Studies on Structure-Function Relationship of Enzymes involved in Lignin Biosynthesis from Leucaena *leucocephala*

A THESIS
SUBMITTED TO THE
UNIVERSITY OF PUNE

FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN
BIOTECHNOLOGY

BY
PARTH PATEL

UNDER THE GUIDANCE OF
Dr. D. C. AGRAWAL & Dr. C. G. SURESH

PLANT TISSUE CULTURE DIVISION
CSIR-NATIONAL CHEMICAL LABORATORY
PUNE- 411008
INDIA
April, 2014

The future destiny of a child is always the work of his mother.

to, My mother

सीएसआयआर–राष्ट्रीय रासायनिक प्रयोगशाला
(वैज्ञानिक तथा औद्योगिक अनुसंधान परिषद)
डॉ. होमी भाभा मार्ग, पुणे – 411 008. भारत
CSIR-NATIONAL CHEMICAL LABORATORY
(Council of Scientific & Industrial Research)
Dr. Homi Bhabha Road, Pune - 411008. India

CERTIFICATE

This is to certify that the work incorporated into the thesis entitled "**Studies on Structure-Function Relationship of Enzymes involved in Lignin Biosynthesis in Leucaena *leucocephala*"** submitted by **Mr. Parth Patel** for the degree of Doctor of Philosophy, was carried out under our supervision at Plant Tissue Culture Division, CSIR-National Chemical Laboratory, Pune. Materials obtained from other sources have been duly acknowledged in the thesis.

Dr. D. C. Agrawal
(Research Guide)

Dr. C. G. Suresh
(Research Co-guide)

Communication Channels
NCL Level DID : 2590
NCL Board No. : +91-20-25902000
EPABX : +91-20-25893300
: +91-20-25893400
FAX
Director's Office : +91-20-25902601
COA's Office : +91-20-25902660
COS&P's Office : +91-20-25902664
WEBSITE
www.ncl-india.org

DECLARATION

I, hereby, declare that the work of the thesis entitled **"Studies on Structure-Function Relationship of Enzymes involved in Lignin Biosynthesis in Leucaena *leucocephala*"** has been carried out at the Plant Tissue Culture Division, National Chemical Laboratory, Pune under the guidance of **Dr. D. C. Agrawal & Dr. C. G. Suresh**. The work is original and has not been submitted in part or full by me for any other degree or diploma to any other university. I further declare that the materials obtained from other sources have been duly acknowledged in the thesis.

(Parth Patel)
Date: April, 2014
Place: Plant Tissue Culture Division,
National Chemical Laboratory (NCL),
Pune-411008,
Maharashtra, India

ACKNOWLEDGEMENT

Pursuing doctoral research is often thought to be an individual project, however, I now realize it is a result of teamwork arising from endless support and help from many people, rather than a one-man-show, and I feel privileged in this regard. My stay at NCL for past half a decade has been quite intellectually demanding, as well as rewarding, apart from being pleasant, mainly due to an endless list of persons who helped me in various capacities. The foremost of them is Dr. B. M. Khan, a mentor who truly made a difference in my professional life. Dr. Khan is one person without whose encouragement, direction, and technical support I would not have developed an interest in plant proteins and their functions. The deep gratitude that I feel for my guide Dr. D. C. Agrawal is beyond the scope of this acknowledgement. I thank him for his kind help and encouragement and providing a welcoming and pleasant working environment together with constructive criticism and advice throughout my research. I would also like to express my sincere gratefulness and appreciation to my co-guide Dr. C. G. Suresh for his support and invaluable advice with a continuous stream of suggestions regarding protein crystallography, feedback and encouragement that have guided me throughout the period of my work. The efforts put into my research by Dr. S. M. Gaikwad are appreciated, especially for her endless discussions and critical feedback especially in structural transition experiments, and for lending her expertise in spectroscopic techniques and structural dynamics.

As I mentioned already, doctoral research is always a teamwork arising from constant feedback, support and troubleshooting from many people. And a student's colleagues stand out in this matter, with whom one also shares a personal bond apart from a professional rapport. Though only a handful of protein biochemists work at NCL (a predominantly chemistry oriented lab), the kind of expertise I received here was beyond my imagination, and a major part of my research takes its current form due to various interactive sessions with these peers. Though there is an endless list of contributors, I would name only a few who had a major impact on this work. And I apologize to anyone who I am unable to mention here, however, their help is duly acknowledged. Help extended towards my research by the following collaborators is appreciated: Dr. M. Fernandes for letting me use the CD spectroscopy facility; Dr. K. Saikrishnan for allowing access to the XRD facility at IISER, Pune; Dr. R. Sureshkumar for collecting diffraction data at SSRL, USA on my behalf; Dr. N. Varshney for introducing me to the methods in protein crystallography, as well as helping me out of various bottlenecks of crystallization and diffraction experiments; Dr. M. J. Kulkarni for allowing me to use HPLC & Mass spectrometry facility; Mr A. Nagpure for his help understanding the working of ICP-OES instrument; Mrs P. Singh for carrying out DLS analysis on various occasions; Dr. U. J. Mehta for allowing access to Gel documentation system.

Mere admiring words cannot do justice to the care and support provided to me by Mr. K. Patel, Dr. S. Singh, Dr. R. Vishwakarma and Mr. P. Sonawane who have been like a family to me. Our daily ritual of scientific discussions over evening tea and dinners had a deep impact on the very foundation of my work and on the way this entire thesis is structured. A word of affection goes to Dr. Neha Gupta-Patel who has always been a guiding star for me, lab-mate, friend, wife, who has always stood by me during my highs and lows. She has been a true inspiration for me to keep on pushing the bar a bit higher, encouraging me to keep on trying after every failed experiment. A special mention of my seniors Dr. P. Kulkarni, Dr. R. Zargar, Dr. P. Sharma-Gupta, Dr. S. Gupta, Dr. S. Omer-Gupta, Dr. S. Jadhav, who helped made me feel at home when I first came to Pune. I would also like to thank Mr. S. Abbassi and Ms. U. Kumari for their kind support during the later stages of my PhD career.

I would like to mention all the graduate lab trainees that worked with me during my tenure: Mr. N. Karthik, Ms. P. Wadate, Ms. M. Manochitra, Ms. A. Shiralkar, Mrs. D. Aswar, Ms. A. Rayjade, Ms. K. Gandhi and Ms. N. Agarwal. I thank all of them for their assistance and patience in working with me, especially handling my annoyance at times of experiments gone wrong. I also appreciate the efforts of Mr. D. Dandekar, Mr. K. Masand, Dr. M. Dixit and Mrs. P. Nagarkar who have been like stress-busters in times when I badly needed a break, as well as providing an outsider's perspective to my research which had helped me in certain extent. I would also like to mention Mr. Dandekar, Mr. K. Pawar, Mr. R. Suryaprasad and all other friends for the incalculable number of chess games we played together. Those were the moments which made me completely forget my experiments, all the troubles, anything that did not matter.

I must express gratitude towards different committee members of my review panel Dr. S. K. Rawal, Dr. A. Banarjee, Dr. H. V. Thulasiram and Dr. A. P. Giri for their critical and periodic assessments made during the progress of my work. I acknowledge the help by CSIR for providing me funding in the form of research fellowship, the Director CSIR-NCL for granting me permission to carry out my research, , as well as all the staff and scientists at CSIR-NCL for being ever so supportive. I thank all my friends with whom I shared accommodation both on and off-campus during my stay at NCL for sharing all the lighter moments and providing me a family away from home.

Talking about family, this acknowledgement is incomplete if not mentioning my mother Mrs. J. Patel, with whose conviction and moral support I was encouraged to take up a career in Science. She has been a single driving force in molding my childhood, cultivating the importance of "Science", and making me realize its true rewarding nature. I would also like to thank my sister Dr. M. Patel, other family members, including my in-laws, for their patience, love, and moral support during my research.

Parth Patel

ABSTRACT

TITLE: Studies on Structure-Function Relationship of Enzymes involved in Lignin Biosynthesis in *Leucaena leucocephala*.

Cinnamyl alcohol dehydrogenases (CADs) carry out the last catalytic reaction in the phenylpropanoid metabolism leading to lignin formation in plants. However, recently this physiological role of CADs has drawn certain ambiguity in angiosperms, particularly for their substrate specificities. Albeit having very high sequence identities, these CADs show a great diversity in their substrate preferences. So what is it that governs such variation amid CADs and in ADHs (alcohol dehydrogenases) in general, if not their conserved sequences? Another feature worth noting in different ADHs is their highly conserved structural architecture, thus raising another question as to what underlies this conserved structural basis amongst ADHs? It would thus be imperative to biochemically characterize CADs from different plant sources in order to not only answer these questions, but also elucidate their precise physiological roles in angiosperms, as well as to take a broader view of the co-ordination between the guaiacyl and syringyl lignins (precursors for lignin in plants), in general. With this outlook, the present work focuses on studying the structural and functional properties of a CAD homologue LlCAD2, from *Leucaena leucocephala,* with distinct catalytic properties than other known CADs.

Here, first the biochemical properties of LlCAD2 were characterized, followed by kinetic insights into its catalytic mechanism. Structural investigations were made by X-ray crystallography along with computational modeling approaches. Also, the active site was investigated following kinetic characterization of several site-directed LlCAD2 mutants. Finally, the structural transitions upon protein unfolding/refolding induced by different agents were portrayed by using CD spectroscopy and fluorimetry techniques. It was learnt from this work that LlCAD2 preferentially catalyzed sinapaldehyde rather than coniferaldehyde (preferred substrate for many other CADs). The molecular mechanism of catalysis is presented in this work, along with other features otherwise not yet investigated in other CADs. Some of these features include the unusual effect of Zn^{2+} on its structure and catalytic ability. Apart from this, atypical effects were observed in the substrate binding pocket, wherein the substrate specificity was dramatically changed upon mutating the co-enzyme binding site. Also, several different folding intermediate states of LlCAD2 structure were identified in this study, with marked structural difference when compared to native LlCAD2. The main objective of studying these intermediate states was to define a complete energy landscape for its folding mechanism and to draw a relationship between the structure of the protein and its sequence.

In conclusion, this work is focused on characterizing a CAD homologue from *Leucaena* sp. Since CADs from only a few model organisms are characterized in terms of their kinetic properties so far, our findings would not only add to the current knowledge of catalytic properties of different CADs, but also hold significance in deriving its physiological role in other economically important plants, and provide a better understanding of lignin metabolism as a whole.

CONTENTS

1

Cinnamyl alcohol dehydrogenase: An introduction

1A

An introduction to phenylpropanoid metabolism

Trees are reservoirs of many economically and biotechnologically significant products. Wood, one such gift of nature, constituted of lignin, hemicelluloses and cellulose. Biochemistry of lignin, being one of the most abundant biopolymers on earth, has been studied extensively, partly due to the significance and interest of such knowledge from industrial point of view. Lignin has far reaching impacts on agriculture, industry and the environment, making phenylpropanoid metabolism; a major route for synthesis of lignin in plants, a globally important part of plant chemistry. Considerable scientific interest is focused on the development of trees with improved wood quality through modification in the phenylpropanoid pathway, which could be important for the enhanced use of wood material. This introductory chapter briefly covers the chemistry of lignin in plants and different phenylpropanoid enzymes, followed by specific involvement of cinnamyl alcohol dehydrogenase in lignin formation.

1.1 Lignin: Occurrence, function and economical importance

Lignin (*Lignum*-Wood), is one of the most abundant natural organic polymer next to cellulose [1]. It is a vital cell wall component of all vascular plants and represents on an average of 25% of the terrestrial biomass, playing a crucial role in structural integrity of cell wall, rigidity and strength of the stem [2, 3]. In addition, lignin waterproofs the cell wall; enabling transport of water and solutes through the vascular system [4]. Lignin is primarily synthesized and deposited in the secondary cell wall of specialized cells such as xylem vessels, tracheids and fibers, and also deposited in minor amounts in the periderm where association with suberin provides a protective role against pathogens. Lignin content is higher in softwoods (27-33%) than in hardwood (18-25%) and grasses (17-24%), and highest amounts of lignin (35-40%) occur in compression wood on the lower part of branches and leaning stems [4]. The amount and composition of lignin vary among taxa, cell types and individual cell wall layer, and are influenced by developmental and environmental cues [5].

Besides its critical role in normal plant health and development, high lignin levels are problematic in the agro-industrial exploitation of various plant species. From an agro-economical point of view, lignin is considered to have a negative impact because it affects the paper manufacture and limits digestibility of forage crops. Both lignin content and composition are known to have impact on pulp and paper because residual lignin in the wood fibers causes a discoloration and a low brightness level

of the pulp [6]. High quantity and low Syringyl (S) to Guaiacyl (G) lignin ratio plays a detrimental role in economy and ecology of paper production. Every unit increase in S/G ratio decreases the cost of paper production by two and half times. Consequently, for the production of high quality paper, lignin has to be removed from cellulose during the pulping process without damaging the polysaccharide component of wood. During chemical (Kraft) pulping, a large amount of Sodium hydroxide (NaOH) and Sodium sulfide (Na_2S) are required to extract lignin from the pulp [7]. Subsequently, the residual lignin is further removed with bleaching agents, such as Chlorine dioxide (ClO_2), Hydrogen peroxide (H_2O_2), Sodium hypochlorite (NaOCl), Oxygen (O_2), or Ozone (O_3) [7-9]. These lignin extraction and bleaching procedures can partly degrade cellulose and consequently, reduce pulp quality and paper strength. Lignin extraction consumes large quantities of chemicals and energy leading to poor environmental image for these industries [10-12].

For this reason, engineering of plants having cell wall structures more susceptible to krafting; thus more amenable to hydrolysis (or which are sufficiently altered so as to shunt the above processes altogether) is an attractive approach to improve pulping efficiency and potentially alleviate some of the negative environmental impacts of the paper making industry. Hence, there is currently intense interest in modifying the content and/or composition of lignin as a means of improving the efficiency of the paper pulping process for forest trees. To engineer plants with agronomically useful lignin related traits, we need to devise strategies that can flexibly and predictably yield alterations in lignin monomer composition. Apart from the great deal of work in the lignin field for improving the pulping process, many examples can be found based on research aimed at altering the lignin content for improving bio-fuel production as well as for improving forage crop digestibility [13-16].

1.2 Phenylpropanoid metabolism

Lignins are complex racemic aromatic heteropolymers synthesized from the dehydrogenative polymerization of monolignols, namely coumaryl, coniferyl and sinapyl alcohol monomers differing in their degree of methoxylation [11, 17], synthesized via the phenylpropanoid pathway in plants [18]. These monolignols produce respectively, p-hydroxyphenyl (H), guaiacyl (G) and syringyl (S) phenylpropanoid units when incorporated into the lignin polymer (Fig. 1.1).

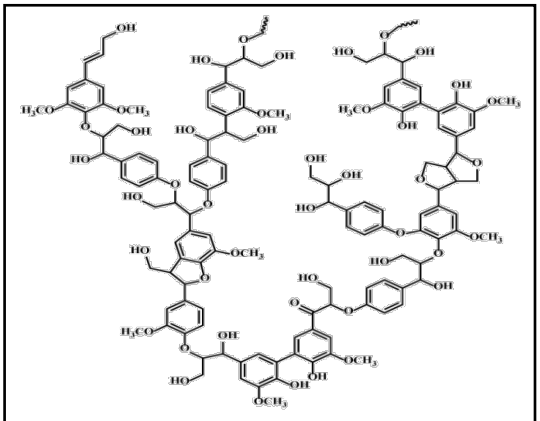


Fig. 1.1: *A hypothetical structure of lignin oligomer. These structures consist of phenylpropanoid units H (hydroxyphenyl), G (guaiacyl) and S (sinapyl). A number of such oligomers cross-polymerizes to form a complex structure of lignin*

Several reviews on the advancement of our understanding of the phenylpropanoid metabolism are available [1, 19-21]. This pathway is responsible for the biosynthesis of a variety of other products that include flavonoids and hydroxycinnamic acid conjugates with many of its intermediates and end products playing important role in plants as phytoalexins, antiherbivory compounds, antioxidants, ultra-violet (UV) protectants, pigments and aroma compounds. Synthesis of lignin represents one of the most energy demanding biosynthetic pathways in plants, requiring large quantities of carbon skeletons which proceed via following steps:

1. Monolignol biosynthesis via phenylpropanoid metabolic pathway
2. Transport of monolignols from the site of synthesis to the site of polymerization
3. Dehydrogenation & polymerization of monolignols.

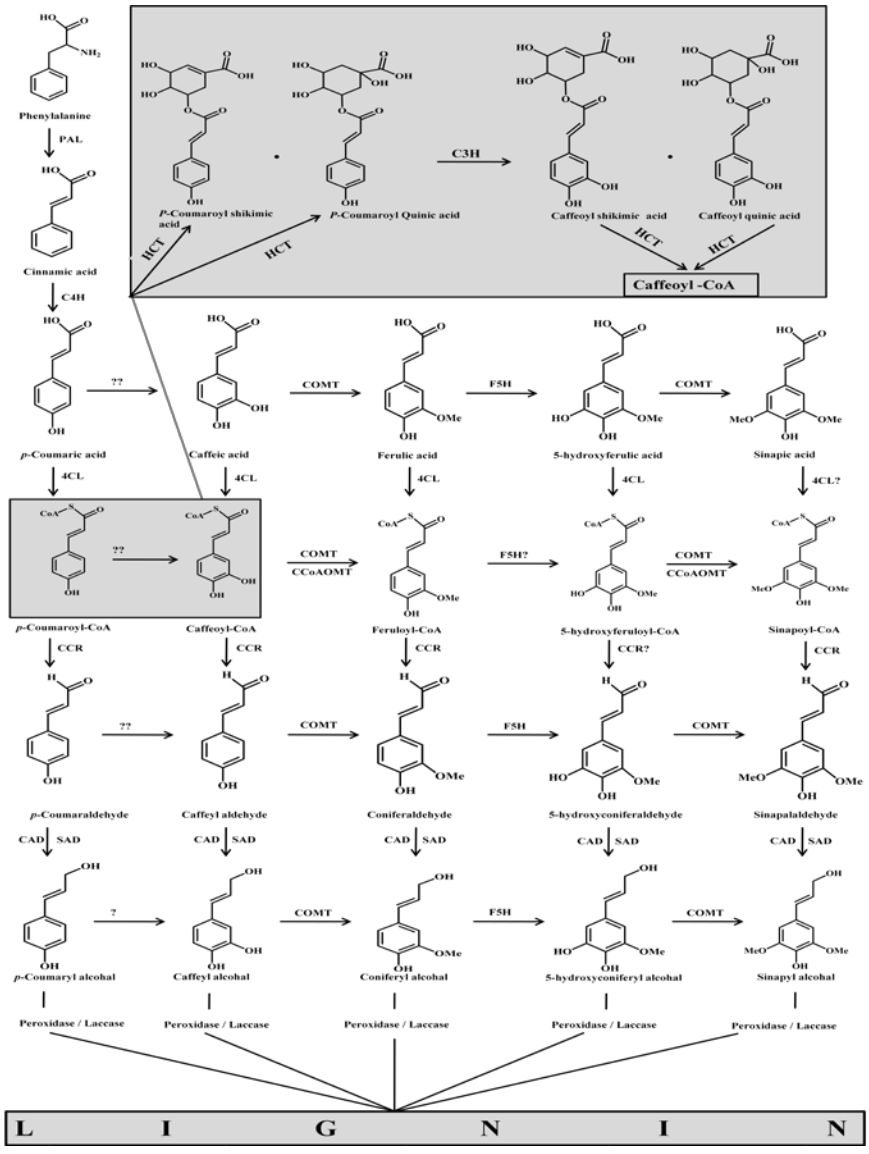


Fig. 1.2: *An overview of monolignol biosynthesis pathway: PAL (Phenylalanine ammonia lyase); C4H (Cinnamic acid 4-hydroxylase); C3H (p-Coumarate 3-hydroxylase); HCT (Hydroxycinnamoyl shikimate/quinate hydroxycinnamoyl transferase); 4CL (4-Coumarate-CoA ligase); CCoAoMT (Caffeoyl-CoA o-methyltransferase); CCR (Cinnamoyl CoA reductase); F5H (Ferulate 5-hydroxylase); CoMT (Caffeate o-methyltransferase); CAD (Cinnamyl alcohol dehydrogenase)*

Immense amount of work has been done in characterizing the monolignol metabolism in past two decades. It is a complex pathway comprising of enzymes with functions like methyltransferase, hydroxylase, reductase, dehydrogenase, etc (Fig. 1.2). The entry point to this pathway starts with deamination of phenylalanine to produce cinnamic acid by Phenylalaline ammonia lyase (PAL), and successive hydroxylations of the aromatic ring by Cinnamate 3 & 4-hydroxylases (C3H, C4H). Consequently, various degrees of methoxylations involving *o*-methyltransferases like Caffeate 3-*o* methyltransferase (C*o*MT) and Caffeoyl CoA 3-*o* methyltransferase (CCoA*o*MT), and hydroxylations by Ferulate 5-hydroxylase (F5H) take place. Co-enzyme A is then transferred to thus formed hydroxicinnamic acids with varying degrees of methoxylations by 4-Coumarate-CoA ligase (4CL) and ultimately the side chain carboxyl groups are reduced to alcohol by Cinnamoyl Co-A reductase (CCR) and Cinnamyl alcohol dehydrogenase (CAD). These hydroxycinnamyl alcohol precursors varying in their extent of methoxylation are then transported to the site of lignin formation in the cell wall. These monolignols are insoluble and toxic to the plant cell, and hence are converted to their respective glucosides by the action of UDP-glycosyltransferases (UDP-GT) before being transported [22]. After transport, lignin is formed through dehydrogenative polymerization of the monolignols [9] by different enzymes such as peroxidases (POX), laccases (LAC), polyphenol oxidases and coniferyl alcohol oxidase [23].

1.3 Cinnamyl alcohol dehydrogenase (CAD)

CAD (EC 1.1.1.195) is a NADP(H) dependent oxidoreductase which catalyzes the last step in the phenylpropanoid metabolic pathway; reversible reduction of various hydroxycinnamyl aldehyde derivatives to corresponding alcohols [24]. These alcohols, in turn, are obligatory precursors of structural cell-wall lignins as well as health-related lignans, podophyllotoxin, matairesinol [25], secoisolariciresinol [26], and mammalian lignans enterolactone and enterodiol [27] as mentioned in Fig. 1.3. This reduction carried out by CAD has been considered to be an indicator of lignin biosynthesis because of its specific role at the end of the monolignol pathway. Bona fide clones encoding *CAD* were described by Knight et al. in *Nicotiana tabaccum* [28] and from *Eucalyptus gunnii* [29]. The expression of CAD is regulated by both developmental and environmental cues. CAD is also expressed in response to stress [30], pathogen elicitors [31] and wounding.

1.4 Down-regulation of CAD

CAD not only serves as an interesting model to understand the regulation of lignin metabolism, but also provides a tool to exploit it in order to alter the lignin content and composition in economically important plants like *Leucaena* sp. Such manipulations achieved by down-regulating CAD in several

plants have been extensively reviewed [20, 32-34] and the findings suggest the significance of CAD in not only achieving desired alterations in plants, but also providing new insights into the regulatory mechanisms of this complex metabolic pathway.

Transgenic plants with reduced CAD activity have been produced in tobacco [35], poplar [36] and alfalfa [37], whereas CAD mutant exists in pine [38], maize [39], Arabidopsis [40] and sorghum [41]. Increased levels of cinnamaldehydes have been detected in the lignin of *CAD* downregulated tobacco [35, 42, 43] and poplar [44], *Arabidopsis* [40] and maize [45] CAD mutants. In tobacco, poplar and *Arabidopsis*, increase in both coniferaldehyde and sinapaldehyde were identified [35, 40, 43, 44]. In CAD antisense poplar and in the maize *bm*1 mutant [39, 46], the S/G ratio is not altered, whereas in CAD antisense tobacco and alfalfa [35, 37], the S/G ratio was reduced. This suggests that in CAD-deficient plants, the deposition of S lignin units is equally or more affected than that of G lignin units. In transgenic poplar and tobacco, the lignin was enriched in free phenolic groups in both S and G units [46, 47]. The increase in free phenolic groups may be important in altering the solubility of lignin which in turn has implications for the ease with which wood from these plants can be pulped [46, 47]. All these reports suggests that the plants are able to circumvent the block created by down-regulating CAD by shipping the hydroxycinnamaldehydes to the cell wall for polymerization, thus showing only marginal reduction in lignin content.

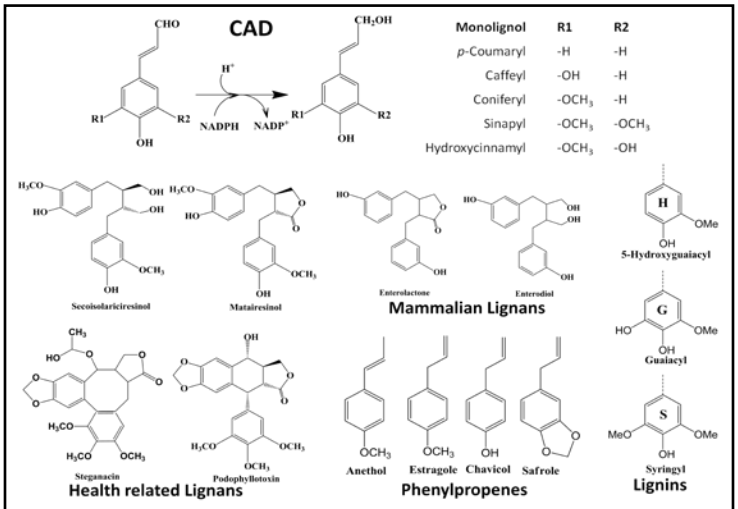


Fig. 1.3: *A figure showing the catalytic reaction carried out by CAD. Different hydroxycinnamyl substrates with different R groups are also mentioned. The end-products of the phenylpropanoid metabolic pathway are synthesized from the hydroxycinnamyl alcohols*

1.5 Biochemical studies of CAD

CADs are dimers and belong to the Zn^{2+}-dependent medium-chain dehydrogenase/ reductase (MDR) super family with two zinc ions per subunit (one catalytic and the other structural). Much of our current understanding of the structure and biochemical properties of CAD stems from comprehensive studies on CAD homologues from *Arabidopsis* [48, 49]; however, these reports suggest some functional redundancy amongst them. There are quite a few reports of CADs from

other plants, which show some similarity to the *Arabidopsis* CAD (AtCAD5), though, their precise biochemical and physiological role is uncertain. CADs from gymnosperms and angiosperms display distinct features amongst them, leading to diverse monolignol compositions in their lignin structure. Several CADs are identified from gymnosperms being coniferaldehyde specific with insignificant activity toward sinapaldehyde, which is consistent with predominant formation of guaiacyl units in these plants [20, 32, 50].

In contrast, multiple CAD homologues have been purified from angiosperms demonstrating comparable catalytic activities for both coniferaldehyde and sinapaldehyde, thus leading to lignin containing both guaiacyl and syringyl units in varying proportions [18, 50-52]. Isoforms of CADs with markedly different substrate affinities were detected in such species as soybean [24] and *Eucalyptus* [53]. Differences in substrate affinities of different CADs from angiosperms and gymnosperms may play a role in controlling the formation of different types of lignin [50]. These studies propose a model wherein the last step toward monolignols in angiosperms is mediated by broad substrate specificity CAD homologues capable of reducing both coniferaldehyde and sinapaldehyde, and can be considered a critical step for lignin biosynthesis [20].

1.6 Sinapyl alcohol dehydrogenase

It has been already proposed that the S and G monolignols are reduced from coniferaldehyde and sinapaldehyde, respectively, via the action of multi substrate CAD isoforms. Though, the discovery of a novel enzyme sinapyl alcohol dehydrogenase (SAD) from *Populus tremuloides*, which preferentially uses sinapaldehyde, raised a basic question to such a hypothesis [54]. This role of SAD has also been reviewed and mentioned in many reports since, though the existence SAD is questioned by several groups. Recently, SAD was also identified in *N. tabacum* [55]. Down-regulation of this *NtSAD* showed no change in the content, composition or structure of lignin in the transgenic plants and the S lignin was observed to be normal. However, additional *CAD* suppression in these plants lead to an increase in sinapaldehyde in lignin and changes in S/G ratio, which was quite similar to plants suppressed in only *CAD*. These results along with many other reports on CAD-suppressed transgenic or mutant plants showed accumulation of both coniferaldehyde and sinapaldehyde in lignin, suggest that CAD and not SAD, is responsible for G as well as S lignin formation in angiosperms.

1.7 Rational of the thesis

L. leucocephala; one of the most important resources for paper and pulp production in India, is so far not much studied for its enzymes taking part in lignin formation. Our group at the CSIR-National

Chemical Laboratory (Pune, India) has been working on genes involved in lignin biosynthesis pathway in *L. leucocephala*, for the past 10 years [34]. Our studies have concentrated on attempts to alter the levels of enzymes involved in the phenylpropanoid pathway, enzymes involved in polymerization like peroxidases, as well as transcription factors in *L. leucocephala*. Though CAD-suppression leads to a small decrease in lignin content, which possibly would result in specific alterations in the monomer architecture of lignin as also observed by others [42, 56, 57]. As mentioned already, CAD belongs to the medium chain dehydrogenase super-family. This family of enzymes also includes several other alcohol dehydrogenases with diverse functions amongst different kingdoms, with a conserved reaction mechanism. AtCAD5 is the most characterized aromatic alcohol dehydrogenases amongst plant kingdom which shows reactivity towards different hydroxycinnamyl aldehydes with varying degrees of specificity. It has been proposed that angiosperms express multi-substrate specific CADs which governs S/G ratio, thus conferring unique composition to lignin in different plant species.

Two isoforms of CAD had been identified from *L. leucocephala* in our lab. Though LlCAD2 (other one being LlCAD) sequence is highly identical to AtCAD4 & 5, and PtCAD (CAD from *P. tremuloides*), its substrate preference is similar to PtSAD, i.e. being sinapaldehyde specific. Another well characterized CAD is reported from the gastric bacterium *Helicobacter pylori* [58, 59], which apart from showing preference for hydroxycinnamyl aldehydes, also shows unique dismutating properties against benzaldehyde. Other than these aromatic alcohol dehydrogenases, there are several aliphatic alcohol dehydrogenases reported from yeast [60], horse [61], human [62], several prokaryotes [63-65] etc., showing substrate preference for primary and secondary aliphatic alcohols. Such diversity in substrate preference of these enzymes leads to a basic question: what governs the substrate preference amongst CADs, if not conserved sequence? Though with modest sequence identities, all these dehydrogenases shows a highly conserved structural architecture. The structural analysis of these enzymes leads to another elementary question: what underlies this conserved structural basis amongst different ADHs. These two questions formed the starting point for the present work as attempts were made to draw a relationship between the structure and the function of LlCAD2.

Though general conclusions cannot be made by analysis of a single enzyme, it would be imperative to biochemically characterize CADs from other sources in order to elucidate their precise physiological roles in angiosperms, and to take a broader view of the co-ordination between the guaiacyl and syringyl lignins, in general. With this outlook, the present work was focused on studying the biochemical and catalytic properties of a CAD homologue from *L. leucocephala.*

1B

Experimental procedures

Different methods and protocols used for this study are described here in detail. Any routine protocols followed commonly in many labs are not mentioned in detail. Protocols adopted from other sources are duly cited. All chemicals were either purchased from Sigma, Invitrogen, GE Healthcare, or Novagen.

1.8 Preliminary experiments

1.8.1 Isolation and cloning

Xylem tissue of mature *L. leucocephala* cv. K-636 grown in NCL campus was used to isolate mRNA using Ambion RNA extraction kit (Life Technologies) and cDNA was prepared using cDNA synthesis kit (Promega). *LlCAD2* was amplified using primers LlCAD2F (catatgggaagcattgaaggagaaag), LlCAD2-R (gtcgacctgatgatcatcaagtttgc) with *Nde*I and *Sal*I sites designed from sequence of *LlCAD* (EU870436), cloned in pGEM-T Easy TA cloning vector and confirmed by nucleotide sequencing. UTRs were amplified by RACE PCR and ORF of *LlCAD2* was amplified using same primers and cloned into pET-30b (+) vector (Novagen) using same primers.

1.8.2 Heterologous expression in E. coli

The recombinant pET-30b (+) vector harboring *LlCAD2* was transformed into *E. coli* BL21 (DE3) cells (Novagen) and maintained as glycerol stocks, and used to inoculate 10 ml LB broth (1% NaCl, 1% Tryptone, 0.5% yeast extract, pH 7.2) containing 50 μg ml^{-1} Kanamycin, grown at 37 °C for 12 h. This culture was used to inoculate 1 L LB broth containing 50 μg ml^{-1} Kan which after active growth (A_{600} 0.5) at 37 °C was induced by 1 mM isopropyl β-D-1-thiogalactopyaranoside (IPTG) and grown at 160 rpm for 16 h at 15 °C. Harvested cells were resuspended in lysis buffer (50 mM Tris, 500 mM NaCl, 20 mM Imidazole, 20% Glycerol, 1 mM DTT, 1.5% Triton-X 100; pH 8), sonicated at 70% amplitude with 5 s pulse with 5 s intervals on XL-2000 sonicator (Mesonix). Lysozyme (0.5 mg ml^{-1}) & streptomycin (10 mg ml^{-1}) were added and incubated for 30 min and centrifuged at 14000 rpm for 15 min. The supernatant was filtered through 0.22 micron syringe filter.

1.8.3 Enzyme purification

Purification of CAD was achieved by a three step protocol using affinity, anion-exchange and gel-filtration chromatography employed on an Akta Explorer FPLC system (GE Healthcare). Cell lysate

was loaded onto Ni-Sepharose 6 FastFlow column (GE Healthcare) pre-equilibrated with Buffer A (50 mM Tris, 500 mM NaCl, 1 mM DTT; pH 8) containing 20 mM imidazole. The column was washed with Buffer A containing 30 mM imidazole till A_{280} reached a baseline and finally the bound protein was eluted by applying Buffer A containing 250 mM imidazole. The elute was concentrated using Amicon Ultracel-10K centrifugal filters with 10,000 MWCO (Millipore) and desalted using HiLoad 16/10 Sephadex G-25 Desalting column pre-equilibrated with Buffer B (20 mM Tris, 1 mM DTT; pH 8) and loaded onto Mono-Q 10/100 GL column equilibrated with Buffer B, and unbound protein was washed for 20 CV (column volume) with same buffer. Bound protein was eluted with a gradient elution of Buffer B containing 1M NaCl (a: 0-250 mM for 20 CV, b: 50-500 mM for 10 CV, c: 1 M for 5 CV). After concentrating the eluted protein on Ultracel, final purification step was performed on HiLoad 16/60 Sephacryl S-100 in Buffer B. Fractions showing CAD activity were pooled and the purity of the enzyme was judged by SDS-PAGE and western blot. Dimeric CAD fractions (peak 2 from Sephacryl S-100) were concentrated and stored as either 250 µl aliquots of 1 mg ml^{-1} (for kinetics studies), or 50 µl aliquots of 10 mg ml^{-1} concentration (for structural studies) in Buffer C (20 mM Tris, 20% Glycerol, 1 mM DTT; pH 8) at -80 °C. The aliquots were thawed on ice for 10 min and used for experiments by diluting to appropriate concentrations in Buffer C. However, for crystallization, fluorescence, and Circular Dichroism (CD) experiments, the samples were buffer exchanged in Buffer D (20 mM Tris; pH 8) before use.

1.8.4 Protein electrophoresis & protein quantification

Acrylamide gels (10%) were used for all separations on SDS-PAGE and 8% for native PAGE, and were stained either by using Coomassie Blue R-250 or silver staining method. Protein was quantified by Bradford method using BSA as standard. Titer plates (96-well) were used for quantifying multiple fractions at a time by adding 180 µl Bradford reagent to each well containing 1-5 µl of the protein sample, and measuring A_{595} using xMark microplate reader (Bio-Rad).

1.9 Protein Characterization

1.9.1 Protein identification by ESI-MS/MS

Protein bands were excised from SDS-PAGE gels and subjected to *in situ* trypsin digestion [66]. The digested peptides with a final concentration 100 ng μl^{-1} were analyzed by using nanoACQUITY UPLC coupled online to a SYNAPT HDMS system (MS^E) (Waters) as described by Cheng et al. [67]. (2009). After MS^E analysis, data were analyzed with Protein Lynx Global Server software (PLGS v 2.4 Waters).

1.9.2 Molecular weight determination

The molecular weight of the purified protein was calculated by running protein samples onto Sephacryl S-100 column. Low molecular weight Gel filtration calibration kit (GE Healthcare) was used for calibrating the column according to kit's instructions. The standard calibration curve is provided in Fig. 1.4.

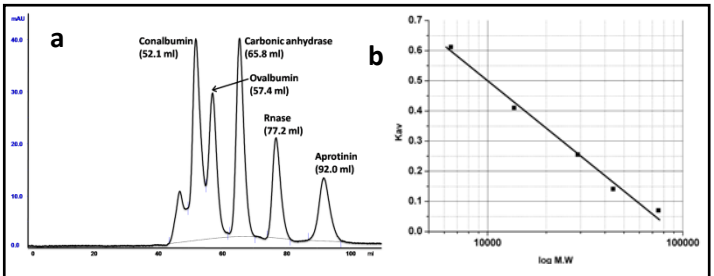


Fig. 1.4: *Molecular weight determination. (a) Elution chromatogram of standard proteins (Aprotinin (6.5 kD), RNase A (13.7 kD), Carbonic anhydrase (29 kD), Ovalbumin (44 kD), Conalbumin (75 kD) on HiPrep 16/60 Sephacryl S-100 HR column with CV 120 ml. Sample volume: 100 µl, Buffer: 50 mM Tris, 150 mM NaCl, pH 8.0, Flow rate: 0.5 ml min^{-1}. (b) Plot of K_{av} vs log Mr.*

1.9.3 Western Blot

Protein samples (20 µg) were run on 10% SDS-PAGE and transferred onto PVDF membrane through iBlot Dry blotting system (Invitrogen) according to manufacturer's protocol. The membrane was hybridized with anti-CAD (polyclonal antibody raised in New Zealand white rabbits at National Toxicology Center, Pune) and anti His-tag antibodies (Merck). The membranes were blocked by BSA, incubated in appropriately diluted antibodies for 2 hours each, with intermittent washings, and then visualized by adding the substrate for alkaline phosphatase [68].

1.9.4 Activity Staining

Protein was run on 8% non-denaturing PAGE followed by gel incubation in 10 ml of 50 mM Tris; pH 8, containing: 1.5 mg nitroblue tetrazolium, 0.1 mg phenazinemethosulfate, 2.5 mg $NADP^+$ and 2.5 mg coniferyl alcohol. Blue colored bands for activity stained bands were visualized within 30 min of incubation in dark [51].

1.9.5 Metal detection by ICP-OES

Protein sample (12 µM) was digested in a mixture of conc. nitric acid/perchloric acid (3:1) at 90 °C for 3 h. The digest was diluted 5 times with distilled water and filtered through 0.22 µm syringe filter. The filtrate was subjected to inductively coupled plasma on a Spectro-Arcos optical emission spectroscope (Spectro Analytical) at a flow rate of 1 ml min^{-1}. The instrument was calibrated using ICP metal standards from Sigma.

1.9.6 Zincon assay

Zn^{2+} was determined with Zincon (2-carboxy-20-hydroxy-50-sulfoformazylbenzene) dye binding assay. Protein samples (1 µM) were first treated with 0.2 M HCl for 5 min and subsequently diluted 10 times with borate buffer (50 mM; pH 9.0) containing 8 M urea. Following incubation for 10 min, Zincon was added to a final concentration of 40 mM, and A_{615} was recorded. Metal concentration was calculated from a standard curve with $ZnCl_2$.

1.9.7 Dynamic Light Scattering (DLS)

Protein solution (2.4 µM) in Buffer D was subjected to particle size analysis at 28 °C on a Brookhaven 90 Plus particle size analyzer (Brookhaven Instruments).

1.9.8 Reaction with 2, 2'-dithiobisnitrobenzoic acid (DTNB)

Free and total cysteine residues were estimated by the method of Habeeb [69]. Molar extinction coefficients of 13,600 M^{-1} cm^{-1} and 12,000 M^{-1} cm^{-1} were used at A_{412} for estimating free and total cysteine, respectively. Blanks containing all reactants except protein solution were subtracted.

1.10 Enzyme Kinetics

1.10.1 Spectrophotometric enzyme assays

The substrates were initially dissolved in 2-methoxyethanol and further diluted with appropriate buffer for obtaining the molar extinction coefficients (Table 1.1). All reagents were consumed within 4-5 h of preparation. Reduction of aldehydes was determined by monitoring the decrease in A_{340} on Lambda 650 spectrophotometer (Perkin Elmer) due to the depletion of NADPH and the aldehydes [24]. The assay was carried out in 0.5 ml reaction mixture containing 100 mM phosphate buffer (pH 6.5), 50 µM aldehyde, 100 µM NADPH and 0.1 µg purified LlCAD2. Oxidation of alcohols was determined similarly by measuring the increase in A_{340} in 100 mM Tris (pH 8.8), 50 µM alcohol, 100 µM $NADP^+$ and 0.5 µg purified protein. All experiments were conducted in triplicates and enzyme activity was expressed as nKat μg^-

Substrate	Extinction coeff (M^{-1} cm^{-1})
Cinnamaldehyde	1423
p-Coumaraldehyde	22347
Coniferaldehyde	26450
Sinapaldehyde	27368
Cinnamyl alcohol	1132
Coniferyl alcohol	1386
Sinapyl alcohol	1590
NADPH	6220
$NADP^+$	1320

Table 1.1: *Molar extinction coefficients*

1.10.2 Determination of kinetic constants

The kinetic parameters for the substrate [B] (similar plots were also prepared for [A], [P] and [Q]) were determined by fitting the data to sequential-order ternary-complex mechanism:

$$v = \frac{V_{max}\,[A][B]}{K_{iA}K_B + K_A[B] + K_B[A] + [A][B]} \qquad (1)$$

where, V_{max} is the maximum velocity, K_A and K_B are Michaelis constants for A and B, K_{iA} is the dissociation constant for A. The turnover number K_{cat} was defined as V_{max}/ [E] with the unit s^{-1}. Catalytic efficiency (K_{cat}/K_m) was defined as $\mu M^{-1}\ s^{-1}$. Several sub-experiments were carried out for determination of apparent kinetic parameters, V^a_{max}, K^a_m for [B] by varying the concentrations of [B] each at different concentrations of [A]. Rearranging and simplifying (1), and defining the apparent parameters as functions of [B]:

$$v = \frac{V^a_{max}\,[A]}{K^a_m + [A]} \qquad (2) \qquad\qquad V^a_{max} = \frac{V_{max}\,[B]}{K_B + [B]} \qquad (3)$$

$$K^a_m = \frac{K_{iA}K_B + K_A[B]}{K_B + [B]} \qquad (4) \qquad\qquad \frac{V^a_{max}}{K^a_m} = \frac{\left(\frac{V_{max}}{K_A}\right)[B]}{\left(\frac{K_{iA}K_B}{K_A}\right) + [B]} \qquad (5)$$

The following plots were prepared for determining V^a_{max}, K^a_m and for visual assessment:

$\boldsymbol{v} \rightarrow [\boldsymbol{B}]$	$\boldsymbol{1/v} \rightarrow \boldsymbol{1/[B]}$	$[\boldsymbol{S}]/\boldsymbol{v} \rightarrow [\boldsymbol{B}]$	Primary plots
(Michaelis-Menten)	(Lineweaver-Burk)	(Hanes)	

Since (3) and (5) resemble M-M equation, true kinetic parameters: V_{max}, K_{iA}, K_A, K_B, were calculated from:

$$\boldsymbol{1/V^a_{max} \rightarrow 1/[A]} \qquad\qquad \boldsymbol{K^a_m/V^a_{max} \rightarrow 1/[A]} \qquad \text{Secondary plots}$$

1.10.3 Activation energies

Energies of activation (E_a) were calculated from modified Arrhenius plot [70]:

$$\boldsymbol{lnV_{max} \rightarrow 1/T} \qquad\qquad (E_a = -slope\ R)$$

(R, gas constant= 8.314×10^{-3} kJ mol^{-1}). The free energy of binding, ΔG was calculated:

$$\Delta G = RTlnK_d \qquad\qquad (K_m = K_d) \qquad (6)$$

1.10.4 Dissociation constants

The molecular dissociation constants (pK_E & pK_{EA}) for free enzyme [E] and that of the enzyme-substrate [EA] complex were calculated by plotting, respectively:

$$\boldsymbol{ln(V_{max}/K_m) \rightarrow pH} \qquad \boldsymbol{ln(V_{max}) \rightarrow pH}$$

1.10.5 Deactivation rate constants and thermodynamic parameters for inactivation

The deactivation rate constant (K_d^*) was determined from [71]:

$$\ln[E_t/E_0] = -K_d^* t \tag{7}$$

where, E_0 and E_t are initial and final enzyme concentrations, respectively. Half-life was estimated by:

$$t_{1/2} = ln2/K_d^* \tag{8}$$

Change in Enthalpy ΔH^* and Entropy ΔS^* for inactivation were calculated from:

$$\boldsymbol{\ln[K_d^*/T] \rightarrow 1/T} \qquad (\Delta H^* = -slope\ R) \qquad (\Delta S^* = R[intercept - \ln(k_b/h)])$$

Free energy change (ΔG^*) for inactivation was calculated by:

$$\Delta G^* = -RT\left[ln\left(\frac{K_d^* T}{K_b h}\right)\right] \tag{9}$$

where, h (Plank constant) = 6.63×10^{-34} J s, K_b (Boltzman constant) = 1.38×10^{-23} J K^{-1}. Activation energy of inactivation (Ea^*) was estimated from:

$$\boldsymbol{\ln[K_d^*] \rightarrow 1/T} \qquad (E_a = -slope\ R)$$

1.10.6 Inhibition kinetics

LlCAD2 (1µg) was assayed as mentioned already, in presence of inhibitors at 30 °C. Determination of the inhibition rate constant, K_i, was performed in triplicate over a range of inhibitor concentration at different concentrations of substrate by fitting the data to:

$$v = \frac{V_{max}\,[S]}{K_m C + [S]U} \tag{10}$$

The inhibition terms C and U in the denominator are as follows:

Type of inhibition	C	U
Competitive	$(1 + [I]/K_{ic})$	1
Mixed	$(1 + [I]/K_{ic})$	$(1 + [I]/K_{iu})$
Non-competitive	$(1 + [I]/K_i)$	$(1 + [I]/K_i)$
Uncompetitive	1	$(1 + [I]/K_{iu})$

K_{ic} and K_{iu} are the inhibition rate constants for competitive and uncompetitive inhibitors, respectively. Type of inhibition & transformation of kinetic data into K_i values were done by the primary and secondary plots mentioned in 1.10.2 and:

$\boldsymbol{1/v \rightarrow [I]}$ (Dixon) and $\boldsymbol{[S]/v \rightarrow [I]}$ (Cornish-Bowden) } Inhibition plots

The data was re-plotted with inhibitors as the variable at fixed [S]. The data was fit into the following equation which can be used to fit inhibition data:

$$i = \frac{I_{max}\,[I]^h}{K_{0.5}^h + [I]^h} \tag{11}$$

This equation was derived from Michealis-Menten equation wherein *i* is the inhibition achieved, which can be defined as the decrease in initial velocity at a particular inhibitor concentration (having same units as the initial velocity *v*), [*I*] is the inhibitor concentration (which replaces [*S*]), I_{max} is the maximum inhibition achieved (analogous to V_{max} with the same units), $k_{0.5}$ replaces K_m ([*I*] at half I_{max}, also called the IC_{50}), and *h* is the Hill coefficient. When *h*=1, this equation is identical to the standard Michaelis-Menten equation. When it is greater than 1, the curve is sigmoidal due to positive allosterity. Though by definition, *h* gives the measure of co-operativity for a protein molecule, here its value can also be used as a measure of the number of interacting binding sites for the inhibitor used. Inhibition was also carried out at temperatures from 30 to 60 °C, and *ΔH* was calculated from:

$$\mathbf{ln}[\boldsymbol{K_a}] \rightarrow \mathbf{1000/}\boldsymbol{T} \qquad (K_a = 1/K_i)$$

ΔG was calculated by using (6). Here, the values of K_i were used for dissociation constant K_d. *ΔS* was calculated as:

$$\Delta G = \Delta H - T\Delta S \tag{12}$$

1.10.7 Substrate inhibition

In all reasonable bi-substrate reactions, there is a probability of at least one of the reactants binding to the wrong form of enzyme, thus leading to enzyme inhibition at higher substrate concentrations with an inhibition constant of K_{si} . Data were analyzed by primary plots, by fitting the data to:

$$v = \frac{V_{max}\,[B]}{K_B + [B](1 + [B]/K_{si})} \tag{13}$$

1.10.8 Product inhibition

Product inhibition studies were employed using different reaction products to elucidate the reaction mechanism of LlCAD2. The primary aim of this experiment was to define the type of inhibition caused by each product on their respective reactions and understand the order of substrate binding and product release. Due to the complexity presented by the rate equation for compulsory-order ternary-complex mechanism, and due to our limited understanding of the equation, no attempts were made to estimate the rate constants. Variable substrate was varied from 0-30 µM fixed

concentrations of other substrate (5, 50 µM). Inhibitor product was varied from 0-30 µM. Type of inhibition was characterized using primary, secondary & inhibition plots shown in 1.10.2 and 1.10.6.

1.10.9 Competition plot

Since there are multiple substrates available for CAD, it was highly imperative to measure enzyme activity in presence of more than one substrate in the same reaction mixture. The concentration of both substrates was chosen such that their initial velocities v_0 would be equal:

$$\frac{V_{max}\,[S]}{K_m + [S]} = \frac{V'_{max}\,[S']}{K'_m + [S']} = v_0 \qquad (14)$$

where, S is first substrate, and K_m and V_{max} are its kinetic parameters, similarly, S' is the second substrate with K'_m and V'_{max} as its parameters. v_0 is the initial velocity. The above equation shows that there are thus two concentrations that lead to the same initial rate (at the same enzyme concentration) for each substrate when the other is absent. A series of mixtures were prepared in which the concentrations are linearly interpolated between zero and these reference concentrations;

$$\left.\begin{aligned} a &= (1-r)[S]_0 \\ a' &= r[S']_0 \end{aligned}\right\} \qquad (15)$$

Total velocity in presence of both substrates v_{tot} was measured at A_{340}, and plotted:

$$\boldsymbol{v_{tot} \rightarrow r} \qquad \text{(Competition plot)}$$

1.10.10 Equilibrium kinetics

The reversible reaction carried out by CAD can be written as:

$$(A + B \leftrightharpoons P + Q) \qquad (16)$$

When the reaction is at equilibrium, the concentrations of the substrates and products can be used to estimate the equilibrium constant K_{eq}:

$$K_{eq} = \frac{[P]_{eq}\,[Q]_{eq}}{[A]_{eq}\,[B]_{eq}} \qquad (17)$$

Here, the reaction was set up as described earlier, and was allowed to equilibrate at 30 ºC for 3 h. A_{340} was measured before adding the enzyme and after incubation, and difference in A_{340} was calculated. Equilibrium concentrations of all reactant and products were estimated according to individual extinction coefficients. Since only the reactants absorbed at 340 nm, their concentrations were calculated by Beer-Lambert's law, while the product concentrations were estimated as:

$$[P]_{eq} = [S]_0 - [S]_{eq} = \Delta[S] \qquad (18)$$

Free energy of change ΔG' was calculated from (6) where $K_d = 1/K_{eq}$.

1.10.11 Equilibrium binding

Quenching of native tryptophan fluorescence of LlCAD2 upon substrate binding was used to measure dissociation constants (K_d) of binding. Protein (0.6 µM) was titrated against substrate solution in small increments and changes in tryptophan fluorescence were monitored using LS55 spectrofluorimeter as mentioned in 1.12.1, 1.12.2. Buffer corrections for dilutions were made. The concentration of bound and free ligand was estimated by:

$$[L]_b = [E]\left(\frac{F}{F_{max}}\right) \quad (19) \qquad [L]_f = [L] - [L]_b \quad (20)$$

where, $[L]_f$ is the free ligand concentration, $[L]_b$ is bound ligand, $[L]$ is total ligand, $[E]$ is total enzyme, F is change in fluorescence intensity upon protein-ligand interaction, F_{max} is change in fluorescence intensity at saturation binding. The data was analyzed by plotting the data to:

$$\boldsymbol{v \rightarrow [L]} \qquad \boldsymbol{[L]_b/[L]_f \rightarrow [L]_b}$$

(Binding plot) (Scatchard plot)

Binding data was fit into any of the equations:

$$v = \frac{n[L]}{K_d + [L]} \quad (21) \qquad v = \frac{n_1[L]}{K_{d1} + [L]} + \frac{n_2[L]}{K_{d2} + [L]} \quad (22) \qquad v = \frac{n[L]^h}{K_d^h + [L]^h} \quad (23)$$

(One-site binding) (Independent two-site binding) (Co-operative binding)

where, v is number of moles of $[L]_b$ per $[E]$, n is number of binding sites on the protein, K_d is the dissociation rate for $[L]$. Free energy changes of association (ΔG) were determined according to (6).

1.11 Structural studies

1.11.1 Protein crystallization

LlCAD2 was purified as already mentioned; stocks of 10 mg ml^{-1} were stored at -80 ºC in 20 mM Tris; pH 8.0. The aliquots were thawed at 4 ºC, centrifuged at 10000 rpm for 10 min and each aliquot was checked on SDS-PAGE gels to confirm their homogeneity and purity before setting up screens. Hanging-drop vapor-diffusion method was employed with 2 µl protein sample added to 2 µl of reservoir solution on 22 mm circular siliconized glass cover slips. Initially several commercially available crystallization screens were used to obtain optimum conditions for crystallization. Subsequently several conditions were optimized in order to get good diffraction quality crystals; pH of solution, temperature, protein concentration, precipitant concentration, salt composition and concentration, additives and detergents screens, etc. Other techniques such as seeding, annealing were also tried for getting betters crystals in order to get good quality diffraction data.

1.11.2 X-ray diffraction and data collection

Preliminary data were collected on frozen *LlCAD2* crystals at in-house X-ray diffraction facility at IISER, Pune on a Rigaku microfocus rotating anode X-ray generator operating at 50 kV, 100 mA and mar345 image plate detector. Crystals were cryoprotected using selected cryosolution. Different cryoprotectants like glycerol, PEG 400, 600, 1000, and 2000, MPD, ethylene glycol, sucrose, isopropanol were tried, of which, 20% glycerol was found to be optimal for diffraction. The crystal was mounted on the goniometer head using a precooled cryoloop and frozen by dipping in liquid nitrogen. A liquid nitrogen jet produced by X-stream (Rigaku) was used to keep the crystal frozen at temperature of 100 K. After mounting the crystal, it was aligned to the center of the X-ray beam and the detector was placed at 200 mm distance. Data was collected for 180º in 0.5° oscillation frames, with exposure time of 300 s each. The determination of the orientation of the crystals, unit cell dimensions and tentative space group, and initial processing of the diffraction data were performed using automar data processing suite. The 3.5 Å resolution LlCAD2 data were collected at the Bl-12-2 beamline at the Stanford Synchrotron Radiation Lightsource (SSRL, USA) using an ADSC Quantum 315r CCD detector. All diffraction data were indexed, integrated, scaled and merged using HKL. Intensities were converted to structure factors using the CCP4 program suite.

1.11.3 Homology Modeling

AtCAD5 from *A. thaliana* (pdb-2CF6) was used to generate a homology model of LlCAD2 due to its highest similarity in BLAST search against PDB database. Initial models of LlCAD2 (300 models) were prepared using the coordinates of 2CF6 and further refined by Modeller 9v9 (SaliLab). Models were selected by assessing their DOPE (Discrete optimized protein energy) scores. The quality of final models was improved by CHARMM27 force field for lipids, proteins and nucleotides to remove bad contacts in the side chains. The model with least energy was further validated by ProCheck to evaluate the stereochemical quality by Ramachandran's plot and WhatCheck. Compatibility of model to its sequence was determined by Verify 3D, environmental profile of non-bonded interactions was analyzed using ERRAT and Z scores were calculated using Dali server. The model was further refined using the loop-modify script in Modeller in order to refine poor-modeled regions. The final energy minimized refined model was used for the active site identification and for docking analysis.

1.11.4 Substrate docking

Structure files for NADPH and different CAD substrates were downloaded from Pubchem database and converted to 3D molecules using LigPrep module, and were used for docking studies on

LlCAD2 model. Protein-ligand complexes were minimized within an RMSD of 0.30 Å with force field OPLS2005 using MacroModel package (Schrodinger). Protein-ligand docking simulations were conducted using AutoDock Vina tool to prepare the systems for calculations. For each ligand, around 100 docking runs with default parameters were performed treating protein as rigid and the ligand as flexible. The results were visualized using PyMol (Schrödinger), wherein all the conformations of each of the ligand were found to be within the cavity of protein indicating that the docking run was free from errors. The conformational clusters with lowest binding energy (Ea) for each ligand were considered for further studies.

1.11.5 Chemical modifications

Modification of amino acid residues was performed using different chemical modifiers [72, 73]. The enzyme solution (1.2 µM) was incubated with varying concentrations of modifiers at 25 ºC and aliquots were removed at suitable time intervals and residual activity was determined under standard assay conditions. K_m and V_{max} of partially modified and inactivated enzyme were determined. Suitable controls without modifier were employed. Pseudo-first-order rate constant for enzyme inactivation K_{app} were estimated from the slopes of:

$$\boldsymbol{logR \rightarrow t}$$

where, *R* is the residual activity after modification, *t* is time. The order of reaction *n* was estimated from:

$$\boldsymbol{logK_{app} \rightarrow log[M]}$$

where, [M] is the modifier concentration. The plot gives a straight line with its ordinate giving the order of the reaction *n* (the number of amino acids modified which are essential for catalysis). In all chemical modification reactions, the protective effect of substrate on modification was studied by incubating the enzyme with excess amount of the substrate (50 µM) for 10-20 min followed by treatment with various modifiers and residual activity was assayed.

1.11.6 Site-directed mutagenesis

Site directed mutagenesis was performed using the Quickchange Lightening Site Directed Mutagenesis Kit (Stratagene) according to manufacturer's instructions. The wild type *LlCAD2* in pET30b (+) was used as template for mutations and primers were designed from the online primer-design tool provided on Stratagene website. A list of designed primers used for these studies are shown in Table 1.2. Following cycling, the product was treated with *Dpn*I and was transformed into

E. coli XL-10 competent cells. Putative positive clones were picked, plasmids were isolated and sequenced. Positive plasmids were then transformed in *E. coli* BL21 (DE3) cells and mutant proteins were expressed, purified and characterized following protocols already mentioned.

Name	Sequence (5′-3′)
S49A-F	cactactgtggaatatgccatgccgatctgcatc
S49A-R	gatgcagatcggcatggcatattccacagtagtg
H52G-F	ggaatatgccattccgatctgggtcagattaagaatgatcttgg
H52G-R	ccaagatcattcttaatctgacccagatcggaatggcatattcc
S211I-F	ggacacaacgtgacagtgataatctcgtcggagaa
S211I-R	ttctccgacgagattatcactgtcacgttgtgtcc
D57A-F	ccgatctgcatcagattaagaatgctcttggcatgtccaatta
D57A-R	taattggacatgccaagagcattcttaatctgatgcagatcgg
S212D-F	cacaacgtgacagtgataagcgattcggagaagaagaagcaggag
S212D-R	ctcctgcttcttcttctccgaatcgcttatcactgtcacgttgtg
S213N-F	cgtgacagtgataagctcgaacgagaagaagaagcaggagg
S213N-R	cctcctgcttcttcttctcgttcgagcttatcactgtcacg
S211I-212D-F	acacaacgtgacagtgataatcgattcggagaagaag
S211I-212D-R	cttcttctccgaatcgattatcactgtcacgttgtgt
S212D-213N-F	aacgtgacagtgataagcgataacgagaagaagaagcaggaggc
S212D-213N-R	gcctcctgcttcttcttctcgttatcgcttatcactgtcacgtt
S211I-212D-213N-F	gacacaacgtgacagtgataatcgataacgagaagaag
S211I-212D-213N-R	cttcttctcgttatcgattatcactgtcacgttgtgtc
E70A-F	gttcctgggcatgcggtggtcggagag
E70A-R	ctctccgaccaccgcatgcccaggaac
L95R-F	gtcggagcaggacgcatcgttgggagc
L95R-R	gctcccaacgatgcgtcctgctccgac
T167A-F	tgcgccggcgtggcggtgtacagtc
T167A-R	gactgtacaccgccacgccggcgca
V192G-F	gggcttggaggaggggacacatgggc
V192G-R	gcccatgtgtccccctcctccaagccc
K216W-F	aagctcgtcggagaagtggaagcaggaggctctg
K216W-R	cagagcctcctgcttccacttctccgacgagctt
P253G-F	atatcatcgatacggtgggagtgggtcaccctcttg
P253G-R	caagagggtgacccactcccaccgtatcgatgatat
G275F-F	gatggcaagctgatcttaatgtttgttatcaacactcctctgca
G275F-R	tgcagaggagtgttgataacaaacattaagatcagcttgccatc
S298A-F	gaggaagacgataacgggagccttcattgggagcataaag
S298A-R	ctttatgctcccaatgaaggctcccgttatcgtcttcctc
I300G-F	gacgataacgggaagcttcggtgggagcataaaggagacg
I300G-R	cgtctcctttatgctcccaccgaagcttcccgttatcgtc

Table 1.2*. List of primers used in Site-directed Mutagenesis*

1.12 Structural transition studies

1.12.1 Fluorescence measurements

Intrinsic fluorescence of the enzyme was measured using a LS55 fluorescence spectrometer (Perkin Elmer) connected to a PTP-1 fluorescence peltier system. Protein solution (0.8 μM) was excited at 295 nm and emission was recorded at 310-400 nm at 25 °C. Excitation & emission monochromator slit widths were set at 5.0 nm each, and averaged spectra of 3 scans were recorded at 50 nm min^{-1}.

Exposure of hydrophobic regions of protein was monitored by 8-anilino-1-naphthalenesulfonic acid (ANS) dye binding assay. The dye was added to protein samples to a final concentration of 50 μM and emission spectra were recorded in the range of 410-550 nm with excitation at 375 nm. Protein aggregation was monitored using static light scattering. Both excitation and emission wavelengths were set at 400 nm, with 5 nm and 2.5 nm slit widths, respectively, and scattering was recorded for 60 s. The appropriate buffer solutions were also scanned in order to eliminate background emissions. Phase diagrams and parameter A were used for detection of protein intermediate states and conformational changes:

$\boldsymbol{I_{365} \rightarrow I_{320}}$ (Phase diagram) $\boldsymbol{I_{365}/I_{320} \rightarrow [M]}$ (Parameter A) where I_{320} and I_{365} are fluorescence intensities at 320 and 365 nm, respectively; [M] is denaturant.

1.12.2 Tryptophan fluorescence quenching

Fluorescence quenching experiments were carried out by adding small aliquots of quenchers (acrylamide, KI, CsCl) from 5 M stock solutions to 0.8 μM protein solution and intrinsic fluorescence were determined. KI stock solution contained 0.2 M sodium thiosulfate to prevent formation of tri-iodide. Fluorescence intensities were corrected for dilutions. The steady-state fluorescence quenching data obtained with different quenchers were analyzed by:

$$\frac{F_0}{F_c} = 1 + K_{sv}[Q] \quad (24)$$

(Stern-Volmer equation)

$$\frac{F_0}{\Delta F} = f_{a-1} + \frac{1}{(K_a\, f_a[Q])} \quad (25)$$

(Modified Stern-Volmer equation)

where F_o and F_c are the relative fluorescence intensities in the absence and presence of the quencher, respectively, *[Q]* is the quencher concentration, K_{sv} is Stern–Volmer quenching constant, ΔF is the change in fluorescence intensity at any point in the quenching titration, K_a is the quenching constant and f_a is the fraction of the total fluorophore accessible to the quencher.

1.12.3 Circular dichroism (CD) measurements

The CD spectra of the enzyme were recorded on a J-815 Spectropolarimeter with a PTC343 Peltier unit (Jasco) at 25 ºC. Each CD spectrum was accumulated from 5 scans at 50 nm min^{-1} with a band pass of 1 nm and a time constant of 1 s for a resolution of 0.5 nm. Far-UV CD spectra of the enzyme (1 μM) were collected in the range of wavelengths of 190-250 nm using a cell path length 1 mm for monitoring the secondary structure. All spectra were corrected for buffer contributions and observed values were converted to mean residue ellipticity (MRE):

$$MRE = \frac{M\,\theta_{\lambda}}{10\,d\,c\,r} \tag{26}$$

where M is the molecular weight of the protein, θ_{λ} is CD signal, d is the path length, c is the protein concentration and r is the average number of amino acid residues in the protein. Secondary structure elements were calculated by using CD-Pro software package. The apparent fractional extent of unfolding, f_{app}, was calculated from ellipticity (θ) values using the equation:

$$f_{app} = \frac{[\theta]_N - [\theta]}{[\theta]_N - [\theta]_U} \tag{27}$$

where, [Θ] represents the observed ellipticity under given conditions, $[\Theta]_N$ and $[\Theta]_U$ are ellipticity values in the N (native) and U (unfolded) states, respectively. The melting temperature (T_m) and melting concentration (C_m) of the protein was determined from:

$$\boldsymbol{MRE \rightarrow T} \qquad\qquad \boldsymbol{MRE \rightarrow [M]}$$

where, *MRE* is mean residual ellipticity and *T* is the temperature, [*M*] is the denaturant concentration. Near-UV CD spectra were monitored for 8.13 μM protein samples in the wavelength 250-320 nm using path length 10 mm.

1.12.4 Parameters used for monitoring structural transitions

Effect of pH (2-12), temperature (20-90 °C), Urea (0-8 M), GnHCl (0-6 M) and EDTA (0-20 mM) on LlCAD2 function and structure were determined by incubating 10 μg enzyme in any of the conditions and monitoring any changes by enzyme activity, fluorescence and CD spectroscopy. Buffers used for different pH are: glycine HCl- pH 2-3; acetate- pH 4-5; phosphate- pH 6-7; tris HCl- pH 8-9; glycine NaOH- pH 10-12. The changes were monitored by enzyme activity, fluorescence and CD spectroscopy. Effect of different metal ions (Mg^{2+},Co^{2+}, Cs^{+}, V^{2+}, Fe^{2+}, Li^{2+}, Cu^{2+}, Mn^{2+}, Ca^{2+}, Hg^{2+}, Zn^{2+}, Ni^{2+}), osmolytes (NaCl, KCl, glycerol, sucrose), reducing agents (DTT, β-ME), chelating agents (1, 10-phenanthroline, EDTA), surfactants (SDS, Tween 80, Triton X-100, CTAB), and their activity was assessed the same way.

For pH refolding experiments, the pH of each sample was adjusted back to pH 8 and incubated at 20 ºC for 2 h. For GnHCl refolding experiments, concentrated protein sample was first denatured in 6 M GnHCl at 20 °C for 16 h and subsequently diluted in buffer containing 0-5.5 M GnHCl and incubated at 25 °C for 2 h. Reactivation of the apoenzyme was determined by incubating the apoenzyme with 2 mM of different metal ions for 1 h and then assaying the enzyme activity.

2

CAD: Function

2A

CAD: Biochemical & functional characterization

The current chapter deals with isolating LlCAD2 from plant source, its expression, purification and some basic physicochemical properties like, catalytic form, metal co-factors, substrate specificities, etc. Also, effects of different chemical reagents and metal ions are determined.

2.1 Cloning, expression and purification of LlCAD2

LlCAD2 was cloned, expressed and purified according to the protocols highlighted in Fig.2.1. A 1.1 kb gene fragment was amplified from cDNA of *Leucaena leucocephala* using the primers LlCAD2-F & LlCAD2-R and cloned in pET-30b (+) expression vector (Fig. 2.2). The nucleotide and protein sequences of the full-length LlCAD2 are given in Fig. 2.3.

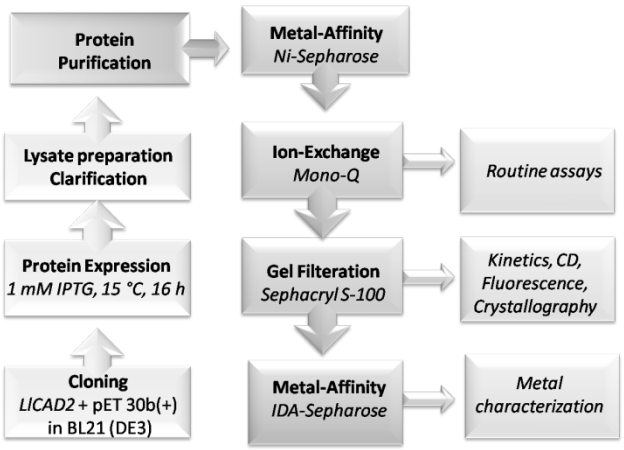


Fig. 2.1: *An overview of cloning, expression purification protocols*

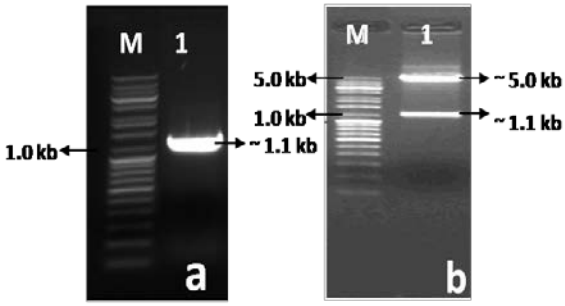


Fig. 2.2: (a) *Agarose gel showing 1.1. kb fragment amplified by primers LlCAD2-F/R from cDNA of L. leucocephala.* ***(b)*** *Restriction digestion pattern of recombinant pET-30b (+) harboring LlCAD2 ORF with enzyme NdeI, SalI*

```
   1 ATGGGAAGCA TTGAAGGAGA AAGAACAACG GTTGGTTGGG CTGCTCGAGA CCCTTCTGGG ACTCTCTCTC
  71 CCTACACTTT CAATCTCAGG AACACGGGAC CTGATGATGT GTATATCAAG GTTCACTACT GTGGAATATG
 141 CCATTCCGAT CTGCATCAGA TTAAGAATGA TCTTGGCATG TCCAATTATC CCATGGTTCC TGGGCATGAG
 211 GTGGTCGGAG AGGTGATCGA GGTGGGTTCG AATGTGACCA AGTTTAAGGC CGGAGAAGTA GTCGGAGCAG
 281 GACTCATCGT TGGGAGCTGC CGGAATTGCA GGGCATGCAA ATCTGATATT GAGCAATATT GCGCCAAGAA
 351 GATCTGGAAT TGCAACGATG TTTACACCGA TGGAAAACCC ACCCAAGGTG GCTTCGCTGA AACCATGGTC
 421 GTCAATCAAA ATTTCGTTGT GAAGATACCA GAGGGGATGT CTCCGGAACA AGTGGCTCCA CTGTTGTGCG
 491 CCGGCGTGAC GGTGTACAGT CCACTGTCAC ACTTTGGACT GAAAGAGAGT GGGCTAAGAG GAGGAATATT
 561 AGGGCTTGGA GGAGTGGGAC ACATGGGCGT GAAGATAGCC AAAGCCATGG GACACAACGT GACAGTGATA
 631 AGCTCGTCGG AGAAGAAGAA GCAGGAGGCT CTGGAACACC TCGGAGCAGA CGATTATGTG GTTAGCTCAG
 701 ACGAAACTCA GATGCAGAAG ATTGCTGATT CACTTGATTA TATCATCGAT ACGGTGCCAG TGGGTCACCC
 771 TCTTGAGCCT TATCTTTCTC TGCTCAAAGT TGATGGCAAG CTGATCTTAA TGGGTGTTAT CAACACTCCT
 841 CTGCAATTCG TCAGCCCCAT GGTCATGCTC GGGAGGAAGA CGATAACGGG AAGCTTCATT GGGAGCATAA
 911 AGGAGACGGA AGAGATGTTG GGGTTCTGGA AAGAGAAGGG GCTGAGTTCA GTGATAGAGG TTGTGAAGAT
 981 GGATTACATC AACAAAGCCT TCGAGAGGTT GGAGAAGAAC GATGTCAGAT ATAGGTTCGT TGTTGACGTT
1051 GCCGGCAGCA AACTTGATGA TCATCAGGTC GACAAGCTTG CGGCCGCACT CGAGCACCAC CACCACCACC
1121 ACTGA
```

```
  1 MGSIEGERTT VGWAARDPSG TLSPYTFNLR NTGPDDVYIK VHYCGICHSD LHQIKNDLGM SNYPMVPGHE
 71 VVGEVIEVGS NVTKFKAGEV VGAGLIVGSC RNCRACKSDI EQYCAKKIWN CNDVYTDGKP TQGGFAETMV
141 VNQNFVVKIP EGMSPEQVAP LLCAGVTVYS PLSHFGLKES GLRGGILGLG GVGHMGVKIA KAMGHNVTVI
211 SSSEKKKQEA LEHLGADDYV VSSDETQMQK IADSLDYIID TVPVGHPLEP YLSLLKVDGK LILMGVINTP
281 LQFVSPMVML GRKTITGSFI GSIKETEEML GFWKEKGLSS VIEVVKMDYI NKAFERLEKN DVRYRFVVDV
351 AGSKLDDHQV DKLAAALEHH HHHH
```

Fig. 2.3: *LlCAD2 cloned in pET 30b (+) with NdeI and SalI sites (underlined), bold nucleotides and amino acids represents sequence from the vector, His-Tag at C-terminal*

LlCAD2 was purified 65 fold by following the purification protocol based on Ni-Sepharose, Q-sepharose and Sephacryl S-100 chromatography, resulting in a homogenous protein which was confirmed by SDS-PAGE and Western Blot analysis (Table 2.1, Fig. 2.4-2.6). The protein was identified by protein-mass-fingerprinting LlCAD2 samples using ESI-MS/MS (1.9.1).

	Total Protein (mg)	Total Activity (µKat)	Specific Activity (µKat mg^{-1})	Fold Purification	Recovery (%)
Cell Lysate	3402	33.0	0.0097	-	100
Ni-Sepharose	58.7	26.4	0.45	46	80
Mono-Q	51.2	24.7	0.48	50	75
Sephacryl S100	37.3	23.5	0.63	65	71

Table 2.1: *Purification procedure for recombinant CAD expressed in E.coli BL21 (DE3) cells. The starting soluble crude lysate was obtained from 1 L of culture broth. Sephacryl S-100 reflects the data for dimeric LlCAD2 (peak 2 of Fig. 4c)*

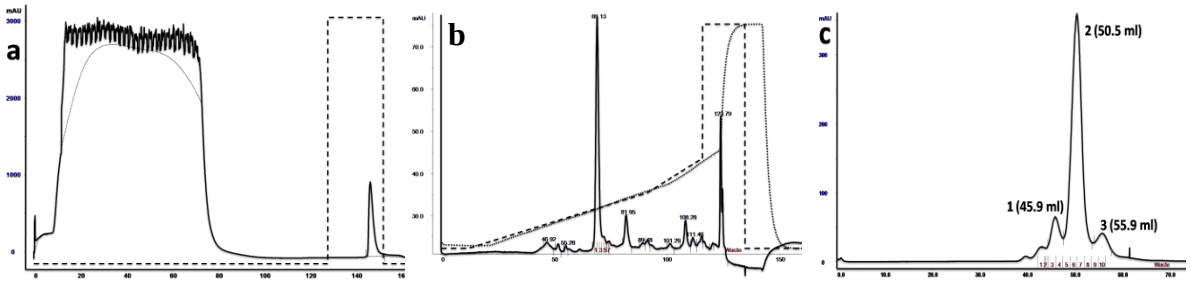


Fig. 2.4: *Purification chromatograms from Akta FPLC.* ***(a)*** *Ni-Sepharose;* ***(b)*** *Mono-Q;* ***(c)*** *Sephacryl S-100*

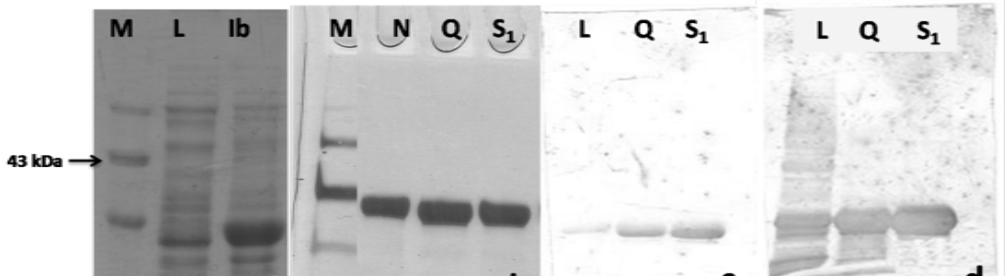


Fig. 2.5: ***(a)*** *10 % SDS PAGE showing LlCAD2 expression;* ***(b)*** *Purification;* ***(c)*** *&* ***(d)*** *Western blots for anti-His tag, and antiCAD antibodies, respectively. Low molecular weight marker (M), Lysate (L), Inclusion bodies (Ib), Ni-Sepharose (N), Mono-Q (Q), Sephacryl S-100 peak 2 (S_1)*

LlCAD2 existed in two forms as seen on gel filtration chromatogram (Fig. 2.4c); a 40 kDa and 80 kDa peaks with hydrodynamic radii of 7 and 13 nm, respectively (Table 2.2, Fig. 2.4, 2.7). Both forms were individually collected and subjected to activity staining and SDS-PAGE. As seen in Fig. 2.6a, activity stained bands were seen only in lanes for the 80 kDa protein and no bands in the 40 kDa sample, while SDS-PAGE showed 40 kDa bands in both the samples. Thus, it can be held that active LlCAD2 is a homo-dimer formed by association of two identical monomers of 40 kDa. In order to understand the association of the subunits, duplicate samples of the dimer were subjected to SDS-PAGE, one boiled in presence of DTT and SDS, while the other incubated at room temperature in presence of SDS only. All the samples showed a single band at 40 kDa (Fig. 2.6b), suggesting that the dimer is formed due to electrostatic or hydrophobic interactions and not by any disulfide linkages. This was further confirmed by modifying the enzyme with DTNB. While the monomer consisted of six cysteine residues under non-denaturing conditions and total eight after reduction of the denatured protein, the dimer showed twelve free and sixteen total cysteine residues. Two disulfide linkages could thus be deduced per dimer from the data; one intra subunit linkage per subunit, and no inter chain linkages. Hence, LlCAD2 can be viewed as a homo-dimer of 40 kDa units associated with non-covalent interactions. This observation was in accordance with several reports of CADs from plants [24, 74] which also showed CADs to be homo-dimers of ~40 kDa subunits. LlCAD2 showed maximum activity between 30-40 ºC, at pH 6.5 for reduction of aldehydes and at pH 9 for oxidation of alcohols, with the rate of reduction almost twice that of oxidation (Fig. 2.8).

	Mol. Wt. (kDa)	**H.Radii** (nm)	**Disulfide linkages**	**Total Cys**	**Free Cys**
Dimer	81	13	2	16	12
Monomer	40	7	1	8	6

Table 2.2: *Molecular weights of monomer and dimer were determined by gel filtration chromatography. Hydrodynamic radii were determined by Dynamic light scattering. The numbers of disulfide linkages were estimated by modifying the enzyme with DTNB*

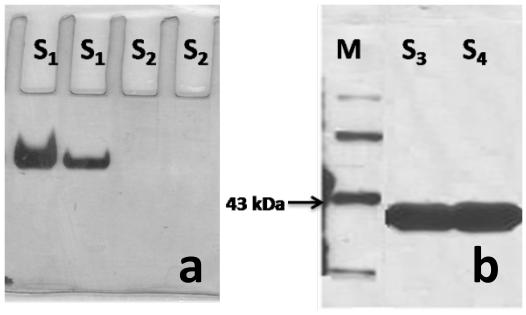


Fig. 2.6: (a) *Activity staining of S_1-dimer & S_2-monomer on native PAGE. The dark blue color shows actively stained protein bands.* ***(b)*** *SDS PAGE of purified dimer, S_3- boiled sample in presence of DTT & SDS, S_4- sample incubated at room temperature. M: Low molecular weight protein marker*

2.3 LlCAD2 is a Zn metalloenzyme

CADs are known to be metalloenzymes, with 2 Zn^{2+} metal ions bound to the protein per subunit, as seen in the crystal structure of AtCAD5 [49]. Thus, the presence of metal was investigated and presence of any metal bound to the protein was confirmed by ICP-OES as described in methods. LlCAD2 contained 4.1 ± 0.06 moles Zn^{2+} per mole of enzyme, with trace amount of Na^{+} (0.3), Cu^{2+} (0.02), Ni^{2+} (0.02), Ca^{2+} (0.03), K^{+} (0.01), Fe^{2+} (0.008) and Mn^{2+} (0.003). The number of cysteine residues in LlCAD2 was estimated by incubating the enzyme with DTNB and measuring A_{419}. While there were no free cysteine residues in the native protein, the metal-chelated apoenzyme showed twelve residues (six per subunit), and showed 16 residues (eight per subunit) after reduction and denaturation (Table 2.2).

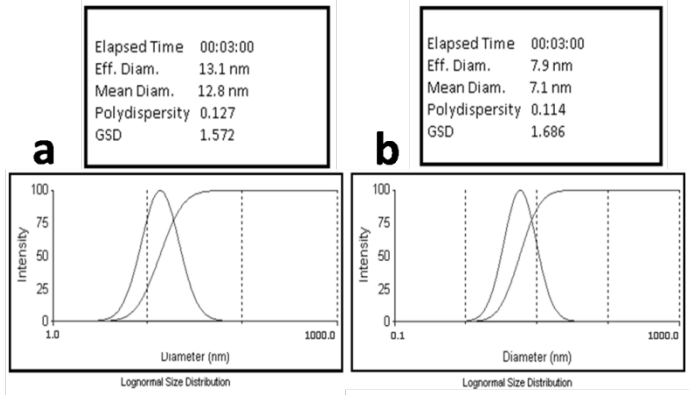


Fig. 2.7: *Lognormal size distribution and summary of DLS analysis of* ***(a)*** *LlCAD2 Dimer,* ***(b)*** *Monomer*

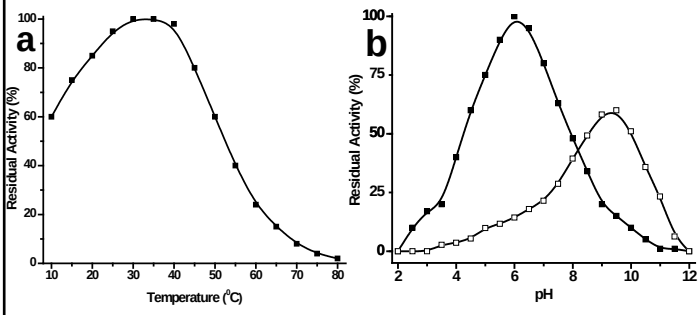


Fig. 2.8: (a) *Temperature optima and* ***(b)*** *pH optima.(■) Reduction reaction, (▲) Oxidation reaction*

The apoenzyme was defined as CAD treated with 10 mM EDTA for 2 h, which showed significant loss of activity as well as the bound metal ions (confirmed by ICP-OES). The treated protein had less than 10% residual activity, and 0.4 ± 0.02 moles of bound metal ions per mole of protein. Reconstitution of loss of activity and structure in the apoenzyme were achieved by incubating the enzyme in 1 mM Zn^{2+}, 10 mM DTT under constant N_2 bubbling for 5 min. The vials were sealed and after an incubation of 30 min at 20°C, enzyme activity was measured. Here, Zn^{2+} was able to reactivate the enzyme by only 15%, thus suggesting that six free cysteine residues may be involved in binding with the two metal ions in the native conformation, and DTT helps in stabilizing these metal-cysteine interactions by protecting the thiols from oxidation.

The number of cysteine residues in LlCAD2 was estimated by incubating the enzyme with DTNB and measuring absorbance at 419 nm. While there were no free cysteine residues in the native protein, the metal-chelated apoenzyme showed twelve residues (six per subunit), and completely denatured and reduced protein (in presence of 8 M urea, 10 mM EDTA, 1 mM DTT) showed 16 residues (eight per subunit). These results along with the observation that Zn^{2+} is able to reactivate

the enzyme only in presence of DTT, suggest that six free cysteine residues may be involved in binding with the two metal ions in the native conformation, and DTT helps in stabilizing these metal-cysteine interactions by protecting the thiols from oxidation.

2.4 LlCAD2 favors hydroxycinnamyl aldehydes

CADs catalyze a reversible reaction; reduction of hydroxycinnamyl aldehydes and oxidation of corresponding alcohols [24]. To begin with, the function of LlCAD2 was tested first by assessing its specificity for different aliphatic and aromatic substrates (benzyl and cinnamyl derived) as summarized in Fig. 2.9. LlCAD2 showed significant activity against cinnamyl substrates, however, was almost inactive against aliphatic and other benzyl compounds. The enzyme was highly co-factor specific and was active only in presence of NADP(H), with insignificant activity with NAD(H) (upto 3 mM). Upon examination of structures of different substrates, and the fact that the enzyme can only catalyze cinnamyl substrates, it was evident that the presence of an aromatic phenyl ring along with a propionic group was essential for favorable interactions with the enzyme and subsequent catalysis. The reason may lie in the size and orientation of the side groups on these substrates. Even though the benzyl derived substrates had the phenyl ring and some of them also contained the methoxy groups (resembling the cinnamyl substrates), they may not be able to orient their oxygen of the proton acceptor (-CHO) group at the active site. This may be due to improper hydrogen bonding with the

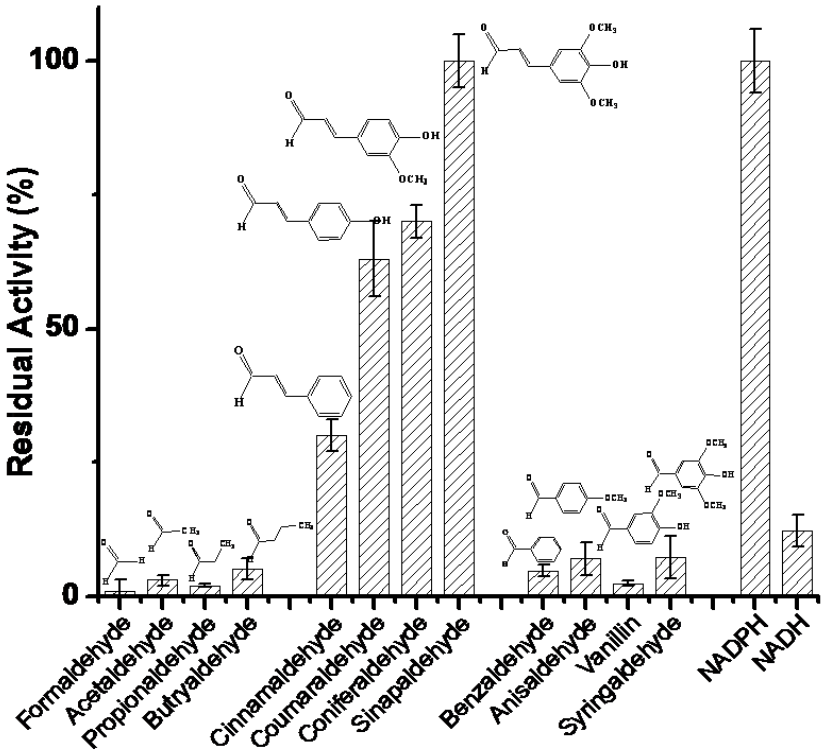


Fig. 2.9: *Enzyme activity of LlCAD2 against different substrates. The residual activity presented here corresponds to a substrate concentration of 100 µM and 0.1 µg of the enzyme for cinnamyl substrates & NADPH, and 2 mM substrate and 20 µg enzyme for all others. The structures shown on each column represents the corresponding substrate*

active site residues due to lack of the propionic group, and thus showed almost negligible activity. While aliphatic substrates had the required aliphatic side chain, however, were unable to catalyze the reaction due to lack of the phenyl ring. The above results in accordance with reports from other plant sources support the hypothesis that though CAD is a broad specificity enzyme, its catalytic capability is confined only to the phenylpropanoid (hydroxycinnamyl) group of metabolites.

2.5 pH dependence of kinetic parameters suggest participation of His in catalysis

The pH activity profile (Fig. 2.10) obtained by plotting log V_{max}/K_m *vs* pH of purified LlCAD2 revealed the participation of two ionizable groups on the enzyme [E] with pK_{E1} of 5.61 and pK_{E2} of 7.49. The pK_E of the amino acid as seen on the acidic limb shows involvement of carboxylate and the one towards basic limb possibly shows the involvement of His (the *pK* values may have slightly changed due to the protein microenvironment). Whereas, plot of log V_{max} *vs* pH gave the ionizable groups pK_{ES1} and pK_{ES2} on the [ES] complex to be 5.78 and 7.91, respectively. Upon comparison of both *pK* values, it seems likely that pK_{E1} and pK_{ES1} bear a resemblance to each other, and on the other hand, pK_{E2} and pK_{ES2} showed a change of about 0.5 pH. A possible explanation of such a shift in the latter may come from any changes induced in the enzyme during binding of the substrate, in turn, altering the ionization of the amino acid involved; histidine in this case. This observation may reflect histidine's role in either substrate binding or catalysis and was confirmed by modifying the enzyme with DEPC (Chapter 3B). As shown in AtCAD5 crystal structure [49], a histidine at position 52 is seen wit hin hydrogen bonding distance with nicotinamide ribose of NADPH, and its putative role in proton relay during catalysis is reported; which may be the reason of the shift in *pK* observed here. However, a possible ionization of the substrate itself during binding with the enzyme may also result in such variations, and thus observed *pK* values must be carefully interpreted.

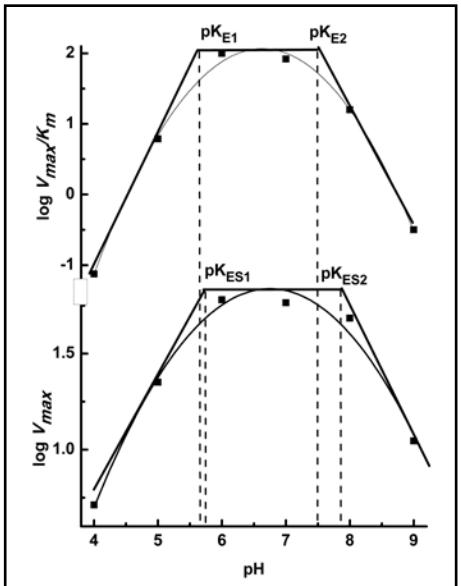


***Fig. 2.10**: pH activity profiles for LlCAD. The slopes of the tangents are 1.9-2.0*

2.6 LlCAD2 cannot be reactivated by other metals

Effect of chelating agents, EDTA and 1, 10-P on LlCAD2 activity was investigated in order to understand the role of Zn^{2+} in catalysis. The enzyme lost 40% activity upon incubating with 1 mM

EDTA for 2 h and 90% loss was observed with 10 mM. Similar result was observed in 1, 10-P, with 50% loss at 1 mM after 2 h. ICP-OES analysis of CAD treated for 2 h with 10 mM EDTA showed loss of the bound Zn^{2+}, with only 0.4 ± 0.02 moles of metal per mol of enzyme. This chelated protein was regarded as the apoenzyme form of CAD (section 2.3).

Subsequently, the effect of different metal ions on activity of LlCAD2 was determined. Mg^{2+}, Co^{2+}, Mn^{2+} & Ca^{2+} showed marginal effects on activity, while almost 50-60% loss of activity was observed with Cs^{+} & Cu^{2+}, respectively (Fig. 2.11). However, Fe^{2+} and V^{2+} showed 80% loss in enzyme activity, while Zn^{2+}, Ni^{2+}, Hg^{2+} and Li^{2+} completely inhibited the enzyme. LlCAD2 apoenzyme was incubated in presence of 2 mM different metals for 1 h to reconstitute the activity, however, none of the above metals were able to regenerate the lost activity, including Zn^{2+}. It was quite surprising that Zn^{2+} did not regenerate any of the activity in the apoenzyme, however, chelating the enzyme in presence of 1mM DTT followed by incubation with Zn^{2+} recovered 15% residual activity. These results demonstrate the geometry of the metal binding site to be very delicate; formation of non-specific disulfide linkages after chelating the metal ions could be a governing factor in disruption of the site, which can be avoided to some extent by addition of DTT.

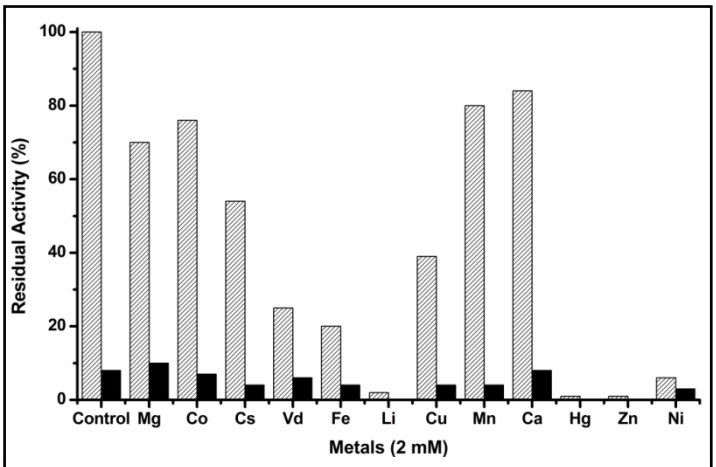


Fig. 2.11: *Effect of metal ions LlCAD2. (▒) represents residual activity of native LlCAD2 & (■) for apo-LlCAD2 after treating with different metals. Positive control corresponds to 100% for native, and 8% for apo-LlCAD2*

2.7 Effects of different chemical reagents on LlCAD2

Osmolytes such as glycerol, sucrose, etc. help in maintaining an isotonic medium with the environment and protect enzymes and other sub-cellular structures against temperature fluctuations. Similarly, salts like NaCl and KCl are routinely used to maintain ionic strengths in protein solutions. Upon studying the effects of different osmolytes, addition of sucrose and glycerol did not affect the activity, however, slight increase in activity in presence of 100 mM NaCl and KCl (15 and 7%, respectively), mainly due to their positive effects towards maintaining the structural integrity of proteins and preventing any non-specific interactions between them. Such stabilizing effects of different sugars, sucrose in particular, and osmolytes on kinetics and thermodynamics of yeast alcohol dehydrogenase have already already been reported [75].

To understand the role of either free cysteine or any disulfide linkages on activity of LlCAD2, effects of DTT and β-ME were studied. Almost 25% increase in residual activity in presence of 1 mM DTT and 8% increase in 10 mM β-ME was observed (Fig. 2.12). Such improved activity of CAD by thiol reagents was also observed in CADs purified from *Populus* where 30% increase in activity was observed with 1 mM DTT, while sulfhydryl reacting compounds showed inhibitory effects [74].

Incubating LlCAD2 with different surfactants resulted in loss of activity in presence of anionic detergent SDS and cationic CTAB, while no such loss was observed in presence of nonionic detergents Triton X-100 and Tween 80. As shown in Fig. 2.6, samples treated with only SDS (without heating) also resulted in dissociation of the dimer into its monomeric 40 kDa subunits. Thus, the inactivation of LlCAD2 in presence of charged detergents like SDS and CTAB resulted in disruption of the electrostatic/hydrophobic interactions at the dimer interface, resulting in dissociation and subsequently inactivation of the enzyme, while non-ionic detergents had no such effect.

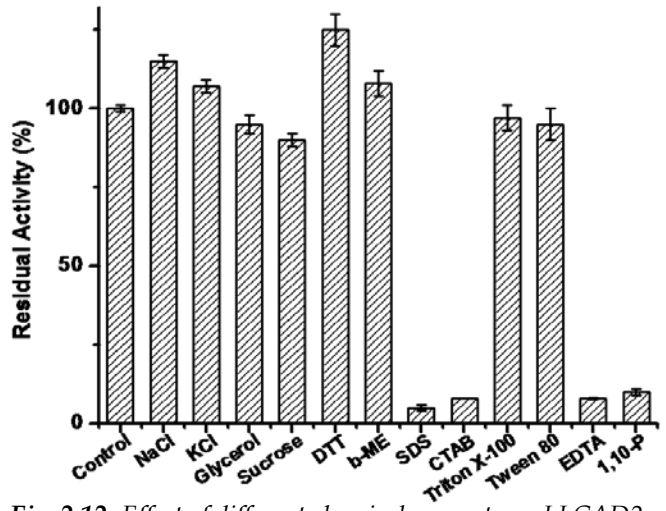


***Fig. 2.12**: Effect of different chemical reagents on LLCAD2 activity. Osmolytes (Nacl 100 mM; KCl 100 mM; Glycerol 20%; Sucrose 20%), Reducing agents (DTT 1 mM; β-ME 10 mM), Surfactants (SDS 1%; Tween 80 1%; CTAB 1%; Triton-X 100 1%), Chelating agents (EDTA 10 mM; 1,10-phenanthroline 10 mM)*

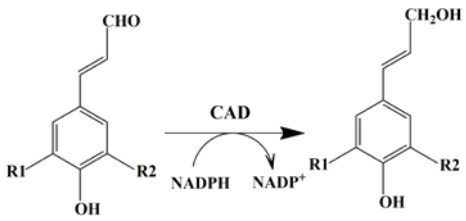
2B

Kinetic & mechanistic insights into catalysis

This section deals with the catalytic properties of this protein. Substrate kinetics takes a major part of this section which includes the kinetic parameters for different CAD substrates, as well as substrate inhibition, equilibrium binding and kinetics. Next, the mechanism of LlCAD2 is also provided based on product inhibition studies. Finally, the inhibitory effects of Zn^{2+} was explored.

2.8 Substrate Kinetics

CADs catalyze a reversible reduction of hydroxycinnamyl aldehydes with concomitant oxidation of NADPH [24].



This can also be written as: $A + B \rightarrow P + Q$

where, A is NADPH, B is hydroxycinnamyl aldehyde, P is hydroxycinnamyl alcohol and Q is $NADP^+$. This reaction can easily be assumed to be pseudo-first order by keeping either of the substrate at a very high saturating concentrations and varying the concentrations of other in order to estimate the kinetic parameters. However, considering the effect of substrate inhibition at high concentrations, and reaction times well below the dead-time of the instruments used, substrate concentrations above 100 µM were not practical, so, much lower concentrations of all the reactants were employed throughout this study. Due to this concentration range, it was not possible to assume that the concentration of the fixed substrate will remain constant throughout the course of an experiment. Thus, instead of using the Michaelis-Menten equation, the kinetic parameters were rather determined following equations where concentrations of both A and B were taken into consideration, essentially, the sequential-order ternary-complex mechanism.

The kinetic parameters for forward reaction V_{max} (maximum velocity), K_A & K_B (Michaelis constants for A & B), K_{iA} (dissociation constant for A), were determined following Equation (1). For reverse reaction, V'_{max}, K_P, K_Q, K_{iQ} were estimated the same way. Initially, several sets of experiments were carried out by varying concentration of one substrate at different fixed

concentrations of the other substrate. The apparent kinetic parameters K_m^{app}, V_{max}^{app} were then calculated by non-linear regression analysis of the Michaelis-Menten plots, while K_{iA} (or K_{iQ}) was determined from Hanes plot. The real kinetic parameters K_m and V_{max} were then estimated from secondary plots.

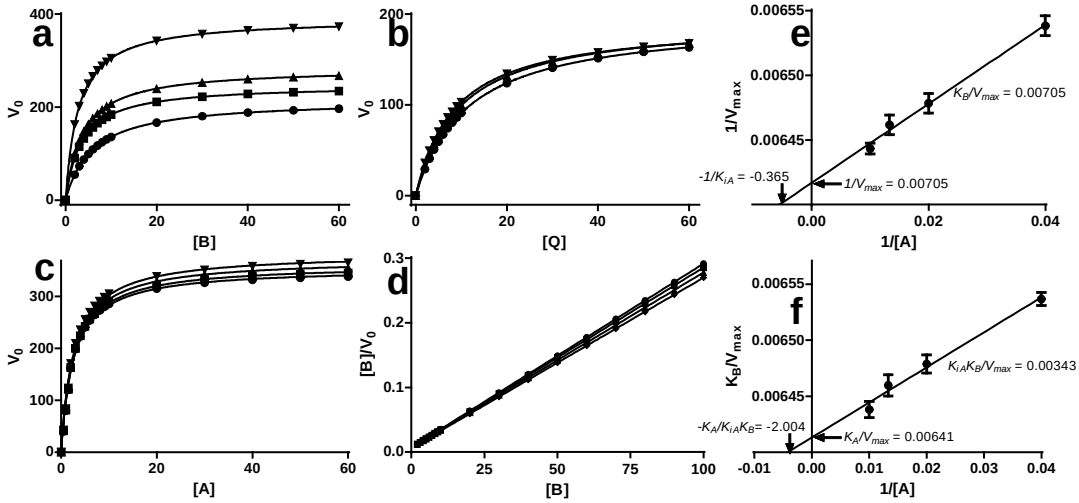


***Fig. 2.13**: **(a, b)** Michaelis-Menten plots for different hydroxycinnamyl **(a)** aldehydes & **(b)** alcohols: Cinnamyl (●), p-Coumaryl (■), Coniferyl (▲), Sinapyl (♦). **(c-f)** Kinetic plots for sinapaldehyde (ignoring substrate inhibition). Sinapaldehyde was varied from 0-60 µM at fixed concentrations of NADPH 25 µM (●), 50 µM (■), 75 µM (▲), 100 µM (♦). **(c)** Dependence of initial rate on sinapaldehyde concentration obeying Michaelis-Menten equation. The apparent kinetic parameters were calculated by non-liner regression analysis in GraphPad Prism software; **(d)** Hane's plot; **(e) & (f)** Secondary plots for sinapaldehyde*

As a case-study, only the kinetic plots for sinapaldehyde as a variable substrate are shown here. Though similar plots for all other substrates were prepared, only comparative Michaelis-Menten plots for different hydroxycinnamyl substrates are provided (Fig. 2.13a, b). Much variation is seen in the plots for the aldehydes, while the alcohols showed almost the same pattern. The primary plots for Sinapaldehyde (Fig. 2.13c, d) show that the initial velocity v_o for the substrate increases with increase in concentration of the fixed substrate NAPDH, though the difference between the data points is not appreciable. Such a small difference in the v_o values as a function of [NADPH] is due to the limitation in deciding the concentration range as mentioned already. A substantial difference could have been observed if much lower [NADPH] were employed, however, the reaction times decreased much below the dead-time of the experimental setup and thus such concentrations were not considered. The secondary plots were used to estimate the real kinetic parameters for different substrates (Fig. 2.13e, f). The dissociation constant K_{iA} for NAPDH was estimated from three individual experiments using sinapaldehyde, coniferaldehyde and NADPH as the variable substrate. It was defined as the negative value of the x-coordinate of the intersection

point of the lines observed in the Hanes plot when NADPH was the variable, while the same coordinate was defined as $-K_{iA}K_B/K_A$ when either sinapaldehyde or coniferaldehyde was varied. The values were in agreement to each other and so, a fixed value of 1.9 µM was considered in calculations for all other substrates. Similarly a fixed value of 4.1 µM was estimated for K_{iQ}. The final kinetic parameters calculated as mentioned here are summarized in Table 2.3. The activation energies, E_a, were calculated from Arrhenius plots (Fig. 2.14), while free energy changes ΔG were calculated as mentioned in 2.3.3.

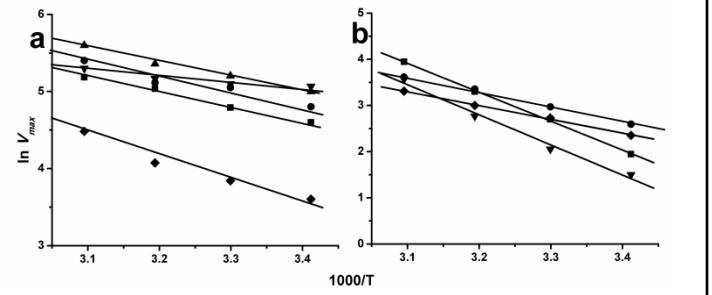


***Fig. 2.14**: Arrnehius plots for **(a)** Reduction; **(b)** Oxidation reactions. (■) Sinapyl, (▲) Coumaryl, (●) Coniferyl, (▼) Cinnamyl, (♦) NADPH/NADP+*

The K_m and K_{cat} values for all the substrates fell in a narrow range suggesting only small differences in their binding efficiencies and catalysis (Table 2.3). The lower E_a, negative ΔG and higher K_{cat}/K_m values for reduction reaction as compared to oxidation of alcohols supports that LlCAD2 prefers formation of alcohols; a reaction physiologically favorable for formation of lignin in plants. Upon comparison, sinapaldehyde was catalyzed at much faster rate than other substrates, with K_m of 2.8 µM and catalytic efficiency of 11.6 x 10^6 M^{-1} s^{-1}, followed by coniferaldehyde and *p*-coumaraldehyde. The additional methoxy group of sinapaldehyde could be playing a role in lowering the energy constraints during the binding and consequent catalysis, thus resulting in the higher catalytic efficiency than other substrates.

Sub	K_B	K_P	K_{si}	K_{Bcat}	K_{Pcat}	K_{Bcat}/K_B	K_{Pcat}/K_P
Cinnamyl	5.9 ± 0.12	11.3 ± 0.7	5.5	18.0 ± 0.18	16.2 ± 0.38	3.1	1.4
Coumyl	3.5 ± 0.26	-	5.7	22.5 ± 0.15	-	6.4	-
Conifyl	3.7 ± 0.19	9.6 ± 0.29	6.8	23.7 ± 0.28	16.3 ± 0.1	6.4	1.7
Sinapyl	2.8 ± 0.2	8.6 ± 0.1	4.1	32.2 ± 0.17	16.0 ± 0.26	11.6	1.9

Sub	K_{iA}	K_A	K_{iQ}	K_Q	K_{Acat}	K_{Qcat}	K_{Acat}/K_A	K_{Qcat}/K_Q
NADPH	1.9± 0.1	2.5 ± 0.09	4.1±0.08	5.7 ± 0.51	28.8 ± 0.11	23.3 ± 0.24	11.5	4.1

Sub	E_{aB}	E_{aP}	ΔG_B	ΔG_P
Cinnamyl	23.82	44.21	-10.6	-8.8
Coumyl	15.13	-	-26.5	-
Conifyl	18.04	32.05	-22.3	-17.8
Sinapyl	14.63	24.69	-27.2	-23.5
NAD(P)H	22.45	24.78	-20.3	-19.7

***Table 2.3:** Steady-state kinetic & thermodynamic parameters of LlCAD2 for different substrates. Cinnamyl, Coumaryl, Coniferyl and Sinapyl are used as abbreviations when both their aldehyde and alcohol forms need to be mentioned*

In a recent report of a different homologue: *LlCAD* isolated from *L. leucocephala* with different substrate specificities [76], showed highest affinity for cinnamyl alcohol, followed by coniferyl and least for sinapyl alcohol, which is in sharp contrast to our findings. It is also worth noting that while other CADs such as AtCAD4 & 5, PtCAD, TaCAD1, etc. showed higher catalytic rates for either coumaraldehyde or coniferaldehyde [48, 54, 77]; LlCAD2 rather preferred sinapaldehyde over the former ones. Although a sinapaldehyde specific alcohol dehydrogenase (PtSAD) was reported [54], its precise physiological role in lignin biosynthesis is not fully understood, and is more likely to play defense related roles. In another report, two CAD forms from *Eucalyptus* (CAD2Pbi & CAD2Pbii) were isolated which showed preference for coniferaldehyde and sinapaldehyde, respectively [51]. However, they demonstrated that CAD2Pbi is, in fact, a hetero-dimer of two subunits of 42 kDa and 44 kDa, and CAD2Pbii is a homo-dimer made up of two identical subunits of 44 kDa. All these reports combined with our findings, reinforce the model of multiple CAD homologues in plants with diverse substrate preferences, and giving us an idea that different combinations of these monomeric units may govern the overall specificity.

2.9 Mixed Kinetics

As mentioned already, sinapaldehyde is catalyzed at a faster rate than coniferaldehyde, which contradicts to previous reports [48]. However, catalytic rates *in vivo* does not necessarily reflect their true rates *in vivo,* where the presence of other competing substrates, inhibiting products and other interacting molecules governs the overall rate of catalysis. In order to understand the catalytic specificity of LlCAD2 for coniferaldehyde and sinapaldehyde, mixed kinetic approach was employed. A competition plot was prepared in order to understand whether the enzyme differentiates between the two substrates at each of its sites. If the two substrates equally react at the same site, the competition plot gives a horizontal straight line, i.e. the v_{tot} is independent of r (refer 2.3.9 for details). Independent reactions at two separate sites give a curve with a maximum; separate reactions with cross-inhibition generate curves with minima. When tested with LlCAD2, the plot was a straight line with no dependence of v_{tot} on r and thus indicated that both substrates competed

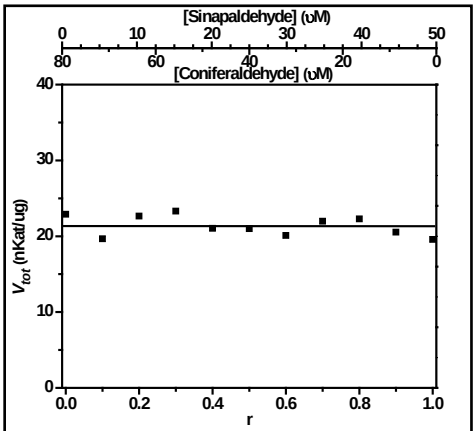


Fig. 2.15*: Competition plot for reduction of coniferaldehyde and sinapaldehyde by LlCAD2. The mixture of both substrates is constructed such that the measured rate V_{tot} is the same for the pure substrates. If the two substrates compete for the same site, the rate is independent of the proportions of substrates, as is the case here*

equally for the same site (Fig. 2.15). These results show that LlCAD2 does not differentiate between its catalytic sites and all the substrates can equally bind to either of them with equal efficiency. Also, mixed kinetic reactions were subject to HPLC analysis in order to estimate rates of individual reactions. However, the results were unreliable and so the data is not shown here.

2.10 Substrate Inhibition

In a bi-substrate reaction, there is a possibility of at least one of the four reactants binding to the wrong form of the enzyme and thus lowering its catalytic rates. This phenomenon is also known as substrate inhibition which takes place at high concentrations of any of the substrates. In order to characterize such inhibition, initial velocity was measured as function of high [S] concentrations. It was observed that aldehydes (and corresponding alcohols) showed inhibition, while NADPH (and $NADP^+$) did not. As an example, inhibitions by sinapyl aldehyde and alcohol are shown (Fig. 2.16). These patterns show that the rate of reaction is inhibited at higher concentrations of both these substrates. The inhibition constant K_{si} was calculated by fitting the data to the substrate inhibition equation (13). The aldehydes showed a proper fit to the model (Fig. 2.16a) & their K_{si} values are shown in Table 2.3. While substrate inhibition by alcohols for reverse reaction did not fit the model, and thus their inhibition constants could not be determined without error (Fig. 2.16d).

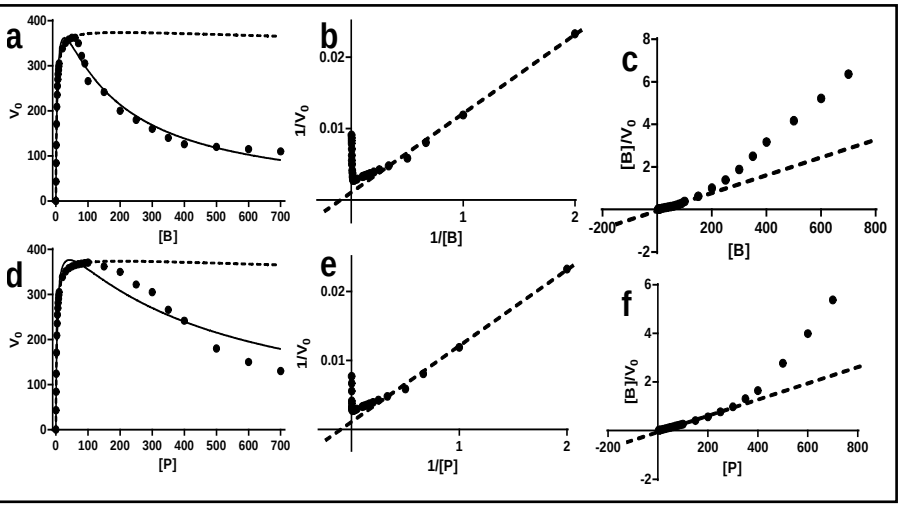


Fig. 2.16**: Substrate inhibition observed at high concentrations of Sinapaldehyde **(a, b, c)** and Sinapyl alcohol **(d, e, f).** (••) represent the data, (——) calculated from the equation, (— —) calculated from MM equation in absence of any substrate inhibition. Michaelis-Menten plot **(a, d)**, LB plot **(b, e)**, Hanes plot **(c, f)

The mechanism for substrate inhibition for LlCAD2 is depicted in Fig. 2.17, where both EBQ and EAP are formed by improper binding of the substrates B and P to the EB and EA complexes, respectively. Note, that during initial velocity measurements when the product has not yet accumulated at high enough concentrations, only one of these complexes will form (depending upon the direction of reaction). Though it is easy to interpret that a complete substrate inhibition occurs, i.e. the initial velocity will tend to zero as substrate concentration tends to infinity which is

due to formation of dead-end complex EBQ (with B as the inhibitory substrate). However, another mechanism resulting in a partial substrate inhibition cannot be neglected wherein higher concentrations of the inhibiting substrate alter the reaction pathway. Though both mechanisms cannot be differentiated based on the data provided here, the possibility of the latter mechanism taking place for LlCAD2 can be corroborated from a previous report of liver alcohol dehydrogenase. This report demonstrates that alcohol substrate showed a partial inhibition mechanism in the direction of alcohol oxidation [78].

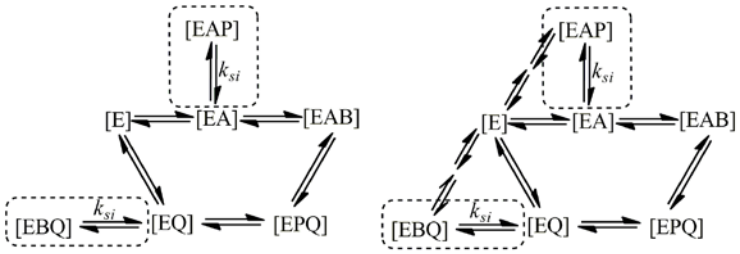


Fig. 2.17*: Mechanism that produces substrate inhibition. The first panel shows a complete inhibition with formation of dead-end complexes EBQ and EAP. The second panel shows a partial inhibition where both EBQ and EAP are still able to release the bound substrate and product to give free enzyme [E]. Note that only B (P in case of a reverse reaction) can give such inhibition in a sequential bi-substrate reaction with formation of EBQ (EAP in case of reverse reaction). Boxes represent complexes formed due to substrate inhibition*

Another point worth noting here is the absence of inhibition by A and Q (NADPH & $NADP^+$) which is due to absence of the complexes [EB] and [EP] during the catalytic cycle. This result indicates that the LlCAD2 catalytic mechanism involves a compulsory order of substrate binding and product release. The substrate inhibition phenomena shown here is common in many bi-substrate reactions, including other CADs from *E. gunnii* [79], *L. leucocephala* [76] and *H. pylori* [58]. However, it is quite surprising that AtCAD5 from *A. thaliana* showed no such inhibition patterns [48].

2.11 Product Inhibition

Since CAD catalyses a bi-substrate reaction, the order of binding of both substrates as well as product release was next investigated. There are several mechanisms by which CAD can bind these substrates: Random order, Compulsory order, Ping-Pong, etc. Product inhibition studies are among the most useful methods for elucidating the mechanism of reaction by measuring the dependence of kinetic parameters in presence of one of the products of the reaction and varying the concentration of one substrate at a fixed concentration of another. Here, the product is considered as a reversible inhibitor of the enzyme, and thus the data can be interpreted as a normal inhibition experiment.

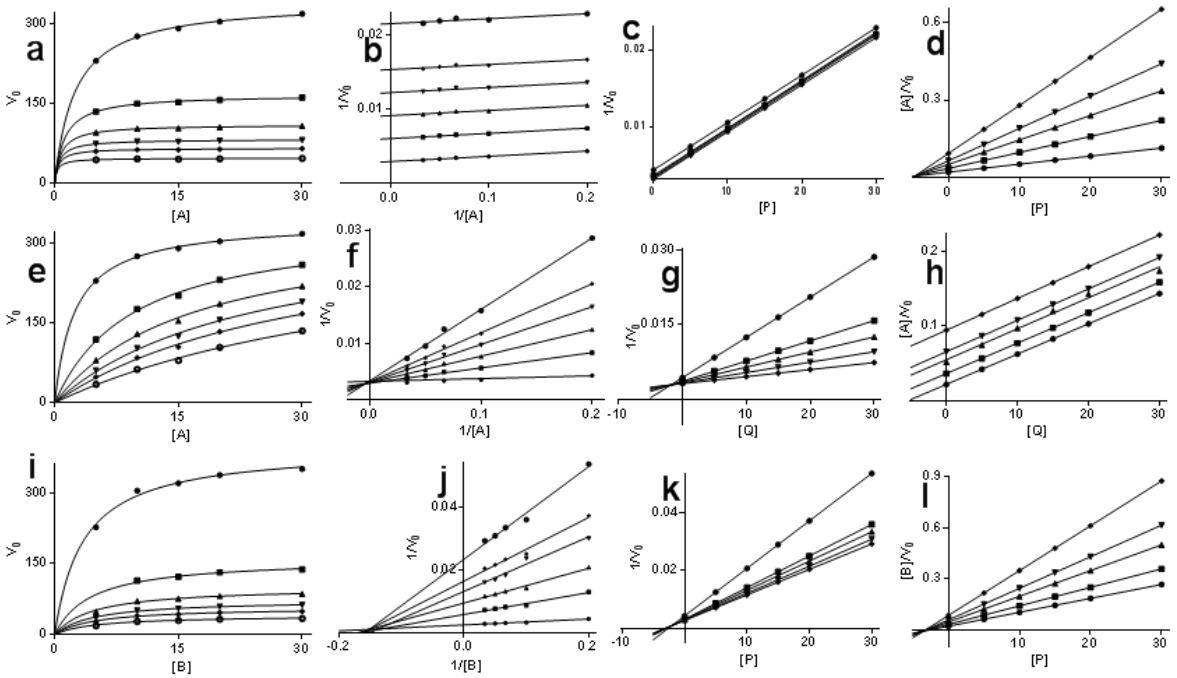


***Fig. 2.18**: Plots for product inhibition experiments. NADPH is the variable substrate with NADP as the inhibitor for plots in **a-d**; sinapyl alcohol as the inhibitor in **e-h;** sinapaldehyde is the variable substrate with sinapyl alcohol as the inhibitor in **i-l.** The data of Michaelis-Menten plots **(a, e, i)** were fit into equations for different inhibition mechanisms in order to estimate the inhibition constants. While, Lineweaver-Burk plots **(b, f, j)**, Dixon plots **(c, g, k)** and Cornish-Bowden plots **(d, h, l)** were used for visual analysis of the data only. The variable $[S]_v$ and inhibitor product $[I]_p$ were varied at the following concentrations: 0, 5, 10, 15, 20, 30 μM at two fixed concentrations of second substrate $[S]_f$ (5, 50 μM)*

		Compulsory ordered						**Random**						**Sequential**					
Variable $[S]_v$	**Inhibitor $[I]_p$**	**Steady state**		**Theorell-Chance**		**Equilibrium**		**Steady state**		**Equilibrium with EBQ**		**Equilibrium with EAP**		**Classic**		**Non classic**		**LICAD2 (this work)**	
		↓	↑	↓	↑	↓	↑	↓	↑	↓	↑	↓	↑	↓	↑	↓	↑	↓	↑
A	**P**	M	U	M	N	N	N	C	N	C	N	M	M	M	N	C	C	**M**	**U**
“	**Q**	C	C	C	C	C	C	C	C	C	C	C	C	C	C	M	N	**C**	**C**
B	**P**	M	M	C	C	N	N	C	N	C	N	C	C	C	C	M	N	**M**	**M**
“	**Q**	M	N	M	N	C	N	M	N	M	N	M	N	M	N	C	C	**M**	**N**

***Table 2.4**: Product inhibition summary. The table shows the type of inhibition observed for different mechanisms [80]. Also, the inhibition data for LICAD2 is provided for each combination of inhibitor product $[I]_p$ and the variable substrate $[S]_v$ at concentrations mentioned in Fig.2.18. A NADPH; B Sinapaldehyde; Q $NADP^+$; P Sinapyl alcohol. Two concentrations of fixed substrate $[S]_f$ were used: sub-saturating ↓ at 5 μM and saturating ↑ at 50 μM. Care was taken not to exceed concentrations above 50 μM due to substrate inhibition effects. Type of inhibition: M Mixed; C Competitive; U Uncompetitive; N No inhibition*

Different combinations of variable and fixed substrates along with an inhibitor product were chosen as shown in Table 2.4. With NADPH as the variable substrate, $NADP^+$ gave a linear competitive inhibition (Fig. 2.18d, e, f), while sinapyl alcohol gave an uncompetitive inhibition (Fig. 2.18a, b, c). With sinapaldehyde as the variable substrate, $NADP^+$ was unable to inhibit the enzyme, while sinapyl alcohol showed a mixed inhibition (Fig. 2.18g, h, i). These results, however, were observed when a saturating concentration of the fixed substrate $[S]_f$ was considered. At sub-saturating concentrations, the inhibition patterns for P with A as the variable, as well as for Q with B as the variable changes to mixed types, while the other two combinations shows the same type of inhibitions. The inhibition constants are mentioned in Table 2.4.

Putting in simple terms, a competitive inhibitor competes for the variable substrate and reversibly binds to the free enzyme form [E], a mixed (generalized form of non-competitive) inhibitor can bind to both [E] as well as [ES], while an un-competitive inhibitor can only bind to [ES]. If a random order mechanism holds true for LlCAD2, then all the products should give a competitive inhibition irrespective of the variable substrate, i.e. all the substrates and products has the ability to bind [E]. In another scenario, if a substituted mechanism takes place where Q releases before binding of P, then P cannot inhibit when A is varied, simply because both A and P will bind to different forms of the enzyme. Also, P will compete for B in this mechanism and thus will give a competitive inhibition pattern.

In a compulsory-order ternary-complex mechanism, A obligatorily binds first to the free enzyme [E] followed by B to form the [EAB] complex. Also, P is released first followed by Q to give the free enzyme [E]. In such a mechanism, only A and Q can bind to the free enzyme, and thus, Q which is structurally analogous to A (oxidized form of A) will behave like a competitive inhibitor when A is varied. Also, P can only bind to the [EA] complex, thus reflecting an un-competitive inhibition pattern. The case complicates a bit when B is varied, however, as there are two complexes where P can bind: [EA] and [EAB], and thus should give a mixed-type inhibition. A schematic of the product inhibitions obtained for LlCAD2 is shown in Fig. 2.19. The results thus effectively show that LlCAD2 follows a compulsory-order ternary-complex mechanism.

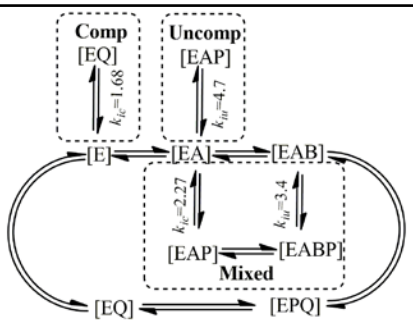


***Fig. 2.19**: Schematic for catalytic mechanism of LlCAD2. Different product inhibitions are also mentioned with inhibition rate constants calculated from Fig. 2.18. Q competes with A for [E] and thus gives a competitive inhibition. P can bind to the wrong form of enzyme [EA] when A is varied and thus gives un-competitive inhibition. P can also bind [EA] as well as [EAB] when B is varied thus giving rise to a mixed form of inhibition. The inhibition rate constants are mentioned which were calculated by fitting data to each inhibition equation models.*

However, this mechanism consists of 3 possible variants. Theorell-Chance mechanism in which the central ternary complex is transient and does not accumulate at all was ruled out based on the inhibition patterns with P as the inhibitor. Here, since the ternary complex [EAB] (and [EPQ]) is not formed, the P directly competes with B for the [EA] complex only, and a competitive inhibition pattern will be observed instead of a mixed inhibition (the one observed for LlCAD2). Another mechanism is the equilibrium-ordered mechanism wherein a rapid equilibrium is established for [EA], thus the rate-limiting step simplifies down to the inter-conversion of ternary complexes [EAB] to [EPQ]. A visual analysis of the Lineweaver-Burk plots for P as an inhibitor, when B was varied, shows that the lines intersected in the 2nd quadrant, which reflects either a steady-state or Theorell-Chance mechanism taking place, thus ruling out any rapid-equilibrium mechanism. Also, when A was varied, parallel lines were obtained (due to $K_A > K_{iA}$, they would intersect in the 2nd quadrant if $K_{iA} > K_A$). For an equilibrium-ordered mechanism these lines would have intersected on either of the axes. Thus, based on the interpretations mentioned above, it can be said that LlCAD2 follows a Steady-State Compulsory-Ordered Ternary-Complex mechanism for catalysis.

Though only inhibition arising from sinapyl reactants (aldehyde and its corresponding alcohol) along with NADPH is shown here, similar experiments with coniferyl were also performed. However, the inhibition patterns as well as the inhibition constants were almost similar (data not shown) which would give the same catalytic mechanism as sinapaldehyde. For detailed explanation and interpretations of different catalytic mechanisms following product inhibition experiments, please refer the work of Cleland [80-82] and Cornish-Bowden [83].

2.12 Zinc Inhibition

As observed in previous chapter, Zn^{2+} was present in the LlCAD2 structure and a prerequisite for catalysis; however, externally supplementing Zn^{2+} resulted in loss of activity. As a step to understand this duality, Zn^{2+} inhibition studies were carried out. It was observed that Zn^{2+} inhibited LlCAD2 at very low concentrations and resulted in complete loss of activity at concentrations above 15 µM. The Michaelis-Menten and Lineweaver-Burk plots for Zn^{2+} showed it to be a mixed inhibitor (Fig. 2.20a, d). However, the secondary plots showed a characteristic hyperbolic pattern (data not shown). Also, Dixon and Cornish-Bowden plots showed a similar curvature (Fig. 2.20b, c). However, the most striking observation was made when the data was re-plotted with Zn^{2+} as the variable at fixed [S]. The data was fit into equation (11) which can be used to fit inhibition data (section 2.3.6). Also, the binding of Zn^{2+} to LlCAD2 was measured by fluorescence quenching studies upon formation of [E-Zn] complex, as well as thermodynamic parameters for Zn inhibition

were calculated as mentioned in experimental section. The binding and inhibition parameters are given in Table 2.5.

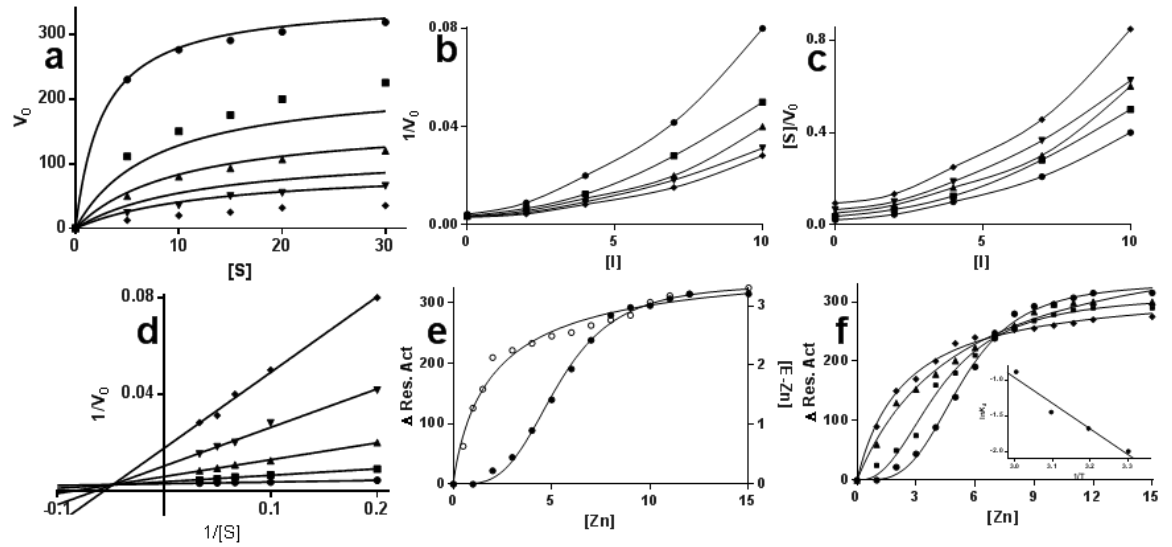


Fig. 2.20*: Plots for Zn^{2+} inhibition (0-10 µM) with sinapaldehyde as the variable substrate (0-30 µM). The data of Michaelis-Menten plots* ***(a)*** *was fit into equations for different inhibition mechanisms in order to estimate the inhibition constants. While, Dixon plot* ***(b)****; Cornish-Bowden plot* ***(c);*** *and Lineweaver-Burk plot* ***(d)*** *were used for visual analysis of the data;* ***(e)*** *Effect of Zn^{2+} on inhibition of initial velocity "i" (•••), and binding curve (o o o). i represents the loss of initial velocity in presence of Zn^{2+}. It can be used as a measure of change in the initial velocity effected by Zn^{2+}. The sigmoidal curve represents the fit to equation (11). The Enzyme-Zinc complex [E-Zn] was determined following a fluorescence titration of LlCAD2 against increasing concentrations of Zn^{2+} and measuring decrease in fluorescence intensity as mentioned in Sections 2.3.6 & 2.5.2;* ***(f)*** *Effect of temperature on i,* ***Inset:*** *Arrhenius plot for Zn inhibition*

It was observed that Zn^{2+} binding followed a hyperbolic pattern and maximum binding was achieved at 15 µM (Fig. 2.20e). This observation shows that binding of Zn^{2+} to LlCAD2 is not co-operative, i.e. binding of Zn^{2+} does not induce any structural change so as to facilitate binding of other Zn^{2+} atoms. However, contradictory results were obtained from the effect of Zn^{2+} on the initial velocity of LlCAD2 reaction. It was observed that Zn^{2+} showed a sigmoidal inhibition pattern when considered as the variable reactant against fixed concentrations of the actual reactant (sinapaldehyde). Based on this observation, it can be said that Zn^{2+} shows a heterotropic allosteric effect on LlCAD2 catalytic reaction (Fig. 2.20e), i.e. it is able to bind to a site other than the active site and induce alterations in the substrate binding pocket, thus, lowering the substrate binding and eventually inhibiting the enzyme. This observation was unusual, as CADs (and ADHs in general) are not known to possess any allosteric sites in their structures. Measuring the inhibition parameters as a function of temperature, it was observed that with increase in temperature, the allosteric behavior decreased, and it attained a true hyperbolic form above 50 °C (Fig. 2.20f). This observation, along with decrease in *h* (Hill's coefficient) suggests that increasing temperature

results in a decrease in the allosteric effects of Zn^{2+}. Decrease in I_{max}, $K_{0.5}$ and ΔG values suggests that Zn^{2+} inhibition is facilitated at lower temperatures, while it decreases at higher temperatures.

	Temp °C	B_{max} (μM)	K_d (μM)	*h*	$\Delta G^{\S 4}$ (KJ mol^{-1})	$\Delta H^{\S 5}$ (KJ mol^{-1})	$\Delta S^{\S 5}$ (J mol^{-1})
Binding[1]	25	3.553	1.846	1.0	4.57	-	-
	***Temp* °C**	I_{max} (nKat ug^{-1})	$K_{0.5}$ (μM)	*h*	$\Delta G^{\S 4}$ (KJ mol^{-1})	$\Delta H^{\S 5}$ (KJ mol^{-1})	$\Delta S^{\S 5}$ (J mol^{-1})
Inhibition[2]	30	331.8	7.349	3.51	-18.52	29.68	-16.1
	40	331.8	4.346	2.42	-11.35		
	50	332.7	4.25	1.0	-11.44		
	60	325.8	2.424	1.0	-6.71		

***Table 2.5**: Zn²⁺ inhibition summary. ¹ calculated from Fig. 2.20e, ² from Fig. 2.20f, ³calculated from equation (5), and ⁵ calculated from Fig. 2.20f inset*

Thus, Zn^{2+} shows some unique properties with respect to LlCAD2. It was seen that the role of metal ion is of dual nature, primarily, facilitating catalysis and maintaining structural integrity of the protein molecule as shown in previous chapter. Removal of this metal not only resulted in loss of activity but also may have induced certain structural changes. These changes are mainly irreversible in nature, as reincubating the chelated protein with metals did not result in active protein, as also suggested in a report of yeast ADH [84]. Secondly, externally supplied Zn^{2+} ions inhibited LlCAD2 activity in a very unique fashion; however, the precise mechanism behind such inhibition needs further investigation. Though one can only speculate based on the results shown here, but still can predict that such duality in the nature of LlCAD2 in response to Zn^{2+} is due to some form of regulatory control exerted by the phenylpropanoid pathway. A similar report on CADH1 and H2 from *Streptomyces* was published recently with different catalytic properties [85]. While CADH1 was a cinnamyl specific CAD with no inhibitory effects of Zn^{2+} and Fe^{2+}; CADH2, a coniferyl specific CAD, showed almost 80-90% inactivation in presence of both the metal ions. Thus, it can be surmised that different CAD isoforms could not only differ in terms of their substrate specificities, but also respond differently to various metals and chemical reagents.

2.13 Equilibrium Kinetics

So far in this chapter, we dealt with different kinetic aspects of LlCAD2, however, all of them were based on the steady-state assumption and initial velocity measurements. However, there is another aspect to enzymatic reactions once equilibrium is achieved amongst its reactants and products. The equilibrium constants for all the substrate reactions were calculated as mentioned in 2.3.10-11. The reactions were set up as already mentioned and were allowed to equilibrate for more than 3 h at 30

°C, and ΔA_{340} was measured. The individual concentrations of substrates and products were calculated based on the extinction coefficients provided (Table 1.1). Also, equilibrium binding of substrates to LlCAD2 was monitored using Trp fluorescence quenching as a measure of binding. Here, a fixed concentration of LlCAD2 was titrated with increasing concentrations of substrates and decrease in fluorescence intensity was measured. Concentrations of free and bound ligands $[L]_f$ and $[L]_b$ were calculated according to Equations (19) & (20). Dissociation rate, K_d, was estimated from data provided in Fig. 2.21. The data was fit into different equations for protein-ligand binding and best fits were obtained with that of a model with two independent binding sites. These two sites may be assumed as the two substrate binding pockets on an LlCAD2 dimer.

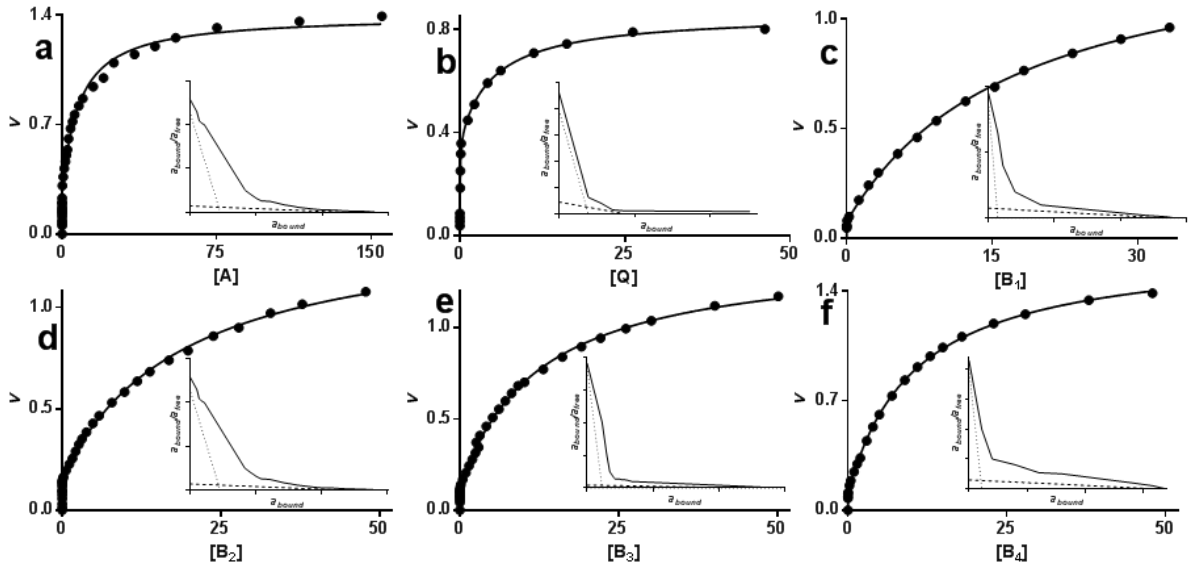


Fig. 2.21*: Binding curves for LlCAD2 with different substrates. v is defined as mol [L] bound per mol [E] for* ***(a)*** *NADPH [A];* ***(b)*** *NADP⁺ [Q];* ***(c)*** *Cinnamaldehyde [B_1];* ***(d)*** *p-Coumaraldehyde [B_2];* ***(e)*** *Coniferaldehyde [B_3];* ***(f)*** *Sinapaldehyde [B_4].* ***Inset:*** *Scatchard plots. The plots show actual data (–), as well as calculated plot for the individual sites - site-1 (- - -), and site-2 (.....)*

	Equilibrium binding[1]					Equilibrium kinetics[2]				
	Site 1		Site 2			pH 6.25		pH 8.8		K_{eq}/K'_{eq}
	K_{d1}	n_1	K_{d2}	n_2	$\Delta G'$	K_{eq}	ΔG	K'_{eq}	$\Delta G'$	
NADPH	10^{-11}	0.38	8.1×10^{-6}	1.58	-22.97	-	-	-	-	-
NADP	4.8×10^{-11}	0.25	5.14×10^{-6}	0.48	-7.57	-	-	-	-	-
Cinnam	1.5×10^{-8}	0.07	1.9×10^{-5}	1.15	-23.53	6.13×10^{5}	-4.5	5.38×10^{5}	-4.2	1.14
Coum	3×10^{-9}	0.11	1.23×10^{-5}	1.27	-24.70	6.87×10^{5}	-4.8	2.78×10^{4}	-2.5	24.7
Conif	8×10^{-10}	0.12	1.77×10^{-5}	1.3	-16.22	8.53×10^{6}	-5.3	2.71×10^{4}	-2.5	314.8
Sinap	5×10^{-10}	0.16	1.05×10^{-5}	1.35	-20.23	9.65×10^{6}	-5.7	2.48×10^{4}	-2.2	389.1

Table 2.6*: Summary of experiments at equilibrium. [1] The values for K_d and n are calculated from Fig. 2.21, [2] Equilibrium kinetic parameters K_{eq} & ΔG (for reduction reaction, at pH 6.25) and K'_{eq} & $\Delta G'$ (for oxidation reaction, at pH 8.8) were calculated as mentioned in 2.3.10*

The dissociation rates, K_{d1}, K_{d2}, are given in Table 2.6. The results shown here propose that one of the sites is in rapid equilibrium with the interacting ligands as suggested with very low values of $K_{d1,}$ while the second site binds with several orders of magnitude lower affinity. Also, the total number of bound ligands is always below 2 (theoretical binding sites). Both the co-substrates bind with much higher affinity than aldehydes, where NADPH shows maximum affinity followed by $NADP^+$ as can be seen from their binding hyperbolic curves in Fig. 2.21. No fluorescence quenching was observed with the alcohol substrates suggesting that they are unable to bind to the free form of LlCAD2. The values for K_d, n and free energies of binding $\Delta G'$ (for site-2) suggests that the binding efficiency for aldehydes increases as the methoxylations in their structures increases; sinapyl group having two methoxy groups shows maximum efficiency for binding to LlCAD2.

The equilibrium kinetics showed that the equilibrium was established far towards the formation of alcohols, suggesting that LlCAD2 carries out a predominant reduction of aldehydes rather than the reverse reaction. The ratio of K_{eq} to K'_{eq} can be used as a measure to predict direction of a reaction. The ratio was almost 1 for cinnamyl reaction, suggesting that this group does not necessarily show any preference from either its reduction or oxidation. While this ratio increased drastically as we move down the substrates (with increasing number of methoxy groups), with sinapaldehyde showing the maximum ratio.

3

CAD: Structure

3A

Crystallization and structure determination

Techniques used to crystallize LlCAD2 and prediction of its 3-dimensional structure is discussed here. The structure was determined using molecular replacement method using AtCAD5 co-ordinates (PDB: 2CF6) and was refined at resolutions 3.5 Å. However, due to high R_{merge}, a homology model using the same template was generated and used instead. The modeled structure of LlCAD2 was then compared to different CADs and ADHs.

3.1 X-ray diffraction gives a low-resolution structure of LlCAD2

LlCAD2 was crystallized as mentioned 2.4.1. Different optimized conditions for crystal formation are summarized in Table 3.1. Microcrystals appeared initially in many different conditions, and further optimization was carried out in order to obtain diffraction quality crystals. Various parameters like precipitant, salt and enzyme concentrations, temperature, pH, additives, detergents, were considered. Also, seeding the crystallization solution with micro-crystals was attempted. LlCAD2 crystallized in many different conditions; however, diffraction quality crystals were obtained in only 3 conditions as mentioned in Table 3.1 & 3.2.

The best crystals of LlCAD2 grew in 0.2M ammonium acetate with 20% PEG 3350, 0.1% NDSB 195 at pH 8.1 after 12 days of growth, and diffracted at a resolution of 3.5 Å. The protein crystallized in monoclinic C_2 space group with the unit cell as mentioned in Table 3.2. Attempts were made to determine the structure by molecular replacement method using coordinates of 2CF6, a zinc-type cinnamyl alcohol dehydrogenase from *A. thaliana* (AtCAD5). However, due to high R_{merge} values during data processing, the structure could not be determined with accuracy, though a low-resolution electron density map was generated; where, only the Cα backbone was predicted, while the side chains could not be determined. The asymmetric unit composed of two molecules and the crystallographic 2-fold axis produces a dimer tightly associated through two 2-fold related β strands. Consequently, the dimer formed an extended 12 stranded β-sheet with 6 strands from each subunit. The crystallographic data and refinement statistics are summarized in Table 3.2.

3.2 Homology model of LlCAD2 is identical to AtCAD5

Initially 300 models were generated using AtCAD5 as a template and their DOPE score was calculated, and the model having lowest DOPE score was considered as the best model which was then energy-minimized and validated using different tools available on SAVEs server. Based upon these analyses the model was further refined using the loop refinement scripts of Modeller until a

satisfactory model was generated. Since LlCAD2 has 77% identity with AtCAD5, the overall backbone structure of the modeled LlCAD2 had no significant changes as shown in Fig. 3.2. The RMSD of alpha carbon (Cα) of the homology model was calculated by structural superimposition of model with template and was found to be 0.117 Å.

No.	Salt	Precipitant	Additive	pH	Temp	Days	Crystals	Res.
1	0.2 M Ammonium acetate	20% PEG 3350	NDSB-195	8.1	22 °C	12		3.5 Å
2	0.2 M Potassium acetate	"	-	8.1	10 °C	13		4.5 Å
3	0.2 M Sodium formate	18% PEG 3350	-	7.2	RT	20		Fiber type
4	"	25% PEG 3350, 25% Glycerol	FOS-Choline-12	"	"	15		4 Å

Table 3.1: *Different conditions for crystal formation*

Condition	Ammonium acetate[1]	Potassium acetate[2]	Sodium formate[3]
Space Group	C_2 monoclinic	P_{222} orthorhombic	P_{222} orthorhombic
Temperature	100K	100K	100K
X-ray source	SSRL BL12-2	Home-source	Home-source
Wavelength (Å)	0.9560	1.5418	1.5418
Unit-cell parameters (Å)	a=99.49, b=119.15, c= 138.64 α = γ = 90°, β= 103.15°	a= 63.95, b=131.49, c= 182.58 α = β = γ = 90°	a=64.18, b=82.24, c=135.90 α = β = γ = 90°
Molecules per AU	2		
Matthews coeff. (Å^3 Da^{-1})	2.50		
Solvent content (%)	51		
Total observations	34207 (5483)		
Unique reflections	16334		
Multiplicity	2.1 (2.2)		
Resolution range (Å)	37.67-3.5 (3.69-3.5)		
Average I/σ(I)	4.2 (1.1)		
Rmerge (%)	21.8 (83.5)		
Rpim(%)	18.7 (69.2)		
Completeness (%)	82.2 (86.1)		

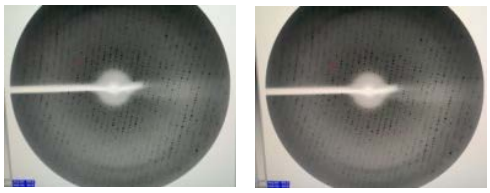


Fig. 3.1*: Diffraction image obtained from crystals of LlCAD2 mentioned in Table 1 (No. 1)*

Table 3.2: *Crystallographic data, phasing and refinement statistics for LlCAD2. [1]Values in parenthesis represents the highest resolution shell. The data was not reliable due to high R_{merge}. [2]The data could not be integrated. [3]Data could not be processed due to high Mosaicity (0.76)*

3.3 Overall structure is quite similar to other medium-chain dehydrogenases (MDRs)

Homodimeric MDRs are generally found in higher eukaryotes, while lower eukaryotes and prokaryotes mostly possess tetrameric MDRs. In order to achieve a basic understanding of the structure and the catalytic mechanism of LlCAD2, its sequence as well as structure was compared with other MDRs. BLASTP search of LlCAD2 amino acid sequence in the NCBI database revealed highest identity (96%) to LlCAD; another isoform from *L. leucocephala* (P. Kulkarni, unpublished) (Fig. 3.3). Other higher eukaryotic plants also showed high identity; however, aspen SAD which shares much of the catalytic specificity with LlCAD2 had a relatively lower identity (54%). Prokaryotic alcohol dehydrogenases like yeast CAD showed 42% identity, while other members of this group were rather low in identity (< 27%). While the 3D structure of LlCAD2 mostly resembled to that of AtCAD5 and PtSAD with Z-scores 61.4 and 50.4, respectively (obtained from Dali search engine), the search also showed additional PDB structures like *S. cerevisiae* CAD (Z-score 45.9), *E. coli* ADH type protein YahK (Z-score 45.7), *H. pylori* CAD (Z-score 45.4), and other prokaryotic ADHs showing highly similar structures, however with much lower amino acid identity.

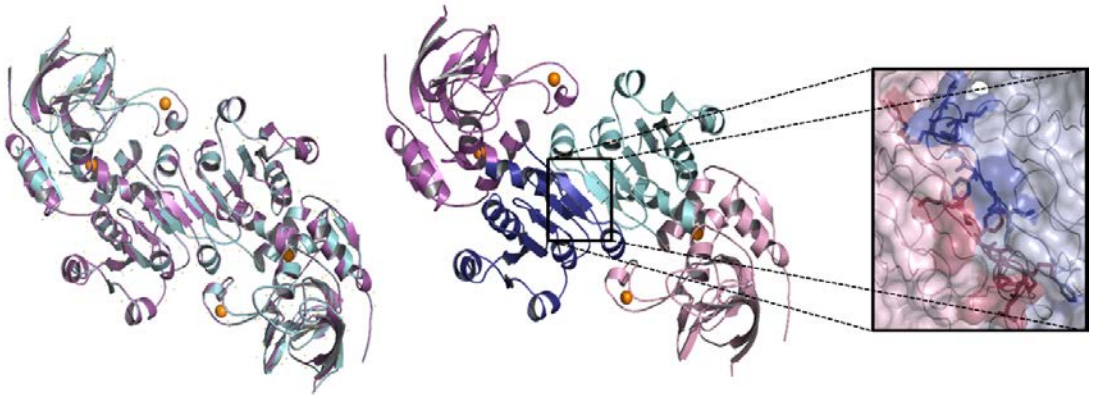


Fig. 3.2: (a) *Modelled structure of LlCAD2 homodimer. The catalytic and nucleotide-binding domains are colored in purple and blue for subunit A, and light purple and cyan for subunit B, respectively. Catalytic and structural Zn2+ are depicted as orange spheres;* ***(b)*** *Superimposition of LlCAD2 (purple) and AtCAD5 (cyan) (RMSD 0.117 Å)* ***(c)*** *The interface of the dimer of LlCAD2 (residues 278-299). Blue colored represents residues from subunit A, and orange color from subunit B*

In spite of its relatively low sequence identity, LlCAD2 folding was similar to different Zn^{2+}-dependent ADH structures belonging to MDR superfamily. Like the horse liver alcohol dehydrogenase (8ADH) and AtCAD5, each monomer is composed of two distinct domains: dinucleotide binding domain having a classic Rossmann fold (residues 163-301), and the catalytic domain (residues 1-162 and 302-359). The nucleotide-binding domain is composed of a β -pleated sheet of six parallel strands flanked by five helices, whereas the catalytic domain consists mainly of a. core of antiparallel β-strands with six helical segments at the surface of the molecule. A deep cleft located between the two domains with a well coordinated catalytic Zn^{2+} ion serves as the active site.

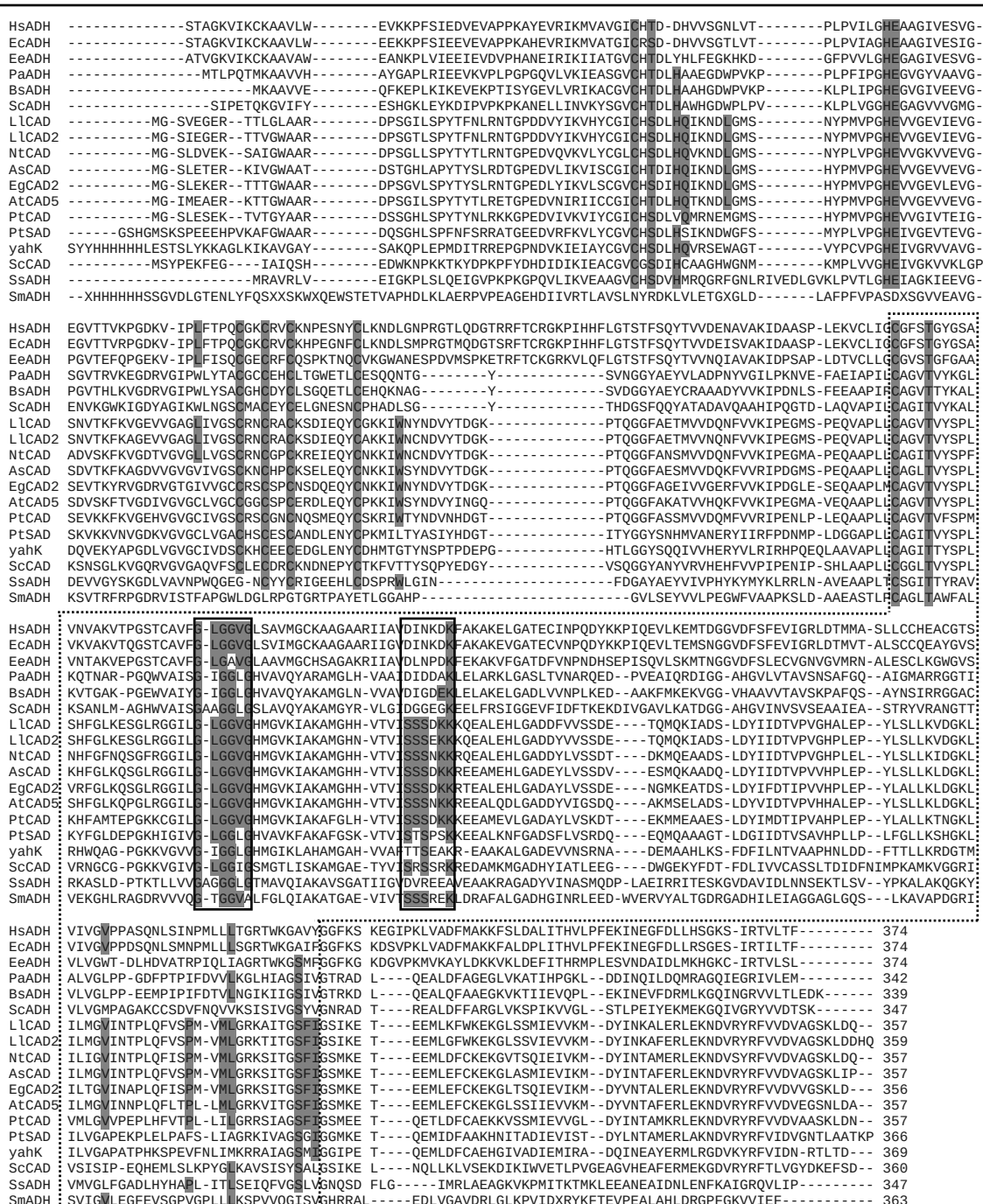


Fig. 3.3: *Multiple sequence alignment of amino acid sequences of different ADHs. The nucleotide-binding domain and conserved motifs are boxed by a thin dotted line and thick solid line, respectively. Residues constituting substrate-binding site are highlighted in grey. HsADH (H. sapiens ADH; 1HDY), EcADH (E. caballus liver ADH; 8ADH), EeADH (G. morhua ADH; 1CDO), PaADH (P. aeruginosa ADH; 1LLU), BsADH (B. stearothermophilus ADH; 1RJW), ScADH (S. cerevisiae ADH1; 2HCY), LlCAD (L. leucocephala CAD1; ACG58885), LlCAD2 (L. leucocephala CAD2; KC907297), NtCAD (N. tabacum CAD; CAA44217), AsCAD (P. taeda CAD; AAF43140), EgCAD2 (E. gunnii CAD; CAA46585), AtCAD5 (A. thaliana CAD5; 2CF6), PtCAD (P. tremuloides CAD; CAA86073), PtSAD (P. tremuloides SAD; 1YQD), YahK (E. coli putative ADH; 1UUF), ScCAD (S. cerevisiae CAD; 1Q1N), SsADH (S. solfataricus ADH; 1R37), SmADH (S. melitoti ADH; 3UOG)*

Residues 278-299 of the nucleotide binding domain of each monomer are involved in dimerization, where the outermost strand of monomer A forms hydrogen bonds with the corresponding β-strand of monomer B, exhibiting an antiparallel arrangement (Fig. 3.2). The resulting 12-stranded twisted β-sheet quaternary structure closely resembles that of AtCAD5 and other ADHs. However, unlike horse liver ADH, substantial deletions of sequences in the loop area between β8- β9 strands in ADHs from higher eukaryotes can be observed in Fig. 3.3. This region is located at the entrance of the substrate binding site, thus presuming its role in allowing entry of larger molecules, which may be restricted in case of LlCAD2.

3.4 Zn centers are similar to other ADHs

Only one Zn^{2+} per subunit was showed in original report of CAD [52], however, LlCAD2 showed two Zn^{2+} per subunits; structural and catalytic, both coordinated in the catalytic domain; like other Zn dependent MDRs. As shown in Fig. 3.4a, the structural zinc is coordinated by four Cys residues (Cys100, Cys103, Cys106 and Cys114) located in a short α-helix containing loop as a protrusion from the core of the catalytic domain. This protrusion is involved in oligomerization of the dimers in many tetrameric ADHs, however in the case of LlCAD2, its functional role remains uncertain. Like AtCAD5 structure, a hydrophobic substrate binding pocket is formed inside a cleft between the two domains, where the catalytic Zn^{2+} is located; coordinated by Cys47, His69, Glu70 and Cys163 (Fig. 3.4b). Thus, taken together the observed structural features, it was established that LlCAD2 belongs to the Zn^{2+}-dependent MDR family [68].

3.5 Cofactor binding site is similar to AtCAD5

Mostly, NAD-dependent Zn^{2+}-ADH structures have been extensively characterized so far [86]. And like NAD enzymes, in NADP enzymes the nicotinamide ring is close to the catalytic Zn^{2+} at the bottom of the substrate-binding pocket (Fig. 3.4c), and the co-factor binding amino acids can be divided into three groups- amide binding, nucleotide binding and phosphate binding. The backbone of Gly275 (part of a conserved region ^{275}GVINTPLQFVSP^{286}) mainly interacts with the amide group of NADP. Leu95 and Thr167 are also seen in hydrogen bonding distance to the amide ring. The binding of the pyrophosphate group of the dinucleotide is accomplished by Leu189 and Val192 located in a glycine rich motif (^{188}GLGGVG^{193}) at the first β-α-β unit of cofactor binding domain. This motif is a characteristic of typical NAD(P) dependent enzymes [87, 88] as well as all Zn-dependent MDRs (Fig. 3.3). Also, Met195 and Thr 251 are within hydrogen-bonding distance to pyrophosphate group.

Another loop region (^{211}SSSEKK^{216}) which interacts with $NADP^+$ is highly conserved amongst bonafide CADs. Lys216 also interacts with the ribose ring. The side chains of conserved serine residues in this region are proposed to play a role in conferring co-substrate specificity (NADP(H) over NAD(H)) [88, 89]. In NADP dependent enzymes, Ser213 located at the C-end of β-B strand hydrogen bonding with 2'-phosphate of NADP, is generally replaced with Asp. Also, Ser211 and Ser212 are located in the loop between β-B & β-C, forming hydrogen bonds with 2'-phosphate of adenine ribose. As has been reported already, these conserved loops interacts with NADP and facilitates the rate-limiting step of many MDRs; the dissociation of the products [27, 88]. Close examination of NADP docked structure demonstrates that residues involved in putative proton relay also forms hydrogen bonds with NADP, supporting the contention that the modeled structure represents a catalytically competent form of the LlCAD2.

3.6 Substrate binding site is close to catalytic Zn

A suitably sized substrate binding pocket can be distinctively deduced to be near the catalytic Zn^{2+}, lined up mostly with hydrophobic residues where hydroxycinnamaldehydes bind via largely shape-dependent Van der Waals interactions [90]. The pocket is made up of 14 residues; where residues of monomer B, in part, define the active site pocket of monomer A, and vice versa. The substrate binding cavity is made up of Ser49, Gln53, Leu58, Met60, Leu95, Trp119, Thr167, Val276, Ser298, Phe299 and Ile300 from one subunit, and Pro286, Met289 and Leu290 from other subunit. The Cα carbons of these residues superimposes well with those of horse liver ADH as well as AtCAD5 structures (data not shown) [88, 91]. Many of these residues are completely conserved, however, a conservative heterogeneity is observed in some of the residues, *e.g.* instead of serine at position 49, several ADHs had threonine, including AtCAD5 and yeast ADH. Unlike horse liver ADH, a major deletion can be seen in all higher ADHs (Fig. 3.3), which might play a role in governing the overall geometry of the substrate-binding pocket. As observed in some Zn^{2+}-dependent MDRs, the modeled substrates ligate to catalytic Zn^{2+} through its –CHO group and results in an extensive hydrogen bonded network, which subsequently allows hydride abstraction and protonation. The modeled structures of different hydroxycinnamyl substrates in the active site of LlCAD2 are shown in Fig. 3.3 d-i. The aldehydic oxygen of the substrates is also within hydrogen-bonding distance to the –OH group of Ser49, where coordination with catalytic Zn^{2+} might facilitate the correct orientation of aldehyde oxygen in order to form hydrogen bond with Ser49 [88]. As already mentioned, both the conserved residues, Ser49 and His52, are within hydrogen-bonding distance of the O2' of the nicotinamide ribose, as well as in close proximity to the catalytic Zn^{2+}, thus presuming their role in allowing hydride transfer from NADP to the substrate in addition to fixing the position of amide

fixing the position of amide ring. The docking energies of all the substrates are mentioned in Table 3.3, which showed only minor differences. These values suggest that the binding affinity of the substrates increases with increase in the side-group size, with sinapaldehyde having the highest affinity for the modeled active site of LlCAD2.

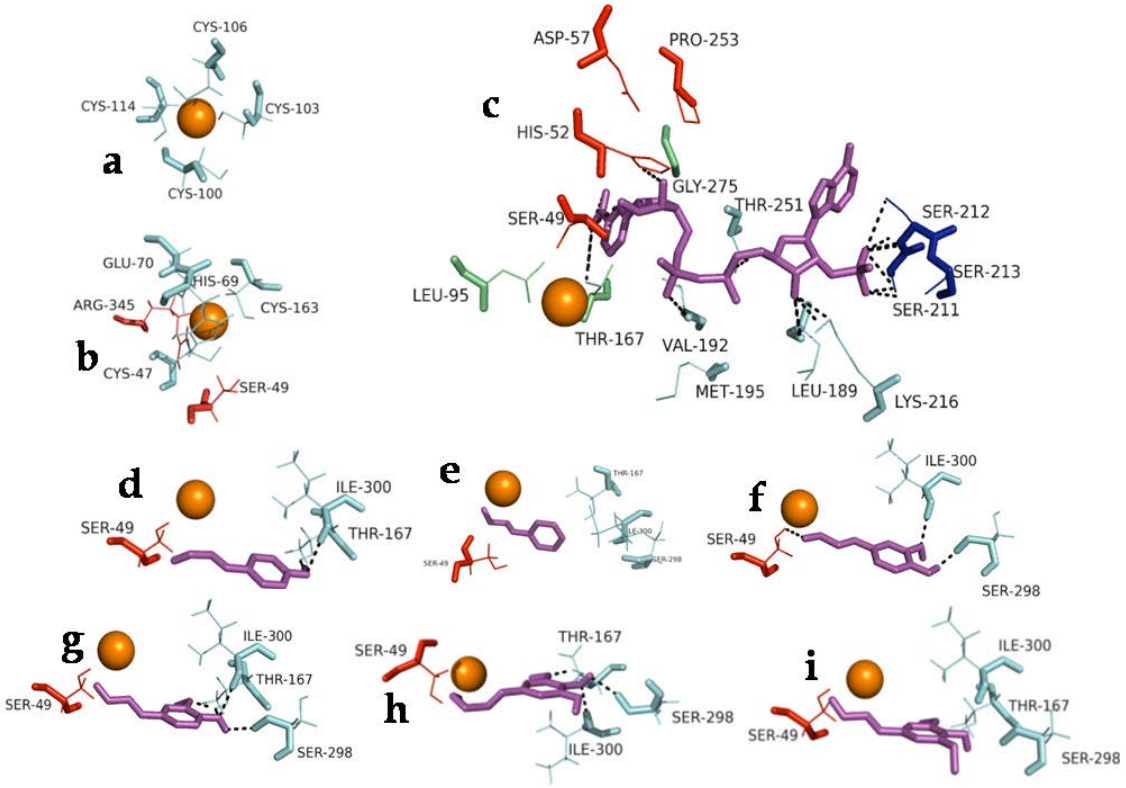


Fig. 3.4: *Modeled metal-binding sites of LlCAD2.* ***(a)*** *Structural Zn center;* ***(b)*** *Catalytic Zn center. Orange- Zn ion, Cyan- residues co-ordinating with* Zn^{2+}*, Red- residues involved in catalysis;* ***(c)*** *NADPH docked to LlCAD2 model. Magenta- NADPH; Red- residues involved in putative proton relay; Green- amide binding; Cyan- nucleotide binding; Blue- phosphate binding residues; Orange- catalytic Zn. Substrate docked to LlCAD2 model:* ***(d)*** *Cinnamyl;* ***(e)*** *p-Coumaryl;* ***(f)*** *Caffeoyl;* ***(g)*** *Coniferyl;* ***(h)*** *5-OH coniferyl;* ***(i)*** *Sinapyl. Magenta- substrates; Cyan- amino acid residues involved in substrate binding*

Substrate	Side Groups		Binding Energy (K J mol^{-1})	
	R1	R2	Aldehyde	Alcohol
Cinnamyl	-H	-H	**-15.1**	**-17.6**
***p*-Coumaryl**	-H	-H	**-21.0**	**-18.4**
Caffeoyl	-OH	-H	**-23.5**	**-21.5**
Coniferyl	$-OCH_2$	-H	**-24.0**	**-23.4**
5-OH coniferyl	$-OCH_2$	-0H	**-25.9**	**-25.9**
Sinapyl	$-OCH_2$	$-OCH_2$	**-26.8**	**-26.1**

Table 3.3: *Binding energies for docked substrates*

3B

Understanding the active-site morphology

Active site directed chemical modification is an important tool for studying structure-function relationship of enzymes. Site-directed mutagenesis is widely used for the same purpose, which need not only the cloned gene but also the structure of the protein. Based on the structural information drawn together in the previous section, attempts were made to understand the nature of the catalytic site. The possible involvement of different amino acids in substrate binding and catalysis was determined. Also, a hypothetical mechanism involving catalysis is proposed here.

3.7 Chemical modifications suggests Ser, His and Cys at active site

LlCAD2 was chemically modified using various amino acid group specific chemical reagents in order to understand their respective roles in catalysis (Table 3.4). NBS, PG, NAI failed to inhibit the enzyme; suggesting no involvement of Trp, Arg and Tyr residues in catalysis, respectively. On the other hand, PMSF (99%), DEPC (77%), pCMB (80%), NEM (85%), DTNB (82%) strongly inhibited LlCAD2 activity. CA (27%) and WRK (21%) also showed marginal inhibitory effects, thus, indicating probable presence of Ser, His, Cys, Lys, Carboxylate at or near the active site. The inactivation rate constants K_{app} and order of the reactions n were determined from Fig. 3.5 and showed that one residue each of His, Lys, Carboxylate, and 2 residues each of Ser and Cys were essential for catalysis. Kinetics of the modified enzyme was studied to understand the role of respective amino acid residue at the active site.

K_{cat}/K_m values decreased for the enzyme modified with all the modifiers. The kinetic parameters for Ser modified enzyme could not be determined as saturation could not be achieved, while the K_m values for carboxylate and Cys modified enzymes showed a modest change, indicating their roles in catalysis. Lys and His modified enzymes showed increase in K_m values indicating decrease in the substrate affinity, along with decrease in K_{cat}/K_m, indicating their roles in substrate binding and holding the reactive groups in proper position. The enzyme was substantially protected by pre-incubating with substrates before modification reactions by CA and WRK, confirming Lys and Carboxylate to be present at the active site. While no such protection was observed in Ser, His and Cys modified enzymes, indicating their roles in catalysis apart from substrate binding.

3.8 Zn coordination site involves Glu70

Based on chemical modifications and interactions observed in docking simulations, several site-directed mutants were generated in order to further understand the functional roles of putative

Residue	Reagent	Conc. (mM)	Activity (%)[1]	Protection (%)[2]	Reaction order n[3]	Sinap[4]		NADPH[4]	
						K_m	K_{cat}/K_m	K_m	K_{cat}/K_m
Control	-	-	100	100	-	2.8	11.6	2.5	11.5
Ser	PMSF	0.5	0.3	0.5	2	n.a.	n.a.	n.a.	n.a.
His	DEPC	1	23	25	1	31.5	0.2	19.6	0.6
Cys	pCMB	0.01	20	17	2	8.34	0.2	12.5	0.3
"	NEM	10	15	16	-	-	-	-	-
"	DTNB	1	18	13	-	-	-	-	-
Lys	CA	5	42	73	1	13.4	1.3	51.2	0.4
COO-	WRK	0.5	45	79	1	7.6	3.1	9.4	2.2
Trp	NBS	5	97	98	-	-	-	-	-
Arg	PG	10	98	97	-	-	-	-	-
Tyr	NAI	3	96	95	-	-	-	-	-

Table 3.4: *List of amino acid residues modified with different chemical reagents. [1] Residual activity calculated as % of total activity in unmodified enzyme. [2] Activity calculated as % of total activity protected by pre-incubation with substrates. [3] Calculated from Fig. 3.5. [4] calculated at IC_{50} of all chemical modifiers. PMSF Phenylmethylsulfonyl fluoride; DEPC Diethyl pyrocarbonate; pCMB pChloromercuricbenxoic acid; NEM N-ethyl maleimide; DTNB Dithionitrobenzoate; CA Citraconic anhydride; WRK Woodwards reagent K; NBS N-bromosuccinimide; PG Phenylglyoxal; NAI N-acetyl imidazole*

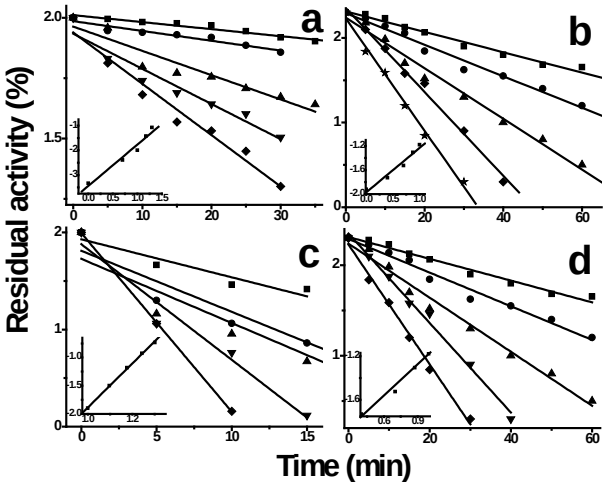


Fig. 3.5: *Pseudo first order plots for inactivation of LlCAD2 by* ***(a)*** *PMSF;* ***(b)*** *DEPC;* ***(c)*** *pCMB; and* ***(d)*** *Citraconic anhydride. Inset: Double logarithmic plots of the first order rate constants (K_{app}) as a function of modifier concentrations*

residues of the active site. A total of 18 different LlCAD2 mutants were generated, cloned, over-expressed and purified to homogeneity before assaying enzyme activity. Kinetic parameters for all the mutants are listed in Tables 3.5 & 3.6. An additional Zn coordination site at Glu70 was observed in the crystal structure of AtCAD5 as well as the modeled LlCAD2. This site was mutated in order to understand its role, and the mutant E70A resulted in complete loss of function (Table 3.5). Zn content of this mutant was estimated by Zincon assay which was found to be 2.7 ± 0.3 mol/mol dimer; much lower than native protein. These findings suggest that Glu70 is responsible for co-ordination with catalytic Zn ion and its mutation leads to the inability of the enzyme to bind to the

metal [92]. Such Glu-Zn coordination was also observed in AtCAD5, TbADH and Rat sorbitol dehydrogenase [49, 93, 94]. However, in some MDRs; including HLADH, PtSAD and ScAdh6p, instead of a Glu70, a water molecule completes the tetrahedral coordination shell [27, 90, 91].

3.9 Ligand binding pocket is identical to AtCAD5

Though a ternary complex of CAD has not been crystallized so far, a binary complex has already been reported from *A. thaliana* [49]. NADPH was found bound near the catalytic Zn ion, and the nicotinamide ring was observed to be in syn- conformation, which is in proper orientation for A-face specific hydride transfer from C4 to the corresponding substrate. Val192 was observed to be within hydrogen bonding distance to the pyrophosphate of $NADP^+$. Ser211-213 and Lys216 are a part of another loop (^{211}SSSEKK^{216}) which binds to NADP. Lys216 particularly interacts with the ribose ring, while the three Ser residues are in close proximity to the phosphate group of NADP and thus may play a role in defining the co-factor specificity. All these residues are highly conserved amongst CADs. Several LlCAD2 mutants were generated from the putative NADPH and substrate binding pockets. V192G, S211I, S212D, S213N and K216W showed an increase in *Km* for NADPH (5-20 folds) indicating lower affinity, with subsequent decrease in *kcat / Km*. Whilst no major changes in kinetic parameters were observed for sinapaldehyde, thus suggesting their roles in NADPH binding and its stabilization in CAD reactions. Val192 is a part of the glycine rich motif (^{188}GLGGVG^{193}) responsible for binding of the pyrophosphate of the dinucleotide as discussed in Chapter 3A. The role of Lys216 has already been established [79, 95]. On the other hand, L95R and G275F (both in proximity to the amide ring) showed no change in activity. The above results show that residues Val 192, Ser211-213 and Lys216 play their respective roles in binding and stabilizing the co-factor at the active site, while Leu95 and Gly275 were not found to be important.

Docking simulations showed 3 amino acid residues to be playing a role in binding of the hydroxycinnamyl substrates. Mutants for all 3 residues were generated and it was observed that T167A, S298A and I300G resulted in 14-22 fold increase in K_m for sinapaldehyde, while K_m for NADPH increased only marginally. These results suggest their specific roles in substrate binding. The loop (^{298}SFI^{300}) is highly conserved in different CADs (Fig. 3, chapter 3A), which play a role in defining the size of the binding pocket and its entrance, thus governing its substrate specificity. SAD from Aspen shows substitutions in this region as well as residues lining the pocket leading to altered substrate versatility [96].

Enzyme	Sinapaldehyde		NADPH		Possible Role[1]
	K_m	K_{cat}/K_m	K_m	K_{cat}/K_m	
Wild type	2.8	11.6	2.5	11.5	
S49A	-	-	-	-	Proton relay
H52D	36.9	0.5	29.1	0.4	"
D57A	6.7	5.3	5.4	5.2	--
P253G	3.7	8.1	4.2	7.4	--
E70A	-	-	-	-	Zn co-ordination
L95R	3.4	8.1	4.6	9.7	--
V192G	5.6	6.4	37.1	2.9	Nucleotide binding
S211I	3.1	11.5	11.2	6.7	Co-substrate specificity
S212D	2.5	24.4	13.9	6.5	"
S213N	2.6	37.5	14.7	7.9	"
K216W	9.2	7.1	49.2	3.4	Ribose binding
G275F	3.7	8.9	5.1	9.6	--
T167A	39.4	3.1	14.5	7.9	Substrate binding
S298A	46.9	5.7	9.1	11.1	"
I300G	61.1	1.9	4.9	9.4	"

Table 3.5: *List of LlCAD2 mutants generated. [1]Predicted from LlCAD2 model, crystal structures of AtCAD5 and HLADH*

3.10 Alterations in enzyme specificity by mutations in Ser rich motif

The ability to differentiate between NADH and NADPH involves a very limited number of residues due to high structural similarity between the two co-factors. It is already reported that a serine rich motif (^{211}SSS^{213}) was observed to interact with the phosphate group of NADP in AtCAD5 [49]. This region is responsible for co-substrate specificity of the enzyme, differentiating between NADH and NADPH [95] as mentioned in 3A. Also, Ser210 mutant in ScAdh6p as well as Ser213 mutant in *E.gunnii* CAD2 established their roles in defining co-factor specificity [27, 79]. In order to better understand their role, single mutants (S211I, S212D, S213N), as well as double (S211I:212D, S212D:213N) and triple mutant (S211I:212D:213N) were generated. This substitution of ^{211}SSS^{213} to ^{211}IDN^{213} was considered as the latter sequence is highly conserved amongst several NAD dependent ADHs [97].

Mutations in this serine rich region resulted in lower binding of NADPH. The wild-type LlCAD2 accepted only NADPH as a co-substrate with a catalytic efficiency of 11.5 $\mu M^{-1}s^{-1}$, while, as already mentioned in Chapter 2A, negligible catalysis in presence of NADH was achieved even with concentrations as high as 2 mM (Table 3.6). It was observed that as we move from single mutants towards the triple mutant, the catalytic efficiency for NADPH decreased while that of NADH gradually increased. Though, single mutants only produced a small change in the catalytic efficiency, the triple mutant exhibited higher catalytic efficiency and lower K_m for NADH in comparison to NADPH, thus showing a complete reversal of co-substrate specificity of the enzyme. In the triple mutant, the extra phosphate group of NAPDH may pose steric and electrostatic hindrances during

binding, while NADH can easily be accommodated. The importance of cooperative effects between two or more adjacent residues can be seen from these results, where double and triple mutations caused greater changes on co-factor kinetics, than individual mutations. Analogous regions in a glucose dehydrogenase from *H. mediaterranei* (^{207}GRR^{209}), as well as, in alcohol dehydrogenase ADH8 from *R. perezi* (^{223}GTH^{225}) were also reported to play similar functions in respective enzymes [89, 98]. A reverse effect with NAD-dependent xylitol dehydrogenase from *S.stipitis* was observed by mutations in ^{207}DIF^{209} leading to increased efficiency for $NADP^+$ [99].

Enzyme	NADPH		NADH		Coniferaldehyde		Sinapaldehyde	
	K_m	K_{cat}/K_m	K_m	K_{cat}/K_m	K_m	K_{cat}/K_m	K_m	K_{cat}/K_m
Wild type	2.5	11.5	-	-	3.7	6.4	2.8	11.6
S211I	11.2	6.7	29.8	0.9	5.4	3.0	3.1	11.5
S212D	13.9	6.5	25.4	1.1	6.7	0.97	2.5	24.4
S213N	14.7	7.9	31.5	2.6	9.5	1.3	2.6	37.5
S211I-212D	24.7	4.8	19.1	5.1	6.1	2.5	2.5	43.0
S212D-213N	21.3	5.3	14.9	5.4	8.9	0.77	1.9	50.7
S211I-212D-213N	39.9	2.4	5.1	8.9	8.7	1.5	2.0	49.3

Table 3.6: *List of LlCAD2 mutants generated for the region ^{211}SSS^{213}*

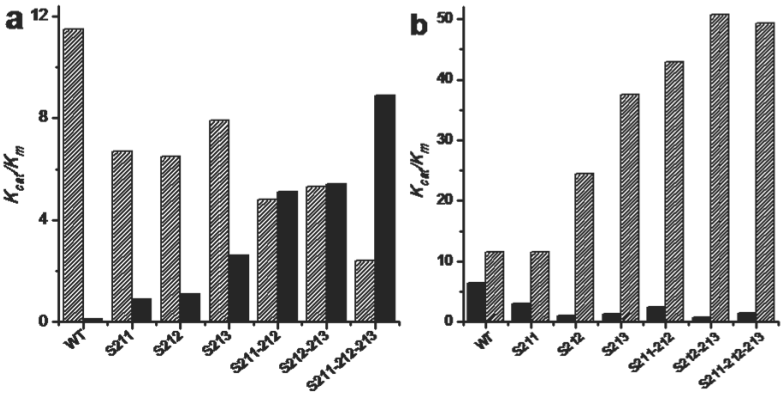


Fig. 3.6: *Graphs showing relative catalytic efficiencies of ^{211}SSS^{213} mutants for* ***(a)*** *Co-substrates,* ***(b)*** *Substrates. K_{cat}/K_m for NADPH (▒), NADH (■), Sinapaldehyde (▒), Coniferaldehyde (■)*

Apart from its role in defining co-substrate specificity, another observation was also made upon assessing the catalytic efficiencies for different substrates. Wild-type LlCAD2 exhibited almost double catalytic efficiency for sinapaldehyde when compared with coniferaldehyde (Chapter 2B). The single mutants S211-213 showed a gradual increase in the catalytic efficiency for sinapaldehyde with S213 showing a 3-fold increase, while that of coniferaldehyde decreased further. The double and triple mutants showed around 4 fold increase in catalytic efficiency for sinapaldehyde along with further decrease in K_m (Fig. 3.6). While, that of coniferaldehyde decreased by more than 75% along with an increase in K_m. These findings suggest that the Ser-rich region not only confers co-substrate

specificity to the enzyme, it also plays a determining role in governing the specificity of different hydroxycinnamyl substrates. The effects of these mutations were not localized totally to NADP binding site only, providing evidence of a physical transmission of structural information between NADP and substrate binding regions of LlCAD2. Similar effects have been reported for *T. flavus* malate dehydrogenase as well as EgCAD2 from from *E. gunni* [79, 100].

3.11 Catalytic mechanism of LlCAD2

Based on the docking studies detailed in Chapter 3A, and mutational data presented here, a putative active-site is presented in Fig. 3.7. Like many Zn-dependent ADHs, LlCAD2 catalyzed reaction is proposed to use a proton shuttle for donation of protons to or abstraction from the substrate [101]. Specifically, in the reductive direction, the dihydronicotinamide C4 hydride attacks the C1 carbonyl carbon of aldehyde, resulting in highly polarized carbonyl oxygen that is stabilized by coordination to the catalytic Zn^{2+}. As proposed in Chapter 3A, the 4 residues probably involved in proton relay Ser49, His52, Asp57 and Pro253 were mutated. It was observed that the mutant S49A was completely inactive, while H52D showed drastic changes in its kinetic parameters; K_m increased by 12-13 folds for both sinapaldehyde and NADPH, while K_{cat}/K_m decreased by 10 fold. While no alteration in the kinetic parameters was observed for D57A and P253G. In the CAD model generated, the –OH group of Ser49 was observed within hydrogen bonding distance to the aldehydic oxygen of the substrate, as well as nicotinamide ring. His52 was also found to be in close proximity to nicotinamide ring. Also, an extended hydrogen bonded system was observed in HLADH between NADPH, substrate, Ser48 and His51, which was not present in the apo-form [102]. These two positions are highly conserved not only in CADs, but also in several other ADHs. As observed in other enzymes, it is proposed that Ser49 shuttles its proton to the alkoxide intermediate which is stabilized by Zn^{2+} which forms the main proton shuttle mechanism along with His52. This proton relay necessitates that the nicotinamide ribose 2' and 3' hydroxyl moieties form hydrogen bonds with both Ser49 and His52, respectively, and with one another. Also, His51 mutant was studied in human HuLADH and shown to facilitate release of $NADP^+$ during catalysis [103]. These results indicate that Ser49 and His52 play a critical role in proton relay.

The overall mechanism of LlCAD2 (Fig 3.8) can thus be assumed similar to that of HLADH. As reported already, co-factor binding brings about a large conformational change in ADHs and goes from apoenzyme to holoenzyme structure [104, 105] which is a rate-limiting step for LADH and human liver ADH [106, 107]. The proton relay pathway includes –OH group of Ser48 and imidazole ring of His51 (equivalent to Ser49, His52 of LlCAD2) and nicotinamide ribose of NADPH [108]. Hydride transfer in HLADH occurs with hydrogen tunneling effect, which is facilitated by Val192

by ensuring a close contact between the reactive C4 of NADPH and C1 of substrate [109]. However, a key difference between the mechanism of AtCAD5 and HLADH was reported recently [110]. Here it was proposed that AtCAD5 does not involve His52 in the proton relay during catalysis which is in sharp contrast to the mechanism of HLADH. Though we observed involvement of His52 in the catalytic events of LlCAD2, its absolute role in proton relay needs to be further investigated.

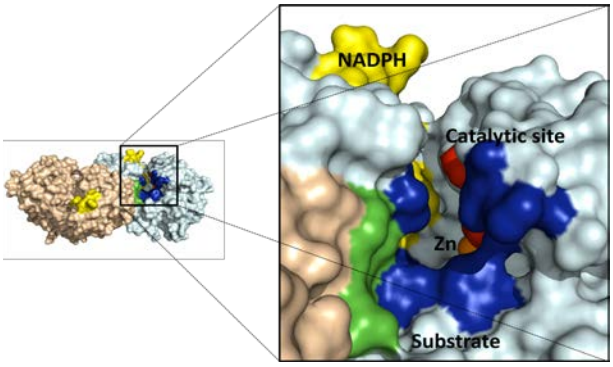


Fig. 3.7: *Putative active-site of LlCAD2. NADPH binding pocket (yellow); Substrate binding pocket (blue); Substrate binding residues from other subunit (green); Catalytic center Ser49 & His52 (Red), Catalytic Zn^{2+} (orange)*

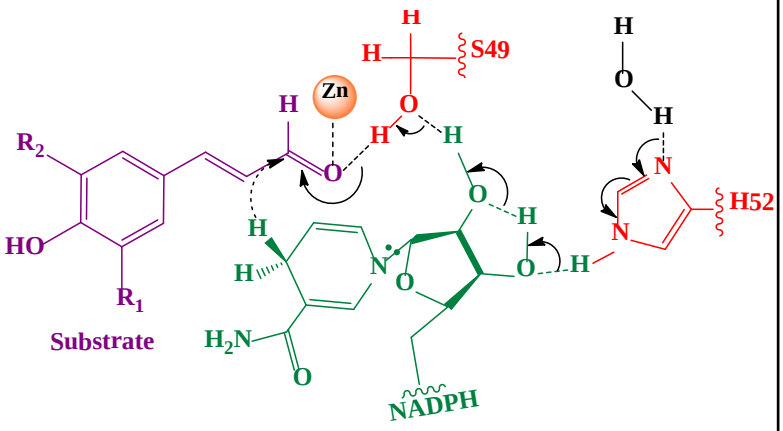


Fig. 3.8: *Proposed molecular mechanism for LlCAD2, showing the substrate (purple), NADPH (green), catalytic Zn (orange) and residues involved in proton relay (red). The proposed proton relay path is depicted in solid black arrows, while hydride transfer is shown in black dashed arrow*

4

CAD: Structural Transitions

4A

Structural investigations into folding dynamics

A protein structure assumes its functional conformation via a folding pathway, which may pass through several intermediate states undergoing a hydrophobic collapse, with many of the native contacts yet to form. Such intermediates were detected using techniques like Circular Dichroism and Fluorescence spectroscopy. This section encompasses the role of pH and temperature on forming such structural transition states of LlCAD2.

4.1 LlCAD2 is inactivated above 60 ºC

LlCAD2 showed maximum activity at 30-40 ºC with rapid loss of activity at higher temperatures, and was completely inactivated at 80 ºC (Fig 4.1a). Catalytic stability was next measured by incubating LlCAD2 at different temperatures and residual activity was measured after appropriate time intervals. As shown in Fig. 4.1a, the enzyme was quite stable upto 40 ºC, with more than 75% activity after 2 h, however, was rapidly inactivated above 60 ºC with complete loss of activity at 65 ºC within 10 min. Stability and thermodynamic parameters like deactivation rate constant (K_d^*), half-life ($t_{1/2}$) and free energy change (ΔG^*) for inactivation were then determined using the stability data. At 20 ºC, the half-life was almost 18 h, which decreased to 2.5 h at 40 ºC, and around 7 min at 60 ºC, while only 7 KJ mol^{-1} increase in ΔG^* was noticed. The thermostability profile can be classified in three phases which may reflect the different transition states of LlCAD2: 1st from 25 to 40ºC, where the enzyme is most active with marginal increase in ΔG^*, however, the half-life drastically fell below 6 h. The 2nd phase upto 60ºC corresponds with 75% loss in enzyme activity along with slight increase in K_d^* & ΔG^*. While in the 3rd phase, there was a 10-fold increase in K_d^* and complete loss of activity. Renaturation or cooling of the heated protein samples to 20 ^{o}C showed that the enzyme was fairly reversible upto 60 ^{o}C, however, higher temperatures lead to irreversible thermal inactivation.

Further to the deactivation study, an investigation of thermodynamic parameters was necessary to understand the behavior of the enzyme. The change in enthalpy (ΔH^*), entropy (ΔS^*) and activation energy (E_a^*) for inactivation of LlCAD2 (calculated from Arrhenius plots; Fig. 4.1c) were 60.28 KJ mol^{-1}, -284.47 J mol^{-1} and 58.39 KJ mol^{-1}respectively, while free energy change (ΔG^*) for inactivation at 30 & 60 ºC were 127.96 & 133.63 kJ mol^{-1}, respectively. The highly negative values of ΔS^* reflects an ordered transition state of LlCAD2, which is due to compaction of enzyme molecules, moreover, such changes could arise from formation of charged particles, associated gains and ordering of solvent molecules [111]. While, the high ΔG^* corresponding to the low ΔH^* suggests

LlCAD2 to be marginally thermo-stable. ΔH^* is well within the range of 20-150 kJ mol^{-1} thus indicating LlCAD2 to be maintaining its rate of reaction even at different temperatures. ΔG^* increased with increasing temperature, however, was comparatively smaller than reported for other enzymes [112]. These parameters may be affected by solvent effect due to presence of surrounding water molecules, as well as structural effect due to conformational changes occurring in the enzyme during reactions.

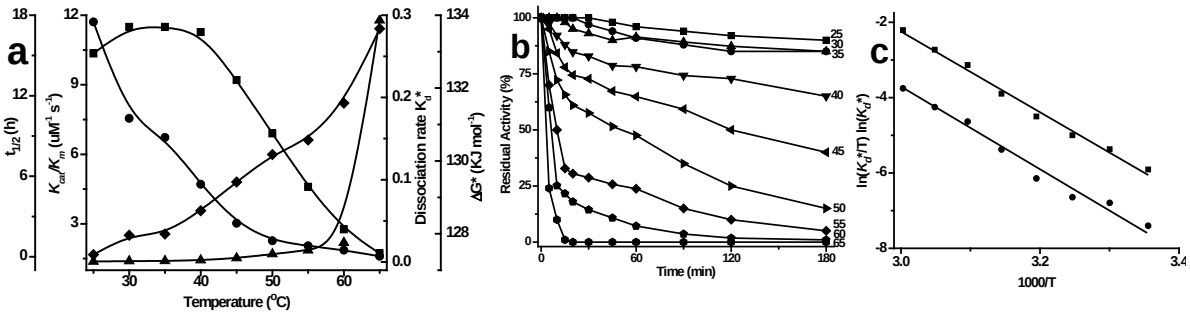


Fig. 4.1: (a) *Stability and thermodynamic parameters for thermal inactivation of LlCAD2. (■) Catalytic efficiency, (▲) Dissociation rate constant K_d^*, (●) Half-life $t_{1/2}$, (♦) Free energy change for deactivation ΔG^*;* ***(b)*** *Effect of temperature on stability of LlCAD2. Enzyme was incubated at 25-65 °C in 50 mM Tris buffer; pH 8.0, and assayed at 30 °C under standard conditions;* ***(c)*** *&* ***(d)*** *Arrhenius plots for thermal inactivation of LlCAD2. (■) For calculating enthalpy ΔH* and entropy ΔS* of inactivation, (●) For calculating activation energy Ea* of inactivation*

4.2 Thermal unfolding takes place above 70 °C

LlCAD2 was subjected to a range of temperature and structural transitions were monitored by biophysical techniques to understand its folding thermodynamics. The native intrinsic fluorescence maximum (λ_{max}) of LlCAD2 was at 351 nm indicating tryptophan residues to be exposed to polar environment (Fig. 4.2a). The observed λ_{max} showed a slight transition between 40-60 °C which remained in the range of 351-353 nm upto 90 °C indicating minor changes in tryptophan environment (Fig. 4.2b). However, the emission intensity decreased with increase in temperature; which could be due to inactivation of the singlet fluorophore, though there is no general response of intrinsic Trp fluorescence intensity to unfolding of a polypeptide chain [113]. No major changes in the light scattering intensity of the protein were observed till 60 °C (Fig. 4.2b). The protein showed an increase in light scattering intensity from 70 °C which drastically increased further at higher temperatures, showing thermal aggregation. The dye, ANS, has been shown to bind the hydrophobic regions of partially unfolded proteins that become exposed to the solvent, however, fully denatured protein do not exhibit such binding [114]. No binding of ANS dye throughout the experiment indicated a quick unfolding or rearrangement of the protein molecule. Phase diagram and Parameter A analysis revealed presence of a folding intermediate in the vicinity of 70 °C (Fig. 4.2b, c). Two

linear sections were observed in the phase diagram intersecting at 70 °C. The T_m calculated for native protein was 81.5 °C.

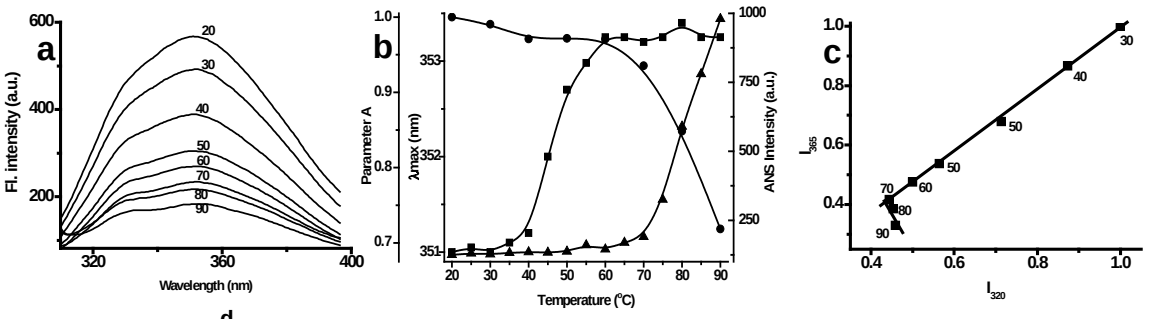
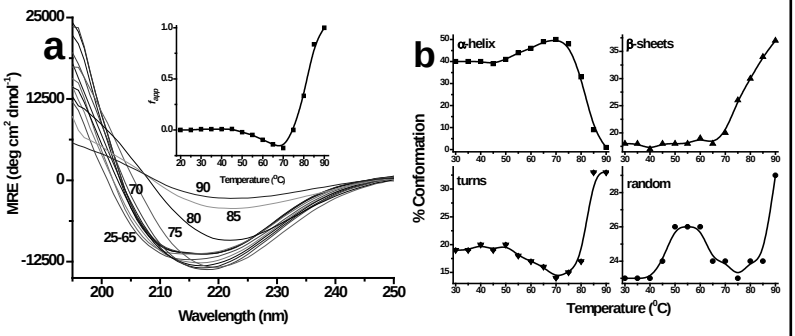


Fig. 4.2: (a) *Intrinsic fluorescence of LlCAD2 after 10 min incubation at different temperatures in 50 mM Tris; pH 8.0;* ***(b)*** *Summary of fluorescence changes; (■) Avg. λ_{max}, (●) Parameter A, (▲) Scattering intensity;* ***(c)*** *Phase Diagram*

Native far-UV CD spectra of LlCAD2 showed a trough between 208-219 nm, and maxima at 195 nm indicating a predominant α-helix in the structure with a lesser amount of β-sheets (Fig. 4.3a). The data was analyzed for secondary structural elements using CDSSTR program. The composition of secondary structure elements as calculated from the far-UV CD spectra are: 41% α-helices, 17% β-sheets, 19% turns and 23% unordered structures. As an indicator of the overall structural change of LlCAD2, changes in content of the secondary structural elements of the protein, as well as fractional unfolding (f_{app}) was employed. Analysis of the spectra showed no particular change in structure upto 40 °C. There was a modest increase in MRE_{219} between 45-70 °C resulting in, unlike the native protein, a single minimum at 219 nm instead of a trough. The f_{app} slightly decreased at 70 °C, and substantial loss of overall structure was observed above 75 °C (Fig. 4.3a Inset). While the enzyme activity was completely lost at temperatures around 60 °C, structural collapse occurred only above 70 °C; indicating thermal inactivation and disruption of the active site of enzyme before any significant conformational changes take place.

Fig. 4.3: (a) *Far-UV CD spectra of LlCAD2 after 5 min incubation at different temperatures in 20 mM Tris buffer; pH 8.0,* ***(inset)*** *Unfolded fraction f_{app} calculated from MRE_{208},* ***(b)*** *Comparison of calculated secondary structures in LlCAD2 from far-UV CD spectroscopy as a function of temperature*



4.3 LlCAD2 maintains quaternary structure at higher temperatures

The thermal inactivation observed is not a result of dissociation of the dimer. DLS analysis of LlCAD2 incubated at 20 ºC and 60 ºC showed consistent dimeric forms, and showed formation of aggregated protein at 90 ºC (data not shown). This result is in accordance to the findings of yeast alcohol dehydrogenase where the dissociation of the tetrameric enzyme was not rate-limiting during thermal inactivation [115], however, a recent report regarding horse liver alcohol dehydrogenase contradicts these results, wherein, the dimeric HLADH dissociated at 46 ºC leading to aggregation of the monomers at 49 ºC and subsequently denaturation at higher temperatures [116]. Also, metal content of temperature denatured LlCAD2 was measured following Zincon dye binding assay. It was observed that the zinc ion content of the enzyme did not change as a function of temperature. This observation indicates that changes in the conformational state of the active site required for catalysis is responsible for the above mentioned thermal inactivation, rather than dissociation of the catalytic zinc ion. Such an observation was also made in yeast alcohol dehydrogenase [115].

4.4 Temperature induced intermediate state of LlCAD2

The present results show that much lower temperatures are required to bring about inactivation than are required to induce any conformational change in the enzyme. Such thermal inactivation of enzyme taking place before any structural changes has been reported in a bile-salt hydrolase [117]. As shown in Fig. 4.3a there is distinct change in the CD signal at 20 ^{o}C and 70 ^{o}C, and complete loss of structure at 90^{o}C. Upon comparison of different secondary structural elements (Fig. 4.3b), it is quite clear that unordered structures increased very vaguely, while β-sheets and turns increased by 15-20% at the expense of α-helix, which considerably declined to less than 1%. Based on the above mentioned fluorescence and CD data, it can be proposed that LlCAD2 forms an intermediate state at 60-70 ^{o}C, in order to explain the thermal inactivation and unfolding observed around these temperatures. This assumption mainly lies on the fact that a considerable fraction of secondary structure still exists at these temperatures, except α-helix, as well as its fairly reversible nature (data not shown) and intact quaternary structure. The overall structural changes due to temperature are summarized in Fig. 4.4.

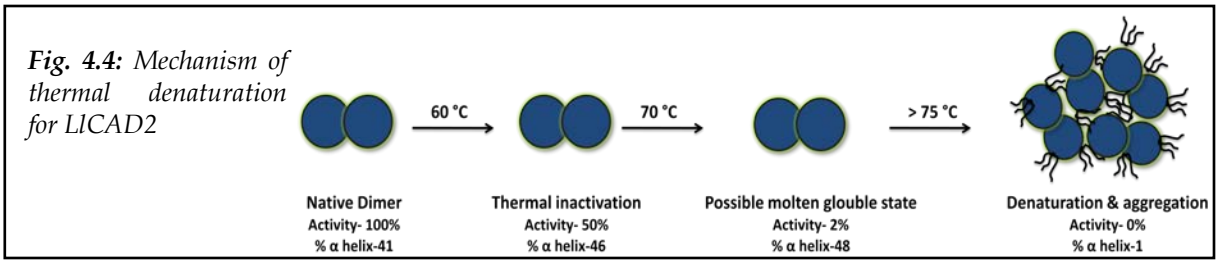


Fig. 4.4: *Mechanism of thermal denaturation for LlCAD2*

4.5 LlCAD2 is most stable at pH 7-9

pH is one of the main factors affecting tertiary and quaternary structures of proteins, and so, effects of pH were also measured on LlCAD2 activity and stability. Optimum pH for reduction and oxidation reaction were 6.5 and 9 respectively (Fig. 4.5a), though reduction optima was virtually double than corresponding oxidation optima. The enzyme was found to be most stable in the pH range of 6-9 with rapid loss outside this window (Fig. 4.5b). At pH 4, less than 25% enzyme was still active, a little above 10% remained at pH 10. Like thermostability profile, the pH stability profile can also be divided into 3 phases. The most stable being between pH 7-9, with $t_{1/2}$ ranging from 15-17 h, least K_d^* of 6.7 x 10^{-4} and highest ΔG^* of 127 kJmol^{-1}. The second phase comprised of pH range 5-7 and 9-10. There is a drop of almost 10 h in the half-lives on both ends and mild decrease in ΔG^* with nearly no change in K_d^*. The third phase below pH 5 and above 10 showed marked increase in K_d^*; 1.1 on the acidic end and 0.6 towards the basic side. A drop in ΔG^* values; 16 kJmol^{-1} and $t_{1/2}$ under1 min was observed. These results show no direct correlation between the optimal activity and stability in terms of pH of the buffer system.

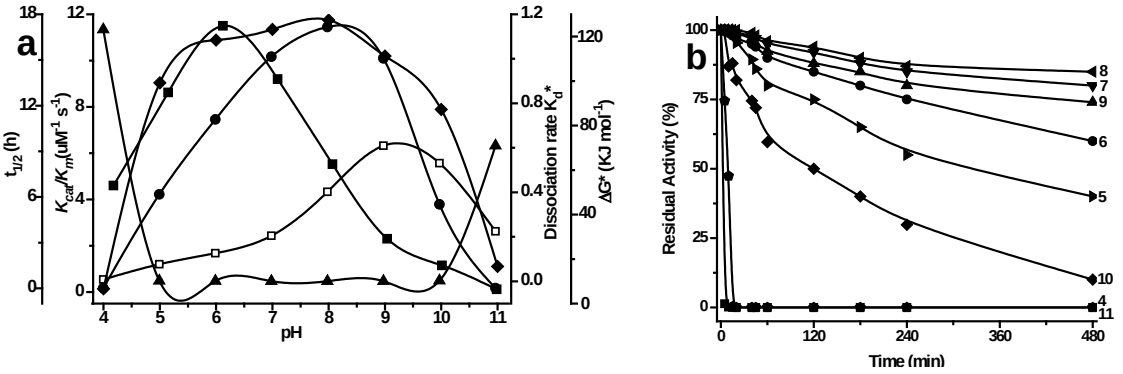


Fig. 4.5: (a) *Effect of pH on initial rate and stability of LlCAD2 for reduction of aldehyde (■), oxidation of alcohol (□). (▲) Dissociation rate constant K_d^*, (●) Half-life $t_{1/2}$, (♦) Free energy change for deactivation ΔG^*;* ***(b)*** *Effect of pH on stability of LlCAD2. Enzyme was incubated in appropriate buffers at different pH as mentioned in experimental section and assayed at 30 °C under standard conditions.*

4.6 Structural changes induced by pH

Steady state fluorescence spectra of LlCAD2 showed maximum fluorescence intensities in the pH range 6-10 (Fig. 4.6a). Fluorescence intensities were significantly reduced for the protein incubated at pH 2-5 and pH 11-12 than at pH 8 due to acid quenching and neutralization of COO- groups on amino acids in the vicinity of Trp residues [118]. Partial unfolding of the protein was observed as λ_{max} red shifted to 353 nm on the acidic side, while no particular change was observed till pH 11 on the alkaline end, however, the λ_{max} increased to 365 nm at pH 12 (Fig. 4.6b).

Effect of pH on ANS fluorescence intensity was also carried out next. Protein incubated at acidic pH showed significant binding of the ANS (highest at pH 2), showing a blue shift in λ_{max} from 520 nm to below 480 nm and more than 10 fold increase in the fluorescence intensity (Fig. 4.6b), signifying unfolding of the protein molecule, and subsequent surface exposure of the hydrophobic core. It was observed that the LlCAD2 conformations that exhibit strong binding to ANS seem to contain non-native α-rich secondary structures and lost tertiary structure (section 4.7), and thus, show properties of a molten globule state [119]. A blue shift in the emission fluorescence peak of ANS reflects increased hydrophobicity of the medium in which ANS molecules are located. One can actually monitor the hydrophobicity of the protein's binding sites by measuring the emission peak positions of the bound ANS molecules [119]. Thus, according to these results, the optimum hydrophobicity of the LlCAD2-ANS binding sites is possessed by the protein intermediate states occurring when the protein undergoes an acid-induced conformational transition. Scattering was not observed towards acidic pH range (Fig. 4.6b) due to rapid resolubilization of any aggregates formed. While marginal increase in scattering intensities was observed at higher pH values indicating unfolding of the protein molecule.

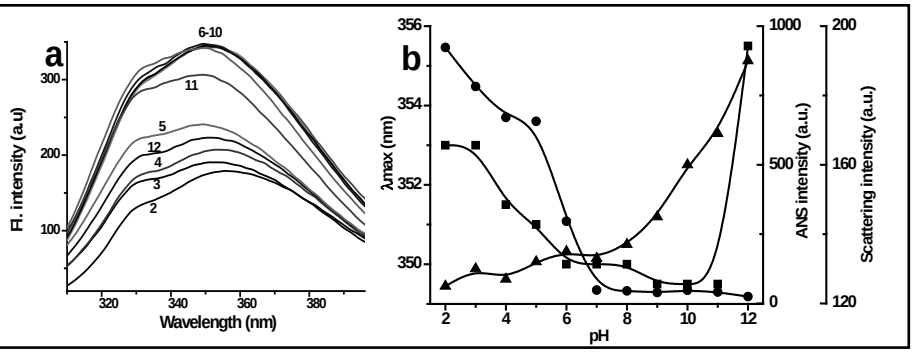


Fig. 4.6: ***(a)*** *Intrinsic fluorescence of LlCAD2 after 3 h incubation at different pH;* ***(b)*** *Weighted Avg. λ_{max}, (■), ANS fluorescence (●),*

The change in secondary structure of LlCAD2 as a function of pH was followed by measuring far-UV CD spectra as described for temperature. No major changes in overall structure were observed between pH 6-10 (Fig. 4.7a). At alkaline pH, the CD spectra showed certain changes; insignificant loss of MRE_{219}, with slight shift in *MRE* below 208 nm. As shown in Fig. 4.7b, there was gradual loss of almost 10% α-helix above pH 9 with simultaneous increase in 8% β-sheet content above pH 10. The protein almost completely lost its helical conformation and adopted a β-sheet structure at higher pH, as computed with CDPro. Conversely, acidic pH below 5 showed major shift in the *MRE* below 208 nm, and a distinctive minima at MRE_{204} was observed, suggesting substantial rearrangement of secondary structure. The α-helix content reduced sharply between pH 4-7, while formation of some non-native α-helix was observed when pH was further lowered, while, the pattern of β-sheets and random structures complemented that of α-helix remarkably. These findings further

suggest definite conformational changes in secondary structure elements. The f_{app} showed major changes in the unfolded fraction with respect to pH where both acidic and alkaline ends showed significantly compact structures when compared to native protein at pH 8.

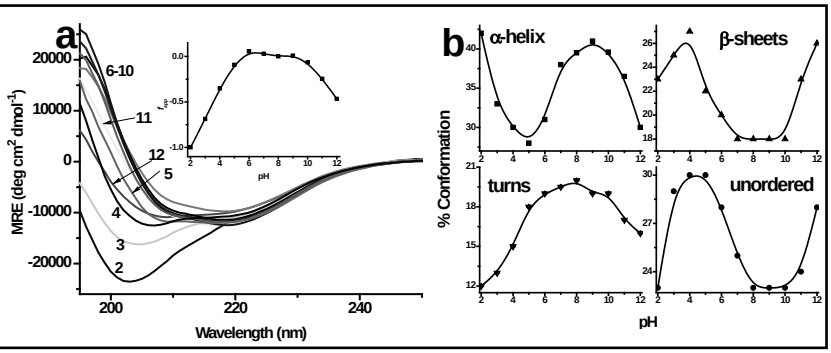


Fig. 4.7: ***(a)*** *Far-UV CD spectra of LlCAD2 at different pH,* ***(inset)*** f_{app} *calculated Avg. MRE;* ***(b)*** *Comparison of calculated secondary structures in LlCAD2 from far-UV CD spectroscopy as a function of pH*

4.7 Acid induced molten globule at pH 2

Studies on conformational transitions of numerous proteins at different conditions like high temperatures, denaturant solutions, acidic pH, etc., have shown that these proteins may undergo transitions through different intermediate states depending on the protein architecture and on the agent inducing a conformational transition. The above findings (a compact rearranged secondary structure, major distortion in the tertiary structure, exposure of hydrophobic patches on surface of protein) suggest existence of an altered conformational state of LlCAD2 at pH 2, with distinct structure than the native protein, which can be considered as an acid induced molten globule-like intermediate state. This molten-globule like structure was further characterized by thermal and chemical denaturation:

4.7.1 Reversibility of molten globule

As shown already, LlCAD2 incubated at pH 2 showed no enzyme activity (Chapter 2A), however, a remarkable feature was observed when the pH was readjusted back to 8, which led to reconstitution of around 25% of the activity. The near UV-CD spectrum of the native enzyme showed two minute positive ellipticity peaks, one at 260 nm and other at 300-320 nm, and a

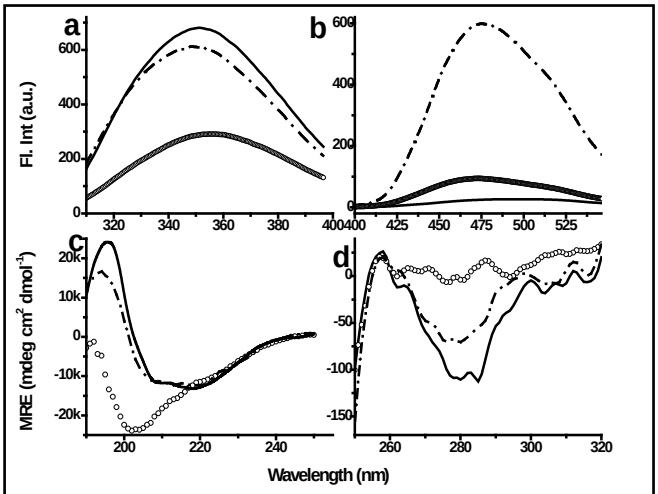


Fig. 4.8: *Structural properties of intermediate at pH 2.* ***(a)*** *Intrinsic fluorescence;* ***(b)*** *ANS binding;* ***(c)*** *Far-UV CD;* ***(d)*** *Near-UV CD for Native (——); unfolded at pH 2 (oooo); and refolded protein (—•—)*

prominent negative ellipticity trough around 277-285 nm indicating ordered tertiary structure of LlCAD2 (Fig. 4.8d). At pH 2 significant decrease in overall ellipticity was observed suggesting major changes in the environment of aromatic amino acids and cysteine residues (Fig. 4.8d). This resulted in loss of the ordered tertiary structure indicating change in the environment of aromatic amino acids, during which the hydrophobic patches must be getting exposed on the surface of the protein [120]. The refolded protein showed near complete refolding of secondary structure as well as major reorganization of tertiary structure with reduced ellipticity signals (Fig. 4.8a, c, d). The composition of the renatured structure was estimated to be: α-helix-44%; β-sheets-19%; turns-16% and unordered structures-21%. ANS binding showed that the exposure of hydrophobic patches in the molten globule has been extensively reversed on renaturation (Fig. 4.8b). Thus, it is evident that the acid induced molten globule was reasonably reversible in nature.

4.7.2 Effect of temperature

It is already mentioned that the native protein shows a modest transition phase between 40-60 ºC marked by an increase in λ_{max} (Fig. 4.2b). On the other hand, intrinsic fluorescence of acid induced molten globule did not show any such transition phase; the λ_{max} and fluorescence intensities showed a gradual change (Fig. 4.9a, b). ANS binding patterns were comparable for both; no significant binding of the dye was observed throughout the temperature range (Fig. 4.9c). Far-UV CD spectra of native protein suggested an intermediate state around 70 ºC and thermally induced denaturation at higher temperatures. The acid induced molten globule on the other hand showed no loss of structure even at temperatures as high as 80 ºC, with insignificant unfolding at 90 ºC as observed from f_{app} (Fig. 4.9d, e). Scattering intensities unlike native protein did not increase with increasing temperatures indicating no significant aggregation of the intermediate state (Fig. 4.9f).

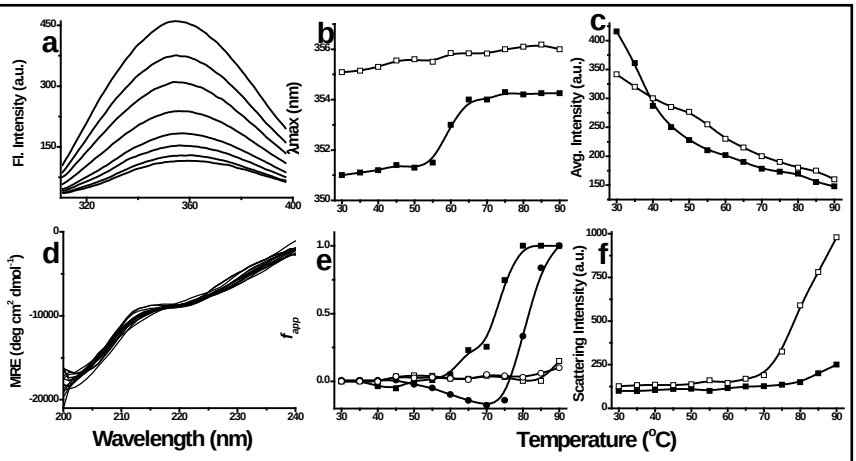


Fig. 4.9: *Effect of temp. on intermediate state at pH 2.* ***(a)*** *Intrinsic fluorescence spectra;* ***(b)*** *Weighted avg. λ_{max};* ***(c)*** *Avg. ANS fluorescence intensity;* ***(d)*** *far-UV CD spectra,* ***(e)*** *f_{app} calculated;* ***(f)*** *Scattering intensities.* ***(■)*** *native LlCAD2; (□) acid-induced molten globule-like state*

Even at higher temperatures, there is significant retention of the non-native α-helical content and overall secondary structure compared to the native protein, signifying the thermo-stable nature of the acid-induced molten globule.

4.7.3 Effect of GnHCl

Both, native protein and the acid induced molten globule were subjected to chemical denaturation in presence of GnHCl. Intrinsic fluorescence of native protein showed a gradual transition in the fluorescence intensity and a red-shift in λ_{max} (Fig. 4.10a, b). Though the molten globule was a comparatively unfolded structure, it followed a more complex path; with an initial blue-shift observed till 0.5 M GnHCl which remained unchanged until 2.0 M concentration, and again red-shifted at higher concentrations. The fluorescence intensities also increased with increasing GnHCl concentrations, rather than decreasing, as observed for the native protein (Fig. 4.10c). Also, ANS binding of the native protein showed dye binding in the vicinity of 1 M GnHCl, while the molten globule showed no such binding (Fig. 4.10f). The far-UV CD spectra showed a gradual change in the secondary structure of the molten globule (Fig. 4.10d). The f_{app} showed a brief collapse in the structure of native LlCAD2 at 1.0 M GnHCl and complete denaturation was achieved above 4.0 M (Fig. 4.10e). The molten globule showed no such local intermediates and followed a two-state unfolding mechanism. The structural transitions observed in the native protein due to GnHCl are discussed in more detail in Chapter 4B.

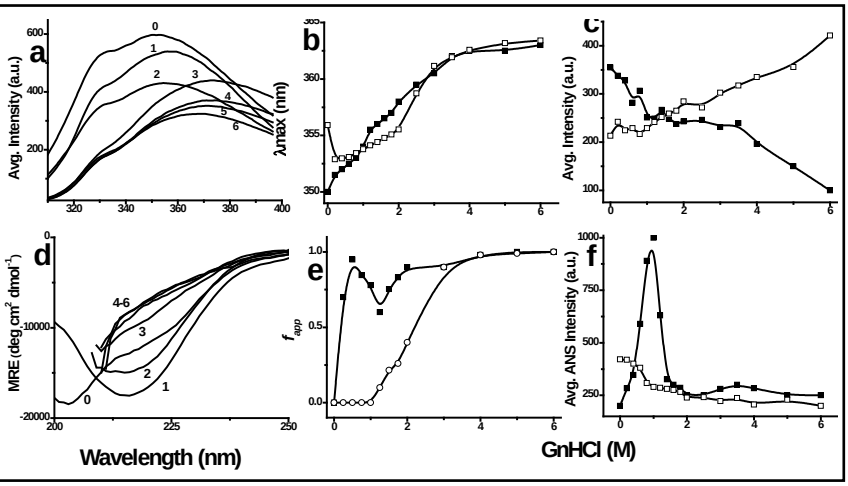


Fig. 4.10: *Effect of GnHCl on intermediate state at pH 2.* ***(a)*** *Intrinsic fluorescence spectra;* ***(b)*** *Weighted avg. λ_{max};* ***(c)*** *Avg. intensity;* ***(d)*** *far-UV CD spectra;* ***(e)*** *f_{app} calculated from MRE_{avg};* ***(f)*** *Avg. ANS fluorescence intensity.* ***(■)*** *native LlCAD2;* ***(□)*** *acid-induced molten globule-like state*

4.7.4 Metal content & molecular size of the molten globule

Zn ions bound to the protein were detected and quantified by using Zincon dye as already mentioned. It was observed that pH changes had no effect on the metal content of the protein, and the native as well as acid induced molten globule contained fairly the same amount of Zn ions (1.95 ± 0.15 mol

Zn ions per mol of the molten-globule monomer), showing that the molten globule intermediate state still retains the Zn ions. Hydrodynamic radii were next measured using DLS and it was observed that the acid induced molten globule showed a hydrodynamic radius of around 10 nm. This value is less than that of the native dimer (13 nm), while it is still higher than that of single monomeric unit (7 nm), suggesting that the acid induced molten globule could be an expanded conformation of LlCAD2 monomer. The molecular size of the molten globule was also measured using gel filtration chromatography in order to understand the dissociation of LlCAD2 as mentioned in Section 2.2.2. Native dimer eluted at 11.9 ml elution volume in 50 mM Tris; pH 8.0 while the retention time of native monomer was 17.6 ml. Acid induced molten globule state eluted at a retention volume of 14.8 ml in Glycine HCl; pH 2.0. upon refolding, the retention time was again shifted to 12.1 ml (Fig. 4.11). It can be deduced from SEC and DLS results that the acid induced molten globule state is, in fact a slightly expanded molecular state of the intermediate resulting from a dissociated monomer of LlCAD2. The formation of acid-induced molten globule intermediate of LlCAD2 is summarized as shown in Fig. 4.12.

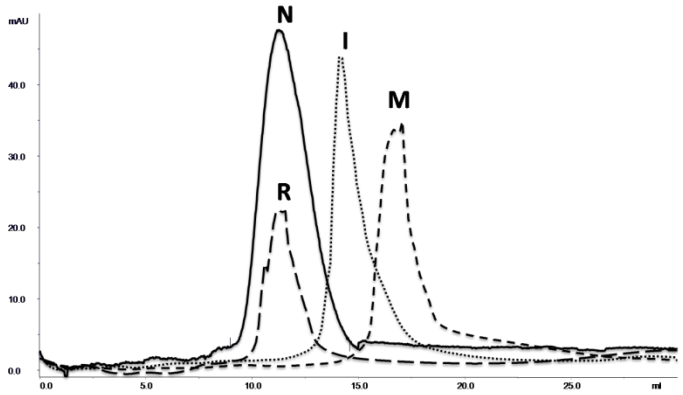


Fig. 4.11: *Sephacryl S-100 chromatogram of Native LlCAd2 dimer (N); Monomer (M); Acid-induced intermediate (I); Refolded intermediate (R)*

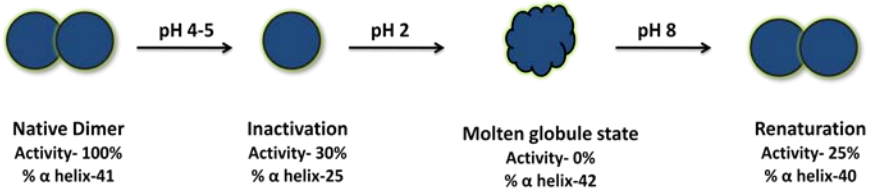


Fig. 4.12: *Mechanism of formation of acid induced molten globule of LlCAD2*

4B

Effects of metals and chemical agents on CAD

Apart from structural transitions induced by pH and temperature, the folding mechanism of the LlCAD2 was further investigated using chaotropic agents like Urea and GnHCl which are known to induce major structural disruptions in proteins. Also, the role of Zn^{2+} in catalysis as well as in overall structure & stability is explored.

4.8 LlCAD2 is more sensitive towards GnHCl than urea

Chaotropic agents like GnHCl and urea were employed to induce unfolding of LlCAD2. The protein was found to be very sensitive towards GnHCl with almost 50% loss of activity at 100 mM within 10 min and complete loss of activity at concentration above 1 M (Fig. 4.13a, b). Such sensitivity can be attributed to the modification of a carboxylate group at the active site, which was also revealed by chemical modification experiments (Chapter 3B), whereas urea was comparatively milder than GnHCl (Fig. 4.13a, c). The enzyme retained 75% activity at 1 M and 20% at 2 M, while complete inactivation took place above 4 M. Effects of both GnHCl and urea on activity of LlCAD2 are summarized in Fig. 4.13d, e. It was observed that K_d^* increased more drastically for GnHCl reaching 0.4 at 0.8 M concentrations, when compared to urea (0.03 at 0.8 M). K_d^* for urea inactivation reached a maximum of 0.1 at 2.5 M concentration. Also, $t_{1/2}$ for both agents showed similar pattern. Change in free energy of LlCAD2 inactivation measured at 30 ºC increased with increase in concentrations of both the agents. GnHCl showed a maximum increase in ΔG^* of 12 KJ mol^{-1} at 0.8 M concentration, while Urea showed an increase of 8 KJ mol^{-1} at 2.5 M. Renaturation was attempted by diluting the protein samples with native buffer but the enzyme did not reactivate even by allowing longer incubations.

4.9 Molten globule state induced by 1 M GnHCl

In order to study the effect of GnHCl on LlCAD2, fluorescence and CD measurements were performed during both unfolding as well as refolding of the protein. As shown in Fig. 4.14a, b, intrinsic fluorescence spectra of LlCAD2 treated with 0-6 M GnHCl showed a gradual increase till 2.5 M, indicating partial unfolding of the protein. While, a decrease in intensity as well as a red-shift in λ_{max} to 363 nm was observed at higher concentrations, indicating total unfolding of the protein above 3 M GnHCl (Fig. 4.14b). Scattering intensity increased with increase in GnHCl concentrations

upto 0.75 M and then decreased. Also, refolding showed similar scattering maxima at 1 M GnHCl (Fig. 4.14c). Parameter A and phase diagrams are powerful tools to analyze structural transitions in protein molecules as well to identify folding intermediates. Parameter A showed a structural transition between 0.5-3 M GnHCl suggesting existence of a folding intermediate in this range (Fig. 4.14e). The phase diagram consisted of three linear parts indicating the existence of three independent transitions linking four different conformations of LlCAD2 (Fig. 4.14d). These four conformations are 1-Native, 2-Intermediate at 1 M GnHCl, 3-A minor intermediate at 3 M GnHCl and 4-Unfolded protein. Both parameter A and the phase diagrams were identical for both unfolding as well as refolding processes, thus showing the transitions to be truly reversible in nature. ANS binding assays showed binding between 0.75-1.75 M GnHCl with a maximum at 1 M (Fig. 4.14f). This binding was found to be more during the unfolding process when compared to refolding experiment; however, a slight binding of ANS was also observed at 3-4 M which was more during the refolding process.

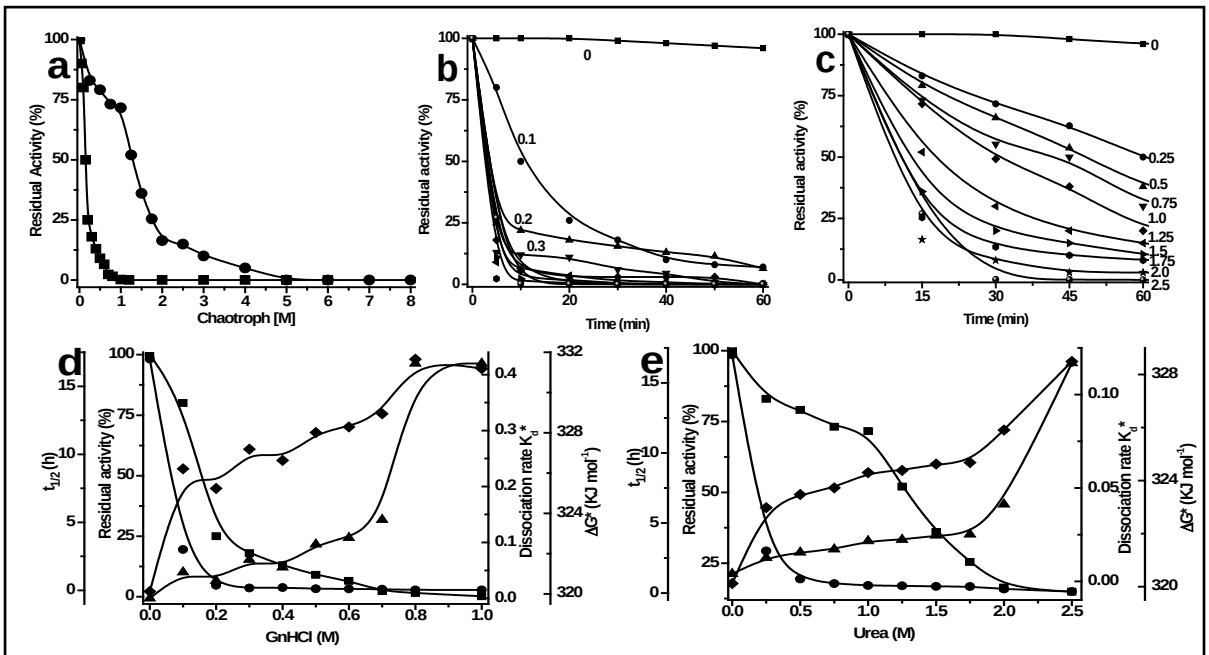


Fig. 4.13:(a) *Effect of chaotropic agents, GnHCl (■) and urea (●), on initial rate of reaction of LlCAD2. LlCAD2 stability in presence of* ***(b)*** *GnHCl;* ***(c)*** *Urea. Stability parameters for inactivation of LlCAD2 for* ***(d)*** *GnHCl;* ***(e)*** *Urea. (■) Residual Activity, (▲) Dissociation rate constant K_d^*, (●) Half-life $t_{1/2}$, (♦) Free energy change for deactivation ΔG*. LlCAD2 was incubated at 25°C in 50 mM Tris; pH 8.0 with either GnHCl (0-1 M) or Urea (0-2.5 M), & assayed at 30 °C under standard conditions*

Far-UV CD spectra for GnHCl treated protein can be divided in three phases (Fig. 4.14g). GnHCl at the concentration between 0-0.75 M induced a steady change in the secondary structure of LlCAD2, wherein the negative ellipticity decreased gradually. There was a decrease in α-helical content and

minor changes were observed in β-sheets, while β-turns and random structure increased (Fig. 4.14j). No major change in CD spectra was observed between 1-2 M GnHCl assuming that the protein might have attained a stable conformation. A major change in CD spectra occurred between 2-3 M GnHCl and remained constant upto 6 M. As shown in Fig. 4.14j, LlCAD2 shows no changes in the secondary structural elements at 3-6 M.

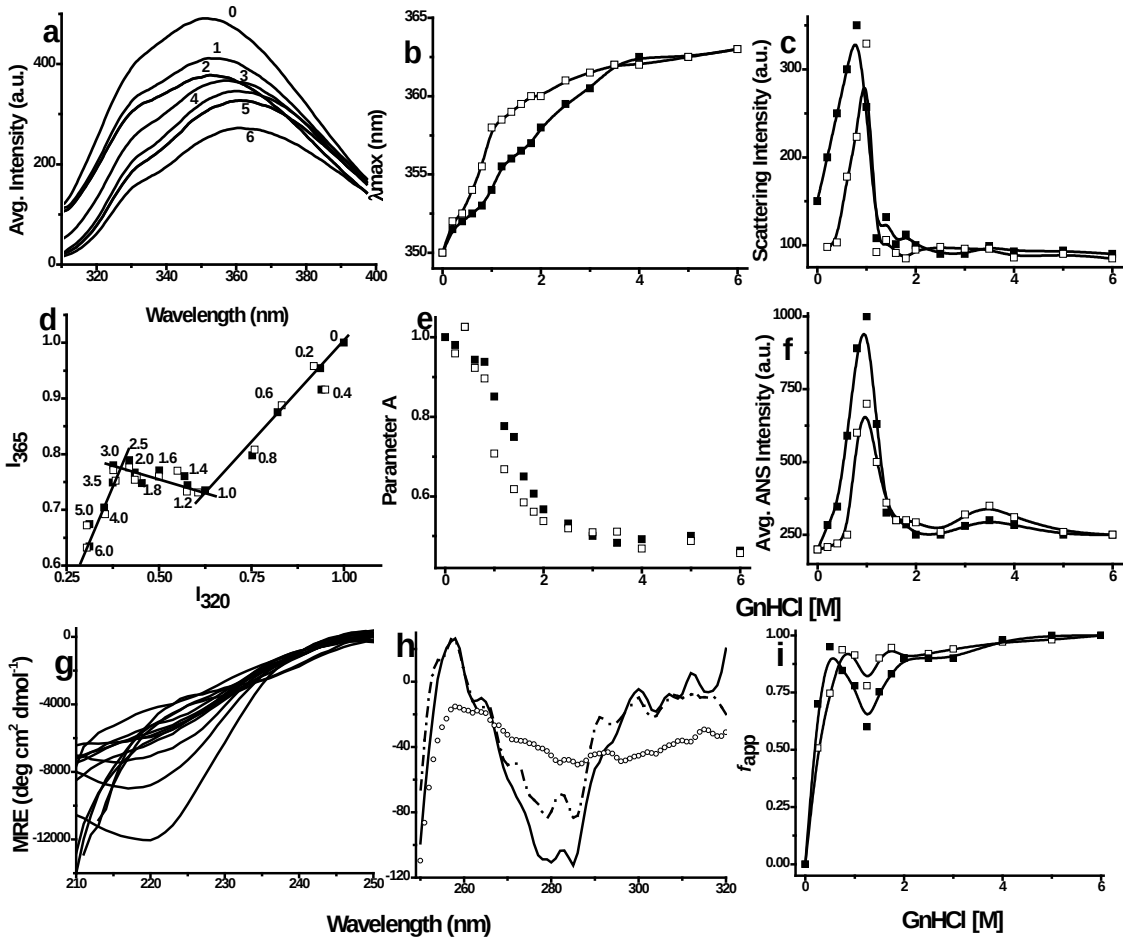


Fig. 4.14: S*tructural changes in LlCAD2 induced by GnHCl measured by fluorescence spectroscopy.* ***(a)*** *Intrinsic fluorescence spectr;,* ***(b)*** λ_{max}*;* ***(c)*** *Scattering intensity;* ***(d)*** *phase diagram;* ***(e)*** *Parameter A;* ***(f)*** *ANS intensity;* ***(g)*** *Far-UV CD spectra of LlCAD2 at different concentrations of GnHCl;* ***(h)*** *Near-UV CD spectra for native protein (——); protein unfolded at 1 M GnHCl (oooo);* ***(i)*** f_{app} *calculated from* MRE_{208} *measured during Unfolding (■) and Refolding (□) of the protein refolded back to 0.2 M GnHCl (—•—). LlCAD2 was incubated for 3 h in 20 mM Tris, pH 8.0 for unfolding, and further 3 h for refolding;* ***(j)*** *Comparison of calculated secondary structures in LlCAD2 from far-UV CD spectroscopy as a function of GnHCl. LlCAD2 was incubated for 3 h in 50 mMTris at pH 8.0 & measured during Unfolding (■), Refolding (□) of the protein.*

It is important to that even at 6 M GnHCl, the CD spectra show a major negative ellipticity, reflecting retention of some structure at higher concentrations. Near-UV CD spectra showed major loss of tertiary structure at 1 M GnHCl, with absence of all the peaks observed in native protein. Upon refolding the protein by buffer dilution, the protein showed reappearance of these peaks (Fig. 4.14h). The f_{app} reveals a complex relationship as a function of GnHCl concentration (Fig. 4.14i). There was an initial unfolding observed upto 0.75 M, and a brief decrease in f_{app} with local minima at 1 M GnHCl which finally achieved complete unfolding at 3 M. DLS analysis of LlCAD2 incubated at 1 M GnHCl showed hydrodynamic radii similar to the dimeric form, while the protein aggregated above 3 M (data not shown). Also, metal content of the molten globule was measured following Zincon dye binding assay which showed the zinc ion content remained unchanged when compared to the native protein. The fluorescence and CD spectra analysis thus confirms at least one folding intermediate in the vicinity of 1 M GnHCl. This molten-globule state can be characterized by a substantial amount of secondary structure present, no native-like tertiary structure, and significant exposure of hydrophobic surface area compared to the native state leading to a propensity for aggregation.

4.10 Urea denaturation also involves intermediate states

Like GnHCl, LlCAD2 was also subjected to urea denaturation in order to understand its folding mechanism. The intrinsic fluorescence spectra showed a gradual decrease in fluorescence intensity, while a major red-shift in λ_{max} was observed till 4 M, which was then stable upto 8 M urea (Fig.4.15a, b). Also, scattering intensity initially decreased till 0.75 M and then increased gradually with further increasing urea concentrations (Fig. 4.15b). Two intermediate states (1.5 M and 3.0 M) were observed from analysis of phase diagram which was also seen from parameter A values as a function of urea concentrations (Fig. 4.15b). Fluorescence intensity was employed to calculate f_{app} in the case of urea, and it was observed that the transition from native to denatured protein was characterized by a single step, resembling a two-state transition model N→U. However, two shoulders were observed in the denaturation curve: one at 1.5-2 M and another at 3-4 M urea (Fig. 4.15b). This might be indicating a non-two-state transition involving intermediate states in the vicinity of 1.5 & 3 M urea.

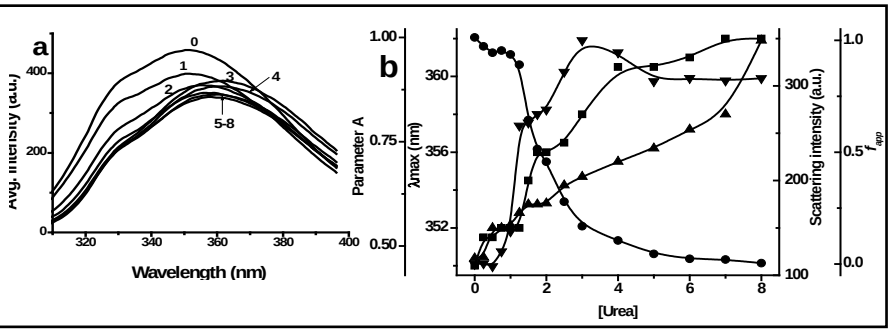


Fig. 4.15: (a) *Urea induced fluorescence in LlCAD incubated for 3 h in 50 mM Tris, pH 8.0;* ***(b)*** *Fluorescence parameters; λ_{max} (■), Parameter A(●), Scattering intensities (▲), f_{app} (▼)*

4.11 Apoenzyme is structurally different than native LlCAD2

As mentioned in chapter 2, LlCAD2 is a Zn-containing metalloprotein, with 4 Zn^{2+} bound per dimer molecule, where its catalytic effects are explained. In order to understand the structural role of Zn in LlCAD2, structural studies were taken up by removing the bound metal ion by EDTA. It was observed that LlCAD2 was very sensitive towards EDTA, with 75% loss of activity at 1 mM EDTA, and complete loss above 10 mM (Fig. 4.16a). Zincon assays showed that Zn content decreased slightly by just 10% with increase in EDTA concentration upto 4 mM, however, it drastically fell below 0.5 Zn^{2+} ions bound per dimer (90% loss) at 10 mM EDTA (Fig. 4.16a). Metal content of LlCAD2 incubated in 10 mM EDTA was also confirmed by ICP-OES which showed 0.4 ± 0.02 ions per protein dimer [68].CAD was found to be very unstable in EDTA solutions as well, with half lives of about 3 h at 0.2 mM, and less than a minute at 1 mM concentrations (Fig. 4.16b, c). Both K_d^* and ΔG^* increased with increase in EDTA upto 1 mM. Furthermore, structural changes induced in LlCAD2 due to removal of metal ions were monitored by fluorescence and CD spectroscopy. The intrinsic fluorescence spectra showed gradual decrease in intensity, while λ_{max} showed a marginal shift of 4 nm upto 20 mM EDTA (Fig. 4.16d, e). Scattering intensity increased with increasing EDTA concentrations (Fig. 4.16e). The phase diagram and parameter A suggested an all-or-none transition between two states- Native and Unfolded (Fig. 4.16f) unlike pH, GnHCL or urea transitions. Also, ANS binding gradually increased with increasing EDTA concentrations, though it was much less than observed in pH and GnHCl studies (Fig. 4.16e).

The far-UV CD spectra showed no change in the spectra at all, thus suggesting no alterations in the secondary structural elements by removal of the metal ions (Fig. 4.16g). However, a major shift was observed in the near-UV region, where the peaks characteristic of the native conformation were either absent or decreased in ellipticity (Fig. 4.16h). This observation showed that major changes took place in the tertiary structure of LlCAD2 upon chelation of the bound metal ions, which were reversed to an extent by externally supplementing Zn^{2+} ions to the chelated protein solution. Here, the native tertiary conformation was almost reconstituted with all the peaks appearing again. Also, unfolding was monitored for the chelated protein as a function of temperature and f_{app} was calculated (Fig. 4.16i). The native protein showed an unfolding transition between the temperature range 70-90 ºC, while the apoenzyme showed lowering in the transition phase between 60-75 ºC. DLS analysis of the apoenzyme showed it to be a dimer (data not shown). The above results suggest that the bound Zn ions help maintain the tertiary structure of the LlCAD2 molecule, as well as provide thermostability to the overall structure. Removal of the bound metals leads to a two-state transition to an unfolded state without formation of any intermediate folded state.

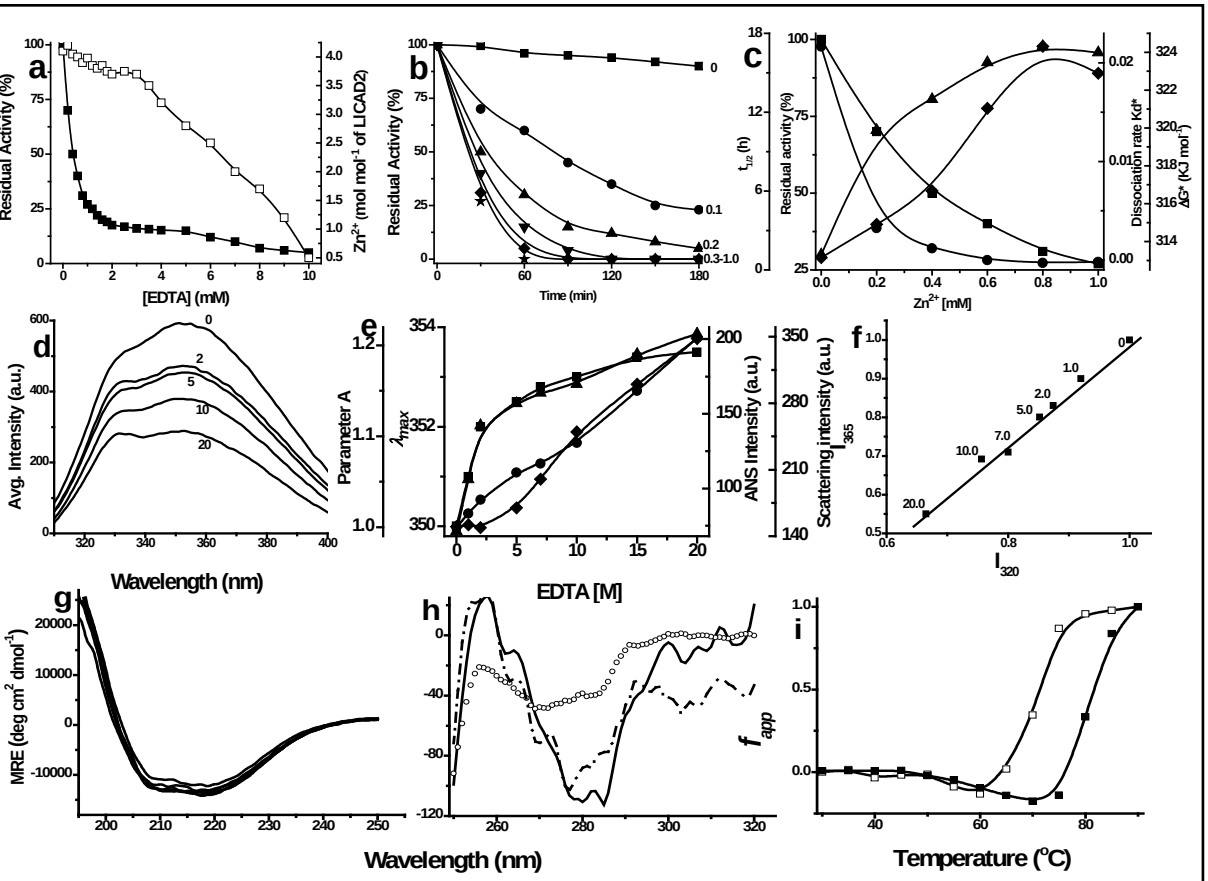


Fig. 4.16: ***(a)*** *Effect of EDTA on Initial rate of LlCAD2 (■), Zn^{2+} content (□);* ***(b)*** *Effect of EDTA on LlCAD2 stability;* ***(c)*** *Stability and thermodynamic parameters for LlCAD2 inactivation. (■) Residual Activity, (▲) Dissociation rate constant K_d^*, (●) Half-life $t_{1/2}$, (♦) Free energy change for deactivation ΔG^*. Enzyme was incubated at 25 °C in 50 mM Tris; pH 8.0 and assayed under standard conditions;* ***(d)*** *Effect of EDTA on LlCAD2 structure Intrinsic fluorescence spectra;* ***(e)*** *λ_{max} (■), Parameter A (●), ANS intensity (▲), Scattering intensity (♦);* ***(f)*** *Phase Diagram;* ***(g)*** *Far-UV CD spectra;* ***(h)*** *Near-UV CD spectra for native protein (——), unfolded at 10 mM EDTA (oooo), and renatured protein (—•—);* ***(i)*** *F_{app} for native (■) and EDTA chelated apoenzyme (□). LlCAD2 was incubated at EDTA concentrations (0-20 mM) for 4 h*

4.12 Solute quenching reveals structural differences amongst LlCAD2 intermediates

The structural differences amongst many conformational states of LlCAD2 were further investigated using Trp fluorescence quenching experiments. The indole nucleus of Trp has a tendency to donate electrons while in the excited state; this property is highly sensitive towards a range of different solute molecules and results in quenching of its fluorescence emission. Here, the neutral quencher acrylamide, as well as positively charged CsCl and negatively charged KI were employed. Acrylamide, due to its small size and ability to penetrate deep inside protein molecule, showed maximum quenching of trp fluorescence for almost all the states, while I^- and Cs^+ ions, cannot penetrate and can only access any surface exposed Trp, showed relatively less quenching. All the three quenchers were fairly accessible to the Trp of the native protein and resulted in higher

quenching efficiencies which can be attributed to surface exposed Trp residues of LlCAD2 (Fig. 4.17).

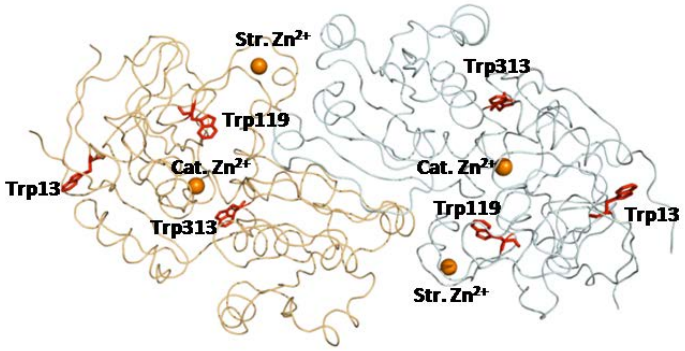


Fig. 4.17: *Positions of trp residue in LlCAD2 structure. All the 6 trp residues are significantly surface exposed. W313 is located in the NADPH binding site in the proximity of catalytic Zn^{2+} while W119 is near the structural Zn^{2+}*

A detailed discussion arising from analyzing the quenching parameters is not provided here, though, some key conclusions made from the data given in Fig. 4.18 and Table 4.1 are provided:

1. Native LlCAD2 dimer was quenched more than 70% by acrylamide, and 43 & 42% by Cs^+& I^-, respectively. Stern-Volmer plots for Acr and I^- were linear, while Cs^+ showed a negative curvature, showing heterogeneity in the microenvironment of the surface exposed Trp with selective quenching of certain Trp which are easily accessible at lower concentrations of the quencher. While, at higher concentrations, relatively inaccessible Trp will dominate the quenching constants.
2. LlCAD2 monomers showed quenching by Cs^+& I^- which was comparable to dimer, while that of Acr increased by 20%. Also the accessible fraction for Acr increased upon monomerization. This result suggests certain changes upon dimerization.
3. Upon denaturation by 6 M GnHCl, quenching efficiency of Acr as well as I^- increased, while that of Cs^+ decreased (now with a linear S-V plot), which was also reflected from the calculated accessible fractions. These quenching parameters (mentioned in Table 4.1) as well as the fluorescence and CD data provided above suggest that LlCAD2 does not fully denature in presence of GnHCl, and still much of the residual structure persists.
4. The acid-induced molten globule MG1 at pH 2 showed major structural differences than the native protein. The quenching of the molten globule by acrylamide showed an upward curvature in the Stern-Volmer plot which arises due to occurrence of both static and collisional quenching. The static mechanism is a consequence of formation of a complex while the dynamic or collisional component involves collisions with Acr during the trp lifetime fluorescence in the excited state.While no such pattern was observed for Cs^+ and I^-, though, the quenching efficiency

for Cs^+ drastically decreased, while that of I^- increased in the molten globule. These observations further confirm the structural changes observed during unfolding.

5. MG1 was further denatured in 6 M GnHCl, which resulted in even higher accessible fractions to Acr, though quenching remained unchanged. Also, only static quenching was observed for Acr. While Cs^+ and I^- did not show any changes in the accessibility of Trp, eventhough, quenching efficiency increased for I^-, and decreased for Cs^+ maybe due to repulsion of Cs^+ ions at pH 2.0. However, the S-V plots showed a negative pattern again reflecting heterogeneity in the trp microenvironment. These results suggest that the protein is still considerably folded, as Cs^+ is still unable to access the partially buried Trp residue.
6. Readjusting pH back to 8 resulted in renaturation of MG1 back to its native state which was also observed in the quenching results. All the quenchers showed parameters similar to native LlCAD2 with Acr showing more accesibility. This may mean that although the structure is fairly reversible, it is still different from that of the native state. This was further explained already, that renaturation leads to only partial recovery of catalytic ability.
7. The quenching parameters of GnHCl induced molten globule MG2 which was obtained at 1 M GnHCl. This revealed a state which was more accessible to Acr and I^- than the native state, while it decreased a bit for Cs^+. These values were still less when compared to denatured state at 6 M GnHCl, suggesting an intermediate state between native and denatured forms.
8. Upon diluting MG2 with buffer in order to renature the protein, it showed comparable quenching efficiency to the native state for Acr. While, S-V plot for Cs^+ was linear, I^- showed a negative curvature, both of which were not observed in the native state. These results suggest rearrangements in the protein structure taking place during unfolding and refolding.
9. Zn-chelated apo form of LlCAD2 was also subjected to fluorescence quenching which resulted in fairly increased quenching by Acr and I^-, while it decreased for Cs^+. Also, the S-V plots showed no major changes, indicating some minor changes in the structure upon removal of the bound metal ions.
10. Upon treating the apoenzyme with 2 mM Zn^{2+} and 10 mM DTT, much of the native tertiary structure was re-established, though catalytic ability was regained only marginally. Quenching studies for this renatured protein showed that Acr had accessibility comparable to the native form, while that of Cs^+ and I^- were still lower. These results along with the near-UV data show that complete refolding of the protein is still not achieved under these conditions, and that removal of the bound metal ions results in irreversible structural changes in LlCAD2.

Protein	Acrylamide				CsCl				KI			
	%Q	Ksv	Fa	Ka	%Q	Ksv	Fa	Ka	%Q	Ksv	Fa	Ka
Dimer	71	4.12	0.93	5.98	43	2.84 0.91	0.42	8.45	42	1.02	0.52	3.79
Monomer	93	25.65	1.16	10.57	41	2.54 0.65	0.42	11.84	52	1.63	0.47	5.91
Denatured	85	15.09	1.5	5.1	37	0.82	0.34	7.38	69	4.33	0.88	6.15
Acid induced MG1	84	5.09 10.2	1.1	4.97	14	0.12	0.22	0.68	70	3.49	0.84	3.48
Denatured MG1	80	7.45	1.98	2.25	33	1.62 0.43	0.29	15.43	74	7.95 3.47	0.9	21.89
Renatured MG1	78	0.98	1.23	4.07	39	2.04 0.73	0.32	11.61	54	2.01	0.47	3.73
GnHCl induced MG2	83	8.77	0.95	10.75	35	2.11 0.62	0.35	5.99	67	3.82	0.84	6.26
Renatured MG2	71	4.26	0.96	4.33	35	0.57	0.36	5.2	51	2.48 1.27	0.74	3.51
Apoenzyme	81	7.59	0.84	9.81	32	1.52 0.57	0.23	11.53	50	1.45	0.62	5.0
Zn treated apoenzyme	75	5.16	0.88	10.6	15	1.14	0.31	7.36	33	2.42	0.47	6.82

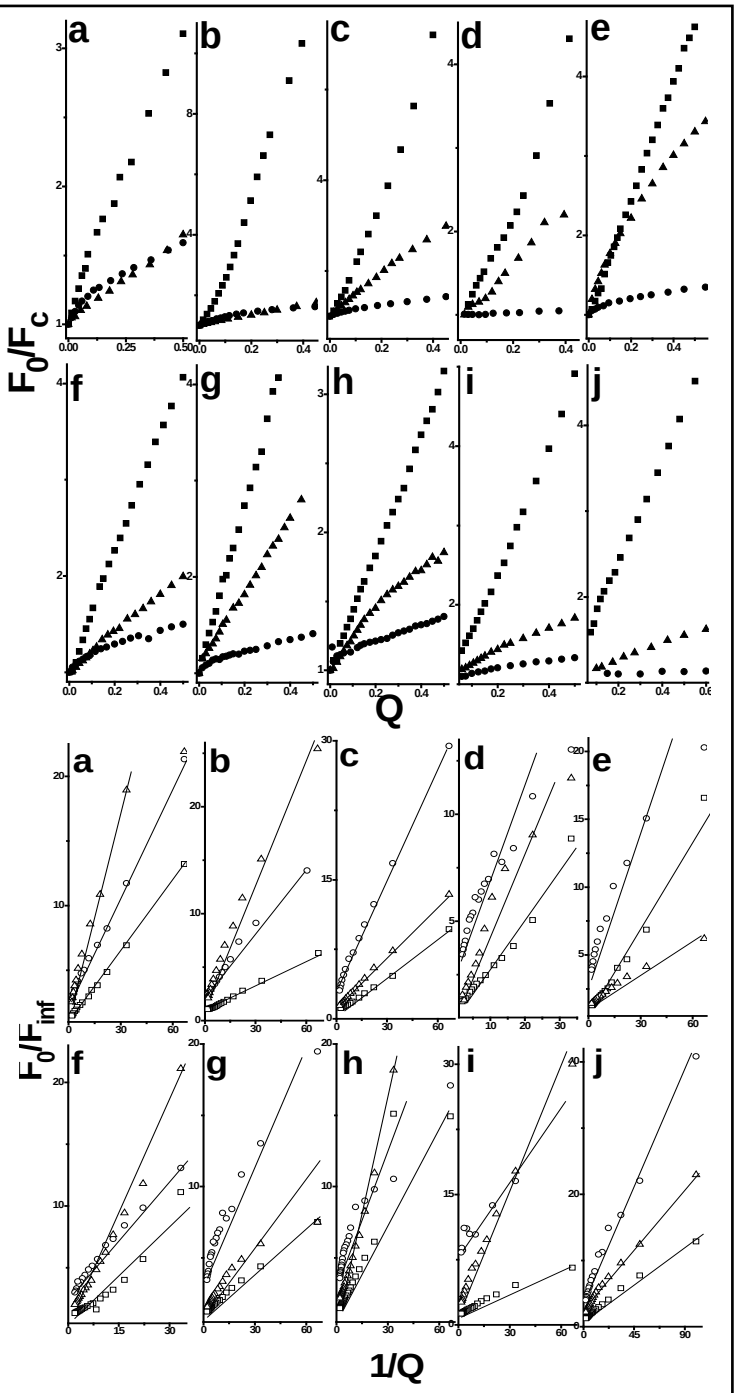


Fig. 4.18: *Fluorescence quenching of different states of LlCAD2.* ***(a)*** *Native dimer,* ***(b)*** *Monomer,* ***(c)*** *Denatured,* ***(d)*** *Acid-induced MG1 (pH 2),* ***(e)*** *Denatured MG1,* ***(f)*** *Renatured MG1,* ***(g)*** *GnHCl induced MG2,* ***(h)*** *Renatured MG2,* ***(i)*** *Apoenzyme,* ***(j)*** *Renatured apoenzyme. Quenchers used were Acrylamide, CsCl, KI represented as ■, ●, ▲(Stern-Volmer plots), and □, ○, Δ (modified Stern-Volmer plots), respectively*

Table 4.1: *Summary of quenching parameters for different states of LlCAD2*

5

CAD Structure-Function Relationship: Elucidated

CAD: Structure-Function Relationship

The genesis of this work can be traced back from the results shown in our previous report [34], where the reduction in total lignin content for L.leucocephala transgenics down-regulated for each individual enzyme of the phenylpropanoid metabolism is provided. The findings suggested that the plant can still manage to side-line various blockades in the pathway in order to form lignin in its cell-wall. Though there are several reports providing evidence to the importance of CAD in lignin metabolism, the above results still questions its true physiological role. Two distinct sequences were isolated from our lab which resembled bona fide CADs; LlCAD & LlCAD2. Other CAD sequences isolated from L.leucocephala available on NCBI database also suggests the presence of these two distinct CAD isoforms. Though LlCAD was seen to catalyze cinnamaldehyde at a much faster rate than sinapaldehyde followed by coniferaldehyde (P. Kulkarni, unpublished), LlCAD2 preferred sinapaldehyde over other substrates [68]. It would have been much interesting to exploit this substrate specificity difference amongst the isoforms, however, this thesis focused only on characterizing LlCAD2. We saw some of the interesting functional and structural properties of LlCAD2 in previous chapters, and here, a correlation between the catalytic mechanism and overall structure is attempted in order to draw certain general conclusions on its physiological role.

5.1 Comparison with other ADHs

CADs represent a class of NADPH-dependent aromatic-alcohol oxidoreductases capable of catalyzing the broad-substrate reaction; converting hydroxycinnamyl aldehydes to corresponding alcohols, which being an obligatory step towards formation of lignin in all plants. Since its discovery, numerous reports have suggested its determining role in lignin biosynthesis; formation of S and G lignins. In a recent report, CADs from different sources were analyzed from an evolutionary perspective and provided evidence that strongly correlated the origin of monolignol biosynthesis pathway, and lignin, to earliest occurrences of CADs [121]. However, even after 4 decades of its discovery, its biochemical and functional properties are still quite puzzling. There are several reports from many diverse species claimed to be CAD-like enzymes, however, improper/ inaccurate/ or partial characterization often raises questions on these reports. The present study thus focused on an in-depth structurally guided kinetic characterization of a CAD homologue from *L. leucocephala* as a means to form a foundation, on which further characterization of other CADs can be made. Such analysis will provide a platform to help understand, as well as exploit, the substrate specificities of lesser characterized CAD-like enzymes. A comparison was made between LlCAD2 with a few well specificities of these CADs vary greatly even when the sequences as well as structural cues are much identical (Table 5.1, Fig. 3.4).

	Enzyme	PDB	Organism	% Identity	RMSD (Å)	Preferred substrate	Co-sub	Subunits (mol. wt.)	Reference
Plants	**LlCAD2**	-	*L. leucocephala*	-	-	Sinapaldehyde	NADP(H)	2 (78)	This work
	LlCAD	-	"	99	0.023	Cinnamaldehyde	"	2 (78)	unpublished
	AtCAD5	2CF6	*A. thaliana*	77	0.117	Coniferaldehyde	"	2 (78)	[49]
	PtSAD	1YQD	*P. tremuloides*	52	*1.136*	Sinapaldehyde	"	2 *(81)*	[90]
Bacteria	**hpCAD**	3TWO	*H. pylori*	39	1.248	Benzaldehyde	"	2 (90)	[59]
	htADH	1RJW	*B. stearothermophilus*	30	1.307	Aliphatic alc.	NAD(H)	4 (147)	[63]
	PaADH	1LLU	*P. aeruginosa*	28	1.343	"	"	4 (147)	[64]
Yeast	**YADH**	2HCY	*S. cerevisiae*	27	1.484	"	"	4 (150)	[60]
Animal	**HLADH**	2OHX	*E. caballus*	21	2.886	"	"	2 (82)	[122]
	β1ADH	1HDY	*H. sapiens*	22	2.799	"	"	2 (82)	[62]
	EcADH	8ADH	*E. caballus*	21	2.543	"	"	1 (40)	[61]
Fish	**EeADh**	1CDO	*G. morhua*	23	2.963	"	"	2 (82)	[123]
Archae	**SsADH**	1R37	*S. solfataricus*	23	2.306	"	"	2 (77)	[65]

Table 5.1: *Different aliphatic and aromatic ADHs from various sources*

Also, one can see much variation in the quaternary structure of these enzymes, which vary from 40 kDa monomeric to 150 kDa tetrameric forms. This comparison shows that though CAD gene family is very diverse, while its basic catalytic nature remains the same.

5.2 Structurally driven kinetic analysis of LlCAD2

5.2.1 Catalytic mechanism

LlCAD2, like HLADH and many other ADHs follows a compulsory-order ternary-complex mechanism for reversibly reducing hydroxycinnamyl aldehydes in plants. It was originally proposed that the formation of ternary complex is not significant, and that release of Q is rate-limiting step [124], however, predicting kinetic mechanism purely on the basis of initial velocity measurements in absence of any products can often lead to errors. Later, extensive product inhibition data was provided by Cleland [82] which gave evidence that the ternary complex is not only formed, but is also kinetically significant in HLADH. Our product inhibition experiments described in Chapter 2B provide evidence of such a mechanism for LlCAD2 as well. Based on the results obtained in our studies, as well as reviewing the available information for AtCAD5 and HLADH, the catalytic cycle for reduction of aldehyde by LlCAD2 can be given as:

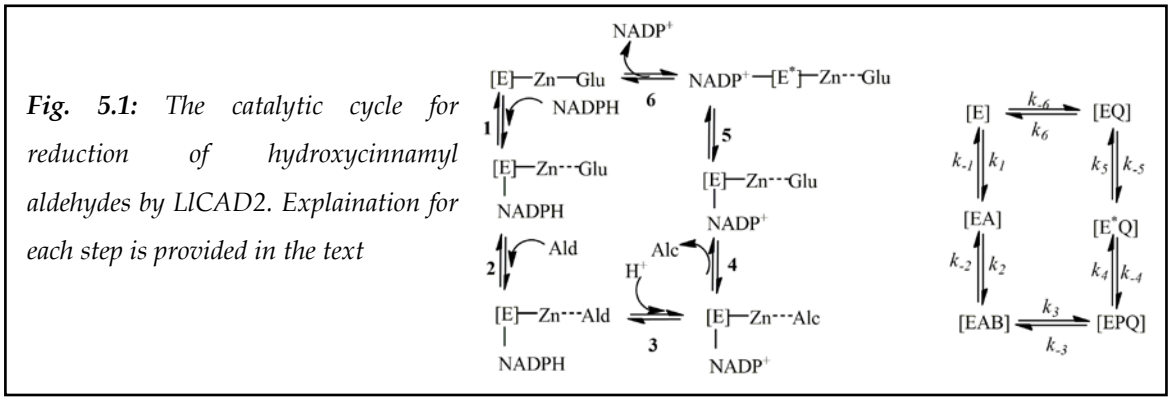


Fig. 5.1: *The catalytic cycle for reduction of hydroxycinnamyl aldehydes by LlCAD2. Explaination for each step is provided in the text*

The steps involved in this catalytic cycle are as follows:

1. Binding of NADPH followed with closing of the catalytic site. This binding follows a conformational change in HLADH which is generally of the order of 10° [61, 104, 105]. Attempts to identify any structural changes occurring in LlCAD2 upon NADPH binding were carried out using CD spectroscopy. However, the results obtained were not reliable enough to predict such changes. However, in AtCAD5, no major conformational change was observed upon cofactor binding, and thus, only a crystal structure of the binary complex help us understand such changes.

2. Upon aldehyde binding, Glu70, previously coordinated to the catalytic Zn, is displaced. The role of Glu70 is discussed below.
3. Hydride transfer then occurs from the nicotinamide ring of NADPH to the formyl group of aldehyde substrate thus forming the alcohol product. It was observed in HLADH that the Val201 (equivalent to Val192 of LlCAD2) is in close proximity to the nicotinamide ring and ensures a close contact between eh reactive C4 of NADPH and C1 of the aldehyde, thus promoting transfer of the hydride [125]. The mutant V192G of LlCAD2 a decrease in the binding of NADPH and thus can be taken as a confirmation of its role in fixing the position of NADPH during hydride transfer. Hydride transfer is succeeded by transfer of a proton from the bulk solvent to the hydroxyl group of the alcohol. The residues Ser49 (Thr49 in AtCAD5, Ser48 in HLADH) and His52 (His51 in HLADH) plays a critical role in this proton relay mechanism. Mutations at both these places resulted in loss of catalytic ability of LlCAD2 as well as in similar studies in HLADH. However, it has been proposed that His52 does not play a role in proton relay in AtCAD5.
4. Upon protonation of the alcohol, Glu70 displaces the former from Zn and results in release of the alcohol.
5. Simultaneous isomerization of E-$NADP^+$ than takes place.
6. Release of $NADP^+$ and opening up of the catalytic domain to repeat the entire catalytic cycle.

5.2.2 Isomerization upon NADP(H) binding

The inhibition experiments along with protein-ligand binding experiments also suggest that the enzyme undergoes isomerization upon forming of [EQ] complex (Enzyme-NADP). This isomerization is critical for the substrates to bind to the enzyme, as, Enzyme-Substrate complex would not lead to a catalytically functional ternary complex, which is evidenced by lack of binding of the hydroxycinnamyl alcohols to free LlCAD2 (2.13). However, the aldehydes were still able to bind to the free LlCAD2. This observation may mislead to interpret that the reaction meachanism is random in terms of substrate binding. However, though binding of aldehyde can take place in absence of NADPH, the E-Substrate complex thus formed can still accommodate NADPH and result in a ternary complex with the mechanism still being ordered. This assumption lies on the fact that dissociation of NADPH from the ternary complex will be much slower than the aldehydes [126]. This isomerization step can be viewed as the transition from an “open” to a “closed” conformation. As evidenced by crystal reports of HLADH, these conformations show a distinct structural difference, where binding of the co-enzyme leads to a catalytically active form. As evidenced form

LADH [127], the dissociation of NADP in LlCAD2 reaction can be assumed to be controlled by the isomerization of the enzyme, and this isomerization is the rate-determining step for the LlCAD2 catalyzed reduction of hydroxycinnamyl aldehydes.

5.2.3 Role of Zn in isomerization of LlCAD2

As observed from the pH activity profiles that His is involved in NADPH binding, and the site-directed His52 mutant showed it to be a part of the proton relay. One interpretation of these two results is that a group on free enzyme shifts its pK from 8 to around 7. This group may possibly be the coordinated Zn ion. It thus seems very likely that the ionization of Zn may control NADPH binding, and that His52 facilitating the protonation of alcohol as depicted in Fig. 3.8. Since the shift in pK value upon NADPH binding is accompanied by protonation, the latter may be assumed to govern the overall conformation change in LlCAD2 upon isomerization.

Looking again at the catalytic cycle of LlCAD2, the substrate can bind to the [EA] complex once the protonation and isomerization takes place. The underlying mechanism is still unclear; however, there are potential candidate theories. One assumes formation of a penta-coordinated Zn complex upon substrate binding, as observed in *T.brokii* ADH [128]. Alternatively, an extended proton relay which would involve the Glu70 which is coordinated to the catalytic Zn. In this mechanism, Glu70 would first displace the hydroxide from the Zn ion, and later will be replaced by the incoming substrate. This mechanism is supported by the fact that the mutant E70A was catalytically inactive; implying that maintaining the Zn-coordination is a prerequisite for catalysis. A crystal structure of a ternary complex of LlCAD2 with both its substrates can shed light to this mechanism.

5.2.4 Substrate Inhibition

Though, one may ask as to why would any enzyme undergo such a conformational change in order to differentiate the binding of its two substrates? The answer to this question is quite simple. Any enzyme showing such isomerization will thus show a compulsory order in binding of the substrates, and subsequent release of the products. It also means that such an enzyme will show less tolerance towards the substrate that binds second. At higher concentrations of the second substrate alcohols (or aldehydes), the abortive Enzyme-NADH-alcohol (or Enzyme-NADP-Aldehyde) complexes can form, thus giving rise to substrate inhibition patterns. It is proposed that the varying degrees of inhibition of LlCAD2 by different substrates have a possible physiological role in governing lignin formation in plants. Different CAD isoforms with their different substrate preferences and inhibition patterns may potentially work in tandem to bring about conversion of only certain substrates in order to form desired lignin traits. This differential inhibition may have an underlying mechanism, by

which the monolignol biosynthetic network is temporally and spatially regulated between different cells and tissues which require distinct lignin compositions. However, further work both *in vitro* and *in vivo* needs to be carried out in order to better understand this regulation.

5.2.5 Substrate specificity

It is already established that LlCAD2 is a memebr of multi-substrate specific isomers responsible for formation of monolignols in plants. The kinetic data provided here shows evidence that LlCAD2 can catalyze different hydroxycinnamyl aldehydes, however, prefers sinapaldehyde over all others. The difference of such preference lies in the different response of LlCAD2 towards sinapaldehyde and other substrates in terms of their binding, rate of catalysis, as well as the substrate inhibition patterns. It can thus be hypothesized that LlCAD2 is the major factor leading to formation of S-lignin in *L.leucephala*.

5.3 Kinetically driven structural analysis of LlCAD2

Another aspect of establishing a structure-function relationship as mentioned above would be to drive structural analysis of enzymes based on their kinetic characterization. The double and triple mutants of LlCAD2 for the phosphate binding motif provide examples for the latter case. These mutants along with their kinetic characterization demonstrate that minor alterations in the active site residues have a major effect on the efficiency of CAD reactions. These mutants can be followed by a suite of similar structurally altered mutants which in turn may lead to engineering of altered or novel substrate specificities. Tailoring such changes in substrate specificities of enzymes involved in lignin chemistry can have a major impact on commercial advancement of plants engineered with such enzymes for forage and paper pulping.

5.4 Thermodynamically driven structural analysis of LlCAD2 (An energy landscape)

It is known that a unique three-dimensional structure is a prerequisite for proteins to be in their biologically active form. However, understanding this process of folding has intrigued researchers for decades. Initially studies were concentrated on characterizing individual small-proteins with less than 100 amino acids in order to make any generalizations about protein folding. However recently, larger proteins and protein families are being investigated in order to understand any patterns in their folding mechanism. It is now known that folding is not a single specific pathway, but a multi-dimensional folding funnel (also known as an energy landscape). This funnel shows that folding of a single polypeptide chain can be very diverse as it can follow one of the many available paths to form a native folded structure. Different intermediately folded transition states may populate each of these

paths which in turn are based on any alterations of the common free-energy profile. Unfolding studies have revealed that smaller proteins would generally unfold via a simple two-state transition without any stable intermediates (though many different kinetic phases can be observed here). While larger multimeric proteins can follow more complex three-state transitions often populated by different folding intermediates. These intermediates might still contain local structured regions which correspond to any domains or sub-domains of the native state.

Structural transitions for unfolding as well as refolding LlCAD2 were measured in this study in order to better understand its folding mechanisms, though at this point generalizations regarding the folding pathway of LlCAD2 cannot be made. The many folding intermediates observed under different denaturing conditions were partially characterized in terms of their kinetic and structural features, though this data is not sufficient to draw an energy landscape for this enzyme. A more detailed analysis regarding the kinetic aspects of unfolding should be taken up in order to understand the true thermodynamic changes upon structural collapse. Such thermodynamically driven kinetic and structural characterization of folding intermediates would help answer the many questions arising as to what folding pathway LlCAD2 would take up under different physiological conditions. All in all, the folding intermediates observed here still provide a plethora of information regarding its structural properties.

5.5 Conclusions

The data presented here is an example as to how can one combine detailed kinetic analysis to structural properties in order to understand enzyme evolution in terms of secondary metabolism in plants. These results suggest that LlCAD2 is most closely related to other CAD-like enzymes involved in lignin formation, and that it works in tandem with other CAD homologues in order to determine the composition and properties of lignin. However, true physiological implications of such structure-function relationship needs to be elaborated further.

References

1. **Boerjan, W., *et al.***, *Annu Rev Plant Biol.* 54 (**2003**) 519-46.
2. **Chabannes, M., *et al.***, *Plant J.* 28 (**2001**) 271-82.
3. **Jones, L., *et al.***, *Plant J.* 26 (**2001**) 205-16.
4. **Sarkanen, K. V., Ludwig, C. H.,** Lignins: Occurrence, Formation, Structure, and Reactions. *Wiley-Interscience*, **(1971)** ISBN: 0471754226
5. **Campbell, M. M., Sederoff, R. R.**, *Plant Physiol.* 110 (**1996**) 3-13.
6. **Chiang, V. L., *et al.***, *Tappi Journals.* 71 (**1988**) 173-6.
7. **Axegard, P., *et al.***, *Papier.* 46 (**1992**) 16-25.
8. **Bierman, C. J.,** Essentials of Pulping and Papermaking, *Academic Press* **(1993)** ISBN: 012097360X
9. **Christensen, J. H., *et al.*,** in Molecular Biology of Woody Plants, 64 **(2000)** 227-67. ISBN: 9789048153381
10. **Bierman, C. J.,** Handbook of Pulping and Papermaking. *Academic Press* **(1996)** ISBN: 0120973626
11. **Higuchi, T.,** in Biosynthesis and biodegradation of wood components, *Academic Press* **(1985)** 141-60. ISBN: 9780323143288
12. **Odendahl, S.**, *Pulp & Paper Canada.* 95 (**1994**) 144-8.
13. **Chen, F., *et al.***, *Plant J.* 48 (**2006**) 113-24.
14. **Chen, F., Dixon, R. A.**, *Nat Biotechnol.* 25 (**2007**) 759-61.
15. **Davison, B. H., *et al.***, *Appl Biochem Biotechnol.* 129-132 (**2006**) 427-35.
16. **Franke, R., *et al.***, *Plant J.* 30 (**2002**) 33-45.
17. **Freudenberg, K., Neish, A. C.,** Constitution and Biosynthesis of Lignin. *Springer-Verlag* **(1968)** ASIN: B0000EGQYB
18. **Gross, G.,** Biosynthesis and metabolism of phenolic acids and monolignols. *Academic Press* **(1985)**
19. **Humphreys, J. M., Chapple, C.**, *Curr Opin Plant Biol.* 5 (**2002**) 224-9.
20. **Whetten, R., Sederoff, R.**, *Plant Cell.* 7 (**1995**) 1001-13.
21. **Whetten, R. W., *et al.***, *Annu Rev Plant Physiol Plant Mol Biol.* 49 (**1998**) 585-609.
22. **Steeves, V., *et al.***, *Phytochemistry.* 57 (**2001**) 1085-93.
23. **Zimmerlin, A., *et al.***, *Biochem J.* 299 (Pt 3) (**1994**) 747-53.
24. **Wyrambik, D., Grisebach, H.**, *Eur J Biochem.* 59 (**1975**) 9-15.
25. **Sibout, R., *et al.***, *Plant Cell.* 17 (**2005**) 2059-76.
26. **Branden, C. I., Eklund, H.**, *Experientia Suppl.* 36 (**1980**) 40-84.
27. **Valencia, E., *et al.***, *J Mol Biol.* 341 (**2004**) 1049-62.
28. **Knight, M. E., *et al.***, *Plant Mol Biol.* 19 (**1992**) 793-801.
29. **Grand, C., *et al.***, *Planta.* 163 (**1985**) 232-7.
30. **Galliano, H., *et al.***, *Plant Mol Biol.* 23 (**1993**) 145-56.
31. **Campbell, M. M., Ellis, B. E.**, *Planta.* 186 (**1992**) 409-17.

32. **Anterola, A. M., Lewis, N. G.**, *Phytochemistry.* 61 (**2002**) 221-94.
33. **Baucher, M.**, ***et al.***, *Crit Rev Biochem Mol Biol.* 38 (**2003**) 305-50.
34. **Khan, B. M.**, ***et al.***, in Genetic Engineering- Basics, New Applications and Responsibilities, *InTech* **(2012)** 93-120. ISBN: 9789533077901
35. **Halpin, C.**, ***et al.***, *Plant Journal.* 6 (**1994**) 339-50.
36. **Baucher, M.**, ***et al.***, *Plant Physiol.* 112 (**1996**) 1479-90.
37. **Baucher, M.**, ***et al.***, *Plant Mol Biol.* 39 (**1999**) 437-47.
38. **Mackay, J. J.**, ***et al.***, *Proc Natl Acad Sci U S A.* 94 (**1997**) 8255-60.
39. **Halpin, C.**, ***et al.***, *Plant J.* 14 (**1998**) 545-53.
40. **Sibout, R.**, ***et al.***, *Plant Physiol.* 132 (**2003**) 848-60.
41. **Pillonel, C.**, ***et al.***, *Planta.* 185 (**1991**) 538-44.
42. **Ralph, J.**, ***et al.***, *Proc Natl Acad Sci U S A.* 95 (**1998**) 12803-8.
43. **Ralph, J.**, ***et al.***, *Phytochemistry.* 57 (**2001**) 993-1003.
44. **Kim, H.**, ***et al.***, *J Biol Chem.* 277 (**2002**) 47412-9.
45. **Marita, J. M.**, ***et al.***, *J Agric Food Chem.* 51 (**2003**) 1313-21.
46. **Lapierre, C.**, ***et al.***, *Plant Physiol.* 119 (**1999**) 153-64.
47. **O'connell, A.**, ***et al.***, *Transgenic Res.* 11 (**2002**) 495-503.
48. **Kim, S. J.**, ***et al.***, *Proc Natl Acad Sci U S A.* 101 (**2004**) 1455-60.
49. **Youn, B.**, ***et al.***, *Org Biomol Chem.* 4 (**2006**) 1687-97.
50. **Kutsuki, H.**, ***et al.***, *Phytochemistry.* 21 (**1982**) 19-23.
51. **Hawkins, S. W., Boudet, A. M.**, *Plant Physiol.* 104 (**1994**) 75-84.
52. **Wyrambik, D., Grisebach, H.**, *Eur J Biochem.* 97 (**1979**) 503-9.
53. **Goffner, D.**, ***et al.***, *Planta.* 188 (**1992**) 48-53.
54. **Li, L.**, ***et al.***, *Plant Cell.* 13 (**2001**) 1567-86.
55. **Barakate, A.**, ***et al.***, *Plant Cell.* 23 (**2011**) 4492-506.
56. **Zhao, Q.**, ***et al.***, *Proc Natl Acad Sci U S A.* 110 (**2013**) 13660-5.
57. **Saathoff, A. J.**, ***et al.***, *PLoS One.* 6 (**2011**) e16416.
58. **Mee, B.**, ***et al.***, *FEBS J.* 272 (**2005**) 1255-64.
59. **Seo, K. H.**, ***et al.***, *FEBS Lett.* 586 (**2012**) 337-43.
60. **Ganzhorn, A. J.**, ***et al.***, *J Biol Chem.* 262 (**1987**) 3754-61.
61. **Colonna-Cesari, F.**, ***et al.***, *J Biol Chem.* 261 (**1986**) 15273-80.
62. **Hurley, T. D.**, ***et al.***, *J Mol Biol.* 239 (**1994**) 415-29.
63. **Ceccarelli, C.**, ***et al.***, *Biochemistry.* 43 (**2004**) 5266-77.
64. **Levin, I.**, ***et al.***, *Protein Sci.* 13 (**2004**) 1547-56.
65. **Esposito, L.**, ***et al.***, *Biochemistry.* 42 (**2003**) 14397-407.
66. **Bhonsle, H. S.**, ***et al.***, *J Proteome Res.* 11 (**2012**) 1391-6.
67. **Cheng, F. Y.**, ***et al.***, *J Proteome Res.* 8 (**2009**) 82-93.
68. **Patel, P.**, ***et al.***, *Int J Biol Macromol.* 63 (**2014**) 254-60.
69. **Habeeb, A. F.**, *Methods Enzymol.* 25 (**1972**) 457-64.
70. **Shashidhara, K. S., Gaikwad, S. M.**, *Int J Biol Macromol.* 44 (**2009**) 112-5.
71. **Gohel, S. D., Singh, S. P.**, *Int J Biol Macromol.* 56 (**2013**) 20-7.

72. **Wen, L.**, ***et al.***, *J Protein Chem.* 18 (**1999**) 677-86.
73. **Gote, M. M.**, *Enzyme and Microbial Technology.* 40 (**2007**) 1312-20.
74. **Sarni, F.**, ***et al.***, *Eur J Biochem.* 139 (**1984**) 259-65.
75. **Miyawaki, O.**, ***et al.***, *Enzyme and Microbial Technology.* 43 (**2008**) 495-9.
76. **Pandey, B.**, ***et al.***, *Protein Expr Purif.* 79 (**2011**) 197-203.
77. **Ma, Q. H.**, *J Exp Bot.* 61 (**2010**) 2735-44.
78. **Dalziel, K., Dickinson, F. M.**, *Biochem J.* 100 (**1966**) 34-46.
79. **Lauvergeat, V.**, ***et al.***, *Biochemistry.* 34 (**1995**) 12426-34.
80. Cook, P. F., Cleland, W. W., Enzyme Kinetics and Mechanism. Garland Science 2007
81. **Wratten, C., Cleland, W. W.**, *Biochemistry.* 4 (**1965**) 2442-51.
82. **Wratten, C. C., Cleland, W. W.**, *Biochemistry.* 2 (**1963**) 935-41.
83. **Cornish-Bowden, A.,** Fundamental of Enzyme Kinetics, *Portland Press* **(2004)** ISBN: 1855781581
84. **Vanni, A.**, ***et al.***, *Int J Biol Macromol.* 30 (**2002**) 41-5.
85. **Nishimura, M.**, ***et al.***, *Biosci Biotechnol Biochem.* 75 (**2011**) 1770-7.
86. **Baker, P. J.**, ***et al.***, *J Mol Biol.* 228 (**1992**) 662-71.
87. **Min, T.**, ***et al.***, *J Biol Chem.* 278 (**2003**) 50714-23.
88. **Youn, B.**, ***et al.***, *J Biol Chem.* 280 (**2005**) 12917-26.
89. **Rosell, A.**, ***et al.***, *J Biol Chem.* 278 (**2003**) 40573-80.
90. **Bomati, E. K., Noel, J. P.**, *Plant Cell.* 17 (**2005**) 1598-611.
91. **Eklund, H., Branden, C. I.,** in Biological Macromolecules and Assemblies, *John Wiley & Sons* **(1987)** 75-142. ISBN: 0471851426
92. **Ryde, U.**, *Protein Sci.* 4 (**1995**) 1124-32.
93. **Korkhin, Y.**, ***et al.***, *J Mol Biol.* 278 (**1998**) 967-81.
94. **Johansson, K.**, ***et al.***, *Chem Biol Interact.* 130-132 (**2001**) 351-8.
95. **Mckie, J. H.**, ***et al.***, *Biochim Biophys Acta.* 1202 (**1993**) 61-9.
96. **Noel, J. P.**, ***et al.***, *Curr Opin Plant Biol.* 8 (**2005**) 249-53.
97. **Sun, H. W., Plapp, B. V.**, *J Mol Evol.* 34 (**1992**) 522-35.
98. **Pire, C., Bonete, M. J.**, *Journal of Molecular Catalysis B: Enzymatic.* (**2009**) 261-5.
99. **Watanabe, S.**, ***et al.***, *J Biol Chem.* 280 (**2005**) 10340-9.
100. **Nishiyama, M.**, ***et al.***, *J Biol Chem.* 268 (**1993**) 4656-60.
101. **Clarke, A. R., Daffron, T. R.,** Nicotinamide Cofactor-Dependent Enzymes. *Academic Press* **(1998)**
102. **Eklund, H.**, ***et al.***, *J Biol Chem.* 257 (**1982**) 14349-58.
103. **Ehrig, T.**, ***et al.***, *Biochemistry.* 30 (**1991**) 1062-8.
104. **Eklund, H., Ramaswamy, S.**, *Cell Mol Life Sci.* 65 (**2008**) 3907-17.
105. **Ramaswamy, S.**, ***et al.***, *Biochemistry.* 33 (**1994**) 5230-7.
106. **Plapp, B. V.**, ***et al.***, *J Biol Chem.* 248 (**1973**) 3470-5.
107. **Stone, C. L.**, ***et al.***, *Biochemistry.* 38 (**1999**) 5829-35.
108. **Agarwal, P. K., Hammes-Schiffer.**, *Journal of American Chemical Society.* 122 (**2000**) 4803.

109. **Nagel, Z. D., Klinman, J. P.**, *Chem Rev.* 106 (**2006**) 3095-118.
110. **Lee, C., *et al.***, *Org Biomol Chem.* 11 (**2013**) 1127-34.
111. **Voordow, G.**, *Biochemistry.* 15 (**1976**) 3719-723.
112. **Gohel, D., Naseby**, *Biochemical Engineering Journal.* 35 (**2007**) 150-7.
113. **Royer, C. A.,** in Protein Stability and Folding, *Humana Press* 40 **(1995)** 65-90. ISBN: 0896033015
114. **Semisotnov, G. V., *et al.***, *Biopolymers.* 31 (**1991**) 119-28.
115. **He, B., *et al.***, *Int J Biochem Cell Biol.* 29 (**1997**) 1021-8.
116. **Moosavi-Movahedi, F., *et al.***, *Int J Biol Macromol.* 58 (**2013**) 66-72.
117. **Kumar, R. S., *et al.***, *IUBMB Life.* 59 (**2007**) 118-25.
118. **Rajaraman, K., *et al.***, *J Biol Chem.* 271 (**1996**) 27595-600.
119. **Poklar, N., *et al.***, *Biochemistry.* 36 (**1997**) 14345-52.
120. **Shashidhara, K. S., Gaikwad, S. M.**, *J Fluoresc.* 20 (**2010**) 827-36.
121. **Guo, D. M., *et al.***, *J Mol Evol.* 71 (**2010**) 202-18.
122. **Al-Karadaghi, S., *et al.***, *Acta Crystallogr D Biol Crystallogr.* 50 (**1994**) 793-807.
123. **Ramaswamy, S., *et al.***, *Protein Sci.* 5 (**1996**) 663-71.
124. **Theorell, H., Chance, B.**, *Acta Chem. Scand.* 5 (**1951**) 1127-44.
125. **Dolega, A.**, *Coord Chem Rev.* 251 (**2010**) 916-37.
126. **Andersson, P., *et al.***, *Eur J Biochem.* 139 (**1984**) 519-27.
127. **Hardman, M. J.**, *Biochem J.* 195 (**1981**) 773-4.
128. **Kleifeld, O., *et al.***, *Nat Struct Biol.* 10 (**2003**) 98-103.