Jaypee University of Information Technology
Waknaghat, Distt. Solan (H.P.)
Learning Resource Center
CLASS NUM: BOOK NUM.:
ACCESSION NO.: SP09132 / SP0913053
This book was issued is overdue due on the date stamped below. if the book is kept over due, a fine will be charged as per the library rules.

Due Date	Due Date	Due Date

Phylogenetic Analysis of Isocitrate Lyase (ICLs) and Pharmacophore Modelling of their Inhibitors to Identify Broad-Specific Inhibitors

Enrollment no: 091510
Name of Student: Yash Gupta
Name of Supervisor: Dr. Chittranjan Rout

November-2013

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE DEGREE OF BACHELOR OF TECHNOLOGY IN BIOINFORMATICS

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY
WAKNAGHAT, SOLAN 173215, HIMACHAL PRADESH, INDIA

Learning Resource Center
JUIT
Waknaghat, HP

CERTIFICATE

This is to verify that the work entitled **"Phylogenetic Analysis of Isocitrate Lyase (ICLs) and Pharmacophore Modelling of their Inhibitors to Identify Broad-Specific Inhibitors"** submitted by **"YASH GUPTA (091510)"** in partial fulfillment for award of **"DEGREE OF BACHELOR OF TECHNOLOGY IN BIOINFORMATICS"** from Jaypee University of Information Technology, Wakhnaghat has been carried out under my supervision. This work has not been submitted partially or wholly to any other University or Institute for award of this or any other degree or diploma.

Signature of Supervisor .....

Name of Supervisor ....Dr. Chittaranjan Rout....

Designation ....Asst. Prof....

Date ....19/11/13....

ACKNOWLEDGEMENTS

Of all the people who have helped us complete this project, I am particularly thankful to our esteemed project supervisor Dr. Chittaranjan Rout for his guidance, patience and encouragement throughout the project. This project would not be possible without his guidance and active support. With deem it privilege to have worked hard under Dr.Chittaranjan Rout, who has endeared himself to his students and scholars. He is one of the best teachers that have had in our life. He is not only our mentor but also role model of our life.

I am indebted to Dr. Yajulu Medury (COO, Jayee education), Brig. (Retd.) Balbir Singh (Director, JUIT), Prof. TS Lamba (Dean, Academic and Researches) and Prof. RS Chauhan (Dean and HOD, BT and BI) for providing all kinds of facilities to carry out our project. The help rendered by all our teachers in one way or other is thankfully acknowledged.

My earnest thanks to Mr. Varun Jaiswal (PhD scholar) for his support and constant encouragement during our project work. I am also thankful to my friends for their cooperation. Needless to say, errors and omissions are mine.

Last but not the least I am thankful to my parents and family members for their support and coorperation.

Yash

Yash Gupta (091510)

CONTENT

LIST OF ABBREVIATIONS

MSA	Multiple sequence alignment
IC50	Inhibition concentration
MIC	Minimum inhibition concentration
ICL	Isocitrate lyase
MCL	2-methylisocitrate lyase
TB	Tuberculosis
FDA	U.S. Food and Drug Administration
INH	Isoniazid
RIF	Rifampin
EMB	Ethambutol
PZA	Pyrazinamide

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Mycobacterium tuberculosis (MTB) is a pathogenic bacterial species in the family Mycobacteriaceae and the causative agent of most acases of tuberculosis (TB). It has been found that ICl is considered to be the major source of carbon in the bacteria; therefore, it is responsible for the survival and the growth of bacteria. Isocitrate lyase or ICL, is an enzyme in the glyoxylate cycle that catalyzes the cleavage of isocitrate to succinate and glyoxylate [2][3]. Together with malate synthase, it bypasses the two decarboxylation steps of the tricarboxylic acid cycle (TCA cycle) and is used by bacteria, fungi, and plants [4]. ICL has found to be important in human, animal, and plant pathogenesis [4]. In this latter case, ICL has been found to be essential for survival in the host [8]. Because of its use by pathogenic fungi and bacteria, specific inhibitors are being sought for ICL and malate synthase [4]. In this project, comparative analysis of ICL sequences, and pharmacophore modeling of known ICL inhibitors were carried out. In comparative analysis, I will be conducting sequence analysis between ICL present in different organisms based on these analysis I will be conducting a Phylogenetic analysis between these species. In present research work, a pharmacophore model of Isocitrate lyase inhibitors in *Mycobacterium tuberculosis* was determined. The pharmacophore was used in ZINCPharmer tool to identify potential active molecules from ZINC database and binding mode of these molecules were determined by using LibDock and Cdocker modules of Discovery Studio software. Various potential inhibitors were predicted. Phylogenetic analysis of the ICL sequences not only provided sequence variations across the proteins but also how that variations correlate with the function. Sequence comparison analyses have shown that active site residues are conserved across different subfamily. Therefore, the molecules identified using developed pharmacophore may act as putative broad-specific inhibitors.

Signature of student: Yash

Name: YASH GUPTA

Date: 19/11/13

Signature of supervisor: Chat

Name: Dr. Chittaranjan Rout

Date: 19/11/13

CHAPTER 1

INTRODUCTION

Mycobacterium Tuberculosis

Tuberculosis (TB) is caused by a bacterium called *Mycobacterium tuberculosis*. The bacteria usually attack the lungs, but TB bacteria can attack any part of the body such as the kidney, spine, and brain. If not treated properly, TB disease can be fatal. TB is spread through the air from one person to another. The TB bacteria are put into the air when a person with TB disease of the lungs or throat coughs, sneezes, speaks, or sings. People nearby may breathe in these bacteria and become infected. TB bacteria become active if the immune system can't stop them from growing. When TB bacteria are active (multiplying in your body), this is called TB disease. People with TB disease are sick. They may also be able to spread the bacteria to people they spend time with every day [9].

Symptoms of TB include:

- Bad cough that last 3 weeks or longer
- Pain in the chest
- Weakness or fatigue
- Weight loss
- Coughing up blood or sputum
- No appetite
- Chills
- Fever
- Sweating in night [9]

Treatment for TB Disease

TB bacteria become active (multiplying in the body) if the immune system can't stop them from growing. When TB bacteria are active, this is called TB disease. TB disease will make a person

sick. People with TB disease may spread the bacteria to people with whom they spend many hours.

TB disease can be treated by taking several drugs for 6 to 9 months. There are 10 drugs currently approved by the U.S. Food and Drug Administration (FDA) for treating TB. Of the approved drugs, the first-line anti-TB agents that form the core of treatment regimens include:

- isoniazid (INH)
- rifampin (RIF)
- ethambutol (EMB)
- pyrazinamide (PZA)

Regimens for treating TB disease have an initial phase of 2 months, followed by a choice of several options for the *continuation phase* of either 4 or 7 months (total of 6 to 9 months for treatment).

It is very important that people who have TB disease finish the medicine, taking the drugs exactly as prescribed. If they stop taking the drugs too soon, they can become sick again; if they do not take the drugs correctly, the TB bacteria that are still alive may become resistant to those drugs. TB that is resistant to drugs is harder and more expensive to treat [10].

Tuberculosis (TB) is the leading cause of death in the world from a bacterial infectious disease. The disease affects 1.8 billion people/year which is equal to one-third of the entire world population.

In the United States TB is on the decline. In 2007 a total of 13,293 cases were reported. The TB rate declined to 4.4 cases per 100,000 populations, the lowest recorded rate since national reporting began in 1953. Despite this overall improvement, progress toward TB elimination has slowed in recent years; the average annual percentage decline in the TB rate slowed from 7.3% per year during 1993--2000 to 3.8% during 2000--2007. Also, since 1993 there has been a gradual decline in the number of TB patients with coinfection with HIV, and the number of cases of multiple drug-resistant TB has gradually dropped [11].

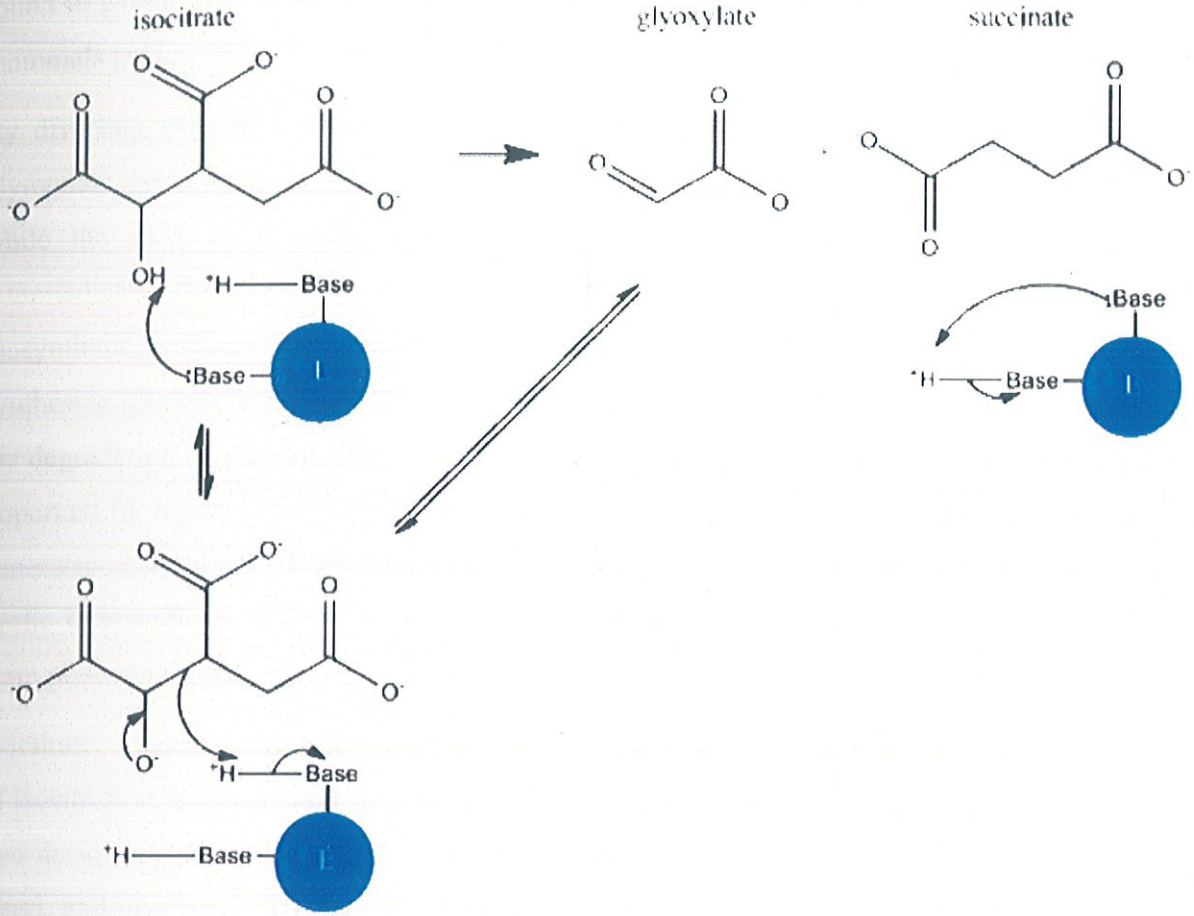
Role of Isocitrate lyase in *Mycobacterium tuberculosis*

The role of isocitrate lyase (ICL) in the glyoxylate cycle and its necessity for persistence and virulence of *Mycobacterium tuberculosis* has been well described. Recent reports have alluded to an additional role for this enzyme in *M. tuberculosis* metabolism, specifically for growth on propionate. A product of beta-oxidation of odd-chain fatty acids is propionyl-CoA. Clearance of propionyl-CoA and the by-products of its metabolism via the methylcitrate cycle is vital due to their potentially toxic effects. Although the genome of *M. tuberculosis* encodes orthologues of two of the three enzymes of the methylcitrate cycle, methylcitrate synthase and methylcitrate dehydratase, it does not appear to contain a distinct 2-methylisocitrate lyase (MCL). Detailed structural analysis of the MCL from Escherichia coli suggested that the differences in substrate specificity between MCLs and ICLs could be attributed to three conserved amino acid substitutions in the active site, suggesting an MCL signature. However, here we provide enzymatic evidence that shows that despite the absence of the MCL signature, ICL1 from M. tuberculosis can clearly function as a MCL. Furthermore, the crystal structure of ICL1 with pyruvate and succinate bound demonstrates that the active site can accommodate the additional methyl group without significant changes to the structure [12].

Mechanism of isocitrate lyase

This enzyme belongs to the family of lyases, specifically the oxo-acid-lyases, which cleave carbon-carbon bonds. Other enzymes also belong to this family including carboxyvinyl-carboxyphosphonate phosphorylmutase which catalyses the conversion of 1-carboxyvinyl carboxyphosphonate to 3-(hydrohydroxyphosphoryl) pyruvate carbon dioxide, and phosphoenolpyruvate mutase, which is involved in the biosynthesis of phosphinothricin tripeptide antibiotics.

During catalysis, isocitrate is deprotonated, and an aldol cleavage results in the release of succinate and glyoxylate. This reaction mechanism functions much like that of aldolase in glycolysis, where a carbon-carbon bond is cleaved and an aldehyde is released [13].



In the glyoxylate cycle, malate synthase then catalyzes the condensation of glyoxylate and acetyl-CoA to form malate so the cycle can continue.

ICL competes with isocitrate dehydrogenase, an enzyme found in the TCA cycle, for isocitrate processing. Flux through these enzymes is controlled by phosphorylation of isocitrate dehydrogenase, which has a much higher affinity for isocitrate as compared to ICL [14]. Deactivation of isocitrate dehydrogenase by phosphorylation thus leads to increased isocitrate channeling through ICL, as seen when bacteria are grown on acetate, a two-carbon compound [14].

Inhibition of Isocitrate lyase

The ICL enzyme has been found to be functional in various archaea, bacteria, protists, plants, fungi, and nematodes[15]. Although the gene has been

found in genomes of nematodes and cnidaria, it has not been found in the genomes of placental mammals [15].

By diverting isocitrate from the TCA cycle, the actions of ICL and malate synthase in the glyoxylate cycle result in the net assimilation of carbon from 2-carbon compounds[16]. Thus, while the TCA cycle yields no net carbon assimilation, the glyoxylate cycle generates intermediates that can be used to synthesize glucose (via gluconeogenesis), and other biosynthetic products. As a result, organisms that use ICL and malate synthase are able to synthesize glucose and metabolic intermediates from acetyl-CoA derived from acetate or from the degradation of ethanol, fatty acids or poly-β-hydroxybutyrate[4]. This function is especially important for higher plants which use oilseeds. In these germinating seeds, the breakdown of oils generates acetyl-CoA. This serves as a substrate for the glyoxylate cycle, which generates other cyclic intermediates and serves as a primary nutrient source prior to the production of sugars from photosynthesis [17].

Isocitrate lyase or ICL, is an enzyme in the glyoxylate cycle that catalyzes the cleavage of isocitrate to succinate and glyoxylate [2][3]. Together with malate synthase, it bypasses the two decarboxylation steps of the tricarboxylic acid cycle (TCA cycle) and is used by bacteria, fungi, and plants [4]. ICL has found to be important in human, animal, and plant pathogenesis [4]. For several agricultural crops including cereals, cucumbers, and melons, increased expression of the gene encoding ICL is important for fungal virulence [4]. For instance, increased gene expression of *icl1* has been seen in the fungus *Leptosphaeria maculans* upon infection of canola. Inactivation of the *icl1* gene leads to reduced pathogenicity of the fungus, which is thought to be a result of the inability of the fungus to use carbon sources provided by the plant [5]. Additionally, upregulation of the glyoxylate cycle has been seen for pathogens that attack humans. This is the case for fungi such as *Candida albicans*, which inhabits the skin, mouth, GI tract, gut and vagina of mammals and can lead to systemic infections of immunocompromised patients; as well as for the bacterium *Mycobacterium tuberculosis*, the major causative agent of tuberculosis[6][7]. In this latter case, ICL has been found to be essential for survival in the host [8]. Because of its use by pathogenic fungi and bacteria, specific inhibitors are being sought for ICL and malate synthase [4]. Although some inhibitors have already been identified,

including itaconate,itaconicanhydride, bromopyruvate,nitropropionate, oxalate and malate, these are non-specific and would also inhibit other enzymes essential for host function [4].

Various ICL1 inhibitors have been discovered until now:

- Hydroquinone[24]
- 3-nitropropioamides[21]
- Munnich base[20]
- Cadiolides[18]
- Novel pthalzinyl derivative[26]
- Pyruvate ionized analogs[19]
- 5-hydroxyindole derivative[27]
- Sargchromanols[22]
- Indole derivates[23]
- Halisulphates[25]

OBJECTIVES

Isocitrate lyase (ICL) is an important enzyme required for the growth of different pathogenic microorganisms. Inhibiting the growth of *Mycobacterium tuberculosis* can be done by finding the chemical compounds which can inhibit isocitrate lyase in *Mycobacterium tuberculosis*.

My project entitled has following topic **"Phylogenetic Analysis of Isocitrate Lyase (ICLs) and Pharmacophore Modelling of their Inhibitors to Identify Broad-Specific Inhibitors"** has following objectives:

- Comparative analysis of Isocitrate lyase in different organisms
- Compiling all inhibitors and their derivatives that inhibit isocitrate lyase in different microorganisms
- Pharmacophore modelling using Discovery Studios3.5
- Docking using Discovery Studio components
    - LibDock
    - Cdocker

Tools Used

Ncbi (http://www.ncbi.nlm.nih.gov/)

The National Center for Biotechnology Information (NCBI) is part of the United States National Library of Medicine (NLM), a branch of the National Institutes of Health. The NCBI is located in Bethesda, Maryland and was founded in 1988 through legislation sponsored by Senator Claude Pepper.

The NCBI houses a series of databases relevant to biotechnology and biomedicine. Major databases include GenBank for DNA sequences and PubMed, a bibliographic database for the biomedical literature. All these databases are available online through the Entrez search engine.

Ncbi-Protein (http://www.ncbi.nlm.nih.gov/protein/)

Protein database were used to extract protein sequences of enzymes in different organisms.

Pubmed (http://www.ncbi.nlm.nih.gov/pubmed/)

PubMed is a free database accessing primarily the MEDLINE database of references and abstracts on life sciences and biomedical topics. The United States National Library of Medicine (NLM) at the National Institutes of Health maintains the database as part of the Entrezsystem of information retrieval.

ClustalW (http://www.genome.jp/tools/clustalw/)

ClustalW is a general purpose multiple alignment program for DNA or proteins.

Multiple alignments are carried out in 3 stages:

1) All sequences are compared to each other (pairwise alignments);

2) A dendrogram (like a phylogenetic tree) is constructed, describing the approximate groupings of the sequences by similarity.

3) The final multiple alignment is carried out using the dendrogram as a guide.

Muscle (http://www.ebi.ac.uk/Tools/msa/muscle/)

MUSCLE is one of the best-performing multiple alignment programs according to published benchmark tests, with accuracy and speed that are consistently better than CLUSTALW. MUSCLE can align hundreds of sequences in seconds. Most users learn everything they need to know about MUSCLE in a few minutes—only a handful of command-line options are needed to perform common alignment tasks.

T-coffee (http://www.ebi.ac.uk/Tools/msa/tcoffee/)

T-Coffee (Tree-based Consistency Objective Function For alignment Evaluation) is a multiple sequence alignment software using a progressive approach. It generates a library of pairwise alignments to guide the multiple sequence alignment. It can also combine multiple sequences alignments obtained previously and in the latest versions can use structural information from PDB files (3D-Coffee). It has advanced features to evaluate the quality of the alignments and some capacity for identifying occurrence of motifs (Mocca). It produces alignment in the aln format (Clustal) by default, but can also produce PIR, MSF, and FASTA format. The most common input formats are supported (FASTA, PIR).

Phylip

PHYLIP (PHYLogeny Inference Package) is a free computational phylogenetics package of programs for inferring evolutionary trees (phylogenies). The name is an acronym for PHYLogenyInference Package. It consists of 35 portable programs, i.e. the source code is written in C and precompiled executables are available for Windows (95/98/NT/2000/me/XP), Mac OS 8 and 9, Mac OS X, and Linux systems. Complete documentation is written for all the programs in the package and is part of the package. The author of this package is Joseph Felsenstein, Professor in the Department of Genome Sciences and the Department of Biology at the University of Washington, Seattle. Methods (implemented by each program) that are available in the package include parsimony, distance matrix, and likelihood methods, including bootstrapping and consensus trees. Data types that can be handled include molecular sequences, gene frequencies, restriction sites and fragments, distance matrices, and discrete characters.

Each program is controlled through a menu, which asks the users which options they want to set, and allows them to start the computation. The data is read into the program from a text file, which the user can prepare using any word processor or text editor (but it is important that this text file not be in the special format of that word processor—it should instead be in flat ASCII or Text Only format). Some sequence analysis programs such as the ClustalW alignment program can write data files in the PHYLIP format. Most of the programs look for the data in a file calledinfile—if they do not find this file they then ask the user to type in the file name of the data file.

Output is written onto files with names like outfile and outtree. Trees written onto outtree are in the Newick format, an informal standard agreed to in 1986 by authors of a number of major phylogeny packages.

Maximum Parsimony

The maximization of parsimony (preferring the simpler of two otherwise equally adequate theorizations) has proven useful in many fields, and this article concerns its application to phylogenetics.Occam's razor, a principle of theoretical parsimony suggested by William of Ockham in the 1320s, asserted that it is vain to give an explanation which involves more assumptions than necessary. When applied to computational phylogenetics, Maximum parsimony describes a particular non-parametric statistical method for constructing phylogenies. In this application, the preferred phylogenetic tree is the tree that supposes the least evolutionary change to explain observed data (hence maximally parsimonious). The basic idea was presented by Walter M. Fitch in 1971.

While evolution is not an inherently parsimonious process, centuries of scientific experience lend support to the aforementioned principle of parsimony (Occam's razor). Namely, the supposition of a simpler, more parsimonious chain of events is preferable to the supposition of a more complicated, less parsimonious chain of events. Hence, parsimony (sensu lato) is typically sought in constructing phylogenetic trees, and in scientific explanation generally.[2] However, complications in both actual evolutionary processes and in the methods used to reconstruct them make the science of phylogenetics difficult, relying more and more on advanced genomics and computational and statistical methods to complement traditional field research and morphological methods. In both principle and practice, parsimony helps guide this work.

Maximum likelihood

In statistics, maximum-likelihood estimation (MLE) is a method of estimating the parameters of a statistical model. When applied to a data set and given a statistical model, maximum-likelihood estimation provides estimates for the model's parameters.

The method of maximum likelihood corresponds to many well-known estimation methods in statistics. For example, one may be interested in the heights of adult female penguins, but be unable to measure the height of every single penguin in a population due to cost or time constraints. Assuming that the heights are normally (Gaussian) distributed with some unknown mean and variance, the mean and variance can be estimated with MLE while only knowing the heights of some sample of the overall population. MLE would accomplish this by taking the mean and variance as parameters and finding particular parametric values that make the observed results the most probable (given the model).

In general, for a fixed set of data and underlying statistical model, the method of maximum likelihood selects the set of values of the model parameters that maximizes the likelihood function. Intuitively, this maximizes the "agreement" of the selected model with the observed data, and for discrete random variables it indeed maximizes the probability of the observed data under the resulting distribution. Maximum-likelihood estimation gives a unified approach to estimation, which is well-defined in the case of the normal distribution and many other problems. However, in some complicated problems, difficulties do occur: in such problems, maximum-likelihood estimators are unsuitable or do not exist.

Neighbour joining

In bioinformatics, neighbor joining is a bottom-up clustering method for the creation of phenetic trees (phenograms), created byNaruya Saitou and Masatoshi Nei.[1] Usually used for trees based on DNA or protein sequence data, the algorithm requires knowledge of the distance between each pair of taxa (e.g., species or sequences) to form the tree.

Complexity: Neighbor joining on a set of r taxa requires $r-3$ iterations. At each step one has to build and search a Q matrix. Initially the Q matrix is size $r \times r$, then the next step it is $(r-1) \times (r-1)$, etc. This leads to an algorithm with a time complexity of $O(r^3)$.

ACD/Chemsketch

ACD/ChemSketch is a chemically intelligent drawing interface that allows you to draw almost any chemical structure including organics, organometallics, polymers, and Markush structures. It can be used to produce professional looking structures and diagrams for reports and publications. The 3D optimization is based on modified molecular mechanics which take into account bond stretching, angle bending, internal rotation, and Van der Waals non-bonded interactions.

Discovery Studios

Discovery Studio is a well-known suite of software for simulating small molecule and macromolecule systems. It is developed and distributed by Accelrys, a company that specializes in scientific software products covering computational chemistry, computational biology, cheminformatics, molecular simulations and Quantum Mechanics.

Discovery Studio is typically used in the development of novel therapeutic medicines, including small-molecule drugs, therapeutic antibodies, vaccines, synthetic enzymes, and even in areas such as consumer products. It is used regularly in a range of academic and commercial entities, but is most relevant to Pharmaceutical, Biotech, and consumer goods industries.

Discovery Studio is a client-server software suite, built on the Pipeline Pilot visual programming product from Accelrys and can be run on both Microsoft Windows clients and servers and also Red Hat and Suse Linux clients and servers. The product suite includes both paid-for licensed versions and free visualization client tools.

The product suite has a strong academic collaboration programme, supporting scientific research and makes use of a number of software algorithms developed originally in the scientific community, including CHARMM, MODELLER, DELPHI, ZDOCK, DMol3 and more.

Libdock

LibDock is an algorithm for docking small molecules into an active receptor site. Initially, a hotspot map is calculated for the receptor active site which contains polar and apolar groups. This hotspot map is subsequently used to rigidly align the ligand conformations to form

favorable interactions. After a final energy-minimization step (allowing the ligand poses to be flexible), the top scoring ligand poses are saved.

Cdocker

The Dock Ligands (CDOCKER) protocol is an implementation of the CDOCKER algorithm. It allows you to run a refinement docking of any number of ligands with a single protein receptor.

CDOCKER is a grid-based molecular docking method that employs CHARMm. The receptor is held rigid while the ligands are allowed to flex during the refinement. For pre-docked ligands, prior knowledge of the binding site is not required. It is possible, however, to specify the ligand placement in the active site using a binding site sphere.

Random ligand conformations are generated from the initial ligand structure through high temperature molecular dynamics, followed by random rotations. The random conformations are refined by grid-based (GRID 1) simulated annealing and a final grid-based or full forcefield minimization.

ZincPharmer (http://zincpharmer.csb.pitt.edu/)

ZINCPharmer is free pharmacophore search software for screening the purchasable subset of the ZINC database. ZINCPharmer can importLigandScout and MOE pharmacophore definitions, as well as identify pharmacophore features directly from structure. The ZincPharmer search technology uses the Pharmer open source pharmacophore search technology to efficiently search a large database of fixed conformers for pharmacophore matches. The search performance scales with the complexity and breadth of the pharmacophore query. Most searches take less than a minute.

Zinc database (http://zinc.docking.org/)

Welcome to ZINC, a free database of commercially-available compounds for virtual screening. ZINC contains over 21 million purchasable compounds in ready-to-dock, 3D formats. ZINC is provided by the Shoichet Laboratory in the Department of Pharmaceutical Chemistry at the University of California, San Francisco (UCSF).

CHAPTER 2

METHODOLOGY

DATA CURATION AND EXTRACTION

To ensure complete collection of data, data sets were collected from various resources

- Protein sequences along with other related information (*i.e.* strain name,fasta sequence ,protein length etc) were collected from protein database of NCBI.
- We have collected ICL sequences in different Prokaryotes and Eukaryotes.
- Different research papers were extracted from Pubmed database of NCBI
- Important active molecules were taken from the research papers for pharmacophore modeling
- Other important information related to ICL sequence in different organisms were taken from research papers.
- Role of ICL in bacterial and fungal pathogenesis were taken from research papers.

Comparative analysis

Phylogenetic analysis

ICL sequences present in different species of microorganism were extracted from the Protein database of NCBI. Using these sequences, phylogenetic analyses were carried out.

108 number of sequence of Isocitrate lyase from different species of bacteria, fungi and plants were extracted from NCBI. Comparison between different subfamilies was also done.

CLUSTALW was used for multiple sequence alignment of all the sequences of ICL. Once the alignment was extracted PHYLIP 3.65 tool was used to draw the phylogenetic tree using neighbor joining program. PROTDIST was used to find out distance between different species. Neighbor tool was used to draw the tree and DRAW GRAM and DRAW TREE tools were used to plot rooted and unrooted tree respectively.

Bootstrap analysis was also done using 100 different bootstrap alignments . Using these bootstrap alignments 100 different unrooted trees were plotted and checked for its significance

Sequence analysis

Based on the phylogenetic analysis, ICl proteins in different species were grouped together and multiple sequence alignment was carried out using following three tools:

- CLUSTALW
- MUSCLE
- T-COFEE

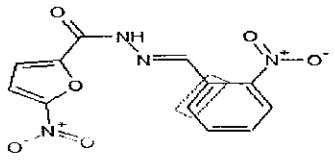
PHARMACOPHORE MODELLING

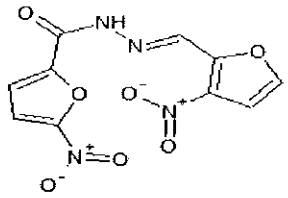
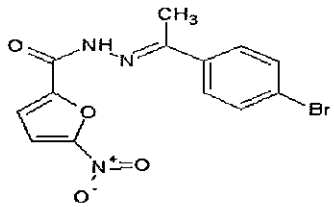
Active molecules

On searching several research papers I developed a library of 274 chemical compounds out of which 69 active compounds were selected based on there IC50 (inhibition concentration), MIC(minimum inhibition concentration) and Inhibition percentage.

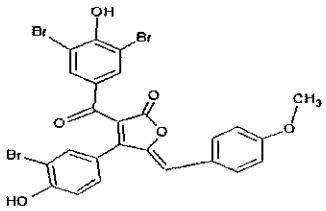
These molecules were built using ChemSketch software and saved in in Mol2 format.

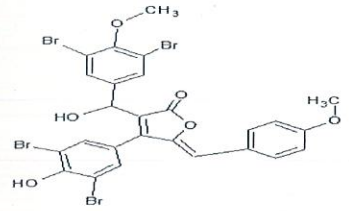
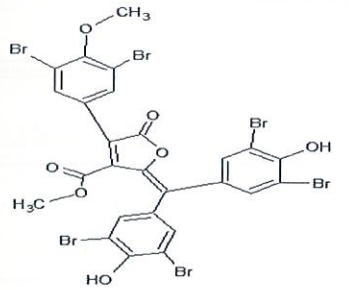
5-Nitro-2-furoic acid hydrazone derivatives[29] (Table 1)

S. No.	Structures	IC50(uM)	MIC(uM)	INHIBITON(%)
1		NA	10.29	48.92

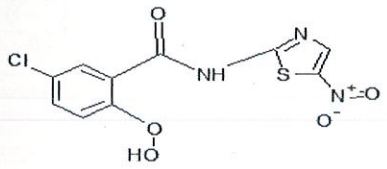
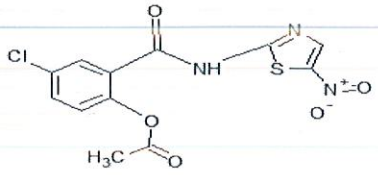
2		NA	2.65	86.8
3		NA	8.88	73.12

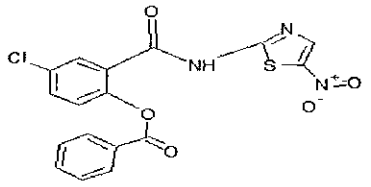
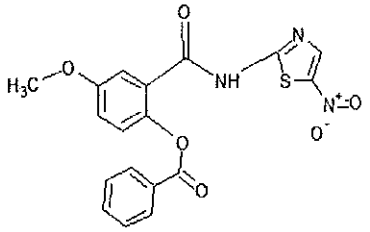
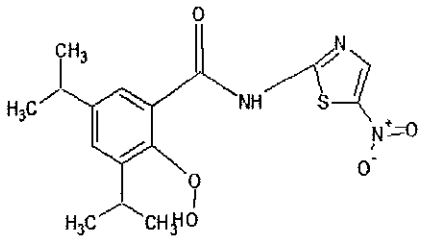
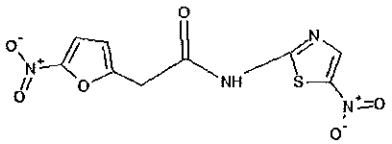
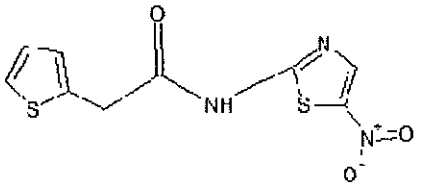
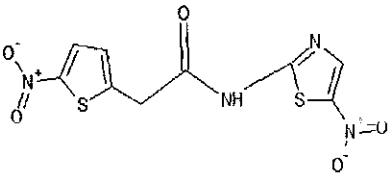
Cadiolides derivatives[18] (Table 2)

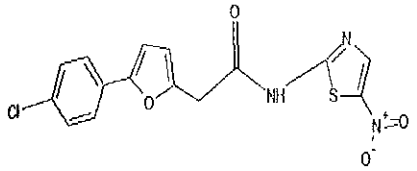
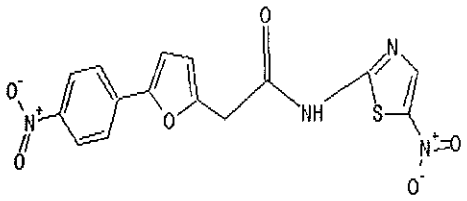
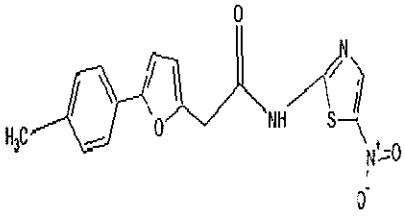
S. No.	Structures	IC50(uM)	MIC(uM)	INHIBITON(%)
1		7.62	12.5	NA

2		17.6	50	NA
3		10.36	50	NT

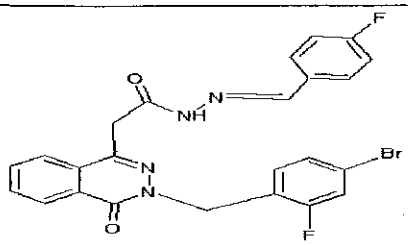
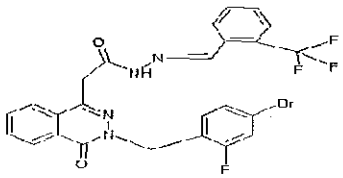
5-Nitro-2,6-dioxohexahydro-4-pyrimidinecarboxamides[30] (Table 3)

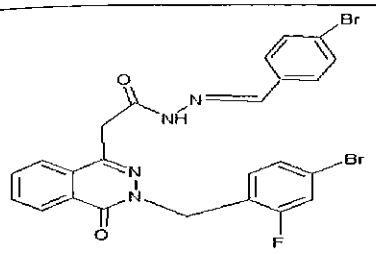
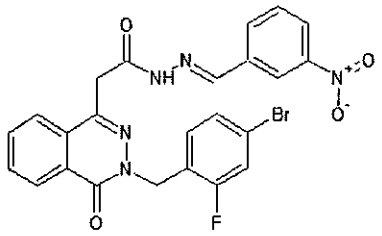
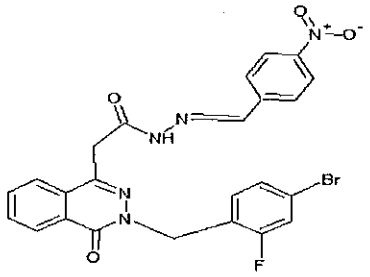
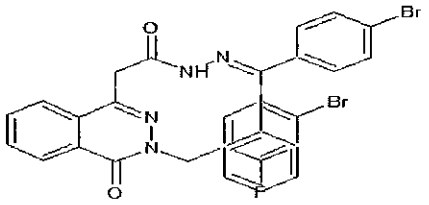
S. No.	Structures	IC50(uM)	MIC(uM)	INHIBITON(%)
1		NA	20.85	NA
2		NT	18.28	NT

3		NA	15.47	NA
4		NA	15.64	NA
5		NA	17.88	NA
6		NA	5.48	NA
7		NA	24.8	NA
8		NA	20.81	NA

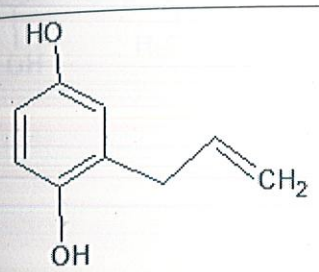
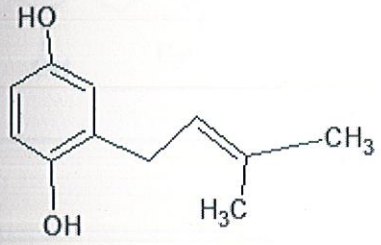
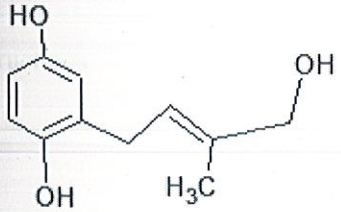
9		NA	17.86	NA
10		NA	17.34	NA
11		NA	18.97	NA

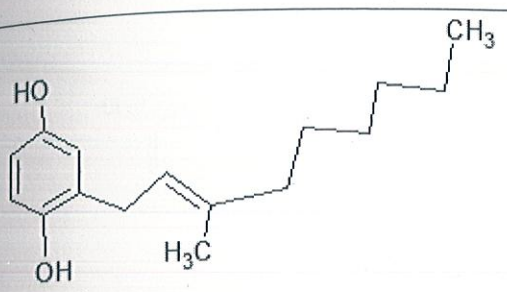
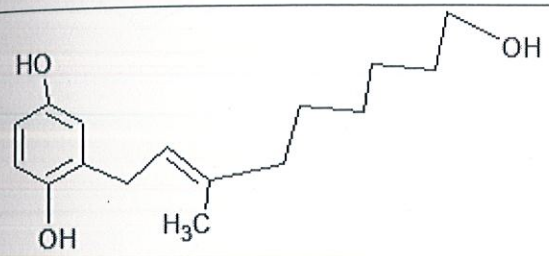
Novel phthalazinyl derivatives[26] (Table 4)

S. No.	Structures	IC50(uM)	MIC(uM)	INHIBITON(%)
1		129.3	0.81	NA
2		117.2	0.36	NA

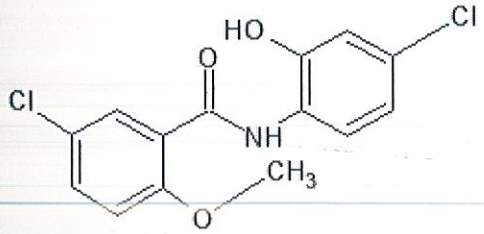
3		114.9	0.72	NA
4		122.5	0.37	NA
5		122.5	0.18	NA
6		115.4	0.35	NA

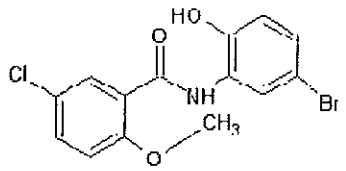
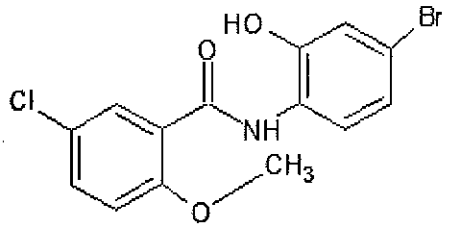
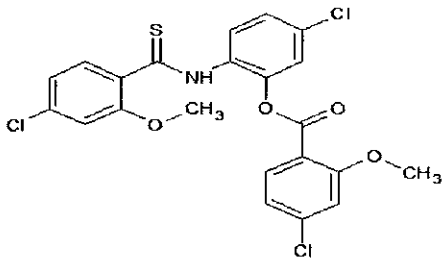
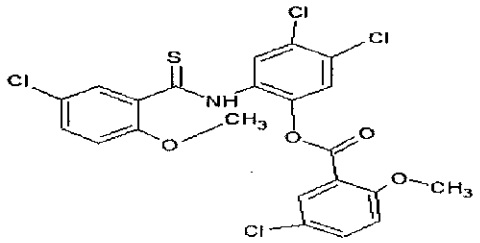
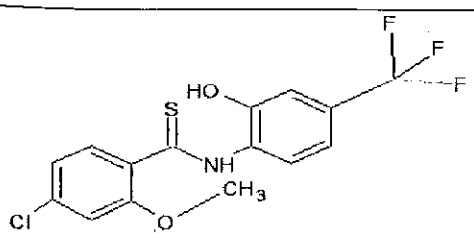
Hydroquinone derivatives[24] (Table 5)

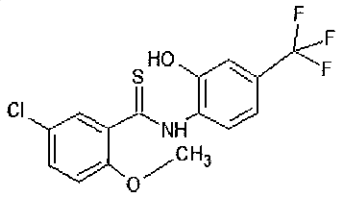
S. No.	Structures	IC50(uM)	MIC(uM)	INHIBITON(%)
1		1020	NA	50
2		610	NA	50
3		290	NA	50

4		400	NA	50
5		280	NA	50

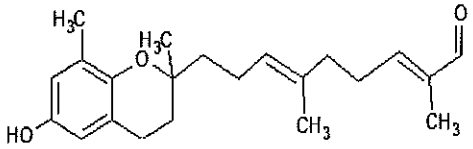
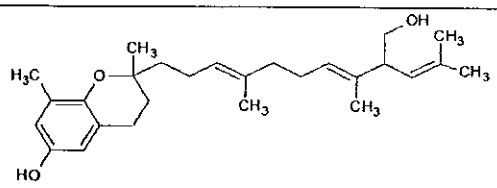
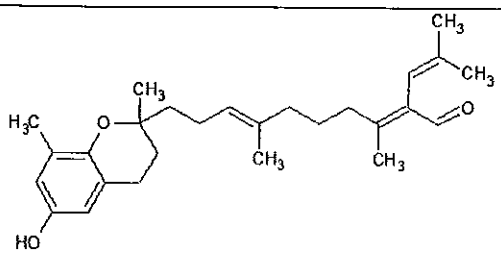
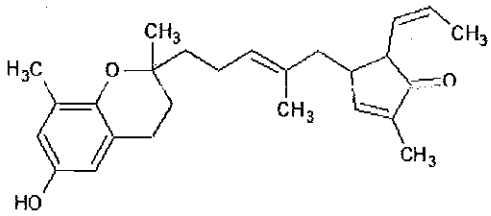
Novel 2-methoxy-2'-hydroxybenzanilides, their thioxo analogues and benzoxazoles[28] (Table 6)

S. No.	Structures	IC50(uM)	MIC(uM)	INHIBITON(%)
1		NA	4	NA
2				

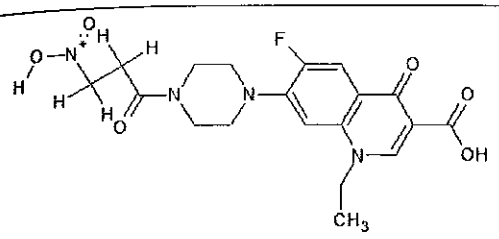
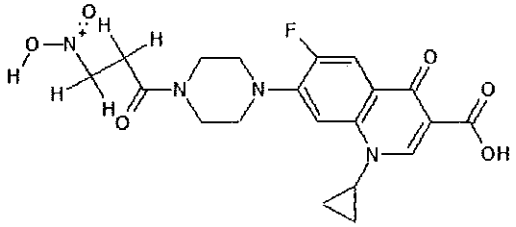
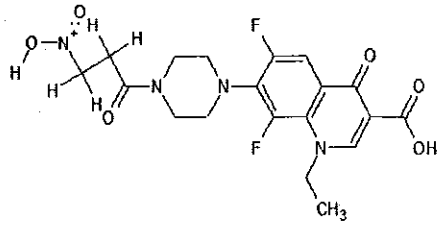
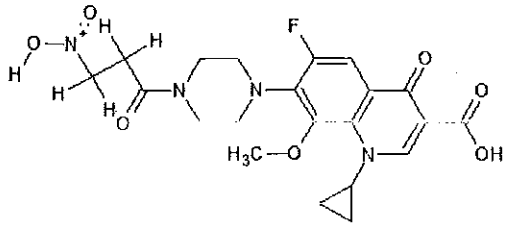
		NA	4	NA
3		NA	8	NA
4		NA	4	NA
5		NA	8	NA
6		NA	32	NA

7		NA	62.5	NA

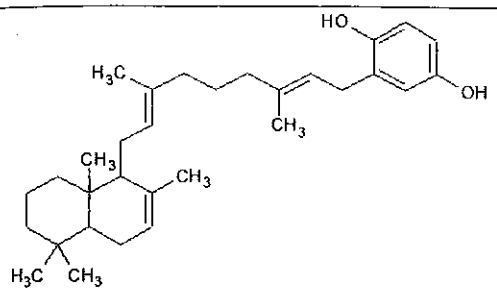
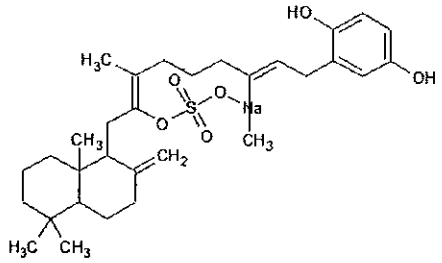
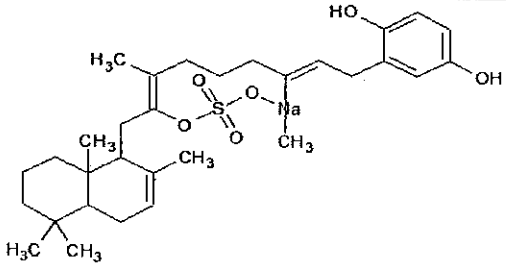
Sargachromanols[22] (Table 7)

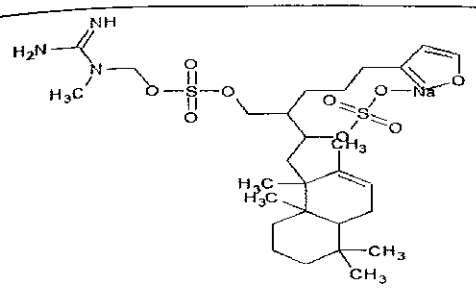
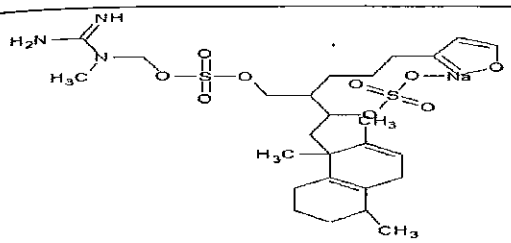
S. No.	Structures	IC50(uM)	MIC(uM)	INHIBITON(%)
1		185	NA	NA
2		118.1	NA	NA
3		172.9	NA	NA
4		141	NA	NA

3-nitropropionamides[21] (Table 8)

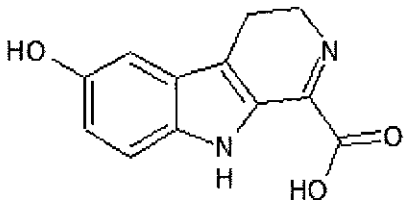
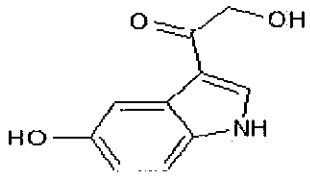
S. No.	Structures	IC50(uM)	MIC(uM)	INHIBITON(%)
1		0.2	1.34	NA
2		0.12	1.39	NA
3		1.46	2.01	NA
4		0.1	3.37	NA

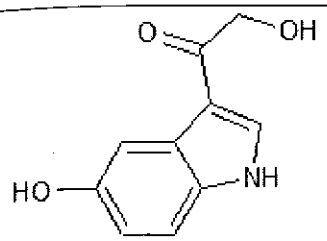
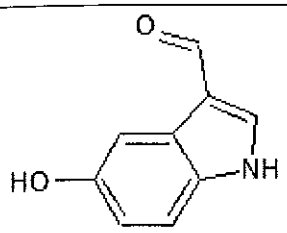
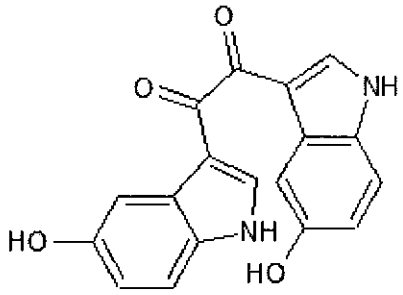
Sesterterpene sulfates[32] (Table 9)

S. No.	Structures	IC50(uM)	MIC(uM)	INHIBITON(%)
1		76.8	222.2	NA
2		71	131.2	NA
3		72	131.2	NA

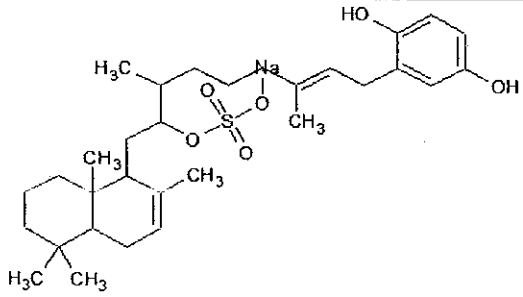
4		33.8	131.2	NA
5		31.3	185.5	NA

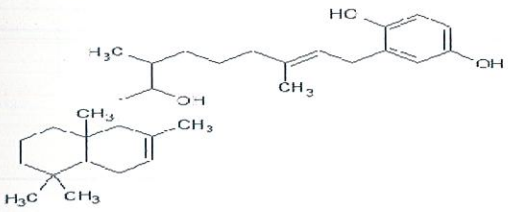
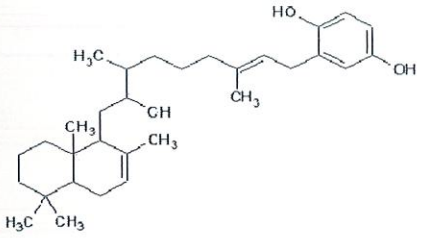
5-Hydroxyindole-type alkaloids[27] (Table 10)

S. No.	Structures	IC50(uM)	MIC(uM)	INHIBITON(%)
1		37.6	185.5	NA
2		30.3	113.19	NA

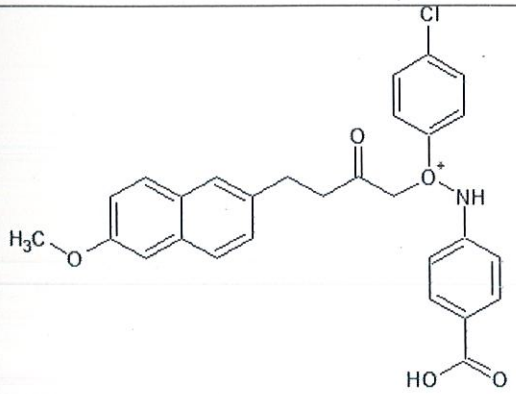
3		31.8	104.67	NA
4		21.7	102.1	NA
5		89	62.1	NA

Halisulfate[25] (Table 11)

S. No.	Structures	IC50(uM)	MIC(uM)	INHIBITON(%)
1		12.6	87	NA

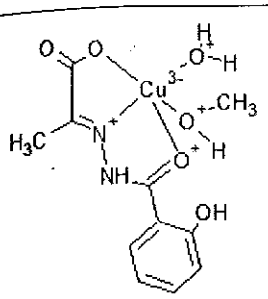
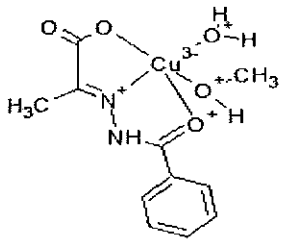
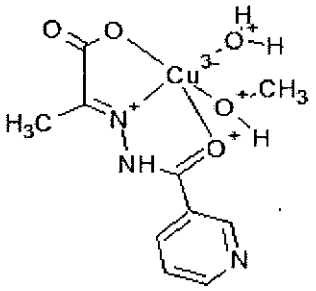
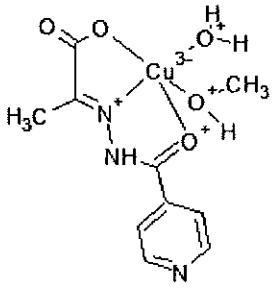
2		67.4	106.3	NA
3		15	105.2	NA

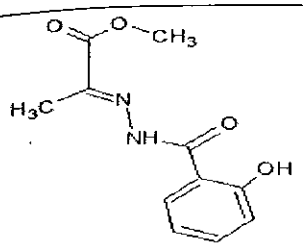
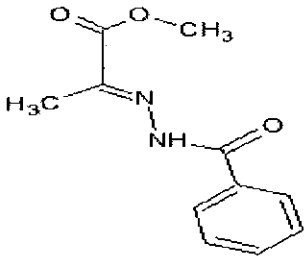
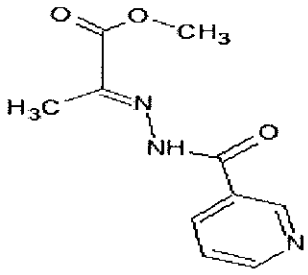
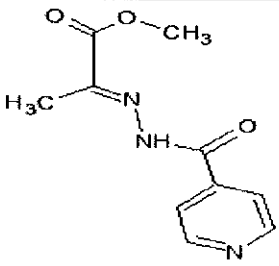
Mannich Base[20] (Table 12)

S. No.	Structures	IC50(uM)	MIC(uM)	INHIBITON(%)
1		108.89	NA	52.6

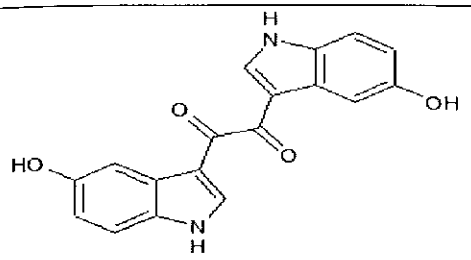
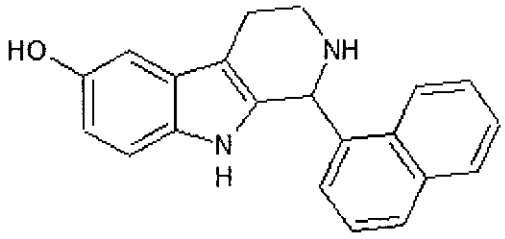
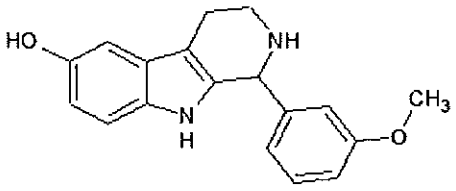
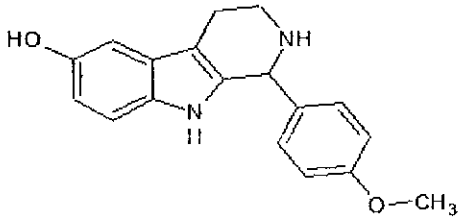
Pyruvate-isoniazid analogs and their copper complexes[19] (Table 13)

S. No.	Structures	IC50(uM)	MIC(uM)	INHIBITION(%)

1		NA	NA	50
2		NA	NA	50
3		NA	NA	50
4		NA	NA	50

5		NA	NA	50
6		NA	NA	50
7		NA	NA	50
8		NA	NA	50

Indole-containing natural products [23] (Table 14)

S. No.	Structures	IC50(uM)	MIC(uM)	INHIBITON (%)
1		89	6.5	NA
2		137	12.5	NA
3		75	6.25	NA
4		80	12.5	NA

PHARMACOPHORE MODELLING USING AUTO DOCK

Using these active molecules I developed a pharmacophore using autodock feature of Discovery Studios 3.5 client.

LINKING WITH ZINCPHARMER

Once the pharmacophore was developed, then it was linked with **ZINCPharmer** to find the molecules containing this pharmacophore in ZINC database. All the ZINC molecules related to the pharmacophore were extracted. These molecules were prepared using **PREPARE LIGAND** tool of discovery studios present in prepare ligand tool of **SMALL MOLECULES** .

DOCKING USING DISCOVERY STUDIOS 3.5

Different ICL structures from different organisms were taken from PDB

Structures→

- 1F61(**Fig 5**)→Apoenzyme structure of ICL in *M. tuberculosis*
- 1F8I(**Fig 6**)→Closed structure of ICL in *M. tuberculosis*
- 1DQU(**Fig 7**)→Structure of ICL in *Aspergillus nidulans*
- Modelled Structure(ACEAa+ ACEAb) of ICL(**Fig 8**)

All the above proteins structures were prepared using **PREPARE PROTEIN** tool of Discovery studios present in **MACROMOLECULES** tool.

Active sites in these structures were recognized

- 1F61, 1F8I,Modelled structure (**Fig5,6,8**)→**His193,Asn313,Ser315,ser317and Thr347[34]**
- IDQU (**Fig 7**)→ **Asp114, Asp168, Asp170 and Glu197 [33]**

Once the active sites were recognized, these active amino acid residues were selected one by one to develop the binding site of these proteins using **Define and Edit Binding Site tool** present in **Receptor-Ligand Interaction** toolbar.

DOCKING

Now these prepared ligands are used to dock on the prepared proteins using **LibDOCK** with default parameters.

Once the docking is complete for all the ICL structures then results were viewed and all the ligands have lipdock score of above 110 were taken. Now these ligands are analyzed to check whether they are fitting properly or not.

Once these ligands were separated, the selected ligands were taken for **CDOCKER** analysis.

CHAPTER 3

RESULTS AND DISCUSSIONS

Phylogenetic analysis

Unrooted tree

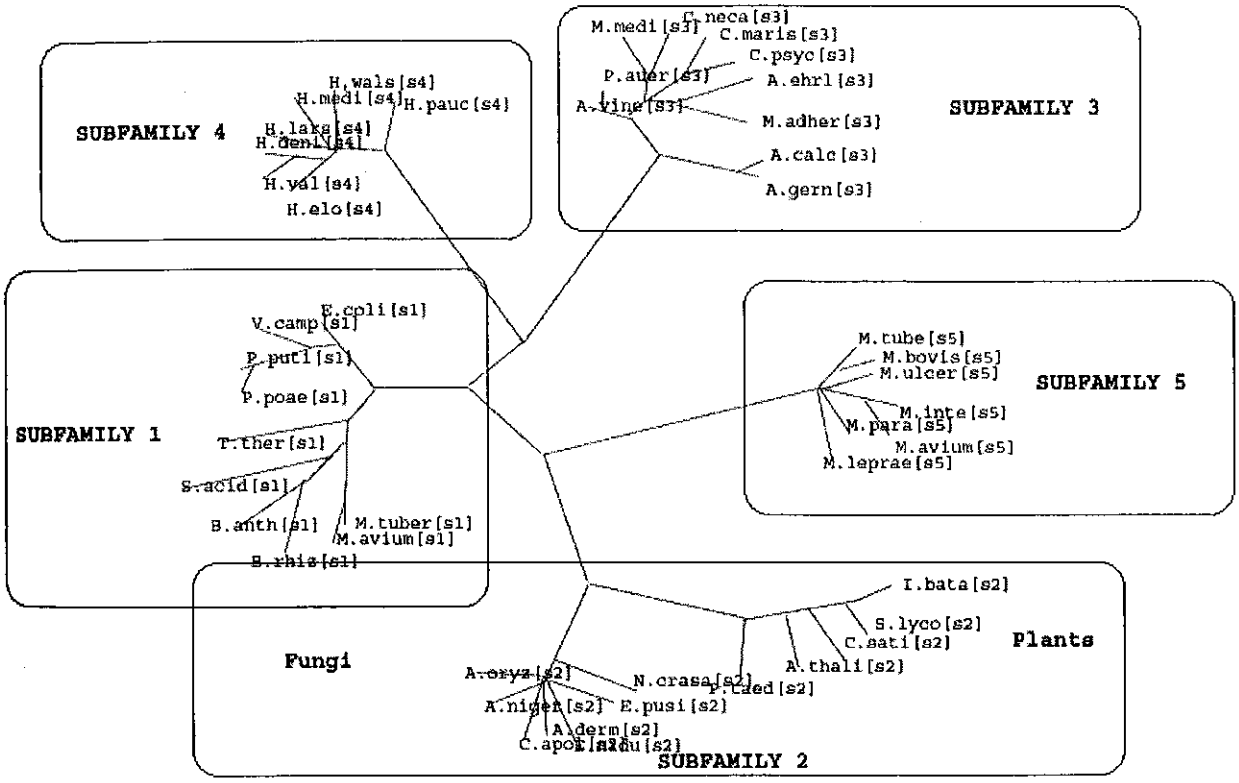


Figure 1- Unrooted phylogenetic tree of ICL subfamilies

This tree showed a significance of 78.3% using bootstrap analysis of 100 different bootstrap alignments.

Rooted tree

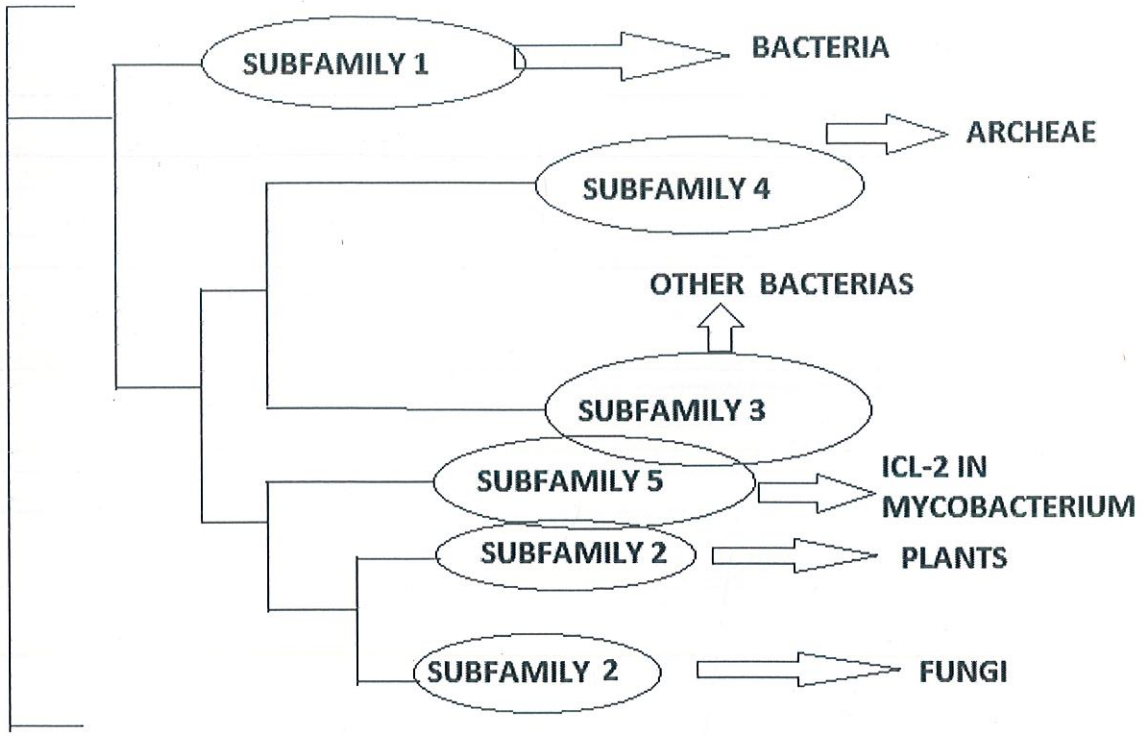


Figure 2- Rooted phylogenetic tree of ICL subfamilies

Phylogenetic analysis showed that ICL family is divided into 5 subfamilies [36]. Members of subfamily 1 are ICL proteins, with a size range of 350–450 aa, primarily present in prokaryotes, except for those from nematodes [36]. Subfamily 2 whose members are larger than 500 aa, contains only members from the Fungi and Plantae kingdoms; this group currently includes isoforms that primarily display MICL activity [36]. Subfamily 3 contains ICL proteins that are larger than 530 aa and are commonly present in prokaryotes [36]. Although the *P. aeruginosa* ICL is in this group, all other ICLs from the genus Pseudomonas are grouped in subfamily 1 except for ICL homologues from *Pseudomonas stutzeri* and *Pseudomonas mendocina* [36]. Two recently described subfamilies agree with our findings, including subfamily 4, which includes enzymes from Archaea, and subfamily 5, which includes proteins larger than 700 aa; these proteins are present in prokaryotic sources, mainly in the genus Mycobacterium [36].

Sequence analysis

Subfamily 1

66 sequences of bacteria were taken and MSA was carried out using CLUSTALW. MSA showed that sequences are highly conserved and may have been evolved from same ancestors.

```
M.tuber      ------------------------------------MSVVGTPKSAEQIQQEW
M.Bovis      ------------------------------------MSVVGTPKSAEQIQQEW
M.smegmat    ------------------------------------MSTVGTPKSPEQIQHDW
M.vanbaal    ------------------------------------MSTVGTPKSPEQIQHDW
M.avium      ------------------------------------MPTSVVGTPKSAEEIQKDW
G.bron       ------------------------------------MSNVGKPRTAAEIQQDW
Arthroba     ------------------------------MTAAFEPTQQTPEQQAAALELEW
S.nass       ------------------------------------MSKPTDLRTRTIEQLELDW
X.cellu      ------------------MTALTDRSASPTAPGVRAGDTVQTAAELAAAW
N.dass       ------------------------------------MGTSARRQEASAALQRDW
A.mirum      ------------------------------------MLSAEQIADQW
T.curvata    ------------------------------------MTDSRLKGAAEELQRQW
C.flav       ----------MTAPTTQIDDLQGTPAAVAGAVPARPGDQTRTAEQLEHDW
M.cilva      ------------------------------------MNLTPEMKLEAEKLRKEW
M.rube       ------------------------------------MSKLTPEMKLEAEKLRREW
T.osha       ------------------------------MEPLKVLSPEMQKEAEALKKEW
T.ther       ------------------------------------MEGLTPEMQREAEALRREW
M.hydr       ------------------------------------MQEAQALARSW
Roseiflex    ------------------------------------MSAVNRDQEIAELERRW
K.race       ------------------------------------MMSSQHAENGNARDLQMAW
K.tusc       ------------------------------------MIDKQEAEALERSW
B.anth       ------------------------------------MKNERIEKLQESW
A.acidoca    ------------------------------------MTYFKTAAELEQHW
S.acid       ------------------------------------MTELNDRIRMIEEQ
S.melilot    ------------------------------------MTDFYKLVPN
S.medicae    ------------------------------------MTDFYKLVPN
B.melit      ------------------------------------MTDFYSLIPS
R.legumin    ------------------------------------MTDFYKLVPN
P.lava       ------------------------------------MSKKTFYDLVPG
Sphing       ------------------------------------MVDFTDLVA
A.exce       ------------------------------------MTTFEHLVPN
M.fulvus     ------------------------------MYDATPTTADASPHAK
D.geoth      ------------------------------MTPADPRLHSPRTHAEILEKTW
B.ambifa     ------------------------------------MSRHQQAQELQQQW
B.vietnam    ------------------------------------MSRQQQALELQKQW
B.cenoce     ------------------------------------MSRQQQAQELQQQW
B.multivo    ------------------------------------MSRQQQAQELQKQW
B.psue       ------------------------------------MSRQQQAQELQKQW
B.thailan    ------------------------------------MSRQQQVQELQKQW
B.rhizox     ------------------------------------MSRQQEAQKLQVQW
B.phytofi    ------------------------------------MSRQEQVKQLQQQW
B.phymatu    ------------------------------------MTRNEQAKQLQQQW
R.pickett    ------------------------------------MSRELEVQKLQKEW
S.maltoph    ------------------------------------MSKLPTAEQIQHDW
X.axonopo    ------------------------------MSTTLQSAEQLQQDW
P.sownens    ------------------------------------MSTLKTAEQIQHEW
S.ente       ------------------------------------MKTRTQQIEELQKEW
E.lign       ------------------------------------MKTRTQQIEELQKEW
E.coli       ------------------------------------MKTRTQQIEELQKEW
Klebsi       ------------------------------MEHLHMKTRTQQIEELNKEW
D.dandati    ------------------------------MSITSRTQQIRQLEQEW
S.plym       ------------------------------MSTSRTQQIQQLEQEW
V.para       ------------------------------MTLTRRQQIEALEKDW
V.campbel    ------------------------------MTNLTRRQQIEALEKDW
p.lieog      ------------------------------MTLTRQQQIEAIEKDW
Shewanella   ------------------------------MTKATQTSRQAQIDAIKKDW
A.hydroph    ------------------------------MALTREQQIQAIEKDW
D.acidovo    MRSSIRLKFDPKADARKPLVSTTSQESNMPQSLTQQLSREQQISALEKDW
C.testo      ------------------------------MPQSLTQQLSREQQIAALEKDW
A.citrul     ------------------------------MPQTFNEQLSREQQIAALEKDW
V.paradox    ------------------------------MPQTLTEQLSREQQIAALEKEW
C.accu       ------------------------------MSTREQQIAELEKDW
P.necessa    ------------------------------MADRKAEIAALQKDW
P.poae       ------------------------------MALTREQQIAALEKDW
P.putida     ------------------------------MALTREQQIAALEKDW
I.loihien    ------------------------------MTTQNLSFAQQVKELQKQW
```

```
M.tuber        DTNPRWKDVTRTYSAEDVVALQGSVVEEHTLARRGAEVLWEQLH------
M.Bovis        DTNPRWKDVTRTYSAEDVVALQGSVVEEHTLARRGAEVLWEQLH------
M.smegmat      DHNPRWKGIKRDYTPEDVVALQGTVVEEHTLARRGAEVLWEQLH------
M.vanbaal      DHNPRWKGITRTYTPADVVALQGHVVEEHTLARRGAEVLWDQLH------
M.avium        DTNPRWKDVTRTYTAEDVVALQGTVVEENTLARRGAEVLWEQLH------
G.bron         DTNPRWDNVTRDYTAEQVAQLQGSVVEEHTLARRGAEILWEGVNQK----
Arthroba       AANPRWEGVTRDYKASDVVRLRGRVSEEHTLARRGSEKLWKQLTEEHK--
S.nass         RTDPRWQNVTRDYTAEQVVKLRGRVQEEHTLARRGAEKLWAQLTGEAA--
X.cellu        ATDPRWTDVERTYTADDVVALRGSVIEEHTLARRGAELLWERLHTS----
N.dass         ETNPRWAGIERTYSADDVVKLRGSVTEEHTLARRGAERLWDLLR------
A.mirum        ATDARWSGIERTYGAEDVVALRGSVVEEHTLARRGAERLWELLH------
T.curvata      DTDPRWKGIKRDYTAEDVVRLRGSIQEEFTLARLGAERLWKLLH------
C.flav         ATDPRWSGIERRHTGADVVRLRGSVREDPTLARRGAQRLWELLHTR----
M.cilva        QTDPRWQGIKRDYTAEEVVRLRPSLQIEHTLAQKGAERLWELLHQR----
M.rube         ETNPRWKGIKRDYTAEDVVRLRPSLLPEQTLAKAGAERLWELLHTR----
T.osha         AENPRWKGVRRDYTAEDVVRLRPSVQVEYTLAKRGAEKLWQLLHER----
T.ther         ETNPRWKGVRRDYRPEDVVRLRPSVMVEYTLAKRGAEKLWKLLHER----
M.hydr         REDARWQGIRRTYTAEEVVRLRPSIQVEYTLARRGAERLWELMQTR----
Roseiflex      ATDPRWQGIRRDYSAADVVRLRGTLKIEYTLANVGARRLWDLLQTE----
K.race         QGETRWRGVERPYLAEDVLRLRGSVHIEYTLARMGAKRLWTLLRSE----
K.tusc         ETDPRWKGIVRTYSAEDVLRLRGSIRIEHTLARMGAERLWHLLHTE----
B.anth         ELDTRWKGITRPYSAEDVIRLRGSIDIEHTLARRGAEKLWASLHTE----
A.acidoca      KTDPRWKGVERAYTAEDVIRLRGSVHIEHTLAQMGAEKLWRGLHED----
S.acid         WRTPRFEGITRGYGAADVAKLQGSVMIEHTLARRGAERLWRLLHER----
S.melilot      APEGRFDGIERPYSAEDVKRLRGSVEIRYSLAEMGANRLWKLIHEE----
S.medicae      APEGRFDGIERPYSAEDVKRLRGSVEIRYSLAEMGANRLWKLIHEE----
B.melit        APKGRFDGIERAHTAEDVKRLRGSVEIKYSLAEMGANRLWKLIHEE----
R.legumin      APTGRFDGIDRPYSADDVRRLRGSVALTHTLAEMGADRLWRLLRQE----
P.lava         APKGRFDNVKRNYSVEDVEKLRGSFPIEYTLATRGANKLWELLHTT----
Sphing         APEGRFDGIRRTYSPADVARLRGSVAIEHTLARRGAEKLWASLRSE----
A.exce         APKGRFNGITRPYTPQEVEKLRGSVWVRNTLAERGANRLWQLLHEE----
M.fulvus       LHAQRFEGIKRNYTQKDVEKLRGSVTISYTLAELGAKKLWELLHTE----
D.geoth        KTEERWQGIRRNYSADEVVRLRGSLPIEHTLARHGAQKLWRLMKEE----
B.ambifa       ETDPRWKGIKRSYTADDVVRLRGSIAVEHTLAKRGAQKLWSLMNEE----
B.vietnam      DTDPRWTGIKRGYTADDVVRLRGSIAVEHTLAKRGAEKLWRLINEE----
B.cenoce       ETDPRWKGIKRSYLAEDVVRLRGSIPIEHTLAKRGAEKLWGLINNE----
B.multivo      ETDPRWKGIKRSYTAEDVVRLRGSIQVEHTLAKRGAEKLWELINNE----
B.psue         ETDPRWKGIKRAFTAEDVVRLRGSIQQEHTLAKRGAEKLWTLINNE----
B.thailan      ETNPRWKGIKRAFTAEDVVRLRGSVQQEHTLAKRGAEKLWTLVNNE----
B.rhizox       ETNPRWKGIKRGYTAEDVIRLRGSVPVEHTLARRGAEKLWMLINEE----
B.phytofi      ETDPRWKGVKRTYTAEDVIRLRGSVQVEHTLAKRGAEKLWESVNNE----
B.phymatu      ESDSRWKGIKRGFTAEDVVRLRGSVQPEHTLARRGAEKLWSDMHQE----
R.pickett      DTNPRWKGIKRGYTAEDVVRLRGSLQIEHTLAKRGAEKLWNLINNE----
S.maltoph      DTNPRWEGIQRNYSAADVVRLRGTVHIEHSLARLGAEKLWASLHER----
X.axonopo      SNNPRWAGIARNYSAADVARLRGTVHVEHSLARLGAEKLWNSLHAT----
P.sownens      NTDPRWAGITRNYTAEDVVRLRGTIPVDHSIARITSGKLWNYLQDL----
S.ente         T-QPRWEGITRPYSAEEVVKLRGSVNPECTLAQLGAAKMWRLLHGEAK--
E.lign         T-QPRWEGIRRPYSAEEVVKLRGSVNPECTLAQMGAAKMWRLLHGEAK--
E.coli         T-QPRWEGITRPYSAEDVVKLRGSVNPECTLAQLGAAKMWRLLHGESK--
Klebsi         T-NPRWEGITRPYSAEEVVKLRGSVNPECTLAQLGAAKMWRLLHGEAK--
D.dandati      K-TPRWEGIVRPYRTEDVINLRGSVNPACTLAQRGAEKLWALLHGESR--
S.plym         K-SARWEGITRPYSAEDVINLRGSVNPVCTLAQNGAAKLWDLLNGKSR--
V.para         ATNPRWKNVKRTYTAEEVVELRGSMVPANTIAQRGADKLWSLVNGSAK--
V.campbel      ATNPRWKNVKRTYTAEEVVELRGSMVPANTIAQRGADKLWSLVNGSSK--
p.lieog        AENPRWKHVKRTYSAEEVINLRGSFAPANTIAQRGADKLWQLVNGEAK--
Shewanella     AENPRWKNVRRPYTAEEVVALRGSIVPENTIAKRGAAKLWDLVNGGAK--
A.hydroph      AENPRWKGIKRGYSAEDVVNLRGSLQPVHTLAQRGADKLWELIHGGAK--
D.acidovo      AQNPRWKGVKRGYSAADVVRLRGSLQPEYTLAQRGAELLWDRINGGAK--
C.testo        AQNPRWKSVKRGYSAADVVRLRGSLQPEYTLAQRGAEVLWDKINGSSK--
A.citrul       AQNPRWKGIKRGYGAADVVRLRGSFQVEHTLARRGAEKLWNLVHNEP---
V.paradox      ATNPRWKGIKRGYSAADVVRLRGSFPIEHTLARRGAEKLWELVNNEP---
C.accu         AENPRWKGIKRGYSAADVVRLRGSFPIEYTVARRGAEKLWNLVNSEP---
P.necessa      DTNPRWKGITRGYTAEDVVRLRGSLKIEHTLAKHGAERLWELVNNEA---
P.poae         AENPRWKGVTRAYSAADVVRLRGSVQPEHTFAKLGAEKLWNLVTQGAKPS
P.putida       AENPRWKGVTRTYTAADVVRLRGSLQPEHTLARQGAEKLWKLVTQGAHPS
I.loihien      DTDPRWKGVERPYSAEDVVRLRGSVQPEYTYARNGADKLWQLIHGKAQEK
                   *:  : * .   :*  *:  .    : *   :  :*  :
```

```
M.tuber        ---DLEWVNALGALTGNMAVQQVRAGLKAIYLSGWQVAGDANLSGHTYPD
M.Bovis        ---DLEWVNALGALTGNMAVQQVRAGLKAIYLSGWQVAGDANLSGHTYPD
M.smegmat      ---DMDFVNALGALTGNMAVQQVRAGLKAIYLSGWQVAGDANLSGHTYPD
M.vanbaal      ---DMDFVNALGALTGNMAVQQVRAGLKAIYLSGWQVAGDANLSGHTYPD
M.avium        ---DLEYINALGALTGNMAVQQVRAGLKAIYLSGWQVAGDANLSGHTYPD
G.bron         ---DGSYINALGALTGNQAVQQVRAGLKAVYLSGWQVAGDANLSGHTYPD
Arthroba       ---TGGYTNALGALTGNQAVQQVKAGLRAIYLSGWQVAADANNSGHTYPD
S.nass         ---SGGYTNSLGALTGNQAVQQVRAGLRAIYLSGWQVAADANLAGQTYPD
X.cellu        -----EFVNALGALTGNQAVQQVKAGLRAIYLSGWQVAGDANAAGQTYPD
N.dass         ---TEDYVNTLGALTGNQAVQQVRAGLKAIYLSGWQVAADANLSGNTYPD
A.mirum        ---DEDHVHALGALTGNQAVQQVRAGLKAIYLSGWQVAADANLSGQTYPD
T.curvata      ---EEDYINALGALTGNQAVQQVKAGLKAIYLSGWQVAGDANLALQTYPD
C.flav         -----THVPALGALTGNQAVQQVRAGLEAIYLSGWQVAADANLSGQTYPD
M.cilva        -----PYINTFGAYTGAQAVQMVKAGLEAIYLSGWQVAADANLAAQTYPD
M.rube         -----PYVNTFGAYTGAQAVQMVKAGLEAIYLSGWQVAADANLAWQTYPD
T.osha         -----PYVHTFGAYTGAMAVEMVRAGLEAIYLSGWQVAADANLAWQTYPD
T.ther         -----PYVHTFGAYTGAMAVEMVRAGLEAIYLSGWQVAADANLAWQTYPD
M.hydr         -----PYVHTFGALTGAQAVQMVKAGLEAIYVSGWQVAADANLAHQTYPD
Roseiflex      -----PYVATFGALTGAQATQMVRAGIKAIYMSGWQVAADANLAGQTYPD
K.race         -----PFVPALGALTGNQAVQQVRAGLKAIYLSGWQVAADANLAGHMYPD
K.tusc         -----PYVPALGALTGNQAVQQVKAGLKAIYLSGWQVAADANLAGQMYPD
B.anth         -----DYINALGALTGNQAMQQVKAGLKAIYLSGWQVAADANLSGHMYPD
A.acidoca      -----GFIRALGALTGNQAVQQVKAGLKAIYVSGWQVAADANLAEHMYPD
S.acid         -----HFVPALGALTGNQAVQQVKAGLEAIYLSGWQVAADANLAGQTYPD
S.melilot      -----DFVNALGALSGNQAMQMVRAGLKAIYLSGWQVAADANTASAMYPD
S.medicae      -----DFVNALGALSGNQAMQMVRAGLKAIYLSGWQVAADANTASAMYPD
B.melit        -----DFVNALGALSGNQAMQMVRAGLKAIYLSGWQVAADANTASAMYPD
R.legumin      -----DFVNALGALSGNQAMQMVRAGLKAIYLSGWQVAADANTASAMYPD
P.lava         -----DYVHSLGALSGNQAMQMARAGLKAIYLSGWQVAADANTAGAMYPD
Sphing         -----PYINALGALTGNQAMQMVRAGLKAIYLSGWQVAADANTAGAMYPD
A.exce         -----PFINALGAVTGNQAMQMVRAGLKAIYLSGWQVAADANTASAMYPD
M.fulvus       -----DYINALGSLTGNQAVQMVRAGLKAIYLSGWQVAADANSAGQMYPD
D.geoth        -----PFVNALGALTGNQAVQQVKAGLKAIYLSGWQVAADANNAGQMYPD
B.ambifa       -----PFVNALGALTGNQAMQQVKAGLKAIYLSGWQVAGDANVAGEMYPD
B.vietnam      -----PFVNALGALTGNQAMQQVKAGLKAIYLSGWQVAGDANVAGEMYPD
B.cenoce       -----PFVNALGALTGNQAMQQVKAGLKAIYLSGWQVAGDANVAGEMYPD
B.multivo      -----PFVNALGALTGNQAMQQVKAGLKAIYLSGWQVAGDANVAGEMYPD
B.psue         -----PFVNALGALTGNQAMQQVKAGLKAIYLSGWQVAGDANVAGEMYPD
B.thailan      -----PFVNALGALTGNQAMQQVKAGLKAIYLSGWQVAGDANVAGEMYPD
B.rhizox       -----PFVNALGALTGNQAMQQVKAGLKAIYLSGWQVAGDANVAGEMYPD
B.phytofi      -----PFVNSLGALTGNQAMQQVKAGLKAIYLSGWQVAGDANVAGEMYPD
B.phymatu      -----PFVNALGALTGNQAMQQVKAGLKAIYLSGWQVAGDANVAGEMYPD
R.pickett      -----PFVNALGALTGNQAMQQVKAGLKAIYLSGWQVAGDANSNGEMYPD
S.maltoph      -----EFVNALGALTGNQAMQQVKAGLKAIYLSGWQVAADANLAGQMYPD
X.axonopo      -----PFVNALGALTGNQAMQQIKAGLKAIYLSGWQVAADANLAGQMYPD
P.sownens      -----DFVNALGALTGNQAMQQVKAGLKAIYLSGWQVAADANLAGQMYPD
S.ente         ----KGYINSLGALTGGQALQQAKAGIEAIYLSGWQVAADANLASSMYPD
E.lign         ----KGYINSLGALTGGQALQQAKAGIEAVYLSGWQVAADANLASSMYPD
E.coli         ----KGYINSLGALTGGQALQQAKAGIEAVYLSGWQVAADANLAASMYPD
Klebsi         ----KGYVNSLGALTGGQALQQAKAGIEAIYLSGWQVAADANLASSMYPD
D.dandati      ----KGYINCLGALTGGQALQQAKAGLEAVYLSGWQVAADANLAANMYPD
S.plym         ----KGYVNCLGALTGGQALQQAKAGVEAIYLSGWQVAADANTATAMYPD
V.para         ----KGYVNCLGALTGGQAVQQAKAGIEAIYLSGWQVAADNNTASTMYPD
V.campbel      ----KGYVNCLGALTGGQAVQQAKAGIEAIYLSGWQVAADNNTASTMYPD
p.lieog        ----KGYVNCLGALTGGQAVQQAKAGIEAIYLSGWQVAADNNTASTMYPD
Shewanella     ----KGYVNSLGALTGGQAVQQAKAGIEAIYLSGWQVAADANLAGTMYPD
A.hydroph      ----KGYVNCLGALTGGQAVQQAKAGIEAIYLSGWQVAADNNLSSTMYPD
D.acidovo      ----KGYVNAFGAISAGQAMQQAKAGLEAVYLSGWQVAADGNTSETMYPD
C.testo        ----KGYVNAFGAISAGQAMQQAKAGLEAVYLSGWQVAADGNTSETMYPD
A.citrul       ------YVNCLGALTGGQAMQQVKAGIKAIYLSGWQVAADNNEYAAMYPD
V.paradox      ------YVNCLGALTGGQAMQQVKAGVKAIYLSGWQVAADANTYSSMYPD
C.accu         ------YVNCLGALTGGQAMQQVKAGVKAIYLSGWQVAADNNSYAAMYPD
P.necessa      ------YVNCLGALTGGQAMQQVKAGVQAIYLSGWQVAADGNSYAAMYPD
P.poae         FRPDKDFVNCMGALTGGQAVQQVKAGIQAIYLSGWQVAADNNSAESMYPD
P.putida       FRPEKDFVNCMGALTGGQAVQQVKAGIQAIYLSGWQVAADNNSAESMYPD
I.loihien      FG--KDYVNALGALTGGQAVQQVKAGLQAIYLSGWQVAADNNSACSMYPD
                     :*: :.   * :  :**:.*:*:******.* *     ***
```

```
M.tuber        QSLYPANSVPQVVRRINNALQRADQIAKIEGD-T---SVENWLAPIVADG
M.Bovis        QSLYPANSVPQVVRRINNALQRADQIAKIEGD-T---SVENWLAPIVADG
M.smegmat      QSLYPANSVPQVVRRINNALLRADEIAKVEGD-T---SVENWLAPIVADG
M.vanbaal      QSLYPANSVPQVVRRINNALLRADEIAKVEGD-T---SVENWLAPIVADG
M.avium        QSLYPANSVPTVVRRINNALLRADQIAKVEGD-T---SVKNWLAPIVADG
G.bron         QSLYPANSVPNVVRRINNALLRADEISRVEGD-D---SVDNWLVPIVADG
Arthroba       QSLYPANSVPTVVRRINNALLRADQIEFSEGI-Q---TVEDWMVPIVADA
S.nass         QSLYPANSVPQVVRRINNALLRADQIEFAEGA-S---SVEDWLVPIVADA
X.cellu        QSLYPANSVPQVVRRINNALLRADQVERLEGT-R---TVEDWLAPIVADA
N.dass         QSLYPANSVPAVVRRINNALMRADQITWAEGD-D---SAPEWLAPIVADA
A.mirum        QSLYPANSVPQVVRRINNALLRADQVEHAEGR-G---GEREWLAPIVADA
T.curvata      QSIYPANSVPTVVRRINNALLRADQIQWSEGK-D---D-IYWLAPIVADA
C.flav         QSLYPANSVPAVVRRINNALMRADQIDVAENG-R---ATREWLAPIVADA
M.cilva        QSLYPANSVPQVVRRINNALMRADQIERAE-G-V---SGRYWYAPIVADA
M.rube         QSLYPANSVPQVVRRINNALQRADQIERAE-G-K---TDRYWYAPIVADA
T.osha         QSLYPANSVPQIVKRINNALMRADMIERSE-G-K---VTRDWYVPIVADA
T.ther         QSLYPYNSVPQIVKRINNALMRADQVERIE-G-R---VTRDWYVPIVADA
M.hydr         QSLYPANSVPTLVRRINNALMRADQLHRAE-G-Q---SEIDWYAPIVADA
Roseiflex      QSLYPSNSVPALVRRINNALMRADQIHASE-G-H---NDIYWYAPIVADA
K.race         QSLYPANSVPQVVKRINQALQRADQIQHSEGK-----GDIYWFAPIVADA
K.tusc         QSLYPANSVPHVVKRINQALQRADQIHHMEGK-----DDIYWFAPIVADA
B.anth         QSLYPANSVPAVVKRINQTLQRADQIQHMEGS-----GDTDYFVPIVADA
A.acidoca      QSLYPSNSVPSLVRRINNALLRADQIQTAEGK-----GDIDWMVPIVADA
S.acid         QSLYPANSVPEVVRRINRALMRADQIDTLEGR-----HGTDWFAPIVADA
S.melilot      QSLYPANAAPELAKRINRTLQRADQIETAEGKGL---SVETWFAPIVADA
S.medicae      QSLYPANAAPELAKRINRTLQRADQIETAEGKGL---SVDTWFAPIVADA
B.melit        QSLYPANAGPELAKRINRTLQRADQIETAEGKGL---SVDTWFAPIVADA
R.legumin      QSLYPANAGPELAKRINRTLQRADQIETSEGQGL---SVDSWFAPIVADA
P.lava         QSLYPANAGPELAKKINRALKRADEIEHSEGG-----AKRDWFLPIVADA
Sphing         QSLYPANAGPELARRINRTLQRADQIEHSEGG-----AKRDWFVPIVADA
A.exce         QSLYPANAAPELCRRINRTLQRADQIEHSEGG-----AKRDWFAPIVADA
M.fulvus       QSLYPVDSVPTVVRKINNALRRADQIDHAEGR-----KDRYWFAPIIADA
D.geoth        QSLYPASSVPDVVRRINNALRRADQIQHSEGK-----DDIDYFVPIVADA
B.ambifa       QSLYPANSVPLVVKRINNTLTRADQIQWSEGKNPGDEGYVDYFAPIVADA
B.vietnam      QSLYPANSVPLVVKRINNTLTRADQIQWSEGKNPGDDGYVDFFAPIVADA
B.cenoce       QSLYPANSVPLVVKRINNTLTRADQIQWSEGKNPGDEGYVDFFAPIVADA
B.multivo      QSLYPANSVPLVVKRINNTLTRADQIQWSEGKNPGDEGYVDFFAPIVADA
B.psue         QSLYPANSVPLVVKRINNTLTRADQIQWSEGKNPGDEGYVDFFAPIVADA
B.thailan      QSLYPANSVPLVVKRINNTLTRADQIQWSEGKNPGDEGYVDFFAPIVADA
B.rhizox       QSLYPANSVPLVVKRINNTLARADQIQWSEGKNPGDESYIDFFAPIVADA
B.phytofi      QSLYPANSVPLVVKRINNTLTRADQIQWSEGKNPGDEGYVDYFQPIVADA
B.phymatu      QSLYPANSVPLVVKRINNTLTRADQIQWSEGKNPGDEGYIDYFAPIVADA
R.pickett      QSLYSVDSVPKVVKKINNTFQRADQIQWSEGKDD-----IDFFAPIVADA
S.maltoph      QSLYPADSVPAVVKRINNTLLRADQLHHAEGKDD-----IDFLQPIVADA
X.axonopo      QSLYPADSVPAVVKRINNTLLRADQLHHAEGNDA-----IDFLQPIVADA
P.sownens      QSLYPADSVPAVVKRINNTLLRADQLHHAEGNDS-----IDFLQPIVADA
S.ente         QSLYPANSVPAVVDRINNTFRRADQIQWASGIEPNDPRYVDYFLPIVADA
E.lign         QSLYPANSVPAVVDRINNTFRRADQIQWSSGIEPNDPRFVDYFLPIVADA
E.coli         QSLYPANSVPAVVERINNTFRRADQIQWSAGIEPGDPRYVDYFLPIVADA
Klebsi         QSLYPANSVPAVVDRINNTFRRADQIQWSVGIEPNDPRYIDYFLPIVADA
D.dandati      QSLYPANSVPAVVQRINGTFRRADQIQWANGIEPGDPRFIDYFLPIVADA
S.plym         QSLYPVDSVPTVVERINNTFRRADQIQWSNKIEPGSKDYTDFFLPIVADA
V.para         QSLYPVDSVPSVVKRINNSFRRADQIQWANGKSPEDEGGIDYFLPIVADA
V.campbel      QSLYPVDSVPSVVKRINNSFRRADQIQWSAGKSPEDEGGIDYFLPIVADA
p.lieog        QSLYPVDSVPAVVKRINNSFRRADQIQWANGGNPED--GVDYFLPIVADA
Shewanella     QSLYPANSVPAVVARINNSFRRADQIQWSNGVNPEEENFVDYFLPIIADA
A.hydroph      QSLYPANSVPSVVERINNSFARADQIQWANKVGPQDKGFIDYFLPIVADA
D.acidovo      QSLYAYDSVPTMVRRINNTFKRADEIQWGRGINPGDKEFIDYFLPIVADA
C.testo        QSLYAYDSVPTMVRRINNTFKRADEIQWGKGINPGDKEFIDYFLPIVADA
A.citrul       QSLYPVDSVPKVVERINNSFTRADEIQWAKGVNPGDAGFIDYHAPIVADA
V.paradox      QSLYPVDSVPTVVERINNTFRRADEIQWSKNVNPGDKGYVDYFAPIVADA
C.accu         QSLYPVDSVPKVVERINNSFRRADEIQFSKNIDAGDKGYIDYFAPIVADA
P.necessa      QSLYPVDSVPKMVERINNSFQRADEIQTAKGIQRGDAGFIEYFAPIVADA
P.poae         QSLYPVDSVPTVVKRINNSFRRADQIQWKAGKGPGDEGYIDYFAPIVADA
P.putida       QSLYPVDSVPTVVKRINNSFRRADQIQWKAGKNPGDEGYIDYFAPIVADA
I.loihien      QSLYPVDSVPTVVERINNSFARADQIQWSKGVTPEDDNYVDYFAPIVADA
               **:*. .: * :  :** :: *** :              :  **:**.
```

```
M.tuber        EAGFGGALNVYELQKALIAAGVAGSHWEDQLASEKKCGHLGGKVLIPTQQ
M.Bovis        EAGFGGALNVYELQKALIAAGVAGSHWEDQLASEKKCGHLGGKVLIPTQQ
M.smegmat      EAGFGGALNVYELQKAMIAAGVAGSHWEDQLASEKKCGHLGGKVLIPTQQ
M.vanbaal      EAGFGGALNVYELQKAMIAAGVAGSHWEDQLASEKKCGHLGGKVLIPTQQ
M.avium        EAGFGGALNVFELQKAMIAAGVAGSHWEDQLASEKKCGHLGGKVLIPTQQ
G.bron         EAGFGGALNVYELQKAMIAAGAAGTHWEDQLASEKKCGHLGGKVLIPTQQ
Arthroba       EAGFGGPLNAYELMKSMIQAGASGVHWEDQLASEKKCGHLGGKVLIPTQQ
S.nass         EAGFGGPLNAYELMKSMITAGASGVHWEDQLASEKKCGHLGGKVLIPTSQ
X.cellu        EAGFGGPLNAYELMRSMITAGAAGVHWEDQLASEKKCGHLGGKVLVPTQQ
N.dass         EAGFGGVLNAYELMRGMIASGAAGVHWEDQLASEKKCGHLGGKVLIPTGQ
A.mirum        EAGFGGVLNAFELMRAMIAAGAAGVHWEDQLASEKKCGHLGGKVLIPTGQ
T.curvata      EAGFGGVLNAFELMKAMIAAGAAGVHWEDQLASEKKCGHLGGKVLIPTGQ
C.flav         EAGFGGPLNAYELMHAMIQAGAAGVHWEDQLAAEKKCGHLGGKVLVPTSQ
M.cilva        EAGFGGPLNAFELMKSMIEAGAAGVHWEDQLSSEKKCGHLGGKVLIPTQA
M.rube         EAGFGGPLNAFELMKAMIEAGAAGVHWEDQLSSEKKCGHLGGKVLIPTSQ
T.osha         EAGFGGALNVFELTRAMIEAGAAGIHYEDQLASEKKCGHLGGKVLVPTAQ
T.ther         EAGFGGALNVFELTKAMIEAGAAGIHYEDQLASEKKCGHLGGKVLVPTGQ
M.hydr         EAGFGGPLHAFELTKAMIEAGAAGVHFEDQLAAEKKCGHLGGKVLVPTSQ
Roseiflex      EAGFGGPLHAFELTKAMIEAGASGVHFEDQLASEKKCGHLGGKVLVPTSQ
K.race         EAGFGGPLNVFELMKAMIEAGAAGVHFEDQLASEKKCGHLGGKVLIPTAS
K.tusc         EAGFGGPLNVYELMKAMIEAGAAGVHFEDQLAAEKKCGHMGGKVLIPASQ
B.anth         EAGFGGQLNVFELMKGMIEAGASGVHFEDQLSSEKKCGHLGGKVLLPTQT
A.acidoca      EAGFGGPLNVFELMKMMIEAGAAGVHFEDQLSSEKKCGHMGGKVLIPTSH
S.acid         EAGFGGPLNVFELTKALIEAGAAGVHLEDQMSSEKKCGHMGGKVLIPTSW
S.melilot      EAGFGGPLNAFEIMKAFIEAGAAGVHYEDQLASEKKCGHLGGKVLIPTAA
S.medicae      EAGFGGPLNAFEIMKAFIEAGAAGVHYEDQLASEKKCGHLGGKVLIPTAA
B.melit        EAGFGDPLDAFEIMKAYIEAGAAGVHFEDQLASEKKCGHLGGKVLIPTAA
R.legumin      EAGFGGPLNAFEIMKAYIEAGAAGVHFEDQLASEKKCGHLGGKVLIPTAA
P.lava         EAGFGGPLNCFEIMKAYIEAGAAGVHYEDQLASEKKCGHLGGKVLIPSRA
Sphing         EAGFGGPLNCFEIMKAYIEAGAAGVHFEDQLASEKKCGHLGGKVLIPTQA
A.exce         EAGFGGPLNSFEIMKAFIEAGAAGVHFEDQLASEKKCGHLGGKVLIPTQA
M.fulvus       EAGFGGPLNAFELMKGMIEAGAAGVHFEDQLASEKKCGHMGGKVLVPTSH
D.geoth        EAGFGGPLNAFELMKAMIEAGAAGVHFEDQLASEKKCGHLGGKVLVPTSQ
B.ambifa       EAGFGGVLNAFELMKAMIEAGASGVHFEDQLASVKKCGHMGGKVLVPTRE
B.vietnam      EAGFGGVLNAFELMKAMIEAGASGVHFEDQLASVKKCGHMGGKVLVPTRE
B.cenoce       EAGFGGVLNAFELMKAMIEAGASGVHFEDQLASVKKCGHMGGKVLVPTRE
B.multivo      EAGFGGVLNAFELMKAMIEAGASGVHFEDQLASVKKCGHMGGKVLVPTRE
B.psue         EAGFGGVLNAFELMKAMIEAGASGVHFEDQLASVKKCGHMGGKVLVPTRE
B.thailan      EAGFGGVLNAFELMKAMIEAGASGVHFEDQLASVKKCGHMGGKVLVPTRE
B.rhizox       EAGFGGVLNAFELMKAMIEAGAAGVHFEDQLASVKKCGHMGGKVLVSTRE
B.phytofi      EAGFGGVLNAFELMKAMIEAGAAGVHFEDQLASVKKCGHMGGKVLVPTRE
B.phymatu      EAGFGGVLNAFELMKAMIEAGAAGVHFEDQLASVKKCGHMGGKVLVPTRE
R.pickett      EAGFGGVLNAFELMKAMIEAGAAGVHFEDQLAAVKKCGHMGGKVLVPTRE
S.maltoph      EAGFGGVLNAFELMKAMIEAGAAGVHFEDQLASVKKCGHMGGKVLVPTRE
X.axonopo      EAGFGGVLNAFELMKAMIEAGAAGVHFEDQLASVKKCGHMGGKVLVPTRE
P.sownens      EAGFGGVLNAFELMKAMIEAGAAGVHFEDQLASVKKCGHMGGKVLVPTRE
S.ente         EAGFGGVLNAFELMKSMIEAGAAAVHFEDQLASVKKCGHMGGKVLVPTQE
E.lign         EAGFGGVLNAFELMKSMIEAGAAAVHFEDQLASVKKCGHMGGKVLVPTQE
E.coli         EAGFGGVLNAFELMKAMIEAGAAAVHFEDQLASVKKCGHMGGKVLVPTQE
Klebsi         EAGFGGVLNAFELMKSMIEAGAAAVHFEDQLASVKKCGHMGGKVLVPTQE
D.dandati      EAGFGGVLNAFELMKSMIEAGAAAVHFEDQLASVKKCGHMGGKVLVPTQE
S.plym         EAGFGGVLNAYELMKAMIEAGAAGVHFEDQLAAVKKCGHMGGKVLVPTQE
V.para         EAGFGGVLNAYELMKSMIEAGAAGVHFEDQLASVKKCGHMGGKVLVPTQE
V.campbel      EAGFGGVLNAYELMKSMIDAGAAGVHFEDQLASVKKCGHMGGKVLVPTQE
p.lieog        EAGFGGVLNAYELMRSMIEAGAAGVHFEDQLASVKKCGHMGGKVLVPTQE
Shewanella     EAGFGGVLNAFELMKSMIDAGAAGVHFEDQLASVKKCGHMGGKVLVPTQE
A.hydroph      EAGFGGVLNAFELMKGMIEAGAAGVHFEDQLASVKKCGHMGGKVLVPSQE
D.acidovo      EAGFGGVLNAFELMKNMIAAGAAGVHFEDQLAAVKKCGHMGGKVLVPTQE
C.testo        EAGFGGVLNAFELMKNMIQAGAAGVHFEDQLAAVKKCGHMGGKVLVPTRE
A.citrul       EAGFGGVLNAYELMKAMIRAGAAGVHFEDQLASVKKCGHMGGKVLVPTQE
V.paradox      EAGFGGVLNGFELMKAMIKAGAGGVHFEDQLASVKKCGHMGGKVLVPTTE
C.accu         EAGFGGVLNAFELMKAMIRAGAAGVHWEDQLASVKKCGHMGGKVLVPTQE
P.necessa      EAGFGGVLNAFELSKALIKQGAAGVHFEDQLSSVKKCGHLGGKVLLPTSE
P.poae         EAGFGGVLNAYELMKSMIEAGAAGVHFEDQLASVKKCGHMGGKVLVPTQE
P.putida       EAGFGGVLNAYELMKNMIEAGAAGVHFEDQLASVKKCGHMGGKVLVPTQE
I.loihien      EAGFGGVLNAYELMKRMIQSGAAGVHFEDQLASVKKCGHMGGKVLVPTQE
               *****. *. :*: :  *  *... * ***::: *****:*****:.:
```

```
M.tuber        HIRTLTSARLAADVADVPTVVIARTDAEAATLITSDVDERDQPFITG---
M.Bovis        HIRTLTSARLAADVADVPTVVIARTDAEAATLITSDVDERDQPFITG---
M.smegmat      HIRTLTSARLAADVADVPTVVIARTDAEAATLITSDVDERDQPFITG---
M.vanbaal      HIRTLTSARLAADVAGVPTVVIARTDAEAATLITSDVDERDRPFITG---
M.avium        HIRTLTSARLAADVCDVPTVVIARTDAEAATLITSDVDERDQPFITG---
G.bron         HVRTLNSARLAADVAGVPTVVIARTDAEAATLITSDVDERDKEFITG---
Arthroba       HVRTLNAARLAADVAGTPSVVIARTDAEAATLITSDVDERDQEFILREGG
S.nass         HVRTLNAARLAADVAGTPTLIVARTDAEAATLLTSDIDERDAEFLTG---
X.cellu        HIRTLNAARLAADVEGVPTVVVARTDAEAATLLTSDVDERDRPFLTG---
N.dass         HVKTLNAARLAADVAGVPSLVIARTDAQAATLITSDVDERDQPFITG---
A.mirum        HVRTLNAARLAADVSGVPTLVVARTDAQAATLLTTDVDERDRPFTTG---
T.curvata      HIKTLNAARLAADICGVPTLIIARTDAEAATLLTTDVDERDRPFLTG---
C.flav         HVRTLSAARLAADVAGAPTVVIARTDSLGADLLTSDVDERDQEFLTG---
M.cilva        HIRTLQAARLAADVLGVPTLIIARTDAEAATLLTSDVDERDRPFVTG---
M.rube         HIRTLNAARLAADVMNVPSIIICRTDAEAATLLTSDIDERDKPFVKEG--
T.osha         HIRTLQAARLAADIMGVPTVIIARTDAEAATLITSDIDERDRPFILKG--
T.ther         HIRTLQAARLAADVMGVPTVIIARTDAEAATLITSDIDERDKPFIVSD--
M.hydr         FIRTLQAARLAADVLGVPTVLIARTDALAATLITSDVDEYDRPFLTG---
Roseiflex      FIRTLTAARLAADVLDVPTVLIARTDAQAATLLLSDADEYDRPFITG---
K.race         AVKNLVAARFAADVMGVPTVLIARTDSNGAQLITSDIDETDKAFLTGER-
K.tusc         AIRNLTAARLAADVMGVPTILVARTDALAATMITSDIDPRDQGFITGER-
B.anth         AVRNLISARLAADVMGVPTIIVARTDADAADLITSDIDPVDKAFITGER-
A.acidoca      AVRNLTAARLAADVLGVPTVIVARTDANGAFLITSDVDDRDARFITGER-
S.acid         AVRNLTAARLAMDVLGVPTILVARTDANGANLLTSDVDPADHPFMTGER-
S.melilot      HIRNLNAARLAADVMGTPTLVIARTDAEAAKLLTSDIDERDRPFVDYDAG
S.medicae      HIRNLNAARLAADVMGTPTLVIARTDAEAAKLLTSDIDERDRPFVDIDAG
B.melit        HIRNLNAARLAADVMGTPTLIVARTDAEAAKLLTSDIDERDQPFVDYEAG
R.legumin      HIRNLNAARLAADVMGVATLVIARTDAEAAKLLTSDIDERDQPFVDYDAG
P.lava         HERNLNAARLAADVSGTPTLIMARTDAESANLITSDIDERDHEFIDFEKG
Sphing         HIRNLQAARLAADVCGVPTVLVARTDAESAKLITSDVDERDHEFLTGER-
A.exce         HERNLIAARLAADVLGVPTLTVARTDAESAQLITSDIDERDHPFIDRDS-
M.fulvus       FIRTLSAARLAADVMGVPTLLVARTDADSAKLLMSDADEYDHAFIDKKAG
D.geoth        FIRTLNAARLAADVSGVPTVLIARTDADAANLLTSDIDDNDKPFTTGER-
B.ambifa       AVAKLSAARLAADVMGTPTVLVARTDAEAADLITSDIDDNDKPFLTGER-
B.vietnam      AVAKLSAARLAADVMGTPTVLVARTDAEAADLITSDIDDNDKPFLTGER-
B.cenoce       AVAKLSAARLAADVMGTPTVLVARTDAEAADLITSDVDDNDKPFLTGER-
B.multivo      AVAKLSAARLAADVMGTPTVLVARTDAEAADLITSDIDDNDKPFLTGER-
B.psue         AVAKLTAARLAADVMGTPTVLVARTDAEAADLITSDIDDNDKPYLTGER-
B.thailan      AVAKLTAARLAADVMGTPTVLVARTDAEAADLITSDIDDNDKPFLTGER-
B.rhizox       AVAKLISARLAADVSGTPTVLVARTDAEAADLITSDIDDNDKLFLTGER-
B.phytofi      NIAKLTAARLAADVSGTPTVLLARTDAEAADLITSDIDENDKPFLTGER-
B.phymatu      NVAKLTAARLAADVLGVPTVLIARTDAEAADLITSDVDDNDKPFLTGER-
R.pickett      AVSKLVAARLAADVMGVPTVLIARTDAEAADLLTTDVDENDKPFCTGER-
S.maltoph      AIEKLNAARLAADVLGVPTLLVARTDAEAADLLTSDIDGNDRPFTTGER-
X.axonopo      AIEKLNAARLAADVLGVPTVLIARTDAEAADLITSDVDANDQPFTTGER-
P.sownens      AVEKLNAARLAADVMGVPTILIARTDAEAADLLTSDIDDNDKPFCTGER-
S.ente         AIQKLVAARLAADVMGVPTLVIARTDADAADLITSDCDPYDSSFITGER-
E.lign         AIQKLVAARLAADVLGVPTLVIARTDADAADLITSDCDPYDSEFITGER-
E.coli         AIQKLVAARLAADVTGVPTLLVARTDADAADLITSDCDPYDSEFITGER-
Klebsi         AIQKLVAARLAADVMGVPTLVIARTDADAADLITSDCDPYDREFITGDR-
D.dandati      AVQKLVAARLAADVLGVPTLLVARTDADAADLLTSDCDSYDQDFVTGER-
S.plym         AIQKLVAARLAADVLGVPTLVIARTDADAADLLTSDCDPYDSEFVSGER-
V.para         AVQKLVAARLAADVSGTTTLVIARTDANAADLLTSDCDPYDKDFIEGER-
V.campbel      AVQKLVAARLAADVAGTTTLVIARTDANAADLLTSDCDPYDKDFIQGER-
p.lieog        AVQKLVAARLAADVAGTTTLVIARTDANAADLLTSDCDPYDKDFIIGDR-
Shewanella     AVQKLVAARLAADVSGVETLVIARTDANAADLLTSDCDPYDRDFVTGER-
A.hydroph      AVQKLIAARLAADVCGTTTLVIARTDANAADLLTTDADPYDAGFLTGER-
D.acidovo      AVEKLISARFAADVMGVPTIVLARTDAEAANLITSDHDANDKPFLTGER-
C.testo        AVEKLIAARFAADVMGVPTIVLARTDAEAANLITSNHDENDIPFLTGER-
A.citrul       AVQKLIAARMAADVYGVPTLVIARTDAEAADLITSDYDENDKPFLTGER-
V.paradox      AVQKLIAARMAADVCGTPTLVIARTDAEAADLITSDYDENDKPFLTGER-
C.accu         AVQKLIAARFAADVCGVPTLVIARTDAEAADLLTSDCDANDTPFVTGER-
P.necessa      SVQKLISARLAADVMGTPTIILARTDAEAADLLTSDYDENDKPFLTGER-
P.poae         AVQKLTAARLAADVAGTPTIILARTDANAADLLTSDCDPYDQPFVTGER-
P.putida       AVQKLVAARLAADVSGVPTIILARTDANAADLLTSDCDPYDQPFVIGER-
I.loihien      AVQKLSAARLAADVAGTPTLIVARTDANAADLLTSDFDPNDKPFITGER-
                   .* :**:* *: .. :: :.***: .* :: :: *  *  :
```

```
M.tuber        --ERTREGFYRTKN--GIEPCIARAKAYAPFADLIWMETGTPDLEAARQF
M.Bovis        --ERTREGFYRTKN--GIEPCIARAKAYAPFADLIWMETGTPDLEAARQF
M.smegmat      --ERTKEGFFRVKN--GLEPCIARAKAYAPYSDLIWMETGTPDLELAKKF
M.vanbaal      --ERTAEGFYRVKN--GLEPCIARAKAYAPYSDLIWMETGTPDLELAKKF
M.avium        --ERTKEGFYRVRN--GLEPCVARAKAYAPYSDLIWMETGTPDLELAAKF
G.bron         --ERTAEGYYHVKN--GIEPCIARAKSYAPYADMIWMETGTPDLELARKF
Arthroba       QPVRTPEGFYKVRN--GIEPCIARAKAYAPYSDLIWMETGTPDLELARKF
S.nass         --ERTSEGYYRVTN--GLPACIARAKAYAPYSDLLWMETGTPDLEVARQF
X.cellu        --GRTAEGFYEVRP--GLETSIARARAYAPYADLLWMETGTPSLEVARAF
N.dass         --ERTAEGFYRFRN--GVEACVARGLAYAPHSDLLWMETATPDLEVARDF
A.mirum        --ERTAEGFYRVRG--GIEPCVARGLAYAPHADLLWMETSLPDLGVAREF
T.curvata      --ERTAEGFYRVRN--GLEPCIARALAYAPYADLIWMETGTPDLEQARKF
C.flav         --ERTAEGYFRVRP--GLDAVVSRATAYAPYSDLIWVETSTPDVALAVEF
M.cilva        --ERTPEGFYRVKN--GLEPCITRALAYAPYADLLWMETSTPDLEVARKF
M.rube         --ERTPEGFYRIRN--GLEACITRSLAYAPYADLLWMETSTPDLEVARKF
T.osha         --ERTPEGFYRVKN--GIEAGIARALAYAPYADVLWMETSKPDLEEARRF
T.ther         --ERTPEGFYRFKN--GIEAGIARALAYAPYADVIWMETSKPDLEEARKF
M.hydr         --ERTPEGYYRVKN--GLEAAIARALAYAPYADLLWFETSTPDLEEARRF
Roseiflex      --ERTPEGFYRVKS--GLDAAIARGLAYAPYADLIWCETAHPDLDEARRF
K.race         ----TPEGFFRVRS--GIEAAIARGLSYAPYCDLIWCETSSPDMGEAERF
K.tusc         ----TSEGFYRVRN--GLEAAIARGLAYAPYADMIWCETSEPNLEEARLF
B.anth         ----TPEGFYRTKA--GLDQAIARGLAYAPYADLVWCETSEPNLEDAKRF
A.acidoca      ----TPEGFYRFRG--GLEAAIARGLAYAPYADMIWCETSEPNIEEAKRF
S.acid         ----TAEGFYRVKS--GLDQAVHRGLAYAPYADMIWCETSEPDLDEAREF
S.melilot      ---RTVEGFYQVKN--GIEPCIARAIAYAPHCDLIWCETSKPDLEQARKF
S.medicae      ---RTVEGFYQVRN--GIEPCIARAIAYAPHCDLIWCETSKPDLEQARKF
B.melit        ---RTAEGFYQVKN--GIEPCIARAIAYAPYCDLIWMETSKPDLAQARRF
R.legumin      ---RTVEGFYQVRN--GIEPCIARAVAYAPYCDLIWCETSKPDLDQARRF
P.lava         ---RTIEGFYRLKPGTGVDHCIKRGLSFAPYADLLWWETSKPNLDDAKRF
Sphing         ----TPEGFFRLKDGTGVDHCIKRGIAFAEHADLLWWETSHPNLEDAKRF
A.exce         ---RTPEGFFRLKPGTGLDHCIARGLAYAKYADLLWWETSHPDLDDARRF
M.fulvus       ---RTAEGFYRLNG--GLDCAIARGLAYAPYADLVWCETSTPDLAQAKKF
D.geoth        ----TPEGFFYVRP--GIEQAISRALAYAPYADVLWCETSVPNLEDARRF
B.ambifa       ----TVEGFFRTKP--GIEQAISRGLAYAPYADLVWCETGKPDLEYAKKF
B.vietnam      ----TVEGFFRTKP--GLEQAISRGLAYAPYADLVWCETGKPDLEYAKKF
B.cenoce       ----TVEGFFRTKP--GLEQAVSRGLAYAPYADLIWCETGKPDLEYAKKF
B.multivo      ----TVEGFFRTKP--GLEQAVSRGLAYAPYADLIWCETGKPDLEYAKKF
B.psue         ----TVEGFFRTKP--GLEQAISRGLAYAPYADLIWCETGKPDLEYAKKF
B.thailan      ----TVEGFFRTKP--GLEQAISRGLAYAPYADLIWCETGKPDLEYAKKF
B.rhizox       ----TVEGFFRTRP--GLEQAISRGLAYAPYADLIWCETGKPDLEYAKKF
B.phytofi      ----TVEGFYRTKP--GLEQAISRGLAYAPYADMIWCETGKPDLEFAKKF
B.phymatu      ----TVEGFFRTKP--GLEQAVSRGLAYAPYADLIWCETGKPDLEYAKKF
R.pickett      ----TVEGFYRTKP--GLQQAISRGLAYAEVADLVWCETGKPDLEYAKKF
S.maltoph      ----TVEGFYKTRN--GLDQAISRGLAYAPYADLVWCETGKPDLEFARRF
X.axonopo      ----TVEGFFKTRN--GLDQAISRGLAYAPYADLIWCETGKPDLEFARAF
P.sownens      ----TVEGFYRVRN--GLEQSISRGLAYAPYADLVWCETGKPDLEFARRF
S.ente         ----TSEGFYRTHAG--IEQAISRGLAYAPYADLVWCETSTPDLELARRF
E.lign         ----TSEGFYRTHAG--IEQAISRGLAYAPYADLVWCETSTPDLALAKRF
E.coli         ----TSEGFFRTHAG--IEQAISRGLAYAPYADLVWCETSTPDLELARRF
Klebsi         ----TSEGFFRTHAG--IEQAISRGLAYAPYADLVWCETSKPDLEQARRF
D.dandati      ----TVEGFFRTRAG--VEQAISRGLAYAPYADLVWCETSTPDLALARRF
S.plym         ----TAEGFFRTHAG--IEQAISRGLAYSPYADMVWCETSTPDLQAAKRF
V.para         ----TQEGFYRVRAG--IDQAISRGLAYAPYADLIWCETATPCLEEARKF
V.campbel      ----TPEGFYRVRAG--IDQAISRGLAYAPYADLIWCETATPCLEEARKF
p.lieog        ----TAEGFYRVRAG--IDQAIARGLAYAPYADLIWCETATPCLEEARKF
Shewanella     ----TSEGFYRVKAG--LDQAISRGLAYAPYADLIWCETAKPDLEEARRF
A.hydroph      ----TSEGFYKVQAG--IEQAISRGLAYAPYADMVWCETAKPDLDEARKF
D.acidovo      ----TQEGFYRVRNG--MEQSISRGVAYAPYADLVWCETGVPDIGFAREF
C.testo        ----TQEGFYRVKNG--LEQSISRGVAYAPYADLVWCETGVPDIGFAREF
A.citrul       ----TAEGFYKTRKG--IDQAISRAVAYADYADLVWCETGTPDLEFARKF
V.paradox      ----TAEGFYKTKKG--IDQAISRAIAYAHYADLVWCETGTPDLEFAKRF
C.accu         ----TSEGFYKTRKG--LDQAISRAVAYAEYADLVWCETGTPDLEFARKF
P.necessa      ----TPEGFYKTRKG--LDQAISRGLAYAAYADMVWCETGTPDLDFARQF
P.poae         ----TQEGFYKVRAG--LDQAIARGLAYAPYADLIWCETAKPDLDEARRF
P.putida       ----TREGFYKVRAG--LDQAISRGLAYAPYADLIWCETAKPDLDEARRF
I.loihien      ----TAEGFYRVKPG--IEQAISRGLAYAPFADLIWCETAKPDLDEARQF
                   * **::        :   : *. :::  .*::* **. * :  *  *
```

```
M.tuber      SEAVKAEYPDQMLAYNCSPSFNWKKHLDDATIAKFQKELAAMGFKFQFIT
M.Bovis      SEAVKAEYPDQMLAYNCSPSFNWKKHLDDATIAKFQKELAAMGFKFQFIT
M.smegmat    AEGVKAEFPDQMLAYNCSPSFNWKKHLDDATIAKFQKELGAMGFKFQFIT
M.vanbaal    AEGVKSEFPDQMLAYNCSPSFNWRKHLDDATIAKFQKELGAMGFKFQFIT
M.avium      AEGVKAEFPDQMLAYNCSPSFNWKKHLDDATIAKFQKELAAMGFKFQFIT
G.bron       AEAVKSEYPDQLLSYNCSPSFNWSKHLDDSTIAKFQNELGAMGFTFQFIT
Arthroba     AEAVKAEFPDQMLSYNCSPSFNWRKHLDDATIAKFQRELGAMGFTFQFIT
S.nass       AEAVKAEFPDQMLAYNCSPSFNWKRHLDNDTIAKFQNELGAMGFTFQFIT
X.cellu      AEGVRSEFPDQLLAYNCSPSFNWRRHLSEAEIATFQRELGAMGYAFQFIT
N.dass       AERIKAEYPDQMLAYNCSPSFNWKRHLDDATIAKFQRELGHMGYKFQFIT
A.mirum      AEAVKARYPEKMLAYNCSPSFNWRKHLDDTTIAKFQRELGHMGYKFQFIT
T.curvata    AEAVKAEYPDQMLAYNCSPSFNWKKHLDDATIAKFQRELGHMGYKFQFIT
C.flav       AERIHEAFPGKLLAYNCSPSFNWRAHLDDAQIARFQRELAGHGYAFQFIT
M.cilva      AEAVHREFPGKMLAYNCSPSFNWKKYLDAETIAKFQRELGAMGYKFQFIT
M.rube       AEAIHAVYPGKMLAYNCSPSFNWKKNLDDETIAKFQRELGAMGYKFQFIT
T.osha       AEAVKREFPDKLLAYNLSPSFNWKKFLDEETIAKFNRELGAMGYKFQFIT
T.ther       AEAVKKEFPDKLLAYNLSPSFNWKKFLDDETIAKFNRELGAMGYKFQFIT
M.hydr       AEAIHREFPGKLLAYNCSPSFNWRKFLGEAEIARFQKELGAMGYKFQFIT
Roseiflex    AEGIHAKFPGKMLAYNCSPSFNWKRYLDEATIATFQRELAAMGYKFQFIT
K.race       AAEIHARFPGKLLAYNCSPSFNWRKNLSETQIAEFQKHLGSLGYAFQFIT
K.tusc       AEAIHAKFPGKLLAYNCSPSFNWKRKLDDAAIADFQKKLGQMGYKFQFVT
B.anth       ADAIHKEHPGKLLAYNCSPSFNWKQKLDEKTIASFQKEIASYGYKFQFVT
A.acidoca    AEAIKAEFPDKLLAYNCSPSFNWRKKLDPKTIEEFQEIIASFGYKFQFVT
S.acid       ARRIHAEFPGKLLAYNCSPSFNWRRRWDERGLERFQIELADMGYKFQFIT
S.melilot    AEGVHKAHPGKLLAYNCSPSFNWKKNLDDATIAKFQRELGAMGYKFQFIT
S.medicae    AEGVHKAYPGKLLAYNCSPSFNWKKNLDDATIAKFQRELGAMGYKFQFIT
B.melit      AEAVHKAHPGKLLAYNCSPSFNWKKNLDDATIAKFQRELGAMGYKFQFIT
R.legumin    AEGVHRVHPGKLLAYNCSPSFNWKKNLDEATIAKFQRELGAMGYKFQFIT
P.lava       AEAIRKEYPDQLLAYNCSPSFNWEANLDKATIAKFQKEIAAMGYKYQFVT
Sphing       AEAVQKAHPGKMMAYNCSPSFNWEAKLDKDTIARFQRELGAMGYKFQFVT
A.exce       AEAVHKVHPGKLLAYNCSPSFNWKAKLDEATIARFQRELGAMGYKFQFVT
M.fulvus     AESLRAKYPNKLLAYNCSPSFNWKKNLDDATIARFQRELGAMGYKFQFVT
D.geoth      AEAVHAKFPGKLLAYNCSPSFNWKKNLDDETIAKFQRELGAMGYKFQFIT
B.ambifa     AEAIHKQFPGKMLSYNCSPSFNWKKNLDDATIAKFQKELGAMGYKFQFIT
B.vietnam    AEAIHKQFPGKMLSYNCSPSFNWKKNLDDATIAKFQKELGAMGYKFQFIT
B.cenoce     AEAIHKQFPGKLLSYNCSPSFNWKKNLDDATIAKFQKELGAMGYKFQFIT
B.multivo    AEAIHKQFPGKLLSYNCSPSFNWKKNLDDATIAKFQKELGAMGYKFQFIT
B.psue       AEAIHKQFPGKLLSYNCSPSFNWKKNLDDATIAKFQKELGAMGYKFQFIT
B.thailan    AEAIHKQFPGKLLSYNCSPSFNWKKNLDDATIAKFQKELGAMGYKFQFIT
B.rhizox     AEAIHKAFPGKLLSYNCSPSFNWKKNLDDATIAKFQKELGAMGYKFQFIT
B.phytofi    ADAIHKEYPDQLLSYNCSPSFNWKKNLDDATIAKFQRELGAMGYKFQFIT
B.phymatu    AEAIHKQFPGKMLSYNCSPSFNWKKNLDDATIAKFQKELGAMGYKFQFIT
R.pickett    AEAIHAKFPGKMLAYNCSPSFNWKKNLDDVTIARFQKELGAMGYKFQFIT
S.maltoph    AEAIHAKFPGKLLAYNCSPSFNWKKNLDDATIAKFQRELGSYGYKFQFIT
X.axonopo    AQAIHARFPGKLLAYNCSPSFNWKKNLDDATIAKFQTELASYGYKFQFIT
P.sownens    AEAIHARYPGKMLAYNCSPSFNWKKNLDDATIAKFQKEIASYGYKFQFIT
S.ente       ADAIHAKYPGKLLAYNCSPSFNWQKNLDDKTIASFQQQLSDMGYKYQFIT
E.lign       ADAIHAKFPGKLLAYNCSPSFNWQKNLDDKTIASFQQQLSDMGYKYQFIT
E.coli       AQAIHAKYPGKLLAYNCSPSFNWQKNLDDKTIASFQQQLSDMGYKFQFIT
Klebsi       AEAIHARFPGKLLAYNCSPSFNWKKNLDDKTIASFQQQLSDMGYKYQFIT
D.dandati    AEAIHARFPGKLLAYNCSPSFNWKKNLDDRAIARFQDELSEMGYKYQFIT
S.plym       ADAIHAKFPGKLLAYNCSPSFNWKKNLDDQTIARFQEELSAMGYKYQFIT
V.para       AEAIHAEYPDQLLAYNCSPSFNWEKNLDAETIAKFQQELSDMGYKYQFIT
V.campbel    AEAIHAEYPDQLLAYNCSPSFNWEKNLDAETIAKFQQELANMGYKYQFIT
p.lieog      AEAIHAQYPDQLLAYNCSPSFNWEKNLDAETIAKFQQELSDMGYKYQFIT
Shewanella   AEAIHAQYPDQLLAYNCSPSFNWKKNLDDATIARFQQELSDMGYKYQFIT
A.hydroph    AEAIKAQYPDKILAYNCSPSFNWKKNLDDATIARFQQELSDMGYKYQFIT
D.acidovo    AQAVHAACPGKLLSYNCSPSFNWKKNLNDSQIASFQEDLSALGYKFQFIT
C.testo      AQAVHAACPGKLLSYNCSPSFNWKKNLNDSQIASFQEDLSALGYKFQFIT
A.citrul     AEAVHAKHPGKLLAYNCSPSFNWKKNLDDATIAKFQRELGAMGYKYQFIT
V.paradox    ADAVHKVHPGKMLAYNCSPSFNWKKNLDDATIAKFQKELGAMGYKYQFIT
C.accu       AEAVRAKHPGKMLAYNCSPSFNWKKNLDDATIARFQKELGAMGYKYQFIT
P.necessa    AEAIRAKFPGKMLAYNCSPSFNWKKNLDDATIAKFQRELGAMGYKYQFIT
P.poae       AEAIKKEYPDQLLSYNCSPSFNWKKNLDDATIAKFQRELSAMGYKHQFIT
P.putida     AEAIKKEYPDQILSYNCSPSFNWKKNLDDATIAKFQRELSAMGYKHQFIT
I.loihien    AEAVHAKYPGKLLAYNCSPSFNWKRNLDDDTIAKFQRELAAMGYKFQFIT
             :  ::   * ::::** ******    .   :  *:  :.  *: .**:*
```

Learning Resource Center
Acc. No. SP09132
SP091305
JUIT, Waknaghat, Solan (H.P.)

```
M.tuber        LAGFHALNYSMFDLAYGYAQN-QMSAYVELQEREFAAEE-RGYTATKHQR
M.Bovis        LAGFHALNYSMFDLAYGYAQN-QMSAYVELQEREFAAEE-RGYTATKHQR
M.smegmat      LAGFHALNYSMFDLAYGYARN-QMSAYVELQEREFAAEE-RGYTATKHQR
M.vanbaal      LAGFHALNYSMFDLAHGYARN-QMTAYVELQEREFAAEE-RGYTATKHQR
M.avium        LAGFHALNYSMFDLAYGYARN-QMTAYVELQEREFAAEE-RGYTATKHQR
G.bron         LAGFHSLNYGMFDLAYGYARE-QMSAFVDLQNREFKAAEERGFTAVKHQR
Arthroba       LAGFHALNYSMFDLAHGYARE-GMSAYVELQEKEFASES-RGYTATKHQR
S.nass         LAGFHSLNHSMFDLAHRYARE-GMTAYVDLQEREFAAED-AGYTATKHQR
X.cellu        LAGFHALNHSMFDLAHGYARE-QMTAYVRLQESEFASES-RGYTATKHQR
N.dass         LAGFHSLNYSMFNLAKGYAQN-GMTSYVELQEAEFAAES-QGYTATRHQR
A.mirum        LAGFHALNHSMFDLARGYAES-GMSAYVELQEREFAAER-DGYTAVKHQR
T.curvata      LAGFHALNYSMFDLAHGYARE-GMTAYVDLQEREFASES-RGYTATRHQR
C.flav         LAGFHALNESMFTLAKGYAER-GMSAYVDLQTAEFAAEA-DGYTATRHQR
M.cilva        LAGWHNLNYRTFELAKGYKER-GMSAFVELQSLEFAAEK-DGFTAVKHQR
M.rube         LAGWHNLNYRTFELARNYKER-GMSAFVELQQLEFAAEK-DGFTAVKHQR
T.osha         LAGWHTLNYYTWELAKGYKAR-GMPAFVELQEKEFLAQA-QGFTAVKHQR
T.ther         LAGWHALNYHIWELAKGYKER-GMPAFVELQAKEFAAQK-EGFTAVKHQR
M.hydr         LAGWHLLNYHTFDLARRYKRE-GMPAYVALQQAEFAAEE-AGYTATRHQR
Roseiflex      LAGWHMINYYAFELAKAYAAE-GMTAYVRLQQAEFAAEQ-HGYTATRHQR
K.race         LAGFHALNYSMFDLARGYREE-GMTAYSRLQQAEFALEK-VGYTATKHQH
K.tusc         LAGFHALNFSMFDLARQYKER-GMAAYAELQQREFASEV-YGYTATKHQR
B.anth         LAGFHSLNYGMFELARGYKER-GMAAYSELQQAEFAAEK-HGYSATRHQR
A.acidoca      LAGFHALNYSMFELAYGYNQR-GMGAYSDLQQQEFAAEA-RGYTATRHQR
S.acid         LAGFHALNYSMFQLAHDYKAN-GMKAYADLQDQEFQAEK-DGYTAVRHQR
S.melilot      LAGFHQLNYGMFELARGYRDR-QMAAYSELQEAEFAAEA-NGYTATKHQR
S.medicae      LAGFHQLNYGMFELARGYKDR-QMAAYSELQEAEFAAEA-NGYTATKHQR
B.melit        LAGFHQLNYGMFELARGYKDR-QMAAYSELQQAEFAAEA-DGYTATKHQR
R.legumin      LAGFHQLNYGMYELARGYRQR-QMSAYSELQQAEFDAEV-NGYTATKHQR
P.lava         LAGFHQLNHGMFELASGYRDR-GMAAYSELQQAEFASES-KGYTATRHQR
Sphing         LAGFHSLNHGMFELARGYKDR-GMAAYSELQQAEFASEA-QGYTATRHQR
A.exce         LAGFHSLNNGMFELASGYRDR-GMAAYSELQQREFANEA-AGYTATRHQR
M.fulvus       LAGFHALNFGMYELARKYKDR-GMAAYSELQQAEFAAEK-DGYTATRHQR
D.geoth        LAGFHSLNYSMFELAYGYARN-QMTAFVELQEKEFAAQE-RGFTAVKHQR
B.ambifa       LAGFHALNYSMFNLAHGYART-QMSAFVELQQAEFAAAD-KGFTAVKHQR
B.vietnam      LAGFHALNYSMFNLAHGYART-QMSAFVELQQAEFEAAD-KGFTAVKHQR
B.cenoce       LAGFHALNYSMFNLAHGYART-QMSAFVELQQAEFAAAD-KGFTAVKHQR
B.multivo      LAGFHALNYSMFNLAHGYART-QMSAFVELQQAEFAAAD-KGFTAVKHQR
B.psue         LAGFHALNYSMFNLAHGYART-QMSAFVELQQAEFAAAD-KGFTAVKHQR
B.thailan      LAGFHALNYSMFNLAHGYART-QMSAFVELQQAEFAAAD-KGFTAVKHQR
B.rhizox       LAGFHSLNYSMFNLAYGYARN-QMSAFVELQQAEFAAAE-KGFTAVKHQR
B.phytofi      LAGFHALNYSMFNLAHGYARN-QMTAFVEMQQAEFAAAE-KGFTAVKHQR
B.phymatu      LAGFHALNYSMFNLAHGYARQ-QMSAFVELQQAEFAAAE-RGFTAVKHQR
R.pickett      LAGFHALNYSMFNLAYGYARN-QMSAFVELQEAEFKAAE-KGFTAVKHQR
S.maltoph      LAGFHALNYGMFNLAHGYARR-QMSAFVELQEAEFEAAE-RGFTAVKHQR
X.axonopo      LAGFHALNHSMFHLAHGYARR-QMSAFVELQQAEFEAAE-MGFTAVKHQR
P.sownens      LAGFHALNYSMFNLAHGYARN-QMSAFVELQEAEFAAAG-KGFTAVKHQR
S.ente         LAGIHSMWFNMFDLAHAYAQGEGMKHYVEKVQQPEFAAAKDGYTFVSHQQ
E.lign         LAGIHSMWFNMFDLAHAYAQGEGMKHYVEKVQQPEFAAGKEGYSFVSHQQ
E.coli         LAGIHSMWFNMFDLANAYAQGEGMKHYVEKVQQPEFAAAKDGYTFVSHQQ
Klebsi         LAGIHSMWFNMFDLAHAYAQGEGMRHYVEKVQQPEFAAAPEGYSFVSHQQ
D.dandati      LAGIHSMWFNMFDLAHAYAQGEGMKHYVEKVQQREFAAISEGYTFSSHQQ
S.plym         LAGIHSMWFNMFDLAHAYAQGEGMKHYVEKVQEPEFAAVARGYTFASHQQ
V.para         LAGIHNMWFNMFELAHAYAQGEGMRHYVEKVQRPEFQAAEKGYTFVAHQQ
V.campbel      LAGIHNMWFNMFELAHAYAQGEGMRHYVEKVQRPEFEAAEKGYTFVAHQQ
p.lieog        LAGIHNMWFNMFELAHAYAQGEGMRHYVEKVQRPEFEAAEKGYTFVAHQQ
Shewanella     LAGIHNMWYNMFDLAYDYARGEGMKHYVEKVQEVEFAAAKKGYTFVAHQQ
A.hydroph      LAGIHNMWFHMYELAYQYARGEGMKHYVEMVQQPEFAAAERGYTFVAHQQ
D.acidovo      LAGIHINWFNTFKFAHAYANGEGMKHYVNMVQEPEFAAREQGYTFVSHQQ
C.testo        LAGIHSNWYNTFKFAHEYARGEGMKHYVEMIQEPEFAARDKGYTFVSHQQ
A.citrul       LAGIHSMWFNMFDLAQDYVKR-GMSAYVEKVQEPEFAARERGYTFVSHQQ
V.paradox      LAGIHSMWFNMFDLAQDYVQR-GMSAYVEKVQEPEFAARDRGYTFVSHQQ
C.accu         LAGIHNMWYHMFDLAQDYVAR-GMTAYVEKVQEPEFAARDRGYSFVSHQQ
P.necessa      LAGIHSMWYNMFDLAQDYMQR-GMTAYIEKVQEPEFAARDRGYTFVSHQQ
P.poae         LAGIHNMWHSMFNLAHDYARN--DMTAYVKLQEQEFADASKGYTFVAHQQ
P.putida       LAGIHNMWHGMFNLAHDYARN--DMTAYVKLQEQEFADASKGYTFVAHQQ
I.loihien      LAGVHNMWYHMFELAHDYARN--DMSAYVKLQQQEFEAADKGYTFVAHQQ
               *** *       : :*  *                      *::    **:
```

```
M.tuber        EVGAGYFDRIATTVDP-NSSTTALTGSTEEGQFH----------
M.Bovis        EVGAGYFDRIATTVDP-NSSTTALTGSTEEGQFH----------
M.smegmat      EVGAGYFDRIATTVDP-NSSTTALAGSTEEGQFH----------
M.vanbaal      EVGAGYFDRIATTVDP-TSSTTALAGSTEEGQFH----------
M.avium        EVGAGYFDRIATTVDP-TSSTTALTGSTEEGQFH----------
G.bron         EVGAGYFDNIATTVDP-NSSTTALKGSTEEGQFH----------
Arthroba       EVGTGYFDDIATALNP-NASTLALVGSTEEGQFH----------
S.nass         EVGTGYFDEIATALNP-SGSTLALAGSTESGQFH----------
X.cellu        EVGTGYFDRVATALNP-TASTLALAGSTESAQF-----------
N.dass         EVGTGYFDSISTTISP-EASTTALTGSTEEDQFSAAH-------
A.mirum        EVGTGYFDRVSTALNP-AGETTALSGSTEQAQFA----------
T.curvata      EVGTGYFDLVSTAISP-NSSTTALKGSTEDEQFFDK--------
C.flav         EVGTGYFDQVATAISP-DSATLALAGSTETAQFQH---------
M.cilva        EVGAGYFDEVLLAITGGQASTAALVGSTEEAQFH----------
M.rube         EVGAGYFDEVLMAITAGQASTAALVGSTEEAQFH----------
T.osha         EVGAGYFDEVVLALTQGEASTLALKGSTEEAQFNEPVH------
T.ther         EVGAGYFDEVVLALTQGQASTLALKGSTEEAQFNEPVH------
M.hydr         EVGTGYFDQVLLTITGGQAATAALVGSTEAEQFH----------
Roseiflex      EVGTGYFDEVSTIISGGLSSTTALAGSTEEEQFH----------
K.race         EVGTGYFDEVTQVIAGGASSTTALTGSTEEAQFH----------
K.tusc         EVGAGYFDEVSMVISGGTSSTTALTGSTEEEQFV----------
B.anth         EVGTGYFDEVAQVITGGTSSTTALKGSTEEAQFTK---------
A.acidoca      EVGTGYFDEVAQVVSGGLSSTTALTGSTEEEQFVEA--------
S.acid         EVGAGYFDEVSLVISGGSSSTIALVGSTEEAQFQS---------
S.melilot      EVGTGYFDAVSVAITGGQSSTTAMKESTEHDQFRPAAE------
S.medicae      EVGTGYFDAVSVAITGGQSSTTAMKESTEHDQFRPAAE------
B.melit        EVGTGYFDAVSLAITGGQSSTTAMKESTETAQFKPAAE------
R.legumin      EVGTGYFDAVSLAITGGRSSTTAMHDSTEHAQFKPAAE------
P.lava         EVGTGYFDMVSIAAAGGESSTTALGDSTEADQFKKAGAH-----
Sphing         EVGTGYFDLVAEAVSGGMASTTALKESTEADQFAVAAE------
A.exce         EVGTGYFDKIAMTITNGQSSTTALKDSTETAQFQTEPAE-----
M.fulvus       EVGTGYFDQVAEVISGGNASTLALTESTEAHQF-----------
D.geoth        EVGTGYFDLVSTAAGGGQSSTTALAGSTEAQQFHGEKEVAAAHD
B.ambifa       EVGTGYFDAVTQTVER-EASTTALHGSTEDEQFFDGKKVA----
B.vietnam      EVGTGYFDAVTQTVER-EASTTALHGSTEDEQFFDGKKVA----
B.cenoce       EVGTGYFDAVTQTVER-EASTTALHGSTEDEQFFDGKKVA----
B.multivo      EVGTGYFDAVTQTVER-EASTTALHGSTEDEQFFDGKKVA----
B.psue         EVGTGYFDAVTQTVER-EASTTALHGSTEDEQFFDGQKVA----
B.thailan      EVGTGYFDAVTQTVER-EASTTALHGSTEDEQFFDGQKVA----
B.rhizox       EVGTGYFDAVTQTIER-EASTTALHGSTEDEQFFDNKKVA----
B.phytofi      EVGTGYFDAVTQTVER-EASTTALHGSTEDEQFFDKKVA-----
B.phymatu      EVGTGYFDAVTQTVER-EASTTALHGSTEDEQFFDKKVA-----
R.pickett      EVGTGYFDAVTQTIER-EASTTALKGSTEDEQFFEETAQAKKVA
S.maltoph      EVGTGYFDAVTQAIQQGQSSTTALTGSTEEEQFHGGRSERAA--
X.axonopo      EVGTGYFDAVTQAIQQGQSSTTALRGSTEEEQFQSDKAA-----
P.sownens      EVGTGYFDAVTQAIQQGQSSTTALKGSTEEEQFHDRSAA-----
S.ente         EVGTGYFDKVTTIIQGGASSVTALTGSTEESQF-----------
E.lign         EVGTGYFDKVTTIIQGGASSVTALTGSTEEAQF-----------
E.coli         EVGTGYFDKVTTIIQGGTSSVTALTGSTEESQF-----------
Klebsi         EVGTGYFDKVTTIIQGGASSVTALTGSTEEDQF-----------
D.dandati      EVGTGYFDKVTTIIQGGSSSVTALTGSTEEQQF-----------
S.plym         EVGTGYFDKVTTIIQGGASSVTALTGSTEEQQF-----------
V.para         EVGTGYFDRMTNTIQGGNSSVTALTGSTEEDQF-----------
V.campbel      EVGTGYFDRMTNTIQGGNSSVTALTGSTEEDQF-----------
p.lieog        EVGTGYFDTMTNTIQGGNSSVTALTGSTEEDQFK----------
Shewanella     EVGTGYFDQVTTVIQGGQSSVTALTGSTEEEQF-----------
A.hydroph      EVGTGYFDKVTTVIQGGQSSVTALTGSTEEDQFH----------
D.acidovo      EVGAGYFDDVTTVIQGGSSSVKALTGSTEEEQFH----------
C.testo        EVGAGYFDDVTTVIQGGSSSVKALTGSTEEEQFH----------
A.citrul       EVGTGYFDEVTTVIQGGKSSVTALTGSTEEEQFH----------
V.paradox      EVGTGYFDEVTTVIQGGKSSVTALTGSTEEEQFH----------
C.accu         EVGTGYFDEVTTVIQGGTSSVTALTGSTEEEQFH----------
P.necessa      EVGTGYFDDVTTVIQGGKSSVTALTGSTEEEQFH----------
P.poae         EVGTGYFDDMTTVIQGGTSSVTALKGSTEEEQFH----------
P.putida       EVGTGYFDDMTTVIQGGASSVTALTGSTEEEQFH----------
I.loihien      EVGTGYFDELTNVIQGGLSSVTALKGSTEEDQFKTSA-------
               ***:**** :          . .  *:  ***   **
```

Subfamily 2

6 and 11 sequences of ICL protein in fungi and plants respectively were taken and MSA was carried out using ClustalW. The sequences were highly conserved and showed that they belong to same subfamily

MSA of ICl in plants

```
I.batatas       LQAIEDNWIAMAELKTFSETVVDAINRLNIGEPEKQRRLNEWMKHSSYEK
S.lycoper       LQAIEDNWLGMAELKTFSQCVTDAIKKMNLAEYEKQRKLNEWMNHSSYEK
A.thaliana      LQSIEDQWLGSAGLMTFSEAVVQAIKRMNLNENEKNQRLSEWLTHARYEN
C.sativus       LQALEDQWISMAQLKTFSECVTDAIMNTNATENEKRRKLDEWMNHSSYEK
P.taeda         LQDVEDKWMGMANLRLFSDVVGDAIRSLAIPQQEKSNRINLWFN---KID
A.capsulatus    LQEIEDKWNAEAGLKLFSEAVIDHIKAG-----SASNKQALIDKFLRDAR
                ** :**:* . * *  **: * : *        .  .:     .

I.batatas       CLSNFQAREIAENLGLNNLFWDWDLPRTREGFYRFKGSVMAAIVRGWAFA
S.lycoper       CLSHEQAREVAERLGLPNLFWDWDLPRTREGFYRFQGSVEAAIVRGWAFA
A.thaliana      CLSNEQGRVLAAKLGVTDLFWDWDLPRTREGFYRFQGSVAAAVVRGWAFA
C.sativus       CISNEQGREIAEKLGLKNLFWDWDLPRTREGFYRFKGSVMAAIVRGWAFA
P.taeda         QLSHFEARELAEKLGVSDVFWDWDLPRTREGFYRFQGSTKAAILRGWAFG
A.capsulatus    GKSNAEARAIAKALTGSEIYFNWDSARTREGYYRYRGGTQCAVNRAVAFA
                  *: :.* :*  *   :::::** .*****:**::*.. .*: *. **.

I.batatas       PHSDLIWMETSSPDLVECTKFAQGVKSVHPEMMLAYNLSPSFNWDASGMN
S.lycoper       EYCDLVWMETSSPDMVECTKFSQGVKTLRPELMLAYNLSPSFNWDASGMN
A.thaliana      QIADIIWMETASPDLNECTQFAEGIKSKTPEVMLAYNLSPSFNWDASGMT
C.sativus       PHADLIWMETSSPDLVECTTFAKGMKSIHPETMLAYNLSPSFNWDASGMS
P.taeda         PHADLIWMETSSPDMVECKRFAEGVKSKHPDLMLAYNLSPSFNWDASRMS
A.capsulatus    PFADLIWMESKLPDYAQAKEFADGVHAVWPEQKLAYNLSPSFNWKAAMPR
                  .*::***:  **  :.. *:.*:::  *:  ***********.*:

I.batatas       DNQMMDFIPGIAKLGYCWQFITLAGFHGNALIIDTFAKDFASRGMLAYVE
S.lycoper       DNQMMDFIPRIAKLGYCWQFITLAGFHADALIVDTFAKDFARRGMLAYVE
A.thaliana      DQQMVEFIPRIARLGYCWQFITLAGFHADALVVDTFAKDYARRGMLAYVE
C.sativus       DKQMEEFIPRIARLGFCWQFITLAGFHADALVVDTFARDYARRGMLAYVE
P.taeda         DEQMMNFIPQIAQLGFCWQFITLAGFHADALVIDTFARDYAKKGMLAYVQ
A.capsulatus    DE-QETYIRRLGQLGYCWQFITLAGLHTTALISDQFARAYAQNGMRAYGE
                *:    :*  :.:**:*********:*  **: * **: :* .** ** :

I.batatas       KIQREERNNGVDTLAHQKWSGANYYDRVLRTVQGGITSTAAMGKGVTEEQ
S.lycoper       KIQREERSNGVDTLAHQKWSGANYYDRVLRTVQGGITSTAAMGKGVTEEQ
A.thaliana      RIQREERTHGVDTLAHQKWSGANYYDRYLKTVQGGISSTAAMGKGVTEEQ
C.sativus       RIQREERNNGVDTLAHQKWSGANYYDRYLKTVQGGISSTAAMGKGVTEEQ
P.taeda         QIQRQERANGVDTLAHQKWSGANYYDTLLKTVQGGISGTAAMGKGVTEDQ
A.capsulatus    LVQEPEMELGVDVVTHQKWSGANYVDELLKMVTGGISSTSAMGKGVTEDQ
                 :*. *   ***.::********* *  *: * ***:.*:********:*

I.batatas       FQESWTRPGTTNMGDGGVVIAKARM-
S.lycoper       FEEKWTRTGATNLGDGSVVIAKARM-
A.thaliana      FKESWTRPGADGMGEGTSLVVAKSRM
C.sativus       FKESWTREGAVNLGEEGNVVVAKSRM
P.taeda         FKADAWSSDGRALDLTNHGVVARARM
A.capsulatus    FK-----------------------
                *:
```

```
I.batatas       LQAIEDNWIAMAELKTFSETVVDAINRLNIGEPEKQRRLNEWMKHSSYEK
S.lycoper       LQAIEDNWLGMAELKTFSQCVTDAIKKMNLAEYEKQRKLNEWMNHSSYEK
A.thaliana      LQSIEDQWLGSAGLMTFSEAVVQAIKRMNLNENEKNQRLSEWLTHARYEN
C.sativus       LQALEDQWISMAQLKTFSECVTDAIMNTNATENEKRRKLDEWMNHSSYEK
P.taeda         LQDVEDKWMGMANLRLFSDVVGDAIRSLAIPQQEKSNRINLWFN---KID
A.capsulatus    LQEIEDKWNAEAGLKLFSEAVIDHIKAG-----SASNKQALIDKFLRDAR
                ** :**:* . * *  **: * : *        .  .:     .

I.batatas       CLSNFQAREIAENLGLNNLFWDWDLPRTREGFYRFKGSVMAAIVRGWAFA
S.lycoper       CLSHEQAREVAERLGLPNLFWDWDLPRTREGFYRFQGSVEAAIVRGWAFA
A.thaliana      CLSNEQGRVLAAKLGVTDLFWDWDLPRTREGFYRFQGSVAAAVVRGWAFA
C.sativus       CISNEQGREIAEKLGLKNLFWDWDLPRTREGFYRFKGSVMAAIVRGWAFA
P.taeda         QLSHFEARELAEKLGVSDVFWDWDLPRTREGFYRFQGSTKAAILRGWAFG
A.capsulatus    GKSNAEARAIAKALTGSEIYFNWDSARTREGYYRYRGGTQCAVNRAVAFA
                  *: :.* :*  *   :::::** .*****:**::*.. .*: *. **.

I.batatas       PHSDLIWMETSSPDLVECTKFAQGVKSVHPEMMLAYNLSPSFNWDASGMN
S.lycoper       EYCDLVWMETSSPDMVECTKFSQGVKTLRPELMLAYNLSPSFNWDASGMN
A.thaliana      QIADIIWMETASPDLNECTQFAEGIKSKTPEVMLAYNLSPSFNWDASGMT
C.sativus       PHADLIWMETSSPDLVECTTFAKGMKSIHPETMLAYNLSPSFNWDASGMS
P.taeda         PHADLIWMETSSPDMVECKRFAEGVKSKHPDLMLAYNLSPSFNWDASRMS
A.capsulatus    PFADLIWMESKLPDYAQAKEFADGVHAVWPEQKLAYNLSPSFNWKAAMPR
                  .*::***:  **  :.. *:.*:::  *:  ***********.*:

I.batatas       DNQMMDFIPGIAKLGYCWQFITLAGFHGNALIIDTFAKDFASRGMLAYVE
S.lycoper       DNQMMDFIPRIAKLGYCWQFITLAGFHADALIVDTFAKDFARRGMLAYVE
A.thaliana      DQQMVEFIPRIARLGYCWQFITLAGFHADALVVDTFAKDYARRGMLAYVE
C.sativus       DKQMEEFIPRIARLGFCWQFITLAGFHADALVVDTFARDYARRGMLAYVE
P.taeda         DEQMMNFIPQIAQLGFCWQFITLAGFHADALVIDTFARDYAKKGMLAYVQ
A.capsulatus    DE-QETYIRRLGQLGYCWQFITLAGLHTTALISDQFARAYAQNGMRAYGE
                *:    :*  :.:**:*********:*  **: * **: :* .** ** :

I.batatas       KIQREERNNGVDTLAHQKWSGANYYDRVLRTVQGGITSTAAMGKGVTEEQ
S.lycoper       KIQREERSNGVDTLAHQKWSGANYYDRVLRTVQGGITSTAAMGKGVTEEQ
A.thaliana      RIQREERTHGVDTLAHQKWSGANYYDRYLKTVQGGISSTAAMGKGVTEEQ
C.sativus       RIQREERNNGVDTLAHQKWSGANYYDRYLKTVQGGISSTAAMGKGVTEEQ
P.taeda         QIQRQERANGVDTLAHQKWSGANYYDTLLKTVQGGISGTAAMGKGVTEDQ
A.capsulatus    LVQEPEMELGVDVVTHQKWSGANYVDELLKMVTGGISSTSAMGKGVTEDQ
                 :*. *   ***.::********* *  *: * ***:.*:********:*

I.batatas       FQESWTRPGTTNMGDGGVVIAKARM-
S.lycoper       FEEKWTRTGATNLGDGSVVIAKARM-
A.thaliana      FKESWTRPGADGMGEGTSLVVAKSRM
C.sativus       FKESWTREGAVNLGEEGNVVVAKSRM
P.taeda         FKADAWSSDGRALDLTNHGVVARARM
A.capsulatus    FK-----------------------
                *:
```

MSA of ICL in fungi

```
C.apollonis      ----------MSPIDAEQQRYEEEVEAVKQWWNDSRWRFTKRPFTAEQIV
E.pusillum|      MVFTYAIPVDMGYLDDEEQIYLEDVAAVKKWWEDSRWRYTKRPFTAEQIV
T.equi           ----------MSYVEQEEQKFLQDVEQVKNWWKDSRWRYTKRPFTAEQIV
T.tonsurans      ----------MSYVEQEEQKFLQDVEQVKNWWKDSRWRYTKRPFTAEQIV
A.dermaditi      ----------MASIEQEEQQYLADVQAVKQWWQDSRWRYTKRPFTAEQIV
A.terreus        ----------MGFFEDEDQKYQQDVEAVKAWWKDDRWRYTKRPYTAEQIV
A.niger          ----------MASFEEEDRKYLDDVQAVQKWWQDSRWRHTKRPFTAEQIV
A.oryzae         ----------MGFLEDEDKKYLDDVQAVKAWWTDSRWRHTKRPFSAEQIV
A.fumigat        ----------MGFLEDEDQKYWDDVQAVKAWWKDSRWRYTKRPYTAEQIV
E.nidul          -----------MSYIEEEDQRYWDEVAVKNWWKDSRWRYTKRPFTAEQIV
N.cra            -MAANNMVNPAVDPALEDELFAKEVEEVKKWWSDSRWRQTKRPFTAEQIV
                                  *:     :   *: ** *.*** ****::*****

C.apollonis      SKRGNLKIQYPSAVMAKKLWNTVEKRFQEKDASYTYGCLDPIMVSQMAKY
E.pusillum|      QKRGNLKIDYPSNAQAKKLWKIVEGRFANKDVSFTYGCLEPTMLTQMAKF
T.equi           AKRGTLTINYPSNAQSKKLWKILEHRFANKDASFTYGCLEPTMLTQMVKY
T.tonsurans      AKRGTLTIDYPSNAQSKKLWKILEHRFANKDASFTYGCLEPTMLTQMVKY
A.dermaditi      AKRGNLKIEYPSNVQSKKLWRLLEDHFKNKTASFTYGCLEPTMLTQMVKY
A.terreus        AKRGNLKIDYPSNVQSKKLWKILESNFQQKVASFTYGCLEPTMVTQMAKY
A.niger          AKRGTLKIEYPSNAMSKKLWKILESNFEKKVPSFTYGCLEPTMLTQMAKY
A.oryzae         AKRGNLKIEYPSNVQSKKLWKILESNFENKVASFTYGCLEPTMVTQMAKY
A.fumigat        AKRGNLKIHYPSNDQSKKLWKILESNFD-KVASFTYGCLEPTMLTQMAKY
E.nidul          AKRGNLKIEYPSNVQAKKLWGILERNFN-KEASFTYGCLDPTMVTQMAKY
N.cra            SKRGNLKIEYASNAQAKKLWKILEDRFAKRDASYTYGCLEPTMVTQMAKY
                  ***.*.*.*.*   :****  :* .*  :  *:*****:* *::**.*:

C.apollonis      LDTVYVSGWQCSSTASSSNEPGPDLADYPMTTVPNKVEHLFMAQLFHDRK
E.pusillum|      LDTVYVSGWQSSSTASSTDEPSPDLADYPMNTVPNKVHHLWMAQLFHDRK
T.equi           LDTVYVSGWQSSSTASSTDEPSPDLADYPMNTVPNKVNQLFLAQLFHDRK
T.tonsurans      LDTVYVSGWQSSSTASSTDEPSPDLADYPMNTVPNKVNQLFLAQLFHDRK
A.dermaditi      LDTVYVSGWQSSSTASSTDEPSPDLADYPMNTVPNKVNQLWMAQLFHDRK
A.terreus        LDTVYVSGWQSSSTASSTDEPSPDLADYPMNTVPNKVNQLFMAQLFHDRK
A.niger          LDTVYVSGWQSSSTASSTDEPSPDLADYPMNTVPNKVNQLFLAQLFHDRK
A.oryzae         LDTVYVSGWQSSSTASSTDEPSPDLADYPMNTVPNKVNQLWMAQLFHDRK
A.fumigat        LDTVYVSGWQSSSTASSTDEPSPDLADYPMNTVPNKVNQLFMAQLFHDRK
E.nidul          LDTVYVSGWQSSSTASSTDEPSPDLADYPMNTVPNKVNHLWMAQLFHDRK
N.cra            LDTVYVSGWQSSSTASSSDEPGPDLADYPYTTCPNKVGHLFMAQLFHDRK
                 **********.******::**.******* .* **** :*::********

C.apollonis      QREERLSTPKANRAGVANTDFLRPMVADADTGHGGLTAVMKLTKMFVEKG
E.pusillum|      QREERLTTPKSERANVANTDFLRPIIADADTGHGGLTAVMKLTKLFVENG
T.equi           QREERLRSPKEQRNSLPNIDYLRPIIADADTGHGGLTAVMKLTKLFIERG
T.tonsurans      QREERLRSPKEQRNSLPNIDYLRPIIADADTGHGGLTAVMKLTKLFIERG
A.dermaditi      QREERLRASKSQRASLANIDYLRPIIADADTGHGGLTAVMKLTKLFIERG
A.terreus        QREERVTTPKEQRGKVANVDYLRPIIADADTGHGGLTAVMKLTKLFVERG
A.niger          QREERITTPKEQRGKVANVDYLRPIIADADTGHGGLTAVMKLTKLFIERG
A.oryzae         QREERITTPKEKRGNVANIDYLRPIIADADTGHGGLTAVMKLTKLFVERG
A.fumigat        QREERITTPKDQRSKLPNIDYLRPIIADADTGHGGLTAVMKLTKLFIERG
E.nidul          QREERMTTPKDQRHKVTNVDYLRPIIADADTGHGGLTAVMKLTKLFVERG
N.cra            QRQERLSVPKDQREKLANIDYLRPIVADADTGHGGLTAVMKLTKLFIEKG
                 **:**:  .* :*  :.* *:***::******************:*:*.*
```

```
C.apollonis      AAGIHIEDQAPGTKKCGHMAGKVLVPISEHINRLVAIRAQADIMG-TDLL
E.pusillum|      AAGIHIEDQAPGTKKCGHMAGKVLVPISEHINRLVAIRAQADIMG-SDLL
T.equi           AAGIHIEDQAPGTKKCGHMAGKVLVPISEHINRLVAIRAQADIMG-TDLL
T.tonsurans      AAGIHIEDQAPGTKKCGHMAGKVLVPISEHINRLVAIRAQADIMG-TDLL
A.dermaditi      AAGIHVEDQAPGTKKCGHMAGKVLVPISEHINRLVAIRAQADIMG-TDLL
A.terreus        AAGIHIEDQAPGTKKCGHMAGKVLVPISEHINRLVAIRAQADIMG-TDLL
A.niger          AAGIHIEDQAPGTKKCGHMAGKVLVPISEHINRLVAIRAQADIMG-TDLL
A.oryzae         AAGIHIEDQAPGTKKCGHMAGKVLVPISEHINRLVAIRAQADIMG-TDLL
A.fumigat        AAGIHIEDQAPGTKKCGHMAGKVLVPISEHINRLVAIRAQADIMGETDLL
E.nidul          AAGIHIEDQAPGTKKCGHMAGKVLVPISEHINRLVAIRAQADIMG-TDLL
N.cra            AAGIHIEDQAPGTKKCGHMAGKVLVPIQEHINRLVAIRAQADIMG-SDLL
                 *****:**********************.****************** :***

C.apollonis      AIARTDSEAATLITSTIDHRDHGFIMGATNPALQPLNDLMIAAEHAGKNG
E.pusillum|      AVARTDSEAATLITSTIDHRDHSFILGATNPSLQPLNDLMVAAERAGKMG
T.equi           AIARTDSEAATLITSTIDPRDHAFIVGSTNSSIEPLNDLMVAAEQAGKNG
T.tonsurans      AIARTDSEAATLITSTIDPRDHAFIVGSTNSSIEPLNDLMVAAEQAGKNG
A.dermaditi      AVARTDSEAATLITSTIDHRDHDFVIGSTNPNIKPLNDLMMAAEQAGKTG
A.terreus        AIARTDSEAATLITSTIDYRDHAFIVGATNPNLQPLNDLMVAAEQAGKNG
A.niger          AIARTDSEAATLITSTIDYRDHAFVVGSTNPNLQPLNDLMVAAEQAGKNG
A.oryzae         AIARTDAEAATLITSTIDHRDHAFIVGSTNPNIQPLVDLMVAAEQAGKQG
A.fumigat        AIARTDSEAATLITSTIDHRDHAFIVGSTNPNLQPLNDLMLAGEQAGKTG
E.nidul          AIARTDSEAATLITSTIDHRDHPFIIGSTNPDIQPLNDLMVMAEQAGKNG
N.cra            CIARTDAEAATLITTTIDPRDHAFILGCTNPDLEPLAHLMMKAEAEGKTG
                 .:****:*******:*** *** *::*.**. ::** .**: .*  ** *

C.apollonis      MELQAIEDQWTAQAGLKLFHEAVADTINAGV-HVNKAELISQFNILSKGK
E.pusillum|      ETLANIEEEWIKSAQLKLFDEAVIDAISAGV-HVDKASLIKKYKDAVNGK
T.equi           AELQQIEDEWTSKAGLKRFQDAAIDRINATASISNKKAAIDKFLADIKGK
T.tonsurans      AELQQIEDEWTSKAGLKRFQDAAIDRINATASISNKKAAIDKFLADIKGK
A.dermaditi      AQLQEIEDKWNAEAGLKLFSEAVVDQIKASS-VSNKQAVIDKFLRDVKGK
A.terreus        DQLQAIEDQWVAQAGLKLFNDAVIDEINKGS-LPNKPAVIADYLKAVKGK
A.niger          NELQAIEDQWVAQAGLKLFSEAVIDTINKSS-SGNKKALIDEYLKKAKGK
A.oryzae         DELQAIEDQWVAQAGLKLFNDAVIEAINKGA-HSNKQSLIDQYLKAAKGK
A.fumigat        EELQAIEDQWIAQAGLKLFDDAVVDTIKAGV-HVNKDALIKEYLTAAKGK
E.nidul          AELQAIEDEWLAKAGLKLFNDAVVDAINNSP-LPNKKAAIEKYLTQSKGK
N.cra            AQLQAIEDDWLAKADLKRFDEAVLDVIAKGK-FSNAKDLAAKYQAAVKGK
                  *  **:.*  .* ** * :*. : *       :      .:    :**

C.apollonis      --SNSEARAIAKGLTGVDIHFDWDSARTREGYYRYQGGCECAINRAIAYA
E.pusillum|      --SNSEARAISKGLTGVDIHWDWDAPRTREGYYRYQGGCQCAVNRAIAYA
T.equi           --SNSEARAIAKQLTGSDIYWNWDSPRTREGFYRYQGGCECAVNRAVAYG
T.tonsurans      --SNSEARAIAKQLTGSDIYWNWDSPRTREGFYRYQGGCECAVNRAVAYG
A.dermaditi      --SNQEARAIAKELTGVDVYFCWDAARTREGYYRYRGGTQCAVNRAVAFA
A.terreus        --SNSEARAIAKGILGVDIYWNWDSPRTREGYYRYQGGTQCAINRAVAYG
A.niger          --SNREARAIAKDITGVDIYWDWDAPRTREGFYRYQGGTQCAVNRAIAYG
A.oryzae         --SNLEARAIAKNITGQDIYFNWDAPRTREGYYRLQGGTQCAINRGLAYA
A.fumigat        --SNSEARAIAKGITGVDIYWDWDAPRTREGYYRYQGGTQCAINRAVAYA
E.nidul          --SNLEARAIAKEIAGTDIYFDWEAPRTREGYYRYQGGTQCAINRAVAYA
N.cra            QISNREARAIARQLLGQEIFFDWESPRTREGYYRLKGGCDCSINRAISYA
                   ** *****:: : * ::.: *::.*****:** :** :*::**.:::.
```

```
C.apollonis      PYADLIWMESKLPDYAQAKEFADGVHAVWPEQKLAYNLSPSFNWKKAMPR
E.pusillum|      PYADMIWMESKMPDYAQAKEFADGVHAVWPEQKLAYNLSPSFNWKAAMSQ
T.equi           PFADLIWMESKLPDYAQAKEFAEGVHAVWPEQKLAYNLSPSFNWKTAMPR
T.tonsurans      PFADLIWMESKLPDYAQAKEFAEGVHAVWPEQKLAYNLSPSFNWKTAMPR
A.dermaditi      PFADLIWMESKLPDYAQAKEFADGVHAVWPEQKLAYNLSPSFNWKAAMPC
A.terreus        PFADLIWMESKLPDYAQAKEFADGVHAVWPEQKLAYNLSPSFNWKKAMPR
A.niger          PFADLIWMESKLPDYAQAKEFADGVHAVWPEQKLAYNLSPSFNWKKAMPR
A.oryzae         PFADLIWMESKLPDYAQAKEFAEGIHAVWPEQKLAYNLSPSFNWKKAMPR
A.fumigat        PFADLIWMESKLPDYAQAKEFADGVHAVWPEQKLAYNLSPSFNWKKAMPR
E.nidul          PFADLIWMESKLPDYKQAKEFADGVHAVWPEQKLAYNLSPSFNWKKAMPR
N.cra            PYCDAIWMESKLPDYAQAEEFAK-GPRVWPEQKLAYNLSPSFNWKTAMGR
                 *:.* ******:*** **:***.    ******************* **

C.apollonis      SEQETYIQRLAKLGYCWQFITLAGLHQSALMADTFSKAYAKQGMRAYGEL
E.pusillum|      SDQETYIARLGQLGYCWQFITLAGLHSTALISHQFSQAFAKQGMRAYGSL
T.equi           DEQETYIRRLGQLGYSWQFITLAGLHTTALISDQFAKAYAKQGMRAYGEM
T.tonsurans      DEQETYIRRLGQLGYSWQFITLAGLHTTALISDQFAKAYAKQGMRAYGEM
A.dermaditi      DEQETYIRRLGELGYCWQFITLAGLHTTALISDQFAKAYAMQGMRAYGEL
A.terreus        DEQETYIKRLGALGYCWQFITLAGLHTTALITDTFAKAYAKQGMRAYGEL
A.niger          DEQETYIQRLGALGYCWQFITLAGLHTTALITDQFAKAYAKQGMRAYGEL
A.oryzae         DEQETYIKRLGALGYCWQFITLAGLHTTALISDQFAKAYSKQGMRAYGEL
A.fumigat        EEQETYIKRLGALGYAWQFITLAGLHTTALISDQFARAYAKQGMRAYGEL
E.nidul          DEQETYIKRLGALGYAWQFITLAGLHTTALISDTFAKAYAKQGMRAYGEL
N.cra            DDQETYIRRLAKLGYCWQFITLAGLHTTALISDQFAKAYSKIGMRAYGEL
                 .:***** **. ***.********** :**::. *::*::  ******.:

C.apollonis      IQEPEMDNKIDVVTHQKWSGANYMDEMLKMVTGGVSSTSAMGKGVTEDQF
E.pusillum|      IQEPEMALKCDVVTHQKWSGANYVDNLLKMVTGGVSSTSAMGKGVTEDQF
T.equi           VQEPEMDNKVDVVTHQKWSGANYVDELLKMVTGGISSTSAMGKGVTEEQF
T.tonsurans      VHESEMDNKGDVVTHQKWSGANYVDELLKMVTGGISSTSAMGKGVTEEQF
A.dermaditi      VQEPEMELGVDVVTHQKWSGANYVDELLKMVTGGISSTSAMGDGVTEDQF
A.terreus        VQEPEMENGVDVVTHQKWSGANYVDNLLKMVTGGVSSTAAMGKGVTEDQF
A.niger          VQEPEMEQGVDVVTHQKWSGANYVDNMLKMVTGGISSTAAMGKGVTEDQF
A.oryzae         VQEPEMEQGVDVVTHQKWSGANYVDNLLKMVTGGVSSTAAMGKGVTEDQF
A.fumigat        VQEPEMEQGVDVVTHQKWSGANYVDNMLKMLTGGVSSTAAMGKGVTEDQF
E.nidul          VQEPEMANGVDVVTHQKWSGANYVDNMLKMITGGVSSTAAMGKGVTEDQF
N.cra            VQEPEIDNGVDVVKHQKWSGATYVDELQKMVTGGVSSTAAMGKGVTEDQF
                 ::*.*:    ***.*******.*:*:: **:***:***:***.****:**

C.apollonis      SEKGAGASDGAS
E.pusillum|      KGH---------
T.equi           K-----------
T.tonsurans      K-----------
A.dermaditi      K-----------
A.terreus        KH----------
A.niger          KH----------
A.oryzae         KH----------
A.fumigat        KH----------
E.nidul          KS----------
N.cra            H-----------
```

Subfamily 3

27 Sequences of ICL in bacteria were taken MSA was carried out using ClustalW. The sequences showed high similarity and the sequences were highly conserved which showed that these may have come from same ancestors.

MSA of ICL in subfamily 3

```
C.maris        MSNYQSAIEAVKAIKEKAGNSWDAINPESVARMRAQNKFKTGLDIAQYTA
C.piezophi     MSNYQSAIEAVKAIKEKAGSSWAAINPESVARMRTQNRFKTGLEIAQYTA
C.psychr       MSNYQSAIEAVQAIKAKAGSSWDAINPESIARMRAQNKFKTGLEIAQYTA
C.necator      MAQYQDDIKAVAGLKENHGSAWNAINPEYAARMRAQNKFKTGLDIAKYTA
P.fulva        MSAYENDIKAVTALKEAAGNSWSAINPESVARMRAQNRFKTGLDIAKYTA
P.aerugi       MSAYQNEIKAVAALKEKNGSSWSAINPEYAARMRIQNRFKTGLDIAKYTA
A.vinel        MSAYQNEIKALAALKEKNGSAWSAINPEYAARMRIQNRFKTGLDVAKYTA
A.ehrli        MSQYRNDIEEVAGLRQAHEGTWDAINPEYVARMRAQNRFKTGLDIAKYTA
M.adhae        -MPYAQDVDQIASLLKQHP-TWNAINPKHAARMRAQNKFKTGLDIAKYTA
S.agarivor     MSQYTNDIAAIAKLRDTQGSKWEAINPEYAARMRAQNRFKTGLDIAKYTA
S.chloroph     -MTYSQEITKAGTTIGGQ-GHWDGIAAESVARMRLQNRFHTGLDIARYTA
N,penta        -MTYQQKIIEAGRTIGTE-PNWTGIEPESVARMSLQNRFNTGLDIARYTA
M.posidoni     MQTYKNKMQQASQLIGQQGTTWAGMNPEYVARMQLQNRFNTGLDIAKYTA
M.medit        MSNYNKDIDDTATLIASQGDTWKAINPEFAARMKAQNRFRTGLDIARYTA
A.gerne        MTTYQTAIDAIRELKAKFGNTWADISPEDAARMQMQNRFKTGLDIAKYTA
Acineto        MTTYQTAIDAIRELKAKFGNTWADISPEDAARMQMQNRFKTGLDIAKYTA
A.oleiv        MTTYQTAIDAIRELKAKFGSTWRDISPEDAARMQMQNQFKTGLDIAKYTA
A.venet        MSTYLTAIDAIRELKAKFGNTWRDISPEDAARMQMQNRFKTGLDIAKYTA
A.bouve        MTTYQTAIDAIRELKAKFGSTWADISPEDAARMQMQNRFKTGLDIAKYTA
A.calcoace     MTTYQTAIDAVRQLKAKFGNTWADISPEDAARMQLQNRFKTGLDIAKYTA
A.tando        MTTYQTAIDAIRELKAKFGNTWADISPEDAARMQLQNRFKTGLDIAKYTA
A.lwoff        MTTYQTAIDAIRELKAKFGNTWADISPEDAARMQIQNRFKTGLDIAKYTA
A.baumanni     MTTYQTAIDAVRELKAKFGNTWADISPEDAARMQLQNRFKTGLDIAKYTA
A.ursin        -MTYQSALEHVRAVKNKLGGTWDAIRPEDAARMIVQNRFHTGLDIAKYTA
A.bayly        -MTYQSALEHVRAVKNKLGGTWDAIRPEDAARMIVQNRFHTGLDIAKYTA
A.junii        -MTYQSALEHVRAVKEKLGGTWGAIRPEDAARMMVQNRFHTGLDIAKYTA
                  *   :                 *  : .:  ***  **:*.***::*:***

C.maris        DIMRADMAAFDADKTKYTQSLGCWHGFIGQQKMISIKKHFDGKTDRRYLY
C.piezophi     DIMRRDMDAYDADSSQYTQSLGCWHGFVGQQKMISIKKHFDGKTDRRYLY
C.psychr       DIMRADMAAFDADKTQYTQSLGCWHGFVGQQKMISIKKHFDGKTDRRYLY
C.necator      KIMRADMAAYDADSSKYTQSLGCWHGFIGQQKMISIKKHFN-STERRYLY
P.fulva        AIMRKDMAEYDADSSVYTQSLGCWHGFIGQQKLISIKKHLK-TTNKRYLY
P.aerugi       AIMRKDMAEYDADSSVYTQSLGCWHGFIGQQKLISIKKHLK-TTNKRYLY
A.vinel        AIMRKDMAEYDADSSVYTQSLGCWHGFIGQQKLISIKKHLK-TTDKKYLY
A.ehrli        KIMRQDMAEYDADPAQYTQSLGCWHGFVGQQKMIAIKRHFG-TTNKRYLY
M.adhae        KIMREDMANYDKDTSQYTQSLGCWHGFIGQQKMLSIKKHFG-TTKRRYLY
S.agarivor     KIMREDMAAYDANTAEYTQSLGCWHGFIGQQKLIAIKKHLG-NTKRRYLY
S.chloroph     GIMRRDMAAYDADPANYTQSLGCWHGFIGQQKMISIKKHFG-TTKGRYLY
N,penta        RIMREDMAAYDADPANYTQSLGCWHGFIGQQKMISIKKHFG-STKRRYLY
M.posidoni     KIMREDMASYDQDQANYTQSLGCWHGFIGQQKMISIKKHFG-TTRSRYLY
M.medit        KIMRADMERYDSDPSQYTQSLGCWHGFIGQQKMISIKKHFN-STDRRYLY
A.gerne        AIMRRDMAEYDADSSKYTQSLGCWHGFIAQQKMIANKKYFG-TTSKRYIY
Acineto        AIMRRDMAEYDADSSKYTQSLGCWHGFIAQQKMIANKKYFG-TTSKRYIY
A.oleiv        AIMRRDMAAYDADSSKYTQSLGCWHGFIAQQKMIANKKYFG-TTERRYIY
A.venet        AIMRRDMAAYDADSSKYTQSLGCWHGFIAQQKMIANKKYFG-TTERRYIY
A.bouve        AIMRRDMAAYDADSSKYTQSLGCWHGFIAQQKMIANKKYFG-TTERRYIY
A.calcoace     AIMRRDMAEYDADSSKYTQSLGCWHGFIAQQKMIANKKYFG-TTSKRYIY
A.tando        AIMRRDMAAYDADSSKYTQSLGCWHGFIAQQKMIANKKYFG-TTERRYIY
A.lwoff        AIMRRDMAAYDADSSKYTQSLGCWHGFIAQQKMIANKKYFG-TTERRYIY
A.baumanni     AIMRRDMAEYDADSSKYTQSLGCWHGFIAQQKMIANKKYFG-TTSKRYIY
A.ursin        AIMRKDMAEYDADNSSYTQSLGCWHGFIAQQKMIANKKYFG-TTNKRYIY
A.bayly        AIMRKDMAEYDADNSSYTQSLGCWHGFIAQQKMIANKKYFG-TTNKRYIY
A.junii        AIMRKDMAEYDADSSKYTQSLGCWHGFIAQQKMIANKKYFG-TTSKRYIY
                *** **  :* : : ***********:.***::: *:::  .*  :*:*
```

```
C.maris        LSGWMVAALRSEFGPLPDQSMHEKTSVASLVAELYTFLRQADARELGGLF
C.piezophi     LSGWMVAALRSEFGPLPDQSMHEKTSVASLVAELYTFLRQADARELGGLF
C.psychr       LSGWMVAALRSEFGPLPDQSMHEKTSVASLVAELYTFLRQADARELGGLF
C.necator      LSGWMVAALRSEFGPLPDQSMHEKTSVSALIRELYTFLRQADARELGGLF
P.fulva        LSGWMIAALRSDFGPLPDQSMHEKTAVAALIEELYTFLRQADARELDLLF
P.aerugi       LSGWMVAALRSDFGPLPDQSMHEKTAVSGLIEELYTFLRQADARELDLLF
A.vinel        LSGWMVAALRSDFGPLPDQSMHEKTAVSSLIEELYTFLRQADARELDLLF
A.ehrli        LSGWMVAALRSEFGPLPDQSMHEKTAVPALIEELYTFLRQADAREMNHLF
M.adhae        LSGWMVAALRSEFGPLPDQSMHEKTAVSGLIEELYTFLRQADAWELNHLF
S.agarivor     LSGWMVAALRSEFGPLPDQSMHEKTSVAALIEELYTFLRQADARELGGLF
S.chloroph     LSGWMVAALRSKFGPLPDQSMHEKTSVPALIEELYTFLKQADARELGMMF
N,penta        LSGWMVAALRSEFGPLPDQSMHEKTSVPALIEELYTFLKQADARELGMMF
M.posidoni     LSGWMVAAMRSEFGPLPDQSMHEKTSVPALIEELYTFLKQADAREMQHLF
M.medit        LSGWMVAALRSEFGPLPDQSMHEKTSVSSLIGELYTFLRQADARELGELF
A.gerne        LSGWMVAALRSEFGPLPDQSMHEKTSVPALIEEIYTFLRQADAKELNDLF
Acineto        LSGWMVAALRSEFGPLPDQSMHEKTSVPALIEEIYTFLRQADAKELNDLF
A.oleiv        LSGWMVAALRSEFGPLPDQSMHEKTSVPALIEEIYTFLRQADAKELNDLF
A.venet        LSGWMVAALRSEFGPLPDQSMHEKTSVPALIEEIYTFLRQADAKELNDLF
A.bouve        LSGWMVAALRSEFGPLPDQSMHEKTSVPALIEEIYTFLRQADAKELNDLF
A.calcoace     LSGWMVAALRSEFGPLPDQSMHEKTSVPALIEEIYTFLRQADAKELNDLF
A.tando        LSGWMVAALRSEFGPLPDQSMHEKTSVPALIEEIYTFLRQADAKELNDLF
A.lwoff        LSGWMVAALRSEFGPLPDQSMHEKTSVPALIEEIYTFLRQADAKELNDLF
A.baumanni     LSGWMVAALRSEFGPLPDQSMHEKTSVPALIEEIYTFLRQADAKELNDLF
A.ursin        LSGWMVAALRSEFGPLPDQSMHEKTAVPKLIAEIYTFLRQADAKELNDLF
A.bayly        LSGWMVAALRSEFGPLPDQSMHEKTAVPKLIAEIYTFLRQADAKELNDLF
A.junii        LSGWMVAALRSEFGPLPDQSMHEKTAVPKLIAEIYTFLRQADAKELNDLF
               *****:**:**.**************:*. *: *:****:**** *:  :*

C.maris        RDLDAASDGD----KAAIQDKIDNHVTHVVPIIADIDAGFGNAEATYIMA
C.piezophi     RDLDKASEGD----KAAIQYKIDNHVTHVVPIIADIDAGFGNAEATYLMA
C.psychr       RELDAAAEGD----KAAIQSQIDNHVTHVVPIVADIDAGFGNAEATYLMA
C.necator      RELDAAQGPA----KAAIQAKIDNHVTHVVPIIADIDAGFGNAEATYLLA
P.fulva        SALDAARESGDRAKQSEIQAQIDNFETHVVPIIADIDAGFGNPEATYLLA
P.aerugi       TGLDAARAAGDKAKEAELLAQIDNFETHVVPIIADIDAGFGNAEATYLLA
A.vinel        SALDDARNAGDKAKEQEILKQIDGFETHVVPIIADIDAGFGNAEATYLLA
A.ehrli        RELDAAREAGDEAKANEILDKIENFETHIVPIIADIDAGFGNEEATYLLA
M.adhae        RALEEAENAGDNAKAEELIKQIDNHETHVVPIIADIDAGFGNAEATYLLA
S.agarivor     RQLDEARAQDDQAKAKEIQSKIDNFETHVVPIIADIDAGFGNEEATYLLA
S.chloroph     RDLDAAREAGDEVKAQKIQHDIDTYETHVVPIVADIDAGFGNAEATYLLA
N,penta        RELDKARDAGDEMEARRLEAAIDNYETHVVPIIADIDAGFGNAEATYLLA
M.posidoni     KELDEARQAGDQVREQAAQESIDQFETHVVPIIADIDAGFGNEEATYLLA
M.medit        RSLDAARDAGDKATEEKVQAEVDNYQTHIVPIIADIDAGFGNEEATYLLA
A.gerne        RALKKAQEAGDTAKAAEITSQIDNFQTHVVPIIADIDAGFGNEEATYLLA
Acineto        RALKKAQEAGDTAKAAEITSQIDNFQTHVVPIIADIDAGFGNEEATYLLA
A.oleiv        RALKKANEAGDTAKAAEITAQIDGFQTHIVPIIADIDAGFGNEEATYLLA
A.venet        RALKKAQEAGDTAKAAEITAQIDGFQTHVVPIIADIDAGFGNEEATYLLA
A.bouve        RALKKANEAGDTAKAAEITAQIDGFQTHVVPIIADIDAGFGNEEATYLLA
A.calcoace     RALKKAQEAGDTAKAAEITAQIDNFETHVVPIIADIDAGFGNEEATYLLA
A.tando        RALKKAQEAGDTAKAAEITAQIDNFETHVVPIIADIDAGFGNEEATYLLA
A.lwoff        RALKKAQEAGDTAKAAEITSQIDNFQTHVVPIIADIDAGFGNEEATYLLA
A.baumanni     RALNKAKEAGDSAKVAEITAQIDNFETHVVPIIADIDAGFGNEEATYLLA
A.ursin        RALQKAEAAGDSAKVKEIEASIDNFETHVVPIIADIDAGFGNEEATYLLT
A.bayly        RALQKAEAAGDSAKVKDIESQIDNFETHVVPIIADIDAGFGNEEATYLLT
A.junii        RALQKAEQAGDSAKAAEIISQIDNFESHVVPIIADIDAGFGNEEATYLLT
                 *. *                 :: . :*:***:********* ****:::
```

```
C.maris        KQMIEAGACALQIENQVADEKQCGHQDGKVTVPHADFHSKIRALRHAFLE
C.piezophi     KQMIEAGACALQIENQVADEKQCGHQDGKVTVPHADFHSKIRALRHAFLE
C.psychr       KQMIEAGACALQIENQVADEKQCGHQDGKVTVPHADFHSKIRALRHAFLE
C.necator      KQFIEAGACCIQIENQVSDEKQCGHQDGKVTVPHEDFLAKIRAIRYAFLE
P.fulva        KKMIEAGACCIQIENQVSDEKQCGHQDGKVTVPHADFLAKINAVRYAFLE
P.aerugi       KKMIEAGACCIQIENQVSDEKQCGHQDGKVTVPHIDFLAKINAVRYAFLE
A.vinel        RKMIEAGACCIQIENQVSDEKQCGHQDGKVTVPHEDFLAKINAVRYAFLE
A.ehrli        KKMIEAGACCIQIENQVADEKQCGHQDGKVTVPHADFLAKVRAVRYAFLE
M.adhae        KQMIEAGACCIQIENQVSDEKQCGHQDGKVTVPHADFLSKINAVRLAFLE
S.agarivor     KKMIEAGACCIQIENQVSDEKQCGHQDGKVTVPHADFLAKIRAVRYAFLE
S.chloroph     KKFIEAGACCIQIENQVSDEKQCGHQDGKVTVPHEDFLAKIRAVRYAFLE
N,penta        KKFIEAGACCIQIENQVSDEKQCGHQDGKVTVPHEDFIAKIRAVRYAFME
M.posidoni     KKMIEAGACCIQIENQVSDAKQCGHQDGKVTVPHEDFLAKINAVRYAFLE
M.medit        KQMIEAGACCIQIENQVSDEKQCGHQDGKVTVPHEDFLAKIRAVRYAFLE
A.gerne        KKMIEAGACALQIENQVSDAKQCGHQAGKVTVPHEDFISKIHALRYAFLE
Acineto        KKMIEAGACALQIENQVSDAKQCGHQAGKVTVPHEDFISKIHALRYAFLE
A.oleiv        RKMIEAGACALQIENQVSDAKQCGHQAGKVTVPHEDFIAKIHALRYAFLE
A.venet        KKMIEAGACALQIENQVSDAKQCGHQAGKVTVPHEDFIAKIHALRYAFLE
A.bouve        KKMIEAGACALQIENQVSDAKQCGHQAGKVTVPHEDFISKIHALRYAFLE
A.calcoace     KKMIEAGACALQIENQVSDAKQCGHQAGKVTVPHEDFIAKIHALRYAFLE
A.tando        KKMIEAGACALQIENQVSDAKQCGHQAGKVTVPHEDFIAKIHALRYAFLE
A.lwoff        KKMIEAGACALQIENQVSDAKQCGHQAGKVTVPHEDFIAKIHALRYAFLE
A.baumanni     KKMIEAGACALQIENQVSDAKQCGHQAGKVTVPHEDFIAKIHALRYAFLE
A.ursin        KQMIEAGACAIQIENQVSDAKQCGHQAGKVTVPHEDFLAKINAVRYAFLE
A.bayly        KQMIEAGACAIQIENQVSDAKQCGHQAGKVTVPHEDFLAKINAVRYAFLE
A.junii        KQMIEAGACAIQIENQVSDAKQCGHQAGKVTVPHEDFLAKINAVRYAFLE
               :::******.:******:* ****** ******* ** :*:.*:* **:*

C.maris        LGIDNGIIVSRTDSEGAGLTKEIAVVKEPGDQGDIYNSFLDVEEIDVSDM
C.piezophi     LGIDNGIIVSRTDSEGAGLTKEIAVVKEPGDQGDIYNSFLDVEEIDVADM
C.psychr       LGIDNGLIVARTDSEGAGLTKEIAVVKEPGDSGDIYNSYLDVEEISAADM
C.necator      LGVDDGIIVARTDSLGAGLTKQIAVTNTPGDLGDQYNSFLDCEELSADAL
P.fulva        LGVDDGVIVARTDSLGAGLTKQIAVTSEPGDLGDQYNSFLDCDEISPAEL
P.aerugi       LGVDDGVIVARTDSLGAGLTKQIAVTNEPGDLGDLYNSFLDCEEISEAEL
A.vinel        LGVDDGVIVARTDSLGAGLTKQIAVTRQEGDLGDKYNAFLDCEEISPAEM
A.ehrli        LGVNDGVIVARTDSLGAGLTKQIAVTDEPGDLGDKYNSFLDGDYIDSAED
M.adhae        LGVDDGVIVARTDSLGAGLTQKIAVTNEPGDLGDQYNSFIDGEVIEKAED
S.agarivor     LGVEDGVIVARTDSLGAGLTKQIAVTKEPGDLGDLYNSFLDAEEVDAADM
S.chloroph     LGVDDGIIVARTDSLGAGLTKQIAYSKEEGDLGDQYNSFLDVEEVTDLGS
N,penta        LGVDDGVIVARTDSLGAGLTKQIAFVREAGDIGDQYNAFLDCEEVDAAKL
M.posidoni     LGVEDGVIVARTDSLGAGLTQKIPVSQSAGDLAEQYSAFIDGEEVTSLDQ
M.medit        LGVEDGVIVARTDSLGAGLTKQIAVTAKEGDLGDQYNSFLDVEEVDASNM
A.gerne        MGLDDGIIVARTDSEGADLTQKIPVVKEPGDIASQYISYLDTQEIDIADA
Acineto        MGLDDGIIVARTDSEGADLTQKIPVVKEPGDIASQYISYLDTQEIDIADA
A.oleiv        MGLDDGIIVARTDSEGADLTQKIPVVKEPGDIASQYISYLDTVEIDIADA
A.venet        MGLDDGIIVARTDSEGADLTQKIPVVKEPGDIASQYISYLDTVEIDIADA
A.bouve        MGLDDGIIVARTDSEGADLTQKIPVVKEPGDIASQYIGFLDTVEIDIADA
A.calcoace     MGLEEGIIVARTDSEGADLTQKIPVVKEPGDIASQYISYLDTQEIDIADA
A.tando        MGLDDGIIVARTDSEGADLTQKIPVVKEPGDIASQYISYLDTVEIDIADA
A.lwoff        MGLDDGIIVARTDSEGADLTQKIPVVKEPGDIASQYISYLETSEIDIADA
A.baumanni     MGVEDGVIVARTDSEGADLTQKIPVVKEPGDIASKYISYLDTQEIDISEA
A.ursin        LGVDEGVIVARTDSEGADLTQKIPVSKEKGDLASQYISYLDTQEIDIADA
A.bayly        LGVDEGVIVARTDSEGADLTQKIPVSKEKGDLASQYISYLDTQEIDIADA
A.junii        LGVDEGVIVARTDSEGADLTQKIPVSKEKGDLASQYISYLDTKEIDISEA
               :*:::*:**:**** **.**::*.     ** .. * .:::   :
```

```
C.maris        AE-GDVCFNRDGKLVRPKRLPSGLYQFRQGTGEERCVFDSIEAIKAGADL
C.piezophi     AE-GDVCFNRDGKLVRPKRLPSGLYQFRQGTGEERCVFDSIEAIKAGADL
C.psychr       AE-GDVCFNRDGKLVRPKRLPSGLYQFRQGTGHERCVFDCIGALNAGADL
C.necator      GN-GDVIIKRDGKLLRPKRLPSNLFQFRAGTGEARCVLDCVTALQNGADL
P.fulva        KN-GDVVINREGKLLRPKRLPSNLFQFRKGTGEDRCVLDSITSLQNGADM
P.aerugi       GN-GDVVIKREGKLLRPKRLASNLFQFRKGTGEDRCVLDCITSLQNGADL
A.vinel        KN-GDVVLNRNGKLLRPKRLASGLFQFRKGTGEDRCVLDCITSLQNGADL
A.ehrli        INNGDVVVKSEGKLLKPKRLASGLYEFRKGTGFDRVVLDCITSLKNGADL
M.adhae        INNGDVVIKQNGQLVRPKRLASGLFQFKPGTGEDRVVLDCITSLQNGADL
S.agarivor     KN-GDVIINRNGKLLRPKRLASGLFQFRKGTGEDRVVLDCITSLQNGADL
S.chloroph     LRG-DVVLARDGKLFRPKRLPSNLFQFRAGTGEDRCVLDCITSLQNGADL
N,penta        GHG-DVLISRDGKLMRPKRLASNLYQFRPGTGEDRCVLDCIQSLQNGADL
M.posidoni     MQSGDMLLKQGGTCIKPTRLANGLVRFKPDTGADRVVLDCITSLQHGADL
M.medit        KNG-DVLLNQKGKLVRPKRLPSNLFQFKKGSGEDRVVLDCITSLQNGADL
A.gerne        QE-DEILIKRDGKLHRPKRLASGLYQFREGTQIDRVVLDCVSSLQNGADL
Acineto        QE-DEILIKRDGKLHRPKRLASGLYQFREGTQIDRVVLDCVSSLQNGADL
A.oleiv        QE-DEILIKRNGKLHRPKRLASGLYQFRADTQIDRVVLDCVSSLQNGADL
A.venet        QE-DEILIKRNGKLHRPKRLASGLYQFRPDTQIDRVVLDCVSSLQNGADL
A.bouve        QE-DEILIKRDGKLHRPKRLASGLYQFRPDTQIDRVVLDCVSSLQNGADL
A.calcoace     QE-DEILIKRDGKLHRPKRLPSGLYQFRADTQIDRVVLDCVSSLQNGADL
A.tando        QE-DEILIKRDGKLHRPKRLASGLYQFREGTQIDRVVLDCVTSLQNGADL
A.lwoff        QE-DEILIKRDGKLHRPKRLASGLYQFREGTQIDRVVLDCVTSLQNGADL
A.baumanni     QE-DEILIKRDGKLHRPTRLASGLYQFREGTQIDRVVLDCITSLQNGADL
A.ursin        QE-DEILIKRDGKLHRPTRLASGLYQFREGTQHDRVVLDCVTSLQNGADM
A.bayly        QE-DEILIKRDGKLHRPTRLASGLYQFREGTQHDRVVLDCVTSLQNGADM
A.junii        SD-DEILIKRDGKLHRPTRLASGLYQFREGTQHDRVVLDCVTSLQNGADM
               :: .   *   :*.**...* .*: .:   * *:*.: ::: ***:

C.maris        LWIETATPNVADITAMMNEVRKEIPDAKLVYNNSPSFNWTLNFRQQAFDT
C.piezophi     LWIETATPNVADITAMMNEVRKEIPNAKLVYNNSPSFNWTLNFRQQAFDA
C.psychr       LWIETAVPTVHEIAGMMDDVRKVHPDAKLVYNNSPSFNWTLNFRQQAYDA
C.necator      LWIETEKPHIAQIGGMVSEIRKVIPNAKLVYNNSPSFNWTLNFRQQAYDA
P.fulva        IWIETEKPHVGQIKAMVDRVREVIPNAKLVYNNSPSFNWTLSFRQQVFDA
P.aerugi       LWIETEKPHVGQIKAMVDRIREVIPNAKLVYNNSPSFNWTLNFRQQVFDA
A.vinel        LWIETEKPHIGQIKGMVDRIREAIPNAKLVYNNSPSFNWTLNFRQQVFDT
A.ehrli        LWIETEKPHVGQIAEMVNAIREEVPNAKLVYNNSPSFNWTLNFRQQVFDA
M.adhae        LWIETEKPHVGQIAAMVNRIKEVVPDAKLVYNNSPSFNWTLNFRQQVFDA
S.agarivor     LWIETEKPHVGQIAAMVNRIKEVVPDAKLVYNNSPSFNWTLNFRQQVFDA
S.chloroph     LWIETEKPHIQQIAGMVDRIREVVPNAKLVYNNSPSFNWTLNFRQQVFDA
N,penta        LWIETEKPHIGQIGGMVSRIREVIPNAKLVYNNSPSFNWTLNFRQQVFDA
M.posidoni     LWIETEKPHIGQIANLVNRIREVVPNAKLVYNNSPSFNWTLNFRQQVFDA
M.medit        LWIETEKPHVQQIAGMVNRIREVVPNAKLVYNNSPSFNWTLNFRQQTFDA
A.gerne        LWIETATPNVEEIAHMVNRVKEVVPNAKLVYNNSPSFNWTLNFRQQAYDR
Acineto        LWIETATPNVEEIAHMVNRVKEVVPNAKLVYNNSPSFNWTLNFRQQAYDR
A.oleiv        LWIETATPNVEEIAHMVNRVKEVVPNAKLVYNNSPSFNWTLNFRQQAYDR
A.venet        LWIETATPNVEEIAHMVNRVRETVPNAKLVYNNSPSFNWTLNFRQQAYDR
A.bouve        LWIETATPNVEEIAHMVNRVKETVPNAKLVYNNSPSFNWTLNFRQQAYDR
A.calcoace     LWIETATPNVEEIAHMVNRVKEVVPNAKLVYNNSPSFNWTLNFRQQAYDR
A.tando        LWIETATPNVEEIAHMVNRVKETIPNAKLVYNNSPSFNWTLNFRQQAYDR
A.lwoff        LWIETATPNVAEIAHMVNRVKETVPNAKLVYNNSPSFNWTLNFRQQAYDR
A.baumanni     LWIETPTPNVQEIAHMVNRVKEVVPNAKLVYNNSPSFNWTLNFRQQAYDR
A.ursin        IWIETPTPDVAGIAGFVNDIKKQVPNAKLVYNNSPSFNWTLNFRQQSYDR
A.bayly        IWIETPTPDVAGIAGFVNDIKKQVPNAKLVYNNSPSFNWTLNFRQQAYDR
A.junii        IWIETPTPDVAGIAGFVNDIKKQVPNAKLVYNNSPSFNWTLNFRQQAYDR
               :****  * :  *  ::. :::  *:***************.**** :*
```

```
C.maris        MVAEGKDVSSYVRADLMNVDYDTTELAAAADDKIRSFQADTAREAGIFHH
C.piezophi     MVAAGEDVSRYVRADLMSVDYDETELSKRADDKIRSFQADTAREAGIFHH
C.psychr       MVEAGQDVSAYVRADLMKAEYDETELSATADERIRTFQADTAREANVFHH
C.necator      LKAAGKDVSAYDRAQLMSVEYDQTELAKLADEKIRTFQADASREAGIFHH
P.fulva        FVAEGKDVSAYDRAKLMSAEYDETELAQVADEKIRTFQRDGSANAGIFHH
P.aerugi       FVAEGKDVSAYDRNKLMSVEYDDTELAKVADEKIRTFQRDGSAHAGIFHH
A.vinel        WAAEGKDVAKYDRAKLMSVEYDETELALEADARIRTFQRDAAAQAGIFHH
A.ehrli        WKEEGKDVSQYDRDNLMSAEYDDSELAAEADRWVQEFQRKASAEAGIFHH
M.adhae        WKEEGKDVSAYDRAKLMSEEYDNTELGQLADEWCRNFQRDGSREAGIFHH
S.agarivor     WKEAGKDVSAYDRANLMSEEYDNTELSKEADARIQTFQKDGSREAGIFHH
S.chloroph     WVAEGKDVSAYDRAKLMSADYDATDLGLEADEKIRTFQKDAAARAGIFHH
N,penta        WSAAGRDVSAYDRAKLMSVDYDGTDLAAEADERIRTFQKDAAREAGIFHH
M.posidoni     WQEAGKDVNEYDRDRLMSKEYDATTLAIEADERIRNFQRDAAAQAGIFHH
M.medit        WSEAGKDLSAYDRANLMSVEYDSSDLAAEADERIRTFQADAAKEAGIFHH
A.gerne        WVAEGKDVSAYDRAKLMSAEYDNTDLAADADEKVRTFQADASREAGVFHH
Acineto        WVAEGKDVSAYDRAKLMSAEYDNTDLAADADEKVRTFQADASREAGVFHH
A.oleiv        WVAEGKDVSAYDRAKLMSAEYDNTELAADADAKIRTFQADASREAGVFHH
A.venet        WVAEGKDVSAYDRAKLMSAEYDNTELAADADEKVRTFQADASREAGVFHH
A.bouve        WVAEGKDVSAYDRAKLMSAEYDATELAADADEKVRTFQADASREAGVFHH
A.calcoace     WVAEGKDVSAYDRAKLMSAEYDNTELAADADDKIRTFQADASREAGVFHH
A.tando        WVAEGKDVSAYDRAKLMSAEYDNSELAAEADEKVRTFQADASREAGVFHH
A.lwoff        WVAEGKDVSAYDRAKLMSAEYDATELAAEADEKVRTFQADAAREAGVFHH
A.baumanni     WVAEGKDVSAYDRAKLMSAEYDNTELAAEADEKIRTFQADAAREAGVFHH
A.ursin        WVAEGKDVSGYDRAKLMSAEYDNSELAADADEKIRTFQADAAREAGVFHH
A.bayly        WAAEGKDVSAYDRAKLMSAEYDNTELAADADEKIRTFQADAAREAGVFHH
A.junii        WVAEGKDVSGYDRAKLMSAEYDNSELAADADEKIRTFQADAAREAGVFHH
                   *.*:  * *  **. :** : *.  **   : ** . : .*.:***

C.maris        LITLPTYHTAALSTDNLAKEYFGDAGMLGYVEGVQRKEIRQGIACVKHQN
C.piezophi     LITLPTYHTTALSVDNLAKEYFGDEAMLGYVKGVQRKEIRQGIACVKHQN
C.psychr       LITLPTYHTTALSVDNLAKEYFGEQGMLGYVKGVQRKEIRQGIACVKHQN
C.necator      LITLPTYHTAALSTDNLAKEYFGDQGMLGYVAGVQRKEIRQGIACVKHQN
P.fulva        LITLPTYHTAALSTDNLAKGYFADLGMLAYVKGVQRQEIRQGIACVKHQN
P.aerugi       LITLPTYHTAALSTDNLAKGYFADEGMLAYVKGVQRQELRQGIACVKHQN
A.vinel        LITLPTYHTAALSTDNLAKGYFAEEGMLAYVKGVQRQEIRQSIACVKHQN
A.ehrli        LITLPTYHTAALSTDNLAKGYFGDLGMLAYVDGVQRQEIRQGVPTVKHQD
M.adhae        LITLPTYHTAALSTDNLAKGYFGDEGMLAYVAGVQRKEIRQGIATVKHQD
S.agarivor     LITLPTYHTAALSTDNLAKGYFGEQGMLAYVEGVQRKEIRQGIACVKHQN
S.chloroph     LITLPTYHTAALSTDNLAKEYFGEAGMLGYVAGVQRKEIRQGIACVKHQN
N,penta        LITLPTYHTAALSTDNLAKEYFGEAGMLGYVKSVQRQEIRQGIACVKHQN
M.posidoni     LITLPTYHTAALSTDNLAKDYFGDQGMLGYVKEVQRKEIREGLACVKHQD
M.medit        LITLPTYHTAALSTDNLAKEYFGEQGMLGYVKGVQRQEIRQGIACVKHQA
A.gerne        LITLPTYHTAALSTHELAQGYFGDQGMLAYVAGVQRKEIRGTISCVKHQA
Acineto        LITLPTYHTAALSTHELAQGYFGDQGMLAYVAGVQRKEIRGTISCVKHQA
A.oleiv        LITLPTYHTAALSTHELAQGYFGSEGMLAYVAGVQRKEIRGGIACVKHQA
A.venet        LITLPTYHTAALSTHELAQGYFGSEGMLAYVAGVQRKEIRGGIACVKHQA
A.bouve        LITLPTYHTAALSTHELAQGYFGDQGMLAYVAGVQRKEIRGTIACVKHQA
A.calcoace     LITLPTYHTAALSTHELAQGYFGSEGMLAYVAGVQRKEIRGGIACVKHQA
A.tando        LITLPTYHTAALSTHELAQGYFGSEGMLAYVAGVQRKEIRGGIACVKHQA
A.lwoff        LITLPTYHTAALSTHELAQGYFGSEGMLAYVAGVQRKEIRGGIACVKHQA
A.baumanni     LITLPTYHTAALSTHELAQGYFGDQGMLAYVAGVQRKEIRGGIACVKHQA
A.ursin        LITLPTYHTAALSTHELAKGYFGEEGMLAYVAGVQRKEIRGGIACVKHQA
A.bayly        LITLPTYHTAALSTHELAKGYFGEEGMLAYVAGVQRKEIRGGIACVKHQA
A.junii        LITLPTYHTAALSTHELAKGYFGEEGMLAYVAGVQRKEIRGGIACVKHQA
               **********:***..:**: **.. .**.**  ***:*:*  :. ****
```

```
C.maris        MSGSDMGDDHKEYFAGENALKAGG-ANNTSNQFS--
C.piezophi     MSGSDMGDDHKEYFAGENALKAGG-ANNTSNQFS--
C.psychr       MSGSDMGDDHKEYFAGENALKAGG-AKNTSNQFS--
C.necator      MSGSDIGDDHKEYFSGEAALKAAG-KDNTMNQF---
P.fulva        MAGSDIGDNHKEYFAGEAALKAGG-KDNTMNQFH--
P.aerugi       MAGSDIGDNHKEYFAGEAALKASG-KDNTMNQFH--
A.vinel        MAGSDIGDNHKEYFAGEAALKAGG-KDNTMNQFH--
A.ehrli        MAGSNIGDDHKAYFSGEAALKAGG-KGNTMNQF---
M.adhae        MAGSNIGDDHKEFFAGEAALKAGG-KDNTMNQFG--
S.agarivor     MAGSDMGDDHKEYFSGDAALKASG-KDNTMNQFG--
S.chloroph     MSGSDIGDDHKEYFAGEAALKAGG-IHNTMNQFAA-
N,penta        MAGSDIGDDHKEYFAGEAALKAGG-EHNTMNQFAA-
M.posidoni     MSGSNIGDDHKEYFSGDNALKASG-KDNTMNQFAAV
M.medit        MAGSNMGDDHKEYFSGDAALKASG-KDNTMNQF---
A.gerne        MAGSDIGDDHKEIFSGDNALKAHDDAKNTMNQFAAH
Acineto        MAGSDIGDDHKEIFSGDNALKAHDDAKNTMNQFAAH
A.oleiv        MAGSDIGDDHKEIFSGDNALKAHDDAKNTMNQFAAH
A.venet        MAGSDIGDDHKEIFSGDNALKAHDDAKNTMNQFAAH
A.bouve        MAGSDIGDDHKEIFSGDNALKAHDDAKNTMNQFSAH
A.calcoace     MAGSDIGDDHKEIFSGDNALKAHDDAKNTMNQFAAH
A.tando        MAGSDIGDDHKEIFSGDNALKAHDDAKNTMNQFAAH
A.lwoff        MAGSDIGDDHKEIFAGDNALKAGDDSKNTMNQFSA-
A.baumanni     MAGSDIGDDHKEIFAGENALKAGDDAKNTMNQFSAH
A.ursin        MAGSDIGDDHKEIFAGEQALKAGDASKNTMNQFAH-
A.bayly        MAGSDIGDDHKEIFAGEQALKAGDASKNTMNQFAH-
A.junii        MAGSDIGDDHKEIFAGENALKAGDASKNTMNQFGG-
               *:**::**:**  *:*: **** .    ** ***
```

Subfamily 4

10 Sequences of ICL in archae having length between 340-360 aa were taken and MSA was carried out using ClustalW. The sequences were highly conserved and showed high similarity with each other which shows that they may have come from the same ancestors

```
H.muco     MNPTELDSDVFTRDIDNPKGRELREMLNTQDFVFAPGMYHALDARLAEMT 50
H.medi     MNPTELDSDVFVRDIDNPKGRELREMLNTQDFVFAPGMYHALDARLAEMT 50
H.elon     MNPTELDDDVFAQDVDNQKGRELREMLNTQDFVFAPGMYHALDARLAEMT 50
H.lars     MNPTELDDDVFAQDVDNQKGRELREMLNTQDFVFAPGMYHALDARLAEMT 50
H.sulf     MNPTELDSDVFAQDVDNQKARELREMLNTQDFVFAPGMYHALDARLAEMT 50
H.deni     MNPTELDSDVFAQDVDNQKARELREMLNTQDFVFAPGMYHALDARLAEMT 50
H.volc     MNPTELDSDVFAQDVDNQKARELREMLNTQDFVFAPGMYHALDARLAEMT 50
H.gibb     MNPTELDSDVFAQDVDNQKARELREMLNTQDFVFAPGMYHALDARLAEMT 50
H.wals     MNPTELDDNVFNRDIDNPQGEKFRNQLNNQDFVFAPGMYHALDARIAEMT 50
H.pauc     ---MISDTDTTTRDIENRAGKEFREMLNSQNYVFAPGLYHALDARLAEMA 47
                 * :.  :*::*  ..::*: **.*::*****:*******:***:

H.muco     GHDAAYMSGYSTVLGQFGFPDLEMVTMTEMVENAKRMVEATNLPIIADCD 100
H.medi     GHDAAYMSGYSTVLGQFGFPDLEMVTMTEMVENAKRMVEATNLPIIADCD 100
H.elon     GHDAAYMSGYSTVLGQFGFPDLEMVTMTEMVENAKRMVEATNLPVIADCD 100
H.lars     GHDAAYMSGYSTVLGQFGFPDLEMVTMTEMVENAKRMVEATNLPVIADCD 100
H.sulf     GHDAAYMSGYSTVLGQFGFPDLEMVTMTEMVENAKRMVEATNLPVIADCD 100
H.deni     GHDAAYMSGYSTVLGQFGFPDLEMVTMTEMVENAKRMVEATNLPVIADCD 100
H.volc     GHDAAYMSGYSTVLGQFGFPDLEMVTMTEMVENAKRMVEATNLPVIADCD 100
H.gibb     GHDAAYMSGYSTVLGQFGFPDLEMVTMTEMVENAKRMVEATNLPVIADCD 100
H.wals     GHDAAYMSGYSTVLGQFGFPDLEMVTMTEMVENAKRMAEATNLPIIADCD 100
H.pauc     GLDAAYMSGYSTVLGQFGFPDLEMVTMTEMVENAKRIVESCNLPVVADCD 97
           * ************************************:.*: ***::****

H.muco     TGYGGLHNVRRAVREYEKAGVAAVHIEDQTTPKRCGHIAGKQIVSREKAR 150
H.medi     TGYGGIHNVRRAVREYEKAGVAAIHIEDQTTPKRCGHIAGKQIISREKAR 150
H.elon     TGYGGIHNVRRAVREYEKAGVAAIHIEDQTTPKRCGHIAGKQIVSREKAR 150
H.lars     TGYGGIHNVRRAVREYEKAGVAAIHIEDQTTPKRCGHIAGKQIVSREKAR 150
H.sulf     TGYGGIHNVRRAVREYEKAGVAAVHIEDQTTPKRCGHIAGKQIVSREKAK 150
H.deni     TGYGGIHNVRRAVREYEKAGVAAVHIEDQTTPKRCGHIAGKQIVSREKAR 150
H.volc     TGYGGIHNVRRAVREYEKAGVAAVHIEDQTTPKRCGHIAGKQIVSREKAK 150
H.gibb     TGYGGIHNVRRAVREYEKAGVAAVHIEDQTTPKRCGHIAGKQIVSREKAR 150
H.wals     TGYGGIHNVRRAVREYEKAGVAAVHIEDQTTPKRCGHIAGKQIISREKAE 150
H.pauc     TGYGGIHNVRRAVREYEKAGVAAVHIEDQTSPKRCGHIAGKQIVSREKAR 147
           *****:******************:******:************:*****.
```

```
H.muco     ARFEAAVDAKQCDDTVVIARTDAYGSANGDWDEHLERGRIYADAGVDLVW 200
H.medi     ARFEAAVDAKQCDDTVIIARTDAYGSANGDWEEHLERGRIYADAGVDLVW 200
H.elon     ARFEAAVDAKQSDDTVIIARTDAYGSANGDWEEHLERGRIYADAGVDLVW 200
H.lars     ARFEAAVDAKQSDDTVIIARTDAYGSANGDWEEHLERGRIYADAGVDLVW 200
H.sulf     ARFEAAVDAKQSEDTVVIARTDAYGSSNGDWDEHVERGRIYADAGVDIVW 200
H.deni     ARFEAAVDAKQSEDTVVIARTDAYGSSNGDWDEHVERGRIYADAGVDIVW 200
H.volc     ARFEAAVDAKQSEDTVVIARTDAYGSSNGDWDEHVERGRIYADAGVDIVW 200
H.gibb     ARFEAAVDAKQSEDTVVIARTDAYGSSNGDWDEHVERGRIYADAGVDIVW 200
H.wals     ARFTAAVDAKQSEDTVVIARTDAYGSANGDWEEHLERGRMYADAGVDLVW 200
H.pauc     SRFKAAVDAKQSDDTVIIARTDAYGSANGDWEEHLERGRIYADAGVDIVW 197
           :** ******* .:***:*********:****:**:****:*******:**

H.muco     PEMPNPSREDAVNYAETIHETHPDLKLAFNYSSSFAWSEEEDPLTFQELG 250
H.medi     PEMPNPSREDAVNYAEKIHETHPDLKLAFNYSSSFAWSEEEDPLTFQELG 250
H.elon     PEMPNPSREDAVNYAETIHETHPDLKLAFNYSSSFAWSEEEDPLTFQELG 250
H.lars     PEMPNPSREDAVNYAETIHETHPDLKLAFNYSSSFAWSEEEDPLTFQELG 250
H.sulf     PEMPNPSREDAVAYAEEIHETHPDLKLAFNYSSSFAWSEEEDPLTFQELG 250
H.deni     PEMPNPSREDAVAYAEEIHETHPDLKLAFNYSSSFAWSEEEDPLTFQELG 250
H.volc     PEMPNPSREDAVAYAEEIHETHPDLKLAFNYSSSFAWSEEEDPLTFQELG 250
H.gibb     PEMPNPSREDAVAYAEEIHETHPDLKLAFNYSSSFAWSEEEDPLTFQELG 250
H.wals     PEMPNPSREDAVNYAETIHETHPDLKLAFNYSSSFAWSEEDDPLTFEELG 250
H.pauc     PEMPDPSREDAVEYAETIHETHPDLKLAFNYSSSFAWSEEEDPLTFEELG 247
           ****:******* *** ************************:*****:***

H.muco     DLGYKYIFITLFGLHSGAHAVYEDFKKLAEADEQGQFDLEQRYLGHPTES 300
H.medi     DLGYKYIFITLFGLHSGAHSVYEDFKKLAEEDEQGQFDLEQRYLGHPTES 300
H.elon     DLGYKYIFITLFGLHSGAHAVYEDFKKLAEQDEEGQFDLEQRYLGHPTES 300
H.lars     DLGYKYIFITLFGLHSGAHAVYEDFKKLAEQDEEGQFDLEQRYLGHPTES 300
H.sulf     DLGYKYIFITLFGLHSGAHSVYEDFKKLAEQDEEGQFDLEQRYLDHPTES 300
H.deni     DLGYKYIFITLFGLHSGAHSVYEDFKKLAEQDEEGQFDLEQRYLDHPTES 300
H.volc     DLGYKYIFITLFGLHSGAHAVYEDFKKLAEQDEEGQFDLEQRYLDHPTES 300
H.gibb     DLGYKYIFITLFGLHSGAHAVYEDFKKLAEQDEEGQFDLEQRYLGHPTES 300
H.wals     DLGYKYIFITLAGLHSGAHAVYEDFKKLAEADETGQFELEDRYLEHPTES 300
H.pauc     NLGYKYIFITLFALHSGAHSVYEDFERLAEDDERAQFDLEGRYLDHPTES 297
           :*********** .******:*****::*** ** .**:** *** *****

H.muco     HHELSFVSRYQDIETQFDSEARRRIEESEGFSEDEADPITSNARANDDD- 349
H.medi     HHELSFVSRYQDIETEFDPEARRRIEESEGFSEDEADPITSNARANDDD- 349
H.elon     HHELSFVSRYQDIETEFDPEARRRIESSEGFSEEQADPITT----NDDD- 345
H.lars     HHELSFVSRYQDIETQFDPEARRRIESSEGFSEEQADPITT----NDDDV 346
H.sulf     HHELSFVSRYQDIETQFDPEARRRIEESEGFSEEQADPITS----NDDD- 345
H.deni     HHELSFVSRYQDIETQFDPEARRRIEESEGFSEEQADPITS----NDDD- 345
H.volc     HHELSFVSRYQDIETEFDPEARRRIEESEGFSEEQADPITS----NDDD- 345
H.gibb     HHELSFVSRYQDIETEFDPEARRRIEESEGFSEEQADPITS----NDDD- 345
H.wals     HHELSFVSRYQDIETKFDPEARRRIEQSEGFSDEETDPITS-----DDD- 344
H.pauc     HHELSFVSRYQDIETQFDPEARERIEESEGFSEDESNPITSNS-EEDEEM 346
           ****************:**.***.***.*****:::::***:     *::

H.muco     --------
H.medi     --------
H.elon     --------
H.lars     --------
H.sulf     --------
H.deni     --------
H.volc     --------
H.gibb     --------
H.wals     --------
H.pauc     VASEQGDD 354
```

Subfamily 5

8 Sequences of ICLs in bacteria having length larger than 700aa were taken and MSA was carried out using ClustalW. The sequences were highly conserved and showed high similarity with each other which shows that they may have come from the same ancestors.

MSA of ICL in subfamily 5

```
M.intracellulare      ------------------------------------MAIIDKDTQVRPSF
M.avium               ------------------------------------MAIIDKETQVRPSF
M.M.leprae            MAGIPATVVQRCNSTNLDDRLHYSPLGDRAKEERNPMAIMDTNTEVHTLF
M.parascrofulace      -----------------------------------MAIIDSDTEVRTSF
M.tuberculosis        ----------------------MGAAEERKAMAIAETDTEVHTPFEQDF
M.bovis               --------------------------------MAIAETDTEVHTPFEQDF
M.canetti             --------------------------------MAIAETDTEVHTPFEQDF
M.ulcerans            ------------------------------------MAIVETDTEIRTPF
                                                          :  : .* ..  *

M.intracellulare      DDEVAATQQYFDDPRFSRITRLYTARQVAEQRGTIPTDYTVARNAAAAFY
M.avium               DDEVAATQRYFDDPRFARITRLYTARQVAEQRGTIRTDYTVARDAAAAFY
M.M.leprae            TQEVAATQQYFDDPRFAGIIRLYTARQVVEQRGTIPTDYTVARDAATAFY
M.parascrofulace      EDDIAATQRYMDSPRFAGITRLYTARQVVEQRGTIPADYIVAREAAAAFY
M.tuberculosis        EKDVAATQRYFDSSRFAGIIRLYTARQVVEQRGTIPVDHIVAREAAGAFY
M.bovis               EKDVAATQRYFDSSRFAGIIRLYTARQVVEQRGTIPVDHIVAREAAGAFY
M.canetti             EKDVAATQRYFDSSRFAGIIRLYTARQVVEQRGTIPVDHIVAREAAGAFY
M.ulcerans            EQEVADTQRYFDSSRFAGIVRLYTARQVVEQRGTIPNDYTVARTAAGVFY
                       .::* **:*:*..**: * ********.******  *: *** ** .**

M.intracellulare      ERLRELFAQKKSITTFGPYSPGQAVAMKRMGIEGIYLGGWATSAKGSTTE
M.avium               ERLRELFAQKKSITTFGPYSPGQAVTMKRMGIEGIYLGGWATSAKGSTTE
M.M.leprae            ARLRELFAAGKSVTTFGPYSPGQAVSLKRMGIEAIYLGGWATSAKGSITE
M.parascrofulace      DRLRELFAAKKSITTFGPYSPGQAVAMKRMGIEGIYLGGWATSAKGSTTE
M.tuberculosis        ERLRELFAARKSITTFGPYSPGQAVSMKRMGIEAIYLGGWATSAKGSSTE
M.bovis               ERLRELFAARKSITTFGPYSPGQAVSMKRMGIEAIYLGGWATSAKGSSTE
M.canetti             ERLRELFAARKSITTFGPYSPGQAVSMKRMGIEAIYLGGWATSAKGSSTE
M.ulcerans            QRLRELYAEKKSVTTFGPYSPGQAVSMKRMGIEAIYLGGWATSAKGSTTE
                      *****:*  **:*************::******.************** **
```

```
M.intracellulare      DPGPDLASYPLSQVPDDAAVLVRALLTADRNQQYQRLQMSEKQRAAAPVY
M.avium               DPGPDLASYPLSQVPDDAAVLVRALLTADRNQQYLRLQMSEQQRAATKEY
M.M.leprae            DPGPDLASYPLSQVPDDAAVLVRALLAADRNQQYLRLHMTEQQRAATPAY
M.parascrofulace      DPGPDLASYPLSQVPEDAAVLVRALLTADRNQQYQRLNMSEAQRASTPAY
M.tuberculosis        DPGPDLASYPLSQVPDDAAVLVRALLTADRNQHYLRLQMSERQRAATPAY
M.bovis               DPGPDLASYPLSQVPDDAAVLVRALLTADRNQHYLRLQMSERQRAATPAY
M.canetti             DPGPDLASYPLSQVPDDAAVLVRALLTADRNQHYLRLQMSERQRAATPAY
M.ulcerans            DPGPDLASYPLSQVPDDAAVLVRALLTADRNQEYLRLQMTDQQRATTTAY
                      ****************:***********:*****.* **:*:: ***::  *

M.intracellulare      DYRPFIIADADTGHGGDPHVRNLIRRFVEVGVPGYHIEDQRPGTKKCGHQ
M.avium               DYRPFIIADADTGHGGDPHVRNLIRRFVEVGVPGYHIEDQRPGTKKCGHQ
M.M.leprae            DYRPFIIADADTGHGGDSHVRNLIRRFVEIGVPGYHIEDQRPGTKKCGHQ
M.parascrofulace      DYRPFIIADADTGHGGDPHVRNLVRRFVEVGVAGYHIEDQRPGTKKCGHQ
M.tuberculosis        DFRPFIIADADTGHGGDPHVRNLIRRFVEVGVPGYHIEDQRPGTKKCGHQ
M.bovis               DFRPFIIADADTGHGGDPHVRNLIRRFVEVGVPGYHIEDQRPGTKKCGHQ
M.canetti             DFRPFIIADADTGHGGDPHVRNLIRRFVEVGVPGYHIEDQRPGTKKCGHQ
M.ulcerans            DFRPFIIADADTGHGGDPHVRNLIRRFVEVGVPGYHIEDQRPGTKKCGHQ
                      *:****************.*****:*****:**.******************

M.intracellulare      GGKVLVPSDEQIKRLNTARFQLDIMKVPGIIVARTDAEAANLLDSRADER
M.avium               GGKVLVPSDEQIKRLNAARFQLDIMRVPGIIVARTDAEAANLLDSRADER
M.M.leprae            GGKVLVPSDEQIKRLNAARFQLDIMRVPGIIVARTDAEAANLIDSRADER
M.parascrofulace      GGKVLVPSDEQIKRLNAARFQLDIMRVPGIIVARTDAEAANLIDSRADER
M.tuberculosis        GGKVLVPSDEQIKRLNAARFQLDIMRVPGIIVARTDAEAANLIDSRADER
M.bovis               GGKVLVPSDEQIKRLNAARFQLDIMRVPGIIVARTDAEAANLIDSRADER
M.canetti             GGKVLVPSDEQIKRLNAARFQLDIMRVPGIIVARTDAEAANLIDSRADER
M.ulcerans            GGKVLVPSDEQIKRLNAARFQLDVMRVPGIIVARTDAEAANLIDSRADER
                      *****************:******:*:*****************:*******

M.intracellulare      DQPFLLGATNLKIPSYKSCFLALVRRFYELGVKDLNGHLLYALPEGEYAE
M.avium               DQPFLLGATNLDIPSYKACFLAMVRRFYELGVKDLNGHLLYALPEAEYAE
M.M.leprae            DQPFLLGATNLKIPSYKACFLALVRCFYELGVKELHGHLLYALGDGEYAA
M.parascrofulace      DQPFLLGATNLNIPSYKACFLALLRRFYELGVKELNGHLLYALDDSEYAA
M.tuberculosis        DQPFLLGATKLDVPSYKSCFLAMVRRFYELGVKELNGHLLYALGDSEYAA
M.bovis               DQPFLLGATKLDVPSYKSCFLAMVRRFYELGVKELNGHLLYALGDSEYAA
M.canetti             DQPFLLGATKLDVPSYKSCFLAMVRRFYELGVKELNGHLLYALGDGEYAA
M.ulcerans            DQPFLLGATKLDIPSYKSCFLAMLRRFYELGVKELNGHLLYALADAEYAA
                      *********:*.:****:****::* *******:*:******* :.***

M.intracellulare      ATAWLERQGIQGLVSDAVNAWREDGQQSIDDLFDQVESRFVAAWEDDAGL
M.avium               ATAWLERQGIQGVISDAVNAWRENGQQSIDDLFDQVESRFVAAWEDDAGL
M.M.leprae            ASAWLDRQGILAQVSGTVNAWQENGKQSIDDLFEQVEYRLLAAWEKDAGL
M.parascrofulace      ATAWLERQGIQGLISEAANALREDGQHSIDDLFDQVESRFVAAWEDDAGL
M.tuberculosis        AGGWLERQGIFGLVSDAVNAWREDGQQSIDGIFDQVESRFVAAWEDDAGL
M.bovis               AGGWLERQGIFGLVSDAVNAWREDGQQSIDGIFDQVESRFVAAWEDDAGL
M.canetti             AGGWLERQGIFGLVSDAVNAWREDGQQSIDGIFDQVESRFVAAWEDDAGL
M.ulcerans            ANGWLERSGIMGLISDAVNSWREDGMQTTDDLFDQVESRFVAAWEDDAGL
                      * .**:*.** . :* :.*: :*:* :: *.:*:*** *::****.****

M.intracellulare      MTYGEAVADVLAFAASE-GEPADMSAEEWREFAATASLYSARAKAKELGV
M.avium               MTYGEAVAEVLEFAASE-GEPADMSADEWRAFAARASLYSAKAKAKELGF
M.M.leprae            MTYGEAVEEMLQFGESE-GELIGMSPEEWRRFVGRASLYAAREKAKELGV
M.parascrofulace      MTYGEAVAEVLEFGESEEDEPRGMSPAEWRKFAERASLYAAREKARQLGV
M.tuberculosis        MTYGEAVADVLEFGQSE-GEPIGMAPEEWRAFAARASLHAARAKAKELGA
M.bovis               MTYGEAVADVLEFGQSE-GEPIGMAPEEWRAFAARASLHAARAKAKELGA
M.canetti             MTYGEAVADVLEFGQSE-GEPIGMAPEEWRAFAARASLHAARVKAKELGA
M.ulcerans            MTYAEAVAEVLAFGQSE-GDPVPMSPEEWRTFAARASLYAARQKAKEMGV
                      ***.*** ::* *. ** .:   *:. *** *.  ***::*: **:::*
```

```
M.intracellulare      DPGWDCELSKTPEGYYQIRGGIPYAIAKSLAAAPFADLLWMETKTADLDD
M.avium               DPGWDCELAKTPEGYYQIRGGIPYAIAKSLAAAPFADILWMETKTADLAD
M.M.leprae            DPGWDCELAKTPEGYYQIRGGIQYAIAKSLAAAPFADILWMETKTADLAD
M.parascrofulace      DPPWDCELAKTPEGYYQIRGGIPYAIVKSLAAAPFADILWMETKTADLAD
M.tuberculosis        DPPWDCELAKTPEGYYQIRGGIPYAIAKSLAAAPFADILWMETKTADLAD
M.bovis               DPPWDCELAKTPEGYYQIRGGIPYAIAKSLAAAPFADILWMETKTADLAD
M.canetti             DPPWDCELAKTPEGYYQIRGGIPYAIAKSLAAAPFADILWMETKTADLAD
M.ulcerans            DPGWDCELAKTPEGYYQIRGGIPYAIAKSLAAAPFADILWMETKTADLAD
                      ** *****:************** ***.***********:*********** *

M.intracellulare      ARQFADAIHAKFPDQMLAYNLSPSFNWDTTGMTDEQMKQFPEELGKMGFV
M.avium               AKQFADAIHAEFPDQMLAYNLSPSFNWDTTGMTDEQMKQFPEELGKMGFV
M.M.leprae            ARQFAEAIHAEFPEQMLAYNLSPSFNWDTTGMSDEEMKRFPEELGKMGFV
M.parascrofulace      ARQFAEAIHAEFPDQMLAYNLSPSFNWDTTGMTDDEMKRFPEELGKMGFV
M.tuberculosis        ARQFAEAIHAEFPDQMLAYNLSPSFNWDTTGMTDEEMRRFPEELGKMGFV
M.bovis               ARQFAEAIHAEFPDQMLAYNLSPSFNWDTTGMTDEEMRRFPEELGKMGFV
M.canetti             ARQFAEAIHAEFPDQMLAYNLSPSFNWDTTGMTDEEMRRFPEELGKMGFV
M.ulcerans            ARQFAEAIHAEFPDQMLAYNLSPSFNWDTTGMTDEEMKRFPEELGKMGFV
                      *:***:****:**:********************:*::*::************

M.intracellulare      FNFITYGGHQIDGVAAEEFATSLQQDGMLALARLQRKMRLVESPYRTPQT
M.avium               FNFITYGGHQIDGVAAEEFATSLQQDGMLALARLQRKMRLVESPYRTPQT
M.M.leprae            FNFITYGGHQIDGVAAEEFATALRQDGMLALARLQRKMRLVESPYRTPQT
M.parascrofulace      FNFITYGGHQIDGVAAEEFAASLKQDGMLALARLQRKMRLVESPYRTPQT
M.tuberculosis        FNFITYGGHQIDGVAAEEFATALRQDGMLALARLQRKMRLVESPYRTPQT
M.bovis               FNFITYGGHQIDGVAAEEFATALRQDGMLALARLQRKMRLVESPYRTPQT
M.canetti             FNFITYGGHQIDGVAAEEFATALRQDGMLALARLQRKMRLVESPYRTPQT
M.ulcerans            FNFITYGGHQIDGVASEEFATALRQDGMLALARLQRKMRLVESPYRTPQT
                      *****************:****::*:****************************

M.intracellulare      LVGGPRSDAALTASSGRTATTKSMGEGSTQHQHLVQTEVPKKLLEEWLAM
M.avium               LVGGPRSDAALAASSGRTATTKAMGEGSTQHQHLVQTEVPKKLLEEWLAM
M.M.leprae            LVGGPRSDAALAASSGRTATTKAMGKGSTQHQHLVQTEVPKKLLEEWLAL
M.parascrofulace      LVGGPRSDAALQASSGRTATTKAMGKGSTQHQHLVQTEVPKKLLEEWLAL
M.tuberculosis        LVGGPRSDAALAASSGRTATTKAMGKGSTQHQHLVQTEVPRKLLEEWLAM
M.bovis               LVGGPRSDAALAASSGRTATTKAMGKGSTQHQHLVQTEVPRKLLEEWLAM
M.canetti             LVGGPRSDAALAASSGRTATTKAMGKGSTQHQHLVQTEVPRKLLEEWLAM
M.ulcerans            LVAGPRSDAALQASSGRTATTKAMGKGSTQHQHLVQTEVPKKLLEERLAL
                      **.******** ***********:**:***************:****** **:

M.intracellulare      WSENYDLGEKLRVQLRPRRAGSDVLELGIYGNDDEQLANVVVDPIKDRHG
M.avium               WSENYHLGEKLRVQLRPRRAGSDVLELGIYGDGDEQLANVVVDPIKDRHG
M.M.leprae            WSEHYQLGEKLRVQLRPRRAGSDVLELGIYGNGDEQLANVIVDPIKDRRG
M.parascrofulace      WSEHYQLGEKLRVSLRPRRPGSDVLVLGIHGNGDEQLANVVFDPIKDRHG
M.tuberculosis        WSGHYQLKDKLRVQLRPQRAGSEVLELGIHGESDDKLANVIFQPIQDRRG
M.bovis               WSGHYQLKDKLRVQLRPQRAGSEVLELGIHGESDDKLANVIFQPIQDRRG
M.canetti             WSGHYQLKDKLRVQLRPQRAGSEVLELGIHGESDDKLANVIFQPIQDRRG
M.ulcerans            WSDHYQLGEKLRVQLRPRRAGSDVLELGIYGNDDEQLANVVFDPIKDRHG
                      ** :*.* :****.***:*.**:** ***:*:.*::****:.:**:**:*

M.intracellulare      RSILQVRDQNTFAEKLRQKRLMTLIHLWLVHRFKADGVIYVTPTEDNLYQ
M.avium               RSILQVRDQNTFAEKLRQKRLMTLIHLWLVHRFKADAVIYVTPTEDNLYQ
M.M.leprae            RSILQVRDQNTFAEKLRQKRLMTLIHFWLVHRFKADTVIYVTPTEDNLYQ
M.parascrofulace      RSILTVRDQNTFAEKLRQKRLMTLIHLWLVHRFKADAVYYVTPTEDNQYQ
M.tuberculosis        RTILLVRDQNTFGAELRQKRLMTLIHLWLVHRFKAQAVHYVTPTDDNLYQ
M.bovis               RTILLVRDQNTFGAELRQKRLMTLIHLWLVHRFKAQAVHYVTPTDDNLYQ
M.canetti             RTILLVRDQNTFGAELRQKRLMTLIHLWLVHRFKAQAVHYVTPTDDNLYQ
M.ulcerans            RSILTVRDQNTFAHKLRQKRLMTLIHLWLVHRFKADAVYYVTPTEDNQYQ
                      *:** *******. :***********:********: * *****:** **

M.intracellulare      TSKMKSHGIFSEVYQEVGEIIVAEVNQPRIAELLKPDRVALRKLITKEG
M.avium               TSKMKSHGIFSEVYQEVGEIIVAEVNRPRIAELLQPDRVALRKLITKEG
M.M.leprae            TSKMKSHGIFSEVYQEVGEIIVAEVNHPRLVELLTPDRVALRRLITKEG
M.parascrofulace      TSKMKSHGIFTEVNQEVGEIIVAEVNRPRIEELLKPDRVALRKLITKED
M.tuberculosis        TSKMKSHGIFTEVNQEVGEIIVAEVNHPRIAELLTPDRVALRKLITKEA
M.bovis               TSKMKSHGIFTEVNQEVGEIIVAEVNHPRIAELLTPDRVALRKLITKEA
M.canetti             TSKMKSHGIFTEVNQEVGEIIVAEVNHPRIAELLTPDRVALRKLITKEA
M.ulcerans            TSKMKSHGIFSEVNQDVGEIIVAEVNRPRIEELLTADRVALRQLITKGD
                      ***********:** *:***********:**: *** .*******:****
```

Subfamily comparisons: Subfamily 1 and subfamily 2

Sequence comparison between subfamily 1 and 2 showed differences in amino acid sequences. It was found that subfamily 1 members contained amino acid sequence of 420-450 aa while on the other hand it was found that subfamily 2 contained amino acid sequence of 530-600 aa. Subfamily 1 was found in bacteria or the prokaryotes while subfamily 2 was found eukaryotes, fungi and plants. It was also found that subfamily 1 and 2 contained domain-1 which is a TIM-barrel structure which is important for catalytic property of the protein while subfamily 2 contained extra 100aa sequence giving rise to domain-2 in eukaryotes. This extra sequence helps the isocitrate lyase to enter the peroxisomes and help the plants and fungi in their growth by accomplishing the glyoxylate bypass. The ICL initiate the breakdown of fatty acids and additionally possess the enzymes to produce intermediate products for the synthesis of sugars by gluconeogenesis. The seeding uses these sugars synthesized from fats until it is mature enough to produce thereby Photosynthesis

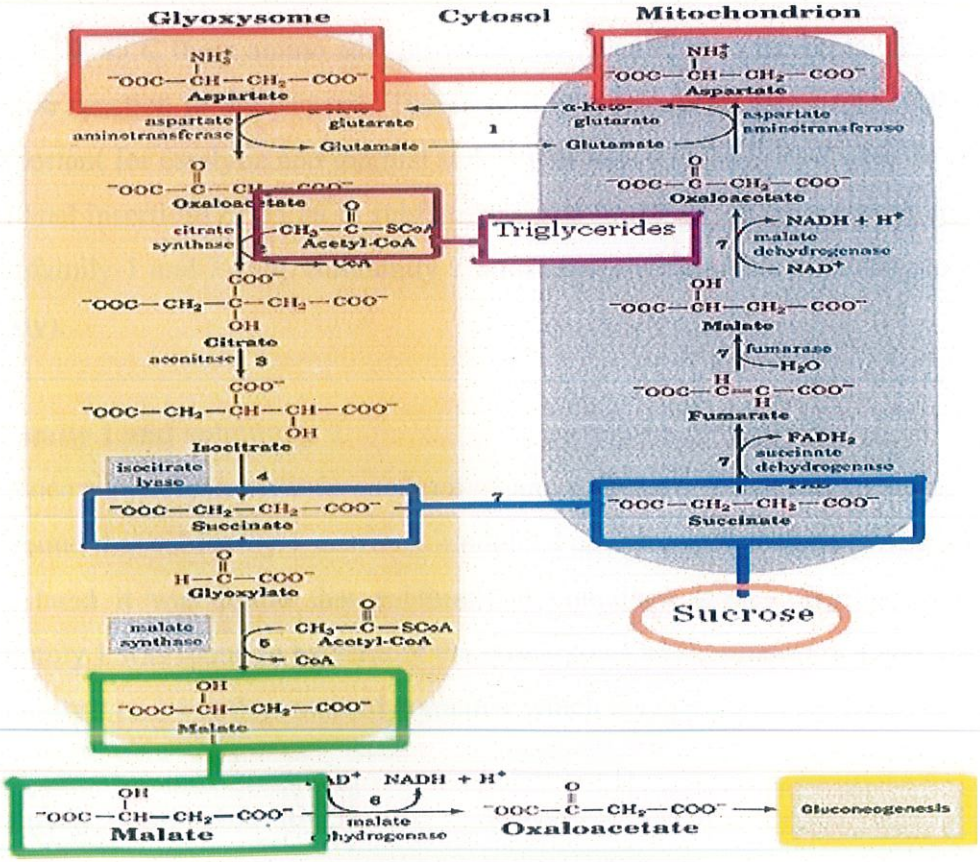


Fig10-ICL activity in plants and fungi

Subfamily 1 and subfamily 3

Sequence comparison between subfamily 1 and 3 showed differences in amino acid sequences. It was found that subfamily 1 members contained amino acid sequence of 420-450 aa while on the other hand it was found that subfamily 3 contained amino acid sequence of 530-560 aa. Subfamily 1 was found in bacteria or the prokaryotes while subfamily 3 was also found bacteria. It was also found that subfamily 1 and 3 contained domain-1 which is a TIM-barrel structure which is important for catalytic property of the protein while subfamily 3 contained extra 46aa sequence giving rise to domain-2 in bacteria [36]. These extra 46 aa sequences in Subfamily -3 is important for catalytic and thermal stability of ICL in the bacteria. Studies have shown that in subfamilty-3 interaction TIM barrel and domain-2 is important for the catalytic activity [36]. It has also been found that Q211, N214, E219 and Q221 are very important for its catalytic function. Studies showed that the protein showed less activity in absence of these amino acids [36]. It was also found that these amino acids were highly conserved in subfamily-3 and absent in other subfamilies. Presence of these amino acids also provides high thermostability to the protein. At 45^0C these amino acid provides thermostability to the protein and help them provide catalytic function to the bacteria. Therefore it was found that motif **QIENQVSDEKQCGHQD** is important for catalytic and thermal stability of ICL in subfamily-3 [36]. It was also found that additional insertions of 35 aa in subfamily-3 was found which gave rise to domain-3 and absent in subfamily-1 and 2 [36]. Subfamily 1 and 3 showed very less similarity to each other (<30% identity).

Subfamily 1 and subfamily 4

Sequence comparison between subfamily 1 and 4 showed differences in amino acid sequences. It was found that subfamily 1 members contained amino acid sequence of 420-450 aa while on the other hand it was found that subfamily 4 contained amino acid sequence of 340-360 aa. Subfamily 1 was found in bacteria or the prokaryotes while subfamily 4 was also found archea. Subfamily 1 and 4 both contained domain-1 which is required for the catalytic activity.

Subfamily 1 and subfamily 5

Sequence comparison between subfamily 1 and 5 showed the differences in amino acid sequences. It was found that subfamily 1 members contained amino acid sequence of 420-450 aa

while on the other hand it was found that subfamily 5 contained ICL-2 members having amino acid sequence of >700 aa. Subfamily 1 was found in bacteria or the prokaryotes while subfamily 5 was also found in mycobacterium. It was found that catalytic function jointly requires ICL-1&2 in *mycobacterium tuberculosis* although ICL1 is more active than ICL2. Group I ICLs comprised small eubacteria ICLs, including *M. tuberculosis* ICL1. Group II ICLs included medium-length plant and fungal ICLs, which contained a central domain (Domain II) absent in group I. Group III ICLs, so far identified only in *Mycobacterium*, had a long C-terminal domain absent in groups I and II; they also contained a central region with significant homology to the central domain of the eukaryotic ICLs and were overall ~ 35% identical to them (Fig. 1b). *In silico* modeling of the three-dimensional structure of ICL2, based on the X-ray crystal structures of the group I ICL1 from *M. tuberculosis* and the group II ICL from *Aspergillus nidulans*, revealed a striking level of structural concordance. Significantly, the active site loops of ICL1 and ICL2, containing the conserved catalytic motif (KKCGH), were nearly super-imposable. Thus, despite the limited homology between ICL1 and ICL2, conservation of tertiary structure and enzymatic activity suggested that these enzymes might provide overlapping biological functions.

ICL domain organization: Domains I and III are present in all ICLs. Domain I contains the conserved catalytic motif KKCGH. Domain II is present in fungal and plant ICLs and in mycobacterial ICL2, but absent in mycobacterial ICL1. Domain IV is unique to mycobacterial ICL2

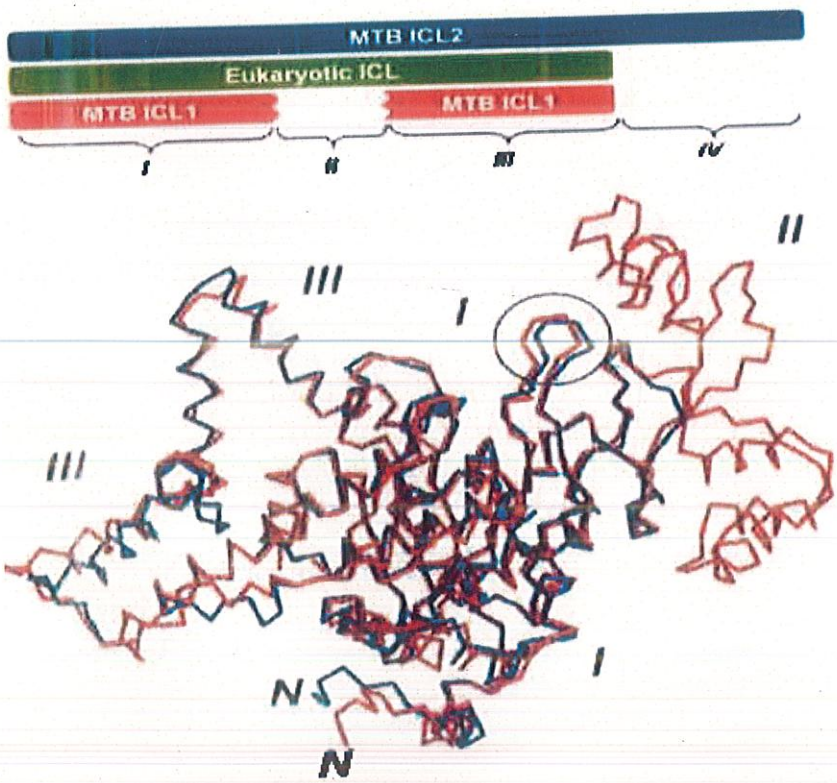


Fig11-Difference in structural features for different ICL

Some other important amino acid indifferent subfamilies→

In *E.coli*, K193, C195, H197 and H365 are the residues involved in catalysis while H184 is related to ICL tetrameric enzyme [37].

In *A. niger* W283, F345, T347 in ICl are involved in substrate recognition [38].

Q207 and Q217 in *C. maris* ICl [39] and A214 in *C. psycherythrae* ICl play important role in cold adaptiveness of the enzyme [40].

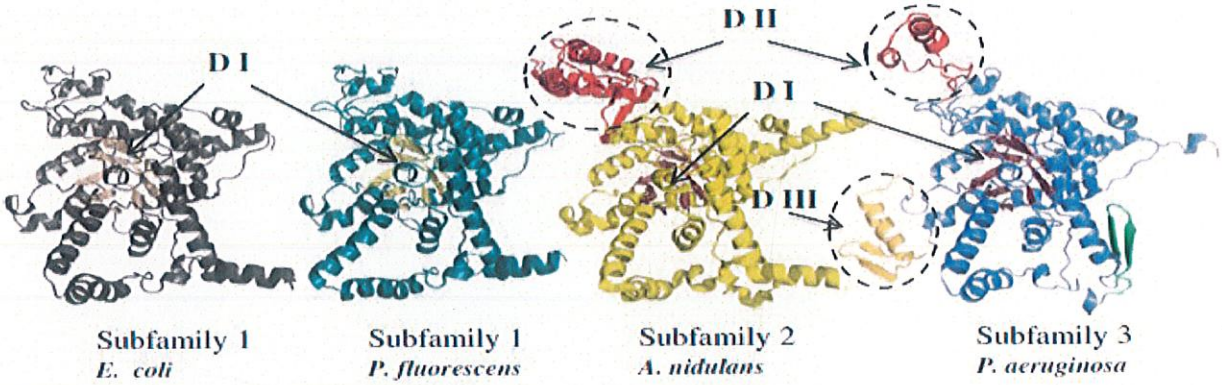


Fig 3[36]→Structure show presence of domain 1 in subfamily-1,2,3. Presence of domain 2 in subfamily 2 and 3. Presence of Domain 3 in subfamily 3 only.

PHARMACOPHORE MODELLING

Generated pharmacophore

Once these 69 active molecules were prepared we got 412 different conformations. These conformations were used to develop the pharmacophore using AUTODOCK.

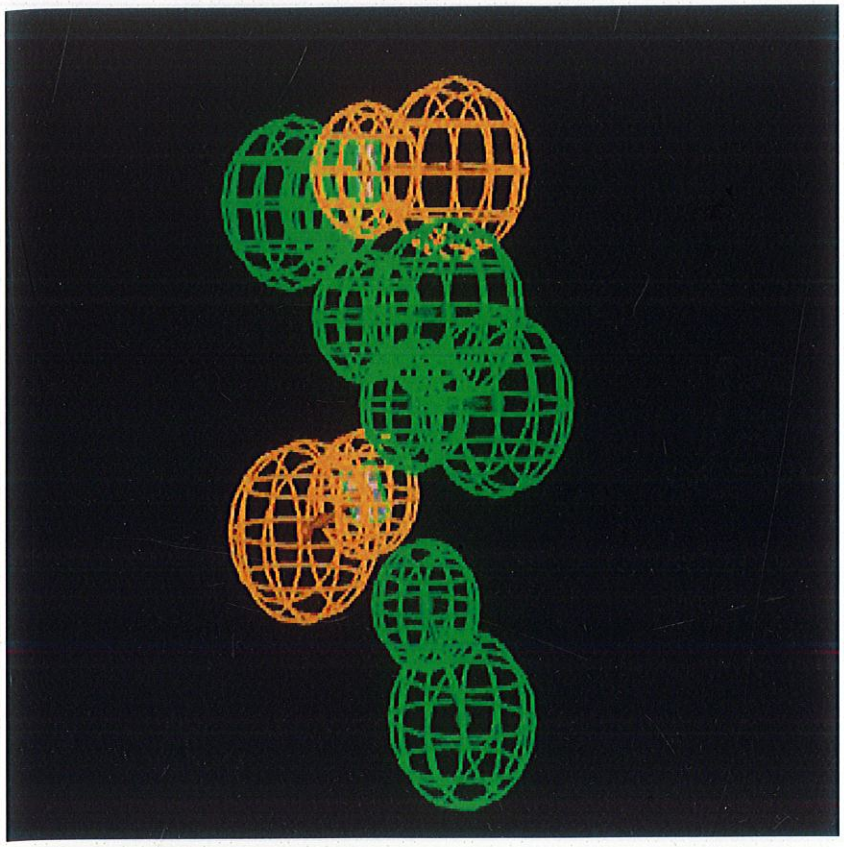


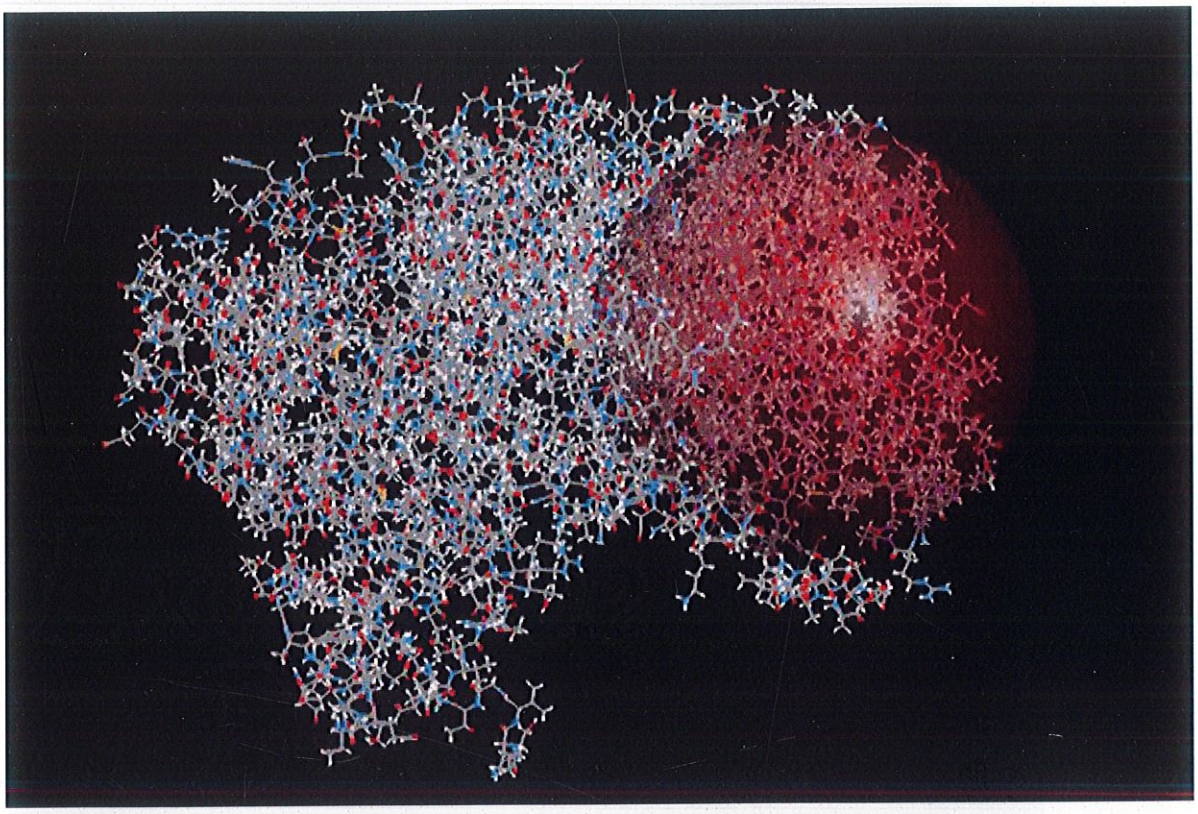
Fig 4→generated pharmacophore

LINKING WITH ZINCPHARMER

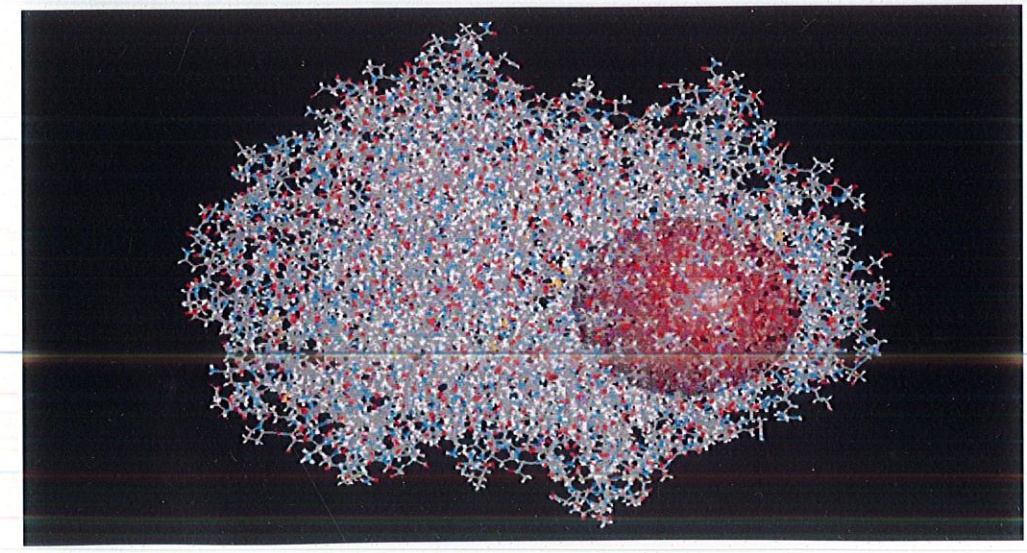
11176 ZINC molecules were related to the above generated pharmacaphore. On preparing these ligands we got 94049 different conformations.

Structure of ICL in different organisms and its binding sites 1F61

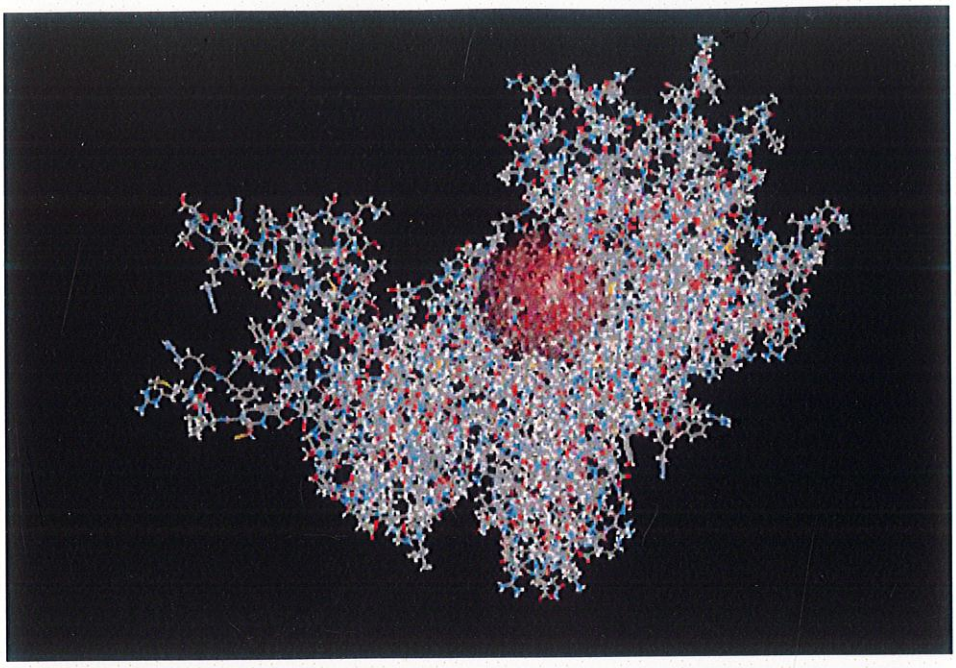
(Fig 5→ Appoenzyme structure of ICL in *M.tuberculosis*)



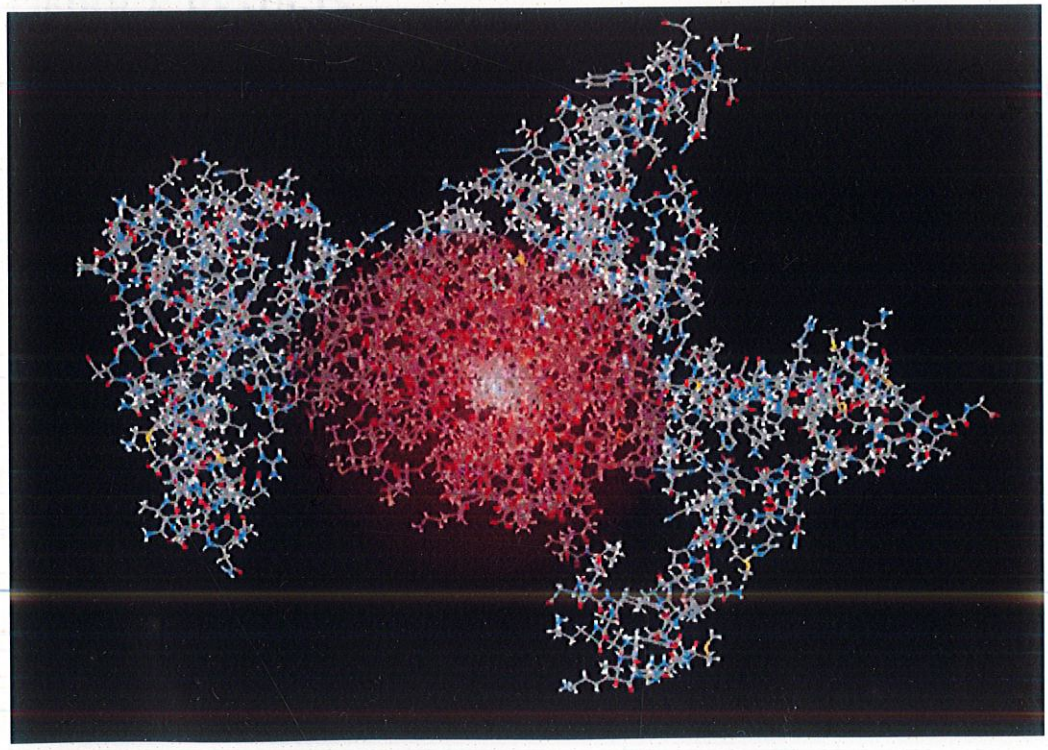
1F8I (fig 6→ Closed structure of ICL in *M.tuberculosis*)

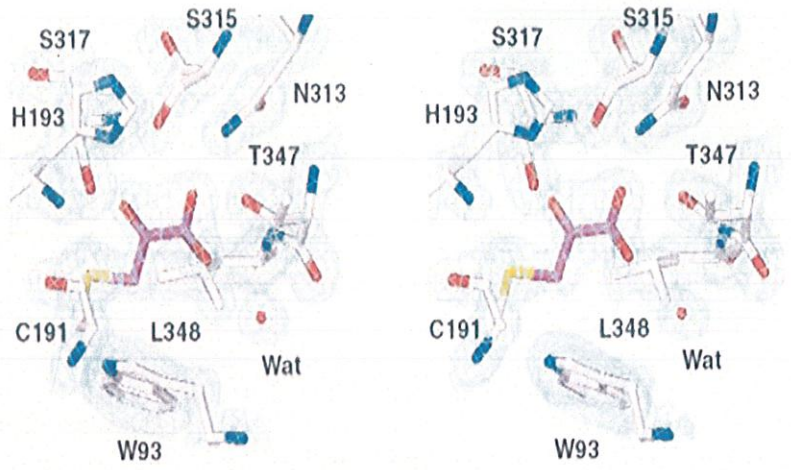


Modelled Structure (Fig 7→ ACEAa+ ACEAb of ICL)



1DQU (Fig 8→structure of *A.nidulans*)





His 193, Asn 313, Ser 315, Ser 317, and Thr347

Fig 9: Active sites in *M. tuberculosis*

DOCKING RESULTS

Using above ligands extracted from ZINCPharmer libdock was carried out.

LibDOCK RESULTS

Lipdock score of ligands that are docked is viewed and evaluated, score greater than 100 is taken as thrashhold.

Table 16

Conformations	1F61	1F8I	1DQU	Modelled Structure
ZINC09295130	117.175			
ZINC09046532	115.733			
ZINC59014529	114.48			
ZINC09349351	109.193			
ZINC06812266	109.172			
ZINC20527885	105.819			
ZINC55889361	105.274			
ZINC23722311	103.707			
ZINC20528189	102.914			

ZINC40978368	102.344			
ZINC02578371	102.3			
ZINC59014489	101.834			
ZINC59039103			150.836	
ZINC13856280			150.714	
ZINC08417660			149.008	
ZINC20910351			146.658	
ZINC05801167			144.427	
ZINC21318960			141.529	
ZINC20029589			141.084	
ZINC04528833			140.766	
ZINC25906896			140.664	
ZINC20911074			140.22	
ZINC72461251			138.521	
ZINC22060683			137.25	
ZINC20424677			136.977	
ZINC58867289			135.714	
ZINC59023068			133.653	
ZINC59021176			133.602	
ZINC08854435			133.47	
ZINC18051106			132.648	
ZINC20527377			131.547	
ZINC20527543			131.514	135.363
ZINC59039099			131.261	
ZINC58995940			131.1	
ZINC20527288			130.687	
ZINC59039159			129.061	
ZINC72461140			129.026	
ZINC09847055			128.853	
ZINC12149666			128.429	134.242
ZINC72461328			128.406	

ZINC20527558			127.378	
ZINC68278601			126.247	
ZINC39969650			126.245	
ZINC03667205			125.207	122.795
ZINC03668951			125.132	
ZINC02855377			123.576	
ZINC12446225			123.295	
ZINC02956477			123.114	
ZINC45078001			122.498	
ZINC59014490			122.122	
ZINC75134564			121.177	
ZINC21253073			120.77	
ZINC01178611			120.254	
ZINC21318752			119.928	
ZINC12847055			119.799	
ZINC59005199			119.762	
ZINC59310837			119.745	
ZINC59310837			119.738	
ZINC77311642			118.549	
ZINC14186214			118.133	
ZINC74347318			117.406	
ZINC40978337			116.902	
ZINC72071554			116.728	
ZINC09403156			116.507	
ZINC14138694			116.488	
ZINC31909993			116.322	
ZINC21318947			116.119	
ZINC58995446			115.392	
ZINC12280074			114.795	
ZINC09836543			114.702	
ZINC71830291			114.626	

ZINC21318954			114.198	
ZINC08748697			114.016	
ZINC80964159			113.355	
ZINC72472534		121.335	113.338	
ZINC72461312			112.615	
ZINC16944493			112.324	
ZINC12599639			111.555	
ZINC15859228			111.276	
ZINC79804199			111.274	
ZINC73872065			111.225	
ZINC00862078			111.185	
ZINC08781927			111.099	
ZINC09391923			110.992	125.46
ZINC15859291			110.562	
ZINC20527541		111.607	110.132	
ZINC72472493		131.864		
ZINC12149584		131.437		
ZINC02890511		128.213		
ZINC16972407		126.279		
ZINC62580977		125.22		
ZINC08614093		124.641		
ZINC78777931		123.037		
ZINC38255675		123.017		
ZINC33467802		122.897		119.21
ZINC38255669		122.634		
ZINC12089289		122.321		
ZINC59018558		122.185		128.804
ZINC12033473		121.219		
ZINC20494512		121.189		115.274
ZINC58996853		120.622		
ZINC20908148		120.513		

ZINC64891185		120.284		
ZINC39572329		119.729		
ZINC59014824		119.716		
ZINC09337008		119.417		
ZINC20906707		118.077		
ZINC59018613		117.723		
ZINC72461307		117.598		
ZINC12698211		116.929		
ZINC39612650		116.24		
ZINC06635961		115.571		
ZINC09129824		114.852		
ZINC13110769		114.466		
ZINC36354025		114.262		
ZINC20291654		113.948		
ZINC20159868		113.836		
ZINC59014783		113.767		
ZINC80975165		113.016		
ZINC67974546		112.742		
ZINC43024792		112.719		
ZINC72472580		112.273		
ZINC57670016		112.034		
ZINC40146169		111.981		
ZINC03340312		111.68		
ZINC05580843		111.585		
ZINC74036709		111.511		
ZINC20527545		111.236		
ZINC32492799		111.164		
ZINC12052564		111.15		
ZINC35311734		111.043		
ZINC20527544		110.974		
ZINC02644866				115.922

ZINC59040046				115.776
ZINC72461140				127.719
ZINC20908717				123.547
ZINC59499066				118.243
ZINC03272189				118.7
ZINC20570144				135.506
ZINC00720812				123.377
ZINC20907183				145.207
ZINC05941620				120.003
ZINC67908086				128.5
ZINC59015423				138.736
ZINC09244933				117.025
ZINC12567583				115.213
ZINC81085075				116.061
ZINC12845135				115.468
ZINC20527375				113.387
ZINC59312823				119.185
ZINC47272712				116.784
ZINC02875345				122.756
ZINC13865529				118.478
ZINC39969064				124.055
ZINC20905283				126.086
ZINC07686964				130.038
ZINC15859263				113.618
ZINC15859211				119.868
ZINC20905282				128.011
ZINC39275571				113.85
ZINC02283727				113.93
ZINC09685895				138.004
ZINC02291586				114.662
ZINC01205371				125.635

ZINC71280257				117.505
ZINC09442139				113.863
ZINC20416397				124.191
ZINC72016158				124.182
ZINC32895043				133.177
ZINC58992458				117.888
ZINC59309436				134.714
ZINC58994644				129.803
ZINC14190344				117.909
ZINC09884521				124.189
ZINC15859257				114.708
ZINC72144372				114.212
ZINC19130084				135.655
ZINC16195910				118.178
ZINC80127853				135.665
ZINC12850950				129.47
ZINC15911761				116.177
ZINC03847180				122.509
ZINC71776281				117.17
ZINC39899216				113.727
ZINC12734423				117.311
ZINC59018560				122.448
ZINC00976883				125.707
ZINC20577103				116.147
ZINC08845479				128.761
ZINC09426169				116.957
ZINC21318870				132.233
ZINC72461191				124.396
ZINC72461154				124.524
ZINC72461150				123.927
ZINC71774753				151.911

ZINC02208918				141.098
ZINC09948493				149.257
ZINC13121296				143.753
ZINC08835919				142.129
ZINC59039761				142.163
ZINC59016659				140.676
ZINC55965387				140.559
ZINC19815399				143.017
ZINC63697962				144.262
ZINC08900744				140.302
ZINC20175756				142.039
ZINC58997418				146.382
ZINC12389511				140.346

CDOCKER RESULTS

Active molecules having libdock score greater than 100 were taken and Cdocker analysis is done. Once we get the result Cdocker score of ligands that are docked is viewed and evaluated, score lesser than -40 is taken as threshold.

Table 17

Conformations	1F61	1F8I	1DQU	Modelled Structure
ZINC09426169	-69.2106			
ZINC20160102	-62.4713			
ZINC29725116	-58.8561			
ZINC23641597	-55.3636			
ZINC14980506	-50.2988			
ZINC04083285	-49.4107			
ZINC03534609	-49.3461			
ZINC77422679	-49.2019			
ZINC80439431	-48.8736			
ZINC23118614	-48.6686			
ZINC02876186	-47.6605			
ZINC79069925	-45.6009			
ZINC23722311	-44.955			
ZINC80440055	-44.5499			
ZINC14775366	-44.5244			
ZINC55200314	-43.5983			
ZINC01778317	-43.0307			
ZINC14775366	-42.7271			
ZINC10895448	-42.1656			
ZINC49944053	-42.004			
ZINC18051106			-77.0035	
ZINC01178611			-74.9935	
ZINC58867289			-72.7029	

ZINC13856280			-67.0299	
ZINC20527288			-59.5124	
ZINC08417660			-57.453	
ZINC09682969		42.4156	-53.8668	
ZINC08781927			-51.0438	
ZINC14138694			-49.2225	
ZINC12149666			-49.1629	
ZINC62580977			-47.703	
ZINC05801167		48.5616	-47.1523	
ZINC74096528			-46.4185	
ZINC21318960			-46.1535	
ZINC73872065			-46.0981	
ZINC12280074			-46.0558	
ZINC74096528			-46.0547	
ZINC31909993			-45.4444	
ZINC20527543			-44.0729	
ZINC68278601			-43.5271	
ZINC39485769			-43.5271	
ZINC21318947			-42.7525	
ZINC48893687			-42.7284	
ZINC21318954			-42.3589	
ZINC21318752			-40.6307	
ZINC72472493		-68.0543		
ZINC12149584		-66.2312		
ZINC02890511		-63.3452		
ZINC70698483		-61.521		
ZINC16972407		-58.9812		
ZINC00729502		-58.1254		
ZINC62580977		-56.4121		
ZINC08614093		-56.3121		
ZINC20424953		-56.0781		

ZINC16972407		-54.5211		
ZINC78777931		-53.2221		
ZINC38255675		-53.2199		
ZINC33467802		-52.9809		
ZINC38255669		-52.5478		
ZINC12089289		-50.512		
ZINC59018558		-49.8801		
ZINC72472534		-49.7998		
ZINC01068830		-48.5021		
ZINC12033473		-48.349		
ZINC20494512		-48.0786		
ZINC19972593		-46.1213		
ZINC58996853		-46.1123		
ZINC20908148		-45.2109		
ZINC64891185		-45.1999		
ZINC39572329		-43.2122		
ZINC59014824		-43.1102		
ZINC09337008		-43.2311		
ZINC08892107		-43.2134		
ZINC59018558		-42.421		
ZINC12698211		-41.7434		
ZINC20906707		-40.989		
ZINC59018613		-40.7689		
ZINC72461307		-40.0987		
ZINC71774753				-71.2134
ZINC09948493				-70.142
ZINC58997418				-70.081
ZINC20907183				-69.5432
ZINC63697962				-67.4321
ZINC13121296				-66.126
ZINC19815399				-66.021

ZINC59039761				-64.001
ZINC08835919				-61.356
ZINC20175756				-59.432
ZINC02208918				-57.645
ZINC59016659				-54.752
ZINC55965387				-52.112
ZINC12389511				-52.098
ZINC08900744				-51.76
ZINC59015423				-50.115
ZINC09685895				-48.33
ZINC80127853				-47.123
ZINC19130084				-46.346
ZINC20570144				-44.213
ZINC20527543				-42.1134
ZINC59309436				-42.424
ZINC12149666				-40.127

CHAPTER 4

CONCLUSION

I have generated the pharmacophore of ICL in *M. tuberculosis*.

Best docking score→

1F61	**1F8I**
ZINC72472493=131.864(libdock)	ZINC09295130= 117.175(libdock)
ZINC72472493=-68.0543(Cdocker)	ZINC09426169=-69.2106(Cdocker)
1DQU	**Modeled Structure**
ZINC59039103=150.836	ZINC20907183=145.207
ZINC18051106=-77.0035	ZINC71774753=-71.2134

On completing the Phylogenetic analysis, it was found that ICL is divided into 5 subfamilies depending upon their sequence dissimilarities. Subfamily 1 contains ICL in bacteria of around 420-450 aa sequence, Subfamily 2 contains ICL in fungi and plants of around 430-600aa sequence, Subfamily 3 contains ICL in other bacteria of around 530-560aa, Subfamily 4 contains ICL in archae of around 340-360 aa and Subfamily 5 contains ICL-2 in Mycobacterium of sequence >700aa.

Sequence comparison analyses have shown that active site residues are conserved across different subfamily. Therefore, the molecules identified using developed pharmacophore may act as putative broad-specific inhibitors.

References

1) Ismael Kassim, Ray CG (editors) (2004). Sherris Medical Microbiology (4th ed.). McGraw Hill. ISBN 0-8385-8529-9 .
2) Beeching JR (1989). "High sequence conservation between isocitrate lyase from Escherichia coli and Ricinus communis". PMID 2696959.
3) Tanaka A, Atomi H, Ueda M, Hikida M, Hishida T, Teranishi Y (1990). "Peroxisomal isocitrate lyase of the n-alkane-assimilating yeast Candida tropicalis: gene analysis and characterization". PMID 2361956.
4) Dunn, MF; Ramirez-Trujill JA, Hernandez-Lucas I (Oct 2009). "Major roles of isocitrate lyase and malate synthase in bacterial and fungal pathogenesis". PMID 19684068.
5) Idnurm, A; Howlett BJ (Oct 2002). "Isocitrate lyase is essential for pathogenicity of the fungus Leptosphaeria maculans to canola (Brassica napus)". PMC 126752. PMID 12455691.
6) Lorenz, MC; Bender JA (Oct 2004). "Fink GR". Eukaryot Cell 3 (5): 1076–1087. PMC 522606. PMID 15470236.
7) Srivastava, V; Jain A, Srivastava BS, Srivastava R (May 2008). "Selection of genes of *Mycobacterium tuberculosis* upregulated during residence in lungs of infected mice". PMID 18054522.
8) Muñoz-Elías, EJ; McKinney JD (Jun 2005). "*Mycobacterium tuberculosis* isocitrate lyases 1 and 2 are jointly required for in vivo growth and virulence". Nat Med 11 (6): 638–644.doi:10.1038/nm1252. PMC 1464426. PMID 15895072.
9) http://www.cdc.gov/tb/topic/basics/ .
10) http://www.cdc.gov/tb/topic/treatment/default.htm.
11) http://textbookofbacteriology.net/themicrobialworld/tuberculosis.html.
12) Gould TA, van de Langemheen H, Muñoz-Elías EJ, McKinney JD, Sacchettini JC (AUG 2006). "Dual role of isocitrate lyase 1 in the glyoxylate and methylcitrate cycles in Mycobacterium tuberculosis".PMID:168979647.
13) Garrett R and Grisham CN (2008). Biochemistry. Brooks Cole. p. 588. ISBN 978-0-495-10935-8 .

14) Cozzone, AJ (1998). "Regulation of acetate metabolism by protein phosphorylation in enteric.PMID 9891796.

15) Kondrashov, FA; Koonin EV, Morgunov IG, Finogenova TV, Kondrashova MN (23). "Evolution of glyoxylate cycle enzymes in Metazoa: evidence of multiple horizontal transfer events and pseudogene formation".. PMC 1630690. PMID 17059607.

16) Kornberg, HL; Krebs HA (18). "Synthesis of cell constituents from C2-units by a modified tricarboxylic acid cycle". PMID 13430766.

17) Eastmond, PJ; Graham IA (Feb 2001). "Re-examining the role of the glyoxylate cycle in oilseeds". PMID 11173291.

18) Ahn CH, Won TH, Kim H, Shin J, Oh KB(JUL 2013)." Inhibition of Candida albicans isocitrate lyase activity by cadiolides and synoilides from the ascidian Synoicum sp". PMID: 23747224.

19) Shingnapurkar D, Dandawate P, Anson CE, Powell AK, Afrasiabi Z, Sinn E, Pandit S, Venkateswara Swamy K, Franzblau S, Padhye S(MAY 2012)." Synthesis and characterization of pyruvate-isoniazid analogs and their copper complexes as potential ICL inhibitors". PMID: 22475559.

20) Ji L, Long Q, Yang D, Xie J(APR 2011)." Identification of mannich base as a novel inhibitor of *Mycobacterium tuberculosis* isocitrate by high-throughput screening". PMCID: PMC3076504. PMID: 21494431.

21) Sriram D, Yogeeswari P, Methuku S, Vyas DR, Senthilkumar P, Alvala M, Jeankumar VU(SEP2011)." Synthesis of various 3-nitropropionamides as Mycobacterium tuberculosis
isocitrate lyase inhibitor". PMID: 21840711.

22) Chung SC, Jang KH, Park J, Ahn CH, Shin J, Oh KB[APR 2011]." Sargachromanols as inhibitors of Na+/K+ ATPase and isocitrate lyase". PMID: 21377877.

23) Lee YJ, Han YR, Park W, Nam SH, Oh KB, Lee HS(DEC 2012)." Synthetic analogs of indole-containing natural products as inhibitors of sortase A and isocitrate lyase". PMID: 21035332.

24) Yang HC, Yu J, Oh KB, Shin DS, Cho WJ, Shin J, Kim S(AUG 2007)." Synthesis and evaluation of hydroquinone derivatives as inhibitors of isocitrate lyase". PMID: 17879748 .

25) hin DS, Lee TH, Lee HS, Shin J, Oh KB(JUL 2007)." Inhibition of infection of the rice blast fungus by halisulfate 1, an isocitrate lyase inhibitor". PMID: 17456183.

26) Sriram D, Yogeeswari P, Senthilkumar P, Dewakar S, Rohit N, Debjani B, Bhat P, Veugopal B, Pavan VV, Thimmappa HM(SEP 2009)." Novel phthalazinyl derivatives: synthesis, antimycobacterial activities, and inhibition of *Mycobacterium tuberculosis* isocitrate lyase enzyme". PMID: 19534678.

27) Lee HS, Yoon KM, Han YR, Lee KJ, Chung SC, Kim TI, Lee SH, Shin J, Oh KB(FEB 2009)." 5-Hydroxyindole-type alkaloids, as Candida albicans isocitrate lyase inhibitors, from the tropical sponge Hyrtios". PMID: 19167886.

28) Kozic J, Novotná E, Volková M, Stolaříková J, Trejtnar F, Wsól V, Vinšová J(OCT 2012)." Synthesis and in vitro antimycobacterial and isocitrate lyase inhibition properties of novel 2-methoxy-2'-hydroxybenzanilides, their thioxo analogues and benzoxazoles". PMID: 22960697.

29) Sriram D, Yogeeswari P, Vyas DR, Senthilkumar P, Bhat P, Srividya M(AUG 2010)." 5-Nitro-2-furoic acid hydrazones: design, synthesis and in vitro antimycobacterial evaluation against log and starved phase cultures". PMID: 20615698.

30) Sriram D, Yogeeswari P, Senthilkumar P, Naidu G, Bhat P(DEC 2010)." 5-Nitro-2,6-dioxohexahydro-4-pyrimidinecarboxamides: synthesis, in vitro antimycobacterial activity, cytotoxicity, and isocitrate lyase inhibition studies". PMID: 20569083.

31) Shingnapurkar D, Dandawate P, Anson CE, Powell AK, Afrasiabi Z, Sinn E, Pandit S,Venkateswara Swamy K, Franzblau S, Padhye S(MAY 2012)." Synthesis and characterization of pyruvate-isoniazid analogs and their copper complexes as potential ICL inhibitors".

32) Lee D, Shin J, Yoon KM, Kim TI, Lee SH, Lee HS, Oh KB(OCT 2008)." Inhibition of Candida albicans isocitrate lyase activity by sesterterpene sulfates from the tropical sponge Dysidea sp". PMID: 18824352.

33) KL Britton, SJ Langridge, PJ Baker, K Weeradechapon, SE Sedelnikova, JR De Lucas, DW Rice ,G Turner." The crystal structure and active site location of isocitrate lyase from the fungus Aspergillus nidulans".PMID:10801489

34) Sharma V, Sharma S, Hoener zu Bentrup K, McKinney JD, Russell DG, Jacobs WR Jr, SacchettiniJC."Structure of isocitrate lyase, a persistence factor of Mycobacterium tub erculosis".
PMID:10932251.

35) Matsuoka, M. & McFadden, B.A. (1988). "Isolation, hyperexpression, and sequencing of the aceA gene encoding isocitrate lyase in EscherichiaColi".

36) Campos-Garcia J, Diaz-Perez C, Diaz-Perez AL(JUN 2013)." Residues Asn214, Gln211, Glu219 and Gln221 contained in the subfamily 3 catalytic signature of the isocitrate lyase from Pseudomonas aeruginosa are involved in its catalytic and thermal properties". PMID: 23338961.

37) Diehl P, McFadden BA (1994)." The importance of four histidine residues in isocitrate lyase from Escherichia coli".

38) Narayanan B, Niu W, Joosten HJ, Li Z, Kuipers RK, Schaap PJ,Dunaway-Mariano D, Herzberg O (2009) ."Structure and function of 2,3-dimethylmalate lyase, a PEP mutase/isocitrate lyase superfamily member".

39) Watanabe S, Takada Y (2004) ."Amino acid residues involved in cold adaptation of isocitrate lyase from a psychrophilic bacterium, Colwellia maris".

40) Sato Y, Watanabe S, Yamaoka N, Takada Y (2008) ."Gene cloning of cold-adapted isocitrate lyase from a psychrophilic bacterium,Colwellia psychrerythraea, and analysis of amino acid residues involved in cold adaptation of this enzyme".