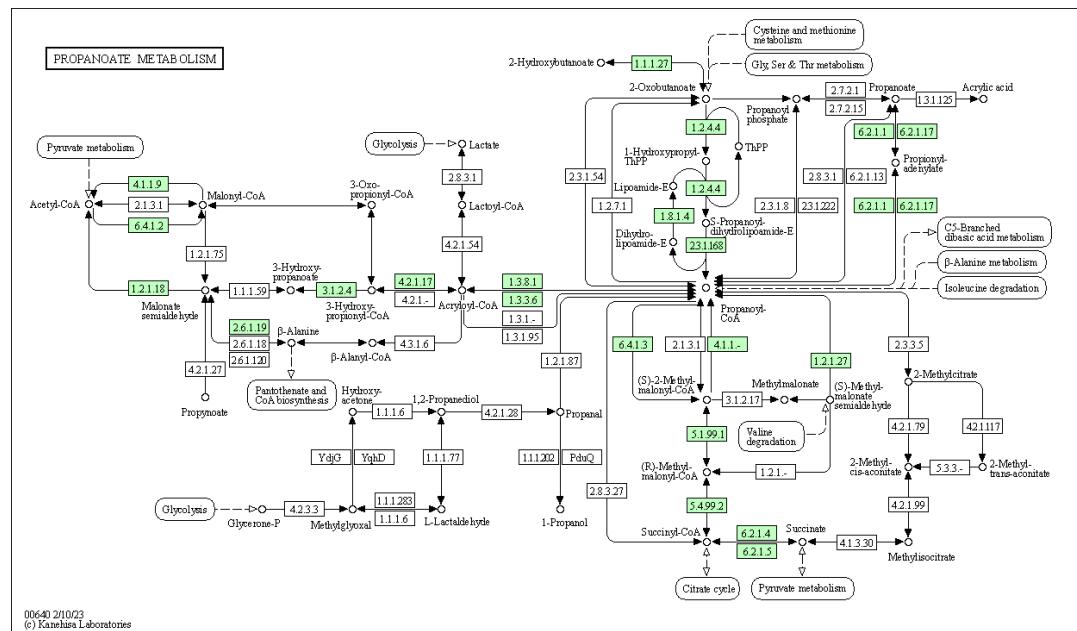


Figure 2.11: The propanoate metabolic pathways in homo sapines. The green pathways are the pathways exist in homo sapines

### 2.1.5 Gene defect analysis using OMIM database

The information of 2-Methyl Citrate Synthase does not exist in the OMIM database.

## 2.2 discussions

a) Q: Compare and contrast the information available on the ENZYME and UniProt pages and explain when you might one to consult one rather than the other.

A: Uniprot gives general-purpose information on the protein, it specifies various aspects like: Cell localization, names and taxonomy, expression, interaction, sequence, relative PDB entries or alphafold-predicted structure. UniProt allows us to have an "Omic" view of the protein. Whereas ENZYME focuses on the enzymatic activity, from a chemical perspective of reactions, products, conditions of temperature and pH. It also provides all the publications and resources relative to claims.

b) Q: In three independent digestions, subject your protein to cleavage with trypsin, chymotrypsin (high specificity) and cyanogen bromide. Use the tool PeptideCutter found on the ExPASy home page (ExPASy → proteomics tools → Protein sequences and Identification → peptidecutter). Turn in the cleavage map with your problem set. For one peptide (10 amino acids or longer) draw out the complete structure, calculate its molecular weight and determine its net charge at pH 2, pH 7 and pH 9. You may use the Peptide Property Calculator.

A: The following figure is the cleavage maps of our protein sequence( corresponding to the UniProt entry P31660 ), the cleavage was performed by Trypsin, High-affinity Chymotrypsin, and Cyanogen-Bromide. The maps allow us to simply visualize where the cut sites are concentrated, and if any sites are cleaved by more than one enzyme. These maps were computed trough the use of Expasy - Peptide Cutter [23]

## 2.2 discussions

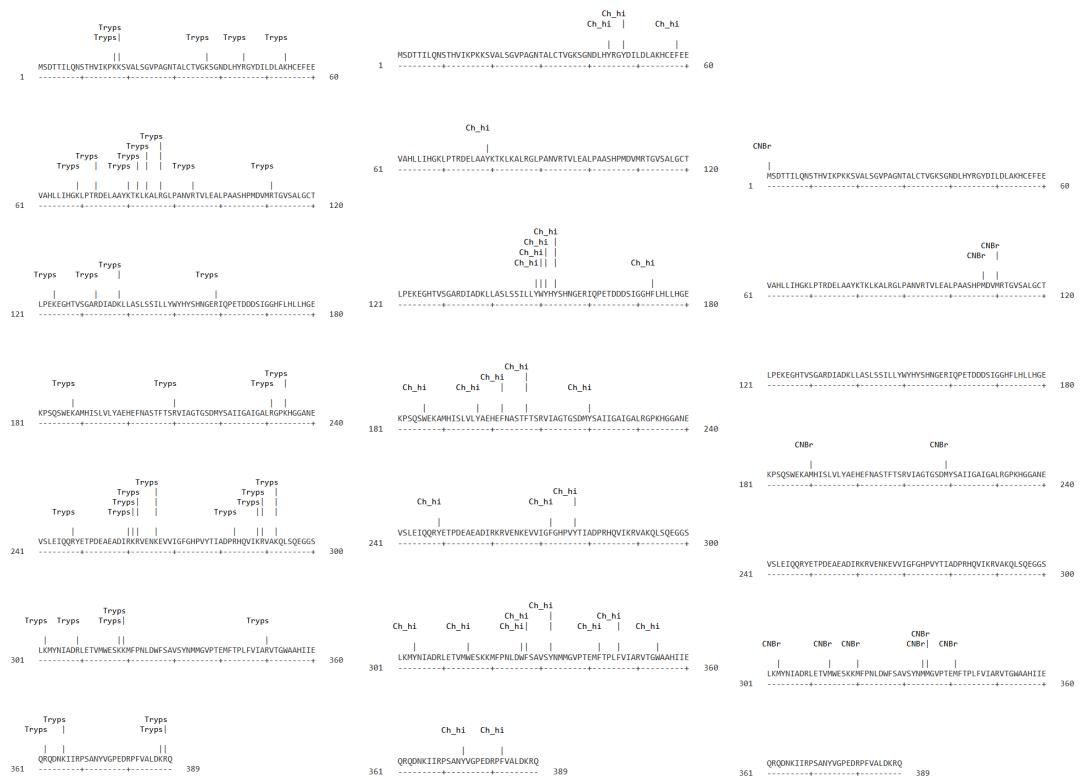


Figure 2.12: Cleavage Sites of trypsin Hi-Aff. chymotrypsin and CyanogenBromide

In the following table we can see the peptide sequence of one of the resulting peptides after enzymatic digestion with trypsin, we also have the molecular weight and Net-charge values at different pH Levels.

<b>Peptide Sequence</b>	MSDTTILQNSTHVIKPK
<b>Molar Weight</b>	1913.19 g/mol
<b>Peptide Net-Charge pH2</b>	3.7
<b>Peptide Net-Charge pH7</b>	1.1
<b>Peptide Net-Charge pH9</b>	0.8

Table 2.1: Peptide Calculations

The following figure displays a continuous plot of the Net-charge over all pH values.

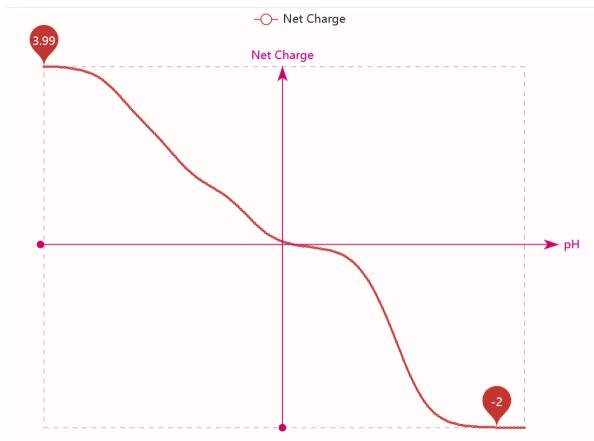


Figure 2.13: Net-charge plot of 2-methylcitrate synthase

c) What chemical reaction does your enzyme catalyze?

A: It is involved in the catabolism of short chain fatty acids (SCFA) via the tricarboxylic acid (TCA)(acetyl degradation route) and via the 2-methylcitrate cycle I (propionate degradation route). Catalyzes the Claisen condensation of propionyl-CoA and oxaloacetate (OAA) to yield 2-methylcitrate (2-MC) and CoA. Also catalyzes the condensation of oxaloacetate with acetyl-CoA to yield citrate but with a lower specificity.

d) Q: Does your enzyme require a cofactor for catalytic activity?

A: The enzyme we investigated requires coenzyme A as a cofactor.

e) Q: In what metabolic pathway does your protein participate?

A: 2-methylcitrate synthase catalyzes the conversion of propionyl-CoA and oxaloacetate into 2-methylcitrate by allowing the transfer of an acyl group which is converted into an alkyl upon transfer. This reaction is a crucial step in the anabolism of propionate, and thus energetic metabolism.

f) Q: Is there a disease associated with a mutation of your protein? If so, what is it? If there are multiple, describe one. Include information on what that disease is, what the mutation is that causes it and whether the disease is related to altered activity, altered expression levels or improper regulation of the activity.

A: Our enzyme is associated with E. Coli, and more generally with bacteria, so a pathological study to see how a mutation affects the protein function , and therefore the organism as a whole, is beyond any reasonable scope, for this reason we assume there are no notable mutations associated with the protein.

# 3 Project 3

## 3.1 results

In the following section we will display the results relative to the three exercises required to better understand our protein from a structural perspective.

### 3.1.1 Space filling vs Cartoons

In figure 3.1 we can see representative views of the crystal structure of 2-methylcitrate synthase from *Salmonella typhimurium*, cultured in E.Coli. On the left we can see a cartoon view displaying in cyan the alpha-helices, in red the antiparallel beta-sheet and in magenta the unstructured coils. This type of depiction allows us to understand the protein from mechanic point of view, alpha-helices are very rigid structures that give the overall protein stability in holding it's shape, it also allows to easily if the beta-sheets are parallel or antiparallel, and finally to visualize the number of inherently disordered regions which can be used for multiple purposes, like ubiquitination or protein transport. On the right we can see a spacefill representation where the surface of the protein is approximated, this type of view allows to better understand the topology of our protein and inspect if there are any pockets where solvents, ions or ligands could be held. In fact we can recognize 3 aminoacids in yellow in what could be recognized as the enzymatic cleft.

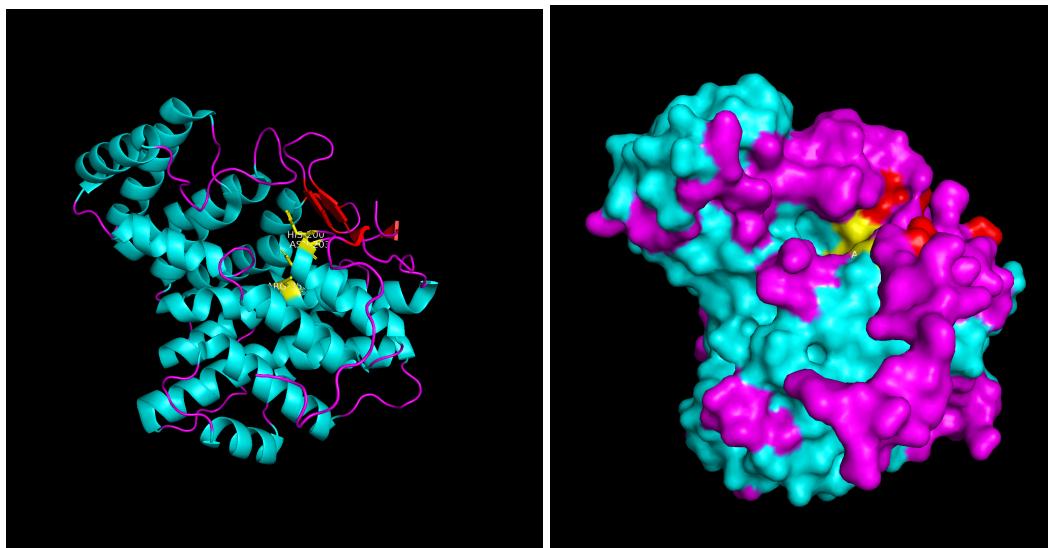


Figure 3.1: Cartoon and Space-fill Representations.

### 3.1.2 Enzymatic Pocket

In Figure 3.2 we can see the enzymatic cleft of our protein around the Glycerol, and we can recognize that the Glycerol has dipole interactions with one water molecule and three distinct aminoacids, in particular Histidine-200, Arginine-350 and Asparagine-203. Particular attention should be given to the role of the R group of Asparagine-203 in the enzymatic reaction, because it is the last aminoacid and it doesn't contribute to the reaction with only the dipole moment of the aminoacid itself, but instead with the one created by the whole alpha-loop[[HDB78]].

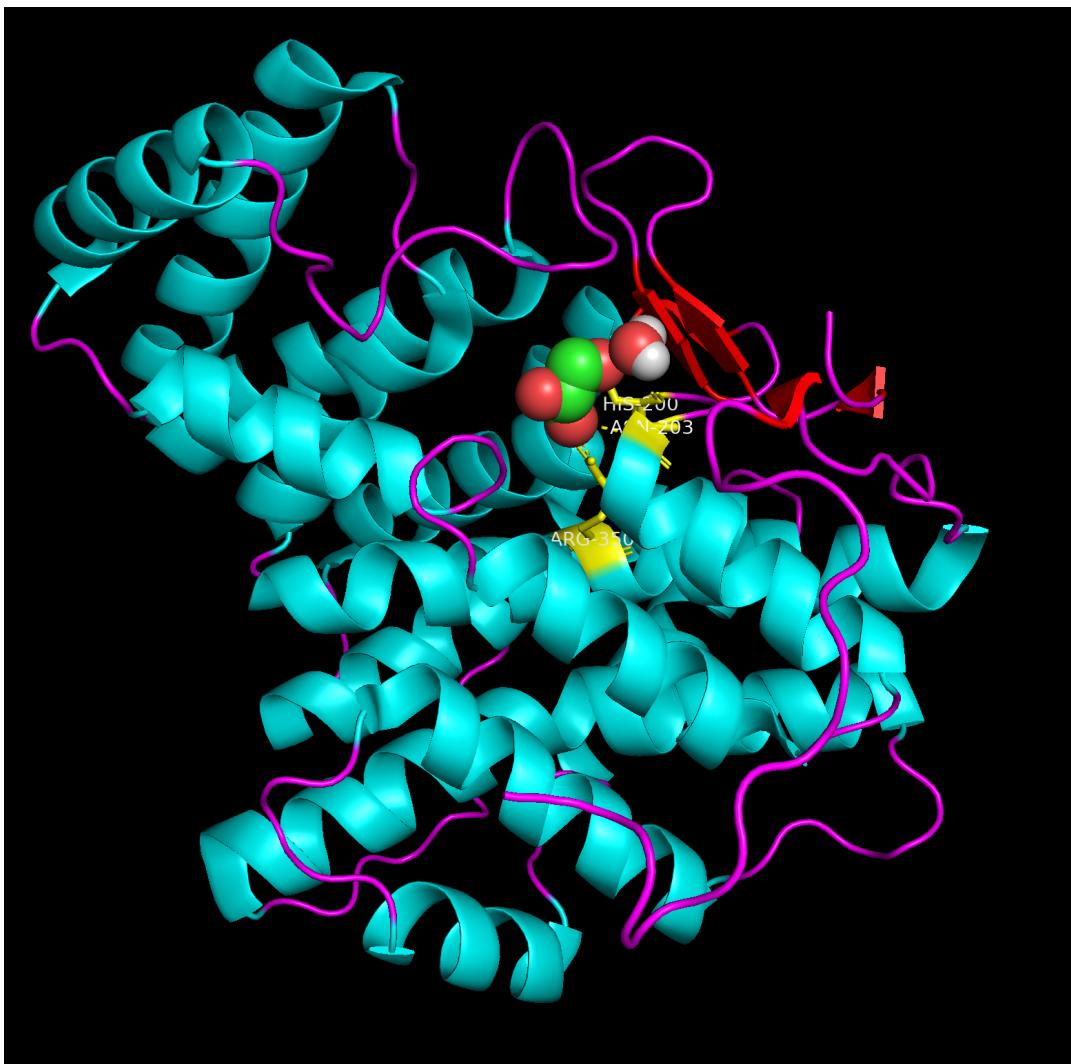


Figure 3.2: Ligand interacting with residues.

### 3.1.3 Protein Description

2-methylcitrate synthase is a key enzyme in the propionate metabolism pathway. It plays a crucial role in converting propionyl-CoA to 2-methylcitrate, this process is important in

using propionate, as a scaffold to synthesizemethylcitrate, and it's particularly relevant in our bacteria, Escherichia coli. The Structure was achieved trough a X-Ray Crystallography, a technique in which a beam of X-rays is directed at a crystal( repeated unit cell ). The crystal acts as a three-dimensional diffraction grating, scattering the X-rays in different directions. By analyzing the resulting diffraction pattern, scientists can mathematically reconstruct the arrangement of atoms within the crystal. because the angles and intensities of the diffracted X-rays provide information about the spatial distribution of electrons in the crystal lattice. The resolution of the unit cell in Figure 3.4 is 0.241 nm ( or 2.41 Angstroms ). In the figure we can recognize that the unit cell contains a complex of 10 subunits, each subunit represents only a single protein, they are displayed in different colours to enable better discrimination between subunits. The small red dots are water molecules complexed with the protein, meaning that they are part of the fixed crystal lattice and should therefore not be considered mobile elements of the solvent, at best it can be used to compute the minimal hydrodynamic radius. The complex is just a result of the energy minimization of steric constraints and hydrophobic regions during the crystallization process, thus it does not indicate the in-vivo conditions of the protein. The ligand

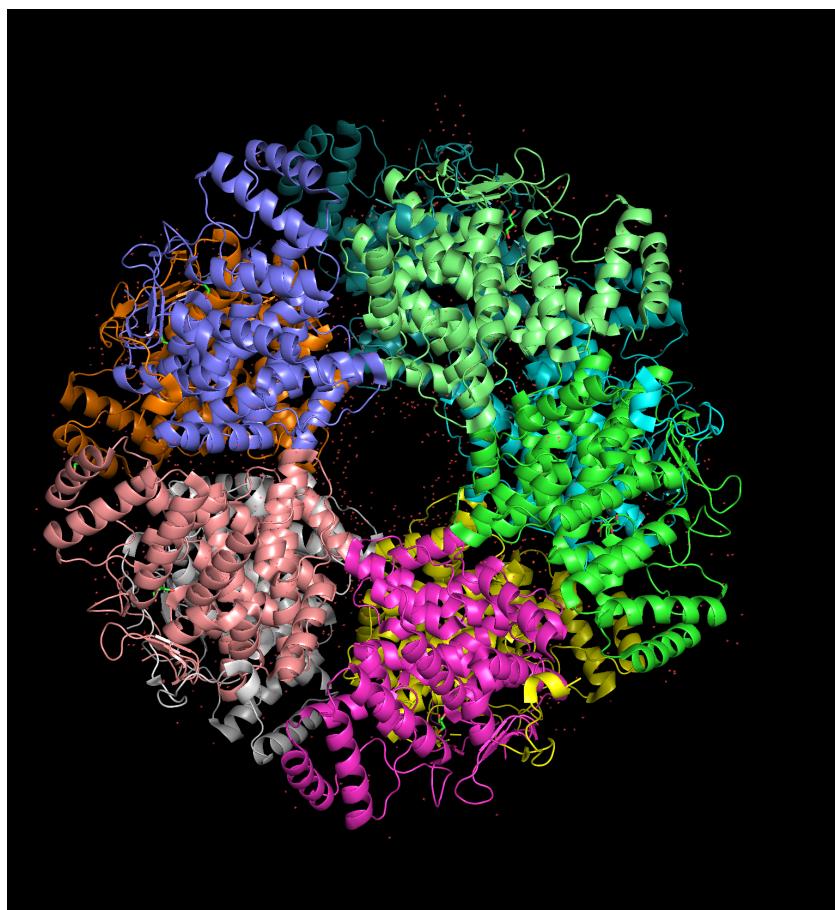


Figure 3.3: Unit-cell of the protein crytsal

A note should be disclosed on the structure we studied: the crystal structure does not

### 3.1 results

correspond to the same protein analyzed in the previous parts of our study. The sequences analyzed in the previous parts of our study gave a putative belonging of our protein to E.Coli, the sequence of which corresponds to the prpC gene, whereas the protein analyzed derives from the same highly conserved (but not perfectly identical in sequence) gene of the organism Salmonella Enterica. The gene was transfected in E.Coli for better culture conditions but the protein sequence is still slightly different from the one we associated with our sample. In Figure 3.3 we display the computed structure of our sequence by AlphaFold [[Jum+21]], we can say from a preliminary inspection that the two structures resemble each other very much, except for a region at the C-terminus, which we can disregard as it seems to not be part of the enzymatic region of our protein.

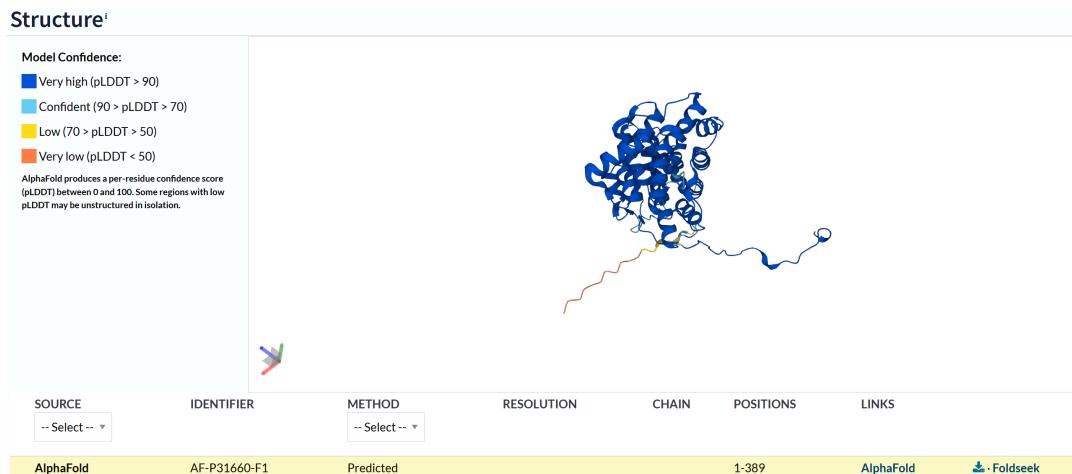


Figure 3.4: Alphafold prediction

In the following Figure 3.5, a summary of the relevant information regarding the crystal structure ,submitted to pdb, we used in our study is displayed.

### 3.1 results

22/01/2024, 15:17

RCSB PDB - 3O8J: Crystal structure of 2-methylcitrate synthase (PrpC) from *Salmonella typhimurium*



Figure 3.5: ProteinDataBank Report

## 3.2 discussions

Considering the above mentioned details regarding the structured of our protein we can say that inspection of protein structures trough computational tools like Pymol [[\[Sch15\]](#)] or Chimera [[\[Pet+04\]](#)] provide great insight to the molecular functioning of proteins. Such tools, and the various models and simulations they enable( which were not investigated in this study because of time constraints) are of great using at the cutting edge of drug development, interactomics and more generally to biophysics as a whole.

# 4 Project 4

This project focuses on exploring the enzymology of our protein through the BRENDA database [Cha+21].

## 4.1 results

The following results were found by searching 2-methylcitrate synthase in the BRENDA database and selecting the only result with an E.C. number of 2.3.3.5.

### 4.1.1 The searching result from BRENDA database

In the following picture, we can see the homepage of BRENDA which allows us to select a variety of information on our protein of interest.

The screenshot shows the BRENDA homepage with the following details:

- Header:** BRENDA Home, Go to enzyme, show all | hide all, No of entries: 57.
- Left sidebar (Nomenclature):**
  - Enzyme Nomenclature: 57
  - Enzyme-Ligand Interactions: 323
  - Diseases: 6
  - Functional Parameters: 170
  - Organism related Information: 130
  - General Information: 14
  - Enzyme Structure: 320
  - Molecular Properties: 44
  - Applications: 4
  - References: 30
  - External Links
- Central Content:**
  - Information on EC 2.3.3.5 - 2-methylcitrate synthase**  
for references in articles please use BRENDA:EC2.3.3.5
  - EC Tree:**
    - 2 Transferases
      - 2.3 Acyltransferases
        - 2.3.3 Acyl groups converted into alkyl groups on transfer
          - 2.3.3.5 2-methylcitrate synthase
  - IUBMB Comments:**

The enzyme acts on acetyl-CoA, propanoyl-CoA, butanoyl-CoA and pentanoyl-CoA. The relative rate of condensation of acetyl-CoA cannot be replaced by glyoxylate, pyruvate or 2-oxoglutarate.
  - Specify your search results:**
    - Mark a special word or phrase in this record:
    - Search Reference ID:
    - Search UniProt Accession:
    - Select one or more organisms in this record:
      - Burkholderia cepacia contig261
      - Citeromyces matritensis
      - Corynebacterium glutamicum
      - Cryptococcus neoformans
      - Cupriavidus metallidurans
      - Cupriavidus necator
    - Submit
    - Show additional data
      - Do not include text mining results
      - Include AMENDA (text mining) results
      - Include BRENDA results (AMENDA + additional results, but less precise)
  - Word Map:** A map showing words related to the enzyme, such as propionyl-coa, 2.3.3.5, propionate, and various microorganisms like *lutzii*, *prpbcd*, *3-hydroxyvalerate*, etc.

Figure 4.1: BRENDA Homepage

## 4.1 results

SYNONYM ▾	ORGANISM ▾	UNIPROT ▾	COMMENTARY ▾ ×	LITERATURE ▾
2-methylcitrate oxaloacetate-lyase	-	-	-	-
citrate synthase II	Escherichia coli	-	-	488171
MCA condensing enzyme	-	-	-	-
MCS	-	-	-	-
Methylcitrate synthase	-	-	-	-
methylcitrate synthetase	-	-	-	-
PrpC	Escherichia coli	P31660	-	755708

Figure 4.2: Synonyms of 2-Methyl citrate synthase

KM VALUE [mM] ▾							SUBSTRATE ▾	ORGANISM ▾	UNIPROT ▾	COMMENTARY ▾ ×	LITERATURE ▾	IMAGE ▾		
0.101	acetyl-CoA	Escherichia coli	-	completed mutant strain W620	488172									
0.005	oxaloacetate	Escherichia coli	-	completed mutant strain W620	488172									
0.017 - 0.037	propionyl-CoA	□ 2 entries												
△ top print hide Go to Specific Activity Search														
SPECIFIC ACTIVITY [μmol/min/mg] ▾							ORGANISM ▾	UNIPROT ▾	COMMENTARY ▾ ×	LITERATURE ▾				
0.33		Escherichia coli	P31660	-		488171								
△ top print hide Go to pH Optimum Search														
pH OPTIMUM ▾	ORGANISM ▾	UNIPROT ▾	COMMENTARY ▾ ×	LITERATURE ▾										
9	Escherichia coli	P31660	-	488171										
△ top print hide Go to Temperature Optimum Search														
TEMPERATURE OPTIMUM ▾	ORGANISM ▾	UNIPROT ▾	COMMENTARY ▾ ×	LITERATURE ▾										
45 - 50	Escherichia coli	P31660	-	488171										

Figure 4.3: Reaction conditions of 2-Methyl citrate synthase

SUBSTRATE ▾	PRODUCT ▾	REACTION DIAGRAM	ORGANISM ▾	UNIPROT ▾	COMMENTARY (Substrate) ▾ ×	LITERATURE (Substrate) ▾	COMMENTARY (Product) ▾ ×	LITERATURE (Product) ▾	REVERSIBILITY
r=reversible ir=irreversible ?=not specified ▾									
acetyl-CoA + H <sub>2</sub> O + oxaloacetate	citrate + CoA	∅	Escherichia coli	-	-	488172	-	-	?
propanoyl-CoA + H <sub>2</sub> O + (2S,3S)-2-hydroxybutane-1,2,3-tricarboxylate + CoA	oxaloacetate	∅	Escherichia coli	P31660	-	755708	-	-	?
propanoyl-CoA + H <sub>2</sub> O + oxaloacetate	2-methylcitrate + CoA	∅	□ 11 entries						
additional information	?	-	□ 2 entries						
△ top print hide 10 entries □ Go to Natural Substrate Search									
NATURAL SUBSTRATE ▾	NATURAL PRODUCT ▾	REACTION DIAGRAM	ORGANISM ▾	UNIPROT ▾	COMMENTARY (Substrate) ▾ ×	LITERATURE (Substrate) ▾	COMMENTARY (Product) ▾ ×	LITERATURE (Product) ▾	REVERSIBILITY
r=reversible ir=irreversible ?=not specified ▾									
propanoyl-CoA + H <sub>2</sub> O + oxaloacetate	2-methylcitrate + CoA	∅	□ 10 entries						

Figure 4.4: Substrates related to the reactions of 2-Methyl citrate synthase

<a href="#">Go to Reaction Type Search</a>				
REACTION TYPE ▲▼	ORGANISM ▲▼	UNIPROT ▲▼	COMMENTARY ▲▼	X LITERATURE ▲▼
condensation	-	-	-	-

<a href="#">Go to Pathway Search</a>	
PATHWAY SOURCE ▲▼	PATHWAYS ▲▼
BRENDA	propionate fermentation
KEGG	Propanoate metabolism
-	-,-

Figure 4.5: Reaction types and pathway sources of 2-Methyl citrate synthase

## 4.2 discussions

a) Q: Use the biochemical literature or biochemistry textbooks to find the mechanism that has been proposed for your enzyme. If you take the mechanism from a source – be sure to include an appropriate reference.

A: The 3 substrates of this enzyme are propanoyl-CoA, H<sub>2</sub>O, and oxaloacetate, whereas its two products are (2R,3S)-2-hydroxybutane-1,2,3-tricarboxylate and CoA. This enzyme belongs to the family of transferases, specifically those acyltransferases that convert acyl groups into alkyl groups on transfer. The systematic name of this enzyme class is propanoyl-CoA:oxaloacetate C-propanoyltransferase (thioester-hydrolysing, 1-carboxyethyl-forming). Other names in common use include 2-methylcitrate oxaloacetate-lyase, MCS, methylcitrate synthase, and methylcitrate synthetase. This enzyme participates in propanoate metabolism.[\[\[Cha+21\]\]](#)

b) Q: What are the ranges of specific activity that have been measured for this enzyme?

A: The measured ranges of activity displayed on BRENDA [\[\[Cha+21\]\]](#) for pH are 5.5 to 10.5, the analyzed temperature range is from 10 to 90 degrees.

c) Q: Are there different specific activities reported for the enzyme from the same organism? If so, why might the specific activity vary?

A: The range of specific activity goes from 48 to 0.002, in particular for E.Coli it is 0.33.

d) Q: Why do you expect the specific activity of the same enzyme isolated from different organisms to be different?

A: The specific activity changes drastically between organisms. The variation in specific activity values for the same enzyme across different organisms can be attributed to several factors. Evolutionary adaptations play a crucial role, as enzymes undergo changes over time to meet the specific needs and environmental conditions of each organism. Substrate specificity is another

factor, with organisms evolving enzymes tailored to their unique metabolic pathways. Optimal conditions, such as pH and temperature also influence enzyme activity, and organisms in different environments may express enzymes optimized for their specific habitats. Variations in metabolic pathways and the utilization of alternative substrates contribute to differences in enzyme specific activity. Additionally, the regulation of enzyme activity through feedback inhibition, allosteric regulation, and genetic control can vary among organisms. Genetic diversity within a species or population further adds to the variability in enzyme properties. Different strains or individuals may express slightly different forms of an enzyme, leading to differences in specific activity. In essence, the diversity in specific activity values reflects the intricate interplay of evolutionary, environmental, and genetic factors shaping biochemical processes in various biological systems.

e) Q: Go to the section on KM. Pick a substrate that has been used in several studies. What are the ranges of KM's that have been measured for that substrate? Explain what leads to variation in the kinetic parameters measured between these different studies.

A: The range of the KM for acetyl-CoA is from 0.00126 to 0.35 1/mMs<sup>-1</sup>

f) Q: What is the pH optimum of your enzyme? You have already explored the structure of your enzyme and the overall mechanism (part a above). Relate the pH optimum to some aspect of the enzyme mechanism.

A: The pH optimum relates to the enzyme mechanism and structure because enzyme structure is directly related to the pH of the solution around it, different pH could cause the enzyme to unfold or to increase the hydrophobicity of the enzyme making it structurally immobile. This results in a change of the efficiency of the enzyme .

g) Q: Identify two competitive inhibitors for the enzyme from the table of inhibitors. Remember that competitive inhibitors usually look like the natural substrate

A: 3,3-thiodipropionic acid and 3-phosphonopropionic acid are competitive inhibitors with a similar structure to the substrate which is oxaloacetate.

# 5 Project 5

## 5.1 results

In this project, we generated five homologs of 2-Methyl citrate synthase from PDB and ENZYME. We applied Cluster Omega to compare the conserved sequences among the homologs, and plotted the 3-D structure of 2-Methyl citrate synthase labeled by conservation possibility, to see the relationship between conserved sequences and protein structure.

### 5.1.1 homologs of 2-Methyl citrate synthases

Searching from the database, we found five homologs of the 2-Methyl citrate synthases in ecoli, which are 2-methylcitrate synthase|Salmonella enterica (90371), Citrate synthase|Pseudomonas aeruginosa PAO1 (208964), Methylcitrate synthase|Mycobacterium tuberculosis (83332), 2-methylcitrate synthase|Coxiella burnetii (777) and 2-methylcitrate synthase, mitochondrial|Neosartorya fumigata (strain CEA10 / CBS 144.89 / FGSC A1163) (451804).

### 5.1.2 Comparing the homologs using ClustalOmega(figure 5.1 and 5.2)

### 5.1.3 The 3D image showing the highly conserved regions(figure 5.3)

## 5.2 discussions

a) Q: What do the “\*”, “.” and “:” indicate on the MSA alignment?

A: An \* (asterisk) indicates positions which have a single, fully conserved residue; a : (colon) indicates conservation between groups of strongly similar properties, roughly equiv-

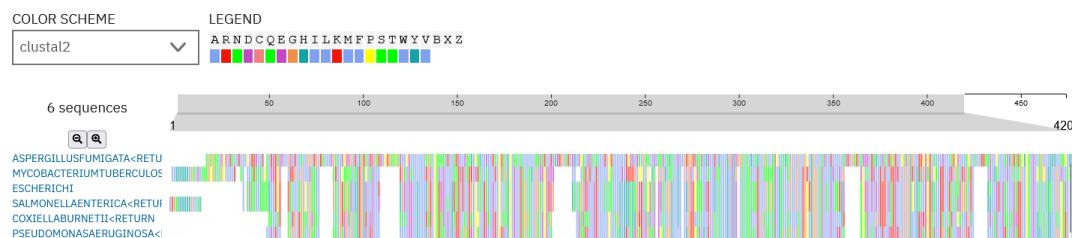


Figure 5.1: The overview of the sequences among the homologs.

## 5.2 discussions

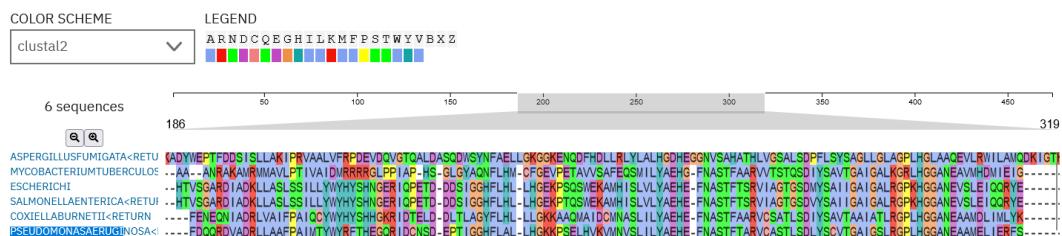


Figure 5.2: The highly conserved sequences area among the homologs

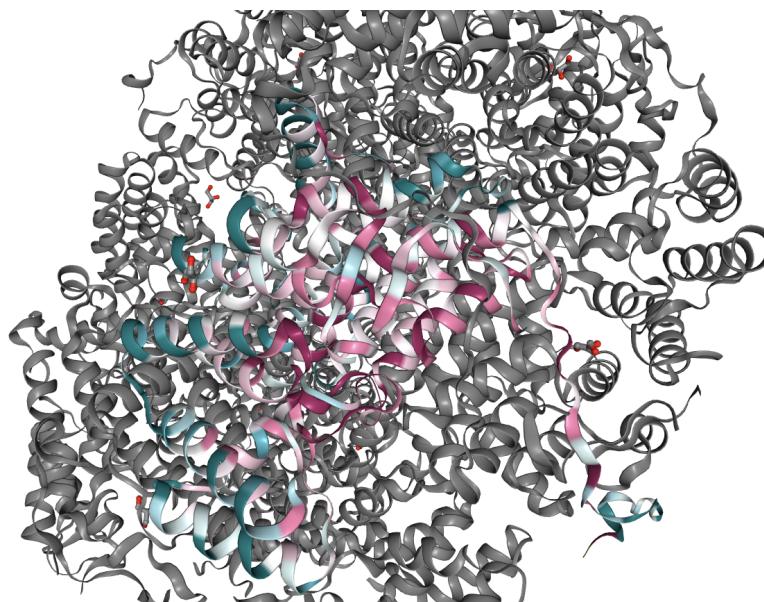


Figure 5.3: The conserved part(red) and the variable part(green) of the enzyme

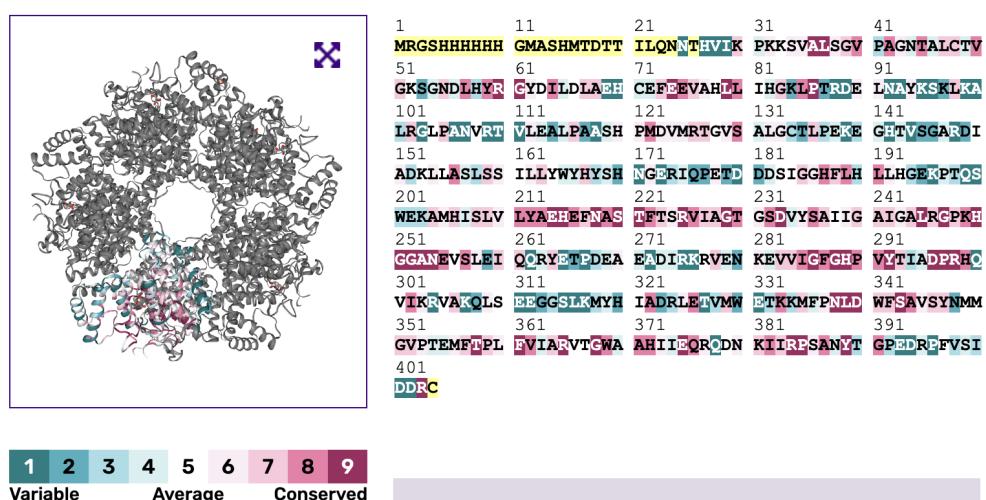


Figure 5.4: The conserved sequences of 2-Methyl citrate synthase chain A, generated by Con-surf

alent to scoring > 0.5 in the Gonnet PAM 250 matrix; A . (period) indicates conservation between groups of weakly similar properties, roughly equivalent to scoring =< 0.5 and > 0 in the Gonnet PAM 250 matrix.

b) Q: How does the alignment generated by ConSurf differ from the one you generated manually using ClustalOmega? Don't focus on the format of the output (which is different only because you selected a certain output style in ClustalOmega), but rather, look at the actual content – the parts that are similar versus different and the sequences that have been used.

A: The results generated from Consurf show the possibility of conservation for each amino-acid(Figure 5.4), while the results of Clustal Omega directly display the sequences of differenet homologs, while the users can find the conserved sequences by themselves. Moreover, the Consurf results only illustrate the possibility on each individual amino acid, while the Clustal Omega computations display the full conserved sequences.

c) Q: When you look at the primary sequence and the amino acid conservation, do you see any patterns or organization?

A: Yes. While observing the primary sequences and the amino acid conservation, we noticed that there are some special amino acid sequences that are highly similar among the homologs, some sequences are similar among several homologs, but not all, while some sequences are highly random between different homologs. Similar sequences appear mainly between the amino acid numbers 200 to 350.

d) Q: What is a conserved sequence motif? With the help of ExPASy, try to identify a specific conserved sequence motif in your alignment.

A: A sequence motif is a set of conserved amino acid residues that are important for protein function and are loacted within a certain distance from each other. The conserved sequence motif means that the sequnece motif is based on conserved sequences.  
An analysis tool called EMBOSS CONs is applied to find the specific conserved sequence motif. With the input of five homologs sequences, the comparing results are shown in the figure 5.5. In the results, for example, sequence 'kiaals' exists in all homologs, which means that it's highly probable for 'kiaals' be a conserved sequence motif.

e) Q: Is there a relationship between the sequence conservation and the overall 3-D structure? To answer this question, compare the printouts you made of the most conserved and the least conserved parts of the sequence.

A: As is shown in the figure 5.3, compared to the more variable parts, most conserved sequences are located at the center part of the chain. Besides, the shape of conserved sequences are basically standard alpha helices and beta sheets, while the variable parts are more random.

```
>EMBOSS0001
MSxxxxxxxxxxmxaqxxxxttALsxxxxkxlssxxxxxxlaNxxxdivxxxxmlxxex
xxxxldxxQxxxlxxRaxxxqx1KxlxxxxexnaxxMxxmllxgsxxmxlxqxxsxsqqx
xxdmmxxivxxikxxxyxxxxhtxxxixxxaDxtxSgxaxxxlxxxkxaxxxxxxMxxsm
xlxxxxxkiaalsxxxxqSxlnaxyxSvxxxAxxsxxxxsxxxxeaxxxxxlYxaxve
xsxeiQxkxxsxexxvalxxrlxxexdxsxxiLxxxxkxxxxxxxxxk1xxxxxaqexxx
mxekxxxxxNexyxsxkayxxvxxkxxxxxxxxMxr vxaxxxaavtqxxxnxxxxxxxxx
yxxexqxaxxtxxxKrexrxvvxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
```

Figure 5.5: the conserved sequence motif generated by EMBOSS CONs.

# Bibliography

- [HDB78] W. G. J. Hol, P. T. van Duijnen, and H. J. C. Berendsen. “The -helix dipole and the properties of proteins”. en. In: *Nature* 273.5662 (June 1978). Number: 5662 Publisher: Nature Publishing Group, pp. 443–446. ISSN: 1476-4687. doi: [10.1038/273443a0](https://doi.org/10.1038/273443a0). URL: <https://www.nature.com/articles/273443a0> (visited on 02/04/2024).
- [Alt+90] S. F. Altschul et al. “Basic local alignment search tool”. eng. In: *J Mol Biol* 215.3 (Oct. 1990), pp. 403–410. ISSN: 0022-2836. doi: [10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
- [Gas+03] Elisabeth Gasteiger et al. “ExPASy: The proteomics server for in-depth protein knowledge and analysis”. eng. In: *Nucleic Acids Res* 31.13 (July 2003), pp. 3784–3788. ISSN: 1362-4962. doi: [10.1093/nar/gkg563](https://doi.org/10.1093/nar/gkg563).
- [Pet+04] Eric F. Pettersen et al. “UCSF Chimera—a visualization system for exploratory research and analysis”. eng. In: *J Comput Chem* 25.13 (Oct. 2004), pp. 1605–1612. ISSN: 0192-8651. doi: [10.1002/jcc.20084](https://doi.org/10.1002/jcc.20084).
- [Sch15] Schrödinger, LLC. “The PyMOL Molecular Graphics System, Version 1.8”. Nov. 2015.
- [Cha+21] Antje Chang et al. “BRENDA, the ELIXIR core data resource in 2021: new developments and updates”. In: *Nucleic Acids Research* 49.D1 (Jan. 2021), pp. D498–D508. ISSN: 0305-1048. doi: [10.1093/nar/gkaa1025](https://doi.org/10.1093/nar/gkaa1025). URL: <https://doi.org/10.1093/nar/gkaa1025> (visited on 02/06/2024).
- [Jum+21] John Jumper et al. “Highly accurate protein structure prediction with AlphaFold”. en. In: *Nature* 596.7873 (Aug. 2021). Number: 7873 Publisher: Nature Publishing Group, pp. 583–589. ISSN: 1476-4687. doi: [10.1038/s41586-021-03819-2](https://doi.org/10.1038/s41586-021-03819-2). URL: <https://www.nature.com/articles/s41586-021-03819-2> (visited on 02/06/2024).
- [Szk+21] Damian Szkłarczyk et al. “The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets”. eng. In: *Nucleic Acids Res* 49.D1 (Jan. 2021), pp. D605–D612. ISSN: 1362-4962. doi: [10.1093/nar/gkaa1074](https://doi.org/10.1093/nar/gkaa1074).
- [Ban+22] Parit Bansal et al. “Rhea, the reaction knowledgebase in 2022”. In: *Nucleic Acids Research* 50.D1 (Jan. 2022), pp. D693–D700. ISSN: 0305-1048. doi: [10.1093/nar/gkab1016](https://doi.org/10.1093/nar/gkab1016). URL: <https://doi.org/10.1093/nar/gkab1016> (visited on 02/06/2024).

- [Kan+23] Minoru Kanehisa et al. “KEGG for taxonomy-based analysis of pathways and genomes”. eng. In: *Nucleic Acids Res* 51.D1 (Jan. 2023), pp. D587–D592. ISSN: 1362-4962. doi: [10.1093/nar/gkac963](https://doi.org/10.1093/nar/gkac963).
- [23] *peptide property*. Accessed: February 5, 2024. 2023. URL: [https://www.novoprolabs.com/tools/calc\\_peptide\\_property](https://www.novoprolabs.com/tools/calc_peptide_property).