STRUCTURAL STUDIES OF PLANT SECOISOLARICIRESINOL DEHYDROGENASE, PLANT VACUOLAR SORTING RECEPTOR AND REDUCTION POTENTIAL OF RUBREDOXIN

By

BUHYUN YOUN

A dissertation submitted in partial fulfillment of
the requirements for the degree of

DOCTOR OF PHILOSOPHY

WASHINGTON STATE UNIVERSITY
School of Molecular Biosciences

DECEMBER 2004

To the Faculty of Washington State University:

The member of the Committee appointed to examine the dissertation of BUHYUN YOUN find it satisfactory and recommended that it be accepted.

_____

Chair

_____

_____

_____

ACKNOELEDGEMENT

I would like to express my gratitude to my advisor, Professor ChulHee Kang, for his support, patience, and encouragement throughout my graduate studies. Fortunately, I met an advisor who always finds the time for listening to the little problems and roadblocks that unavoidably crop up in the course of performing research. His technical advice and never-ending enthusiasm to students were essential to the completion of this dissertation and has taught me innumerable lessons and insights on the workings of academic research in general. My thanks also go to the members of my major committee, Dr. Norman G. Lewis, Dr. John C. Rogers and Dr. Lisa M. Gloss for numerous helpful suggestions and support. I would like to thank all other past and present X-RAY lab members and many undergraduates for all their contribution they made to my research and education. The friendship with my classmate and Korean people in Pullman is much appreciated and has led to many interesting and good-spirited discussions relating to this research and life.

Last, but not least, I would like to thank my wife, SangJeong Jeong for her understanding and love during the past few years. Her support and encouragement were in the end what made this dissertation possible. My parents, JungYul Youn and JungJa Choi, my parents-in-law, HoYoung Jeong and JungSuk Wang, my sisters, Musook Youn and Dr. Heysook Youn, and my daughter Celine Youn receive my deepest gratitude and love for their dedication and enormous support.

STRUCTURAL STUDIES OF PLANT SECOISOLARICIRESINOL DEHYDROGENASE, PLANT VACUOLAR SORTING RECEPTOR AND REDUCTION POTENTIAL OF RUBREDOCXIN

Abstract

by BuHyun Youn, PhD
Washington State University
December 2004

Chair: ChulHee Kang

The knowledge of accurate molecular structures is a prerequisite for rational drug design and for structure based functional studies to aid the development of effective therapeutic agents and pharmaceutical drugs. In this regard, crystallography can reliably provide the answer to many structure related questions, from global folds to atomic details of bonding.

Here, three chapters are concentrated on the description of structure determinations and analyses of three different proteins, *Podophyllum* secoisolariciresinol dehydrogenase (SDH), an enzyme involved in formation of health-protecting lignans and in plant defense (Chapter I), *Clostridium* rubredoxin, a simple iron-sulfur (FeS_4) protein involved in oxidation-reduction reactions (Chapter II) and *Arabidopsis* vacuolar sorting receptor (VSR), a sequence specific lytic vacuole targeting protein (Chapter III).

Final product of a time-consuming structure solving process and a complete understanding of solved structures were able to provide in-depth ideas of how these macromolecules handle their tasks in the cell.

TABLE OF CONTENTS

Page

Dedication

This dissertation is dedicated to my parents and my wife.

CHAPTER ONE

CRYSTAL STRUCTURES OF *PODOPHYLLUM* SECOISOLARICIRESINOL DEHYDROGENASE (SDH)

Preface

Lignans and isoflavonoids represent very important class of phenylpropanoids. They have major roles in plant defense, helping hosts discriminate between biological pathogens and symbiotic non-pathogens and pharmacological roles, especially in cancer prevention. Despite the importance of these compounds in plant biology and in human health, systematic studies on the enzymes involved have not been done. In this chapter, we clarify the reaction mechanism of secoisolariciresinol dehydrogenase (SDH), which catalyzes the enantiospecific conversion of (−)-secoisolariciresinol into (−)-matairesinol. This is the last step in vascular plants to the precursor of the "mammalian" lignan, enterolactone, as well as serving as a later stage intermediate to the antiviral lignan, podophyllotoxin, and its semi-synthetic derivative, teniposide, etoposide and Etophos®.

Following manuscript was submitted to *Journal of Biological Chemistry* in November 2004. Due to size limitations, the reader is referred to the Supplement for detailed concept and mathematical background for the crystallographic experiment used in this chapter.

Crystal structures of apo-form, binary and ternary complexes of *Podophyllum* secoisolariciresinol dehydrogenase (SDH), an enzyme involved in formation of health-protecting lignans and in plant defense

Buhyun Youn[‡], Syed G.A. Moinuldin[§], Laurence B. Davin[§], Norman G. Lewis[§], and ChulHee Kang[‡]

From the [‡]School of Molecular Biosciences, Washington State University, Pullman, WA 99164-4660 and the [§]Institute of Biological Chemistry, Washington State University, Pullman, WA 99164-6340

Running Title: X-Ray Structure of Secoisolariciresinol Dehydrogenase

SUMMARY

(–)-Matairesinol is a central intermediate in the biosynthesis of numerous 8-8′ lignans, including that of the important antiviral agent, podophyllotoxin, and its semi-synthetic anticancer derivatives, etoposide, teniposide and Etopophos®. It is also a plant-derived precursor of the cancer-preventative "mammalian" lignan, enterolactone, formed in the gut following ingestion of, for example, high-fiber dietary foodstuffs. This “phytoestrogen” is formed by the action of secoisolariciresinol dehydrogenase, an NAD(H)-dependent oxidoreductase that catalyzes in *Podophyllum* species the enantiospecific conversion of (–)-secoisolariciresinol into (–)-matairesinol. This enzymatic step is also involved in the pathway to plicatic acid, an important plant defense molecule in western red cedar (*Thuja plicata*) heartwood. In order to understand the molecular basis of this reaction, the crystal structures of the apo-form, binary and ternary complexes of secoisolariciresinol dehydrogenase from *Podophyllum peltatum* were determined at 1.6, 2.8 and 2.0 Å resolution, respectively. The enzyme is a homo-tetramer, consisting of an α/β single domain monomer containing seven parallel β-strands flanked by eight α-helices on both sides. Its overall monomeric structure is very similar to that of the NAD(H)-dependent short-chain dehydrogenase/reductase family, with a conserved Asp47 forming a hydrogen bond with both hydroxyl groups of the adenine ribose of NAD(H), and thus specificity toward NAD(H) instead of NADPH. The highly conserved catalytic triad, consisting of Ser153, Tyr167 and Lys171, is adjacent to both the NAD^+ and substrate molecules, where Tyr167 functions as a general base.

FOOTNOTES AND ABBREVIATIONS

This research was supported in part by the National Institute of General Medical Sciences (to CHK and NGL), the United State Department of Agriculture (99-35103-8037), McIntire-Stennis, and the Murdock Charitable Trust. The cost of publication of this article was defrayed in part by the payment of page charges. This article must therefore be hereby-marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The abbreviations used are:

PCR; polymerase chain reaction

PDB; protein data bank

PLR;pinoresinol/lariciresinol Reductase

Pp; *Podophyllum peltatum*

r.m.s.d.; root mean square deviation

SDH; secoisolariciresinol dehydrogenase

SDR; short-chain dehydrogenases/reductases

INTRODUCTION

The 8-8′-linked lignans represent an abundant class of plant natural products, with important roles in human health protection, pharmacological applications, as well as in plant defense (1,2). Of these, the lignans, matairesinol (**1**), secoisolarici-resinol (**2**) (Fig. 1) or derivatives thereof can accumulate in high fiber foodstuffs, and are metabolized in humans following dietary ingestion to afford the cancer-preventative "mammalian" lignans, enterolactone (**3**) and enterodiol (**4**) (3,4). Matairesinol (**1**) is also metabolized further in various vascular plant species, such as *Podophyllum*, to generate the antiviral podophyllotoxin (**5**) (5), with the latter being used as a starting material in the semi-synthetic preparation of the anti-cancer compounds, teniposide (**6**), etoposide (**7**) and Etopophos® (**8**) (6). Additionally, matairesinol (**1**) derived lignans have important protective functions against various pathogenic organisms e.g., as part of the chemical arsenal present in western red cedar heartwood (7-11).

The biochemical pathway to (–)-matairesinol (**1a**) in *Podophyllum* and other plant species begins with dirigent protein mediated stereoselective coupling of two achiral molecules of *E*-coniferyl alcohol (**9**) to afford (+)-pinoresinol (**10a**) (Fig. 2) (12-15). The latter can undergo sequential enantiospecific reductions, catalyzed by the NADPH-dependent bifunctional pinoresinol-lariciresinol reductase (PLR)[1] (10,16,17), to afford initially (+)-lariciresinol (**11a**) and then (–)-secoisolariciresinol (**2a**). The molecular basis of this enantiospecificity was previously established by determination of the X-ray crystal structure (2.5 Å resolution) (18) and by comparison with an enantiospecifically distinct

PLR which catalyzes the opposite conversion, namely (−)-pinoresinol **10b**/(−)-lariciresinol **11b** into (+)-secoisolarici-resinol (**2b**) (18).

This study was directed toward establishing the crystal structure of secoisolariciresinol dehydrogenase (SDH), which catalyzes the enantiospecific conversion of (−)-secoisolarici-resinol (**2a**) into (−)-matairesinol (**1a**) (Fig. 2) (5,19). This is the last enzymatic step in vascular plants to the precursor of the "mammalian" lignan, enterolactone (**3**), as well as being involved in the pathway to the antiviral lignan, podophyllotoxin (**5**) (5), and thus to its semi-synthetic derivatives, teniposide (**6**), etoposide (**7**) and Etopophos® (**8**).

EXPERIMENTAL PROCEDURE

Metabolite Isolation—(–)Matairesinol (**1a**) was isolated from *Forsythia intermedia* stems as described in Umezawa *et al.* (20).

Expression and Purification of SDH— *SDH_Pp7*, cloned into an Invitrogen pTrcHis2-TOPO® TA vector, was transformed into TOP10 *E. coli* cells as previously described (19). Expression of SDH_Pp7 was induced by addition of isopropyl β-D-thiogalactopyranoside to 1 mM final concentration at mid-log phase (A_{600} = 0.5–0.7). The induced cell suspension cultures were grown for 12 h at 37 °C, with shaking at 250 rpm. The cells were then harvested by centrifugation (3,000 × *g* for 20 min). The SDH_Pp7-derived pellet was suspended in Buffer A [20 mM Tris-HCl (pH 8.0) containing EDTA (3 mM) and DTT (3 mM)], sonicated (5 × 10 s, Model 450 sonifier®, Branson Ultrasonics Co.), and centrifuged (20,000 × *g* for 40 min). The resulting supernatant was next subjected to ammonium sulfate precipitation, with the recombinant SDH_Pp7 precipitating between 70–80% ammonium sulfate. This fraction was reconstituted in 40 ml of Buffer A, desalted over a PD-10 column (Amersham Biosciences), then subjected to anion exchange column chromatography (Self Pack™ POROS® 10HQ, Applied Biosystems), pre-equilibrated in Buffer B [50 mM Tris-HCl (pH 7.5) containing EDTA (3 mM) and DTT (1 mM)], at a flow rate of 5 ml min^{-1}. Recombinant SDH_Pp7 was eluted using a NaCl step gradient (to 0.1, 0.2, 0.5 and 2.0 M, 50 ml each), and the corresponding fractions of interest (eluting at 0.1 M NaCl) were desalted and concentrated into Buffer B by ultra-filtration in an Amicon

8050 cell with a 10 kDa cut-off membrane (Millipore). The concentrated SDH_Pp7 was next applied to a MonoQ® HR16/10 (Amersham Biosciences) anion exchange column equilibrated in Buffer B at a flow rate of 3 ml min^{-1}, and eluted using a NaCl step gradient (to 0.05, 0.1, 0.2, 0.4 and 2.0 M; 20 ml each); the corresponding catalytically active fractions were eluted at 0.05 M NaCl. This SDH_Pp7 enriched fraction was then concentrated as above, the buffer exchanged to 5 mM Na phosphate, pH 6.8, and applied to a CHT-10 hydroxyapatite (BioRad) column (1 × 10 cm), equilibrated in the same buffer, at a flow rate of 3.5 ml min^{-1}. The column was eluted with a linear Na phosphate gradient (5–500 mM in 200 ml). SDH_Pp7 did not bind to the matrix and the flow-through fraction containing it was then concentrated as above with the buffer exchanged to Buffer B. The resulting SDH_Pp7 enriched protein fraction was then applied to a MonoQ™ GL10/100 anion exchange column (Amersham Biosciences) equilibrated in Buffer B at a flow rate of 2 ml min^{-1}, and eluted with a NaCl step gradient (0.05, 0.1, 0.2, 0.4 and 2.0 M; 20 ml each); the catalytically active SDH_Pp7 fraction eluted at 0.05 M NaCl. The SDH_Pp7 so obtained was then concentrated, with a final purity >99% as estimated by SDS-PAGE (Coomassie blue staining).

Molecular Mass Determination by Size Exclusion Chromatography and Multiangle Laser Light-Scattering— A TSK G3000SW (TosoHaas) column was pre-equilibrated in Buffer B followed by loading of the SDH_Pp7 solution (100 µl, 1 or 2 mg ml^{-1}) in the same buffer. The chromatography was performed at a flow rate of 1 ml min^{-1} and temperature of 22°C using an Acuflow series IV pump (Analytical Scientific Inst.) with the resulting eluant

passed in tandem through an UV detector (Gilson), a refractometer (Optilab DSP, Wyatt Tech.), and a multi-angle laser light-scattering (MALLS) detector (Dawn EOS, Wyatt Tech.), respectively, with the scattering data analyzed using the software ASTRA (Wyatt Tech.). Relative weight-average molecular masses were determined from the scattering data using the Zimm fitting method, in which $K^*c/R(Q)$ is plotted against $sin^2(Q/2)$, where Q is the scattering angle, $R(Q)$ is the excess intensity (I) of scattered light at the angle Q, c is the concentration of the sample, and K^* is a constant equal to $4\pi^2n^2(\mathrm{d}n/\mathrm{d}c)^2/\lambda_0^4N_A$ (where n = solvent refractive index, $\mathrm{d}n/\mathrm{d}c$ = refractive index increment of scattering sample, λ_0 = wavelength of scattered light and N_A = Avogadro's number). Extrapolation of the Zimm plot to zero angle gave an estimate of the weight-averaged molecular mass (Mw), where the latter is defined as:

$$Mw = \frac{\sum(ciMi)}{\sum ci}$$

for c moles of i different species with individual molecular weights Mi.

Crystallization of SDH_Pp7—For crystalli-zation of the apo and complex forms, a solution of purified SDH_Pp7 (52 mg ml^{-1}) in 20 mM Tris-HCl (pH 8.0) containing 1 mM EDTA and 1 mM DTT, was prepared. Crystallization trials were performed using the hanging drop vapor diffusion method at two temperatures (277 and 293 K). Apo-SDH_Pp7 crystals were obtained by mixing the above protein solution (1.5 µl) with an equal volume of reservoir solution containing 30% (w/v) PEG 4000, 0.1 M Tris-HCl (pH 8.5) and 0.2 M sodium acetate trihydrate. Diffraction quality crystals usually appeared after ten days and

larger crystals with dimensions of approximately 0.3 × 0.4 × 0.7 mm were obtained after two weeks. The crystal of apo-SDH belongs to the orthorhombic space group, $C222_1$ (a = 107.34, b = 133.56, c = 69.35 Å), with two molecules in the asymmetric unit. The binary complex (SDH_Pp7/NAD^+) and ternary complex [SDH_Pp7/NAD^+/(–)-matai-resinol (**1a**)] crystals were also produced under the same conditions except for addition of 10 mM NAD^+ and 5 mM NAD^+/2 mM (–)-matairesinol (**1a**), respectively. Both binary and ternary complexes crystallized in an orthorhombic space group, F222 with corresponding unit cells of a = 58.51, b = 118.91, c = 132.00 Å and a = 57.35, b = 118.74, c = 131.25 Å, respectively. The apo-form (1.6 Å resolution), the binary complex (2.8 Å) and the ternary complex (2.0 Å) data were collected from the Berkeley Advanced Light Source (ALS, beam line 8.2.1), Rigaku R-AXIS IV^{++}/RU H2R and Rigaku Saturn 92/FR-E, respectively, at a temperature of 100 K.

Structural Solution and Refinement—The structure of apo-SDH_Pp7 was solved by the molecular replacement method using a coordinate of the Rv2002 Gene Product from *Mycobacterium tuberculosis* (PDB code 1NFF) (21) and the software package AmoRe (22). The rigid-body refinement of the initial position was carried out using 15.0 to 3.0 Å resolution data, and gave an R-value of 42%. After several cycles of positional and temperature factor refinements, and a series of simulated annealing omit maps, most residues were fitted against the electron density although the electron density corresponding to ten N-terminal residues was not visible from an early stage of refinement. The binary and ternary complexes of SDH_Pp7 were again solved by the molecular replacement

method but now using the apo-SDH_Pp7 coordinates. The final R-factors (Table I) for the apo form, as well as the binary and ternary complexes of SDH_Pp7, were 19.7% (R_{free} = 22.6% for the random 5% data), 20.1% (R_{free} = 22.1% for the random 5% data) and 20.0% (R_{free} = 22.9% for the random 5% data), respectively. The number of reflections above 2□level for the apo-form were 63,119 (99.7% completeness) between 10.0 and 1.6 Å resolution. The crystals of the binary and ternary complexes did not diffract as well as the apo-form and gave reflection numbers of 4,585 (99.7% completeness) between 10.0 and 2.8 Å resolution and 11,716 (79.1% completeness) between 10.0 and 2.0 Å resolution. The root mean square deviations (r.m.s.d.) (from ideal geometry) of the final coordinates corresponding to apo, binary and ternary complexes are 0.04, 0.03 and 0.05 Å for bonds and 2.9, 3.2 and 3.3º for angles, respectively. All SDH_Pp7 coordinates have been deposited in the PDB.

RESULTS AND DISCUSSION

Overall Structures—Recombinant *P. peltatum* secoisolariciresinol dehydrogenase (SDH_Pp7) was crystallized in its apo, binary and ternary complex forms, with crystals of the latter two complexes obtained by mixing SDH_Pp7 with NAD^+ and NAD^+/(–)-matairesinol (**1a**), respectively, instead of diffusing ligands into the crystals; (–)-matairesinol (**1a**) was isolated from *F. intermedia* as previously described (20).

The structure of the apo-form of SDH_Pp7 was determined at 1.6 Å resolution by molecular replacement using coordinates of the Rv2002 Gene Product from *Mycobacterium tuberculosis* (1NFF), which has the highest sequence similarity to SDH_Pp7 of sequences in PDB (21). In turn, the binary and ternary complex structures of SDH_Pp7 were determined at 2.8 Å and 2.0 Å resolution, respectively, using the coordinates of the deduced apo-form SDH_Pp7 structure. The apo-form was found to have two tightly associated molecules in its asymmetric unit, and a crystallographic symmetry operation assembled these two molecules into a tetrameric unit in the crystal lattice; both molecules in an asymmetric unit were virtually superimposable with a r.m.s.d. of 0.87 Å between the corresponding Cα carbons. Additionally, the oligomeric status of SDH_Pp7 as a tetramer was verified in solution by a multi-angle laser light scattering (MALLS) experiment; i.e., MALLS analyses of solutions of purified SDH_Pp7 (1 and 2 mg ml^{-1}, respectively) revealed that it was mainly of intrinsic tetramer character (Fig. 3). A different crystal form of the P1 space group was also obtained, which contained the same type of tetramer as the

asymmetric unit, but this was not examined further since the corresponding diffraction data was of relatively low resolution (~3.0 Å).

The SDH_Pp7 monomer contains a single α/β domain structure with a characteristic NAD(H) binding motif (Fig. 4*A*), which consists of seven parallel β-strands (βA–βG) with both sides surrounded by eight α-helices (αA–αH). These connect the strands like those typically observed in the Rossman fold, i.e. for proteins having NAD(P)H and FMN binding signatures, such as oxidoreductases and transferases. Of these, the two α-helices (αF and αG) between the 6th and the 7th β-strands (βF and βG) ensure the overall topology to be βαβαβαβαβαβααβα. Furthermore, since this topology also showed a close resemblance to those of short-chain dehydrogenases/reductases (SDR), a detailed comparison using a Dali search (23) was carried out to identify structural homologues. The highest match was to the 3-hydroxyacyl-CoA dehydrogenase from rat brain (1E6W) with a *Z*-score of 32.0, this being followed by a 3-α,20-β-hydroxysteroid dehydrogenase from *Streptomyces exfoliatus* (2HSD) with a Z-score of 31.8 and a glucose dehydrogenase from *Bacillus megaterium* (1GCO) of 29.1. On the other hand, an amino acid sequence analysis (24) of the PDB revealed that the Rv2002 gene product from *M. tuberculosis* (1NFF) shows the highest similarity (33.8%) to SDH_Pp7, followed by the 3-α,20-β-hydroxysteroid dehydrogenase from *S. exfoliatus* (2HSD; 31.8%), a R-specific alcohol dehydrogenase from *Lactobacillus brevis* (1NXQ; 30%) and the rat brain 3-hydroxyacyl-CoA dehydrogenase described above (1E6W; 27.4%) (Fig. 5). In terms of overall topology of the secondary structural elements, however, the 3-α, 20-β-hydroxysteroid dehydrogenase from *S.*

exfoliatus (2HSD) is most similar to that observed in the structure of SDH_Pp7, although the location and size of the αE was rather different in both structures.

Furthermore, the sequence alignment of the various dehydrogenases and reductases (Fig. 5) revealed that they have several deletions and insertions when compared to SDH_Pp7, with some missing the αH region. Especially the highly disordered region containing ten N-terminal amino acids in SDH_Pp7 cannot be aligned relative to the above listed dehydrogenases and reductases. The longest region of highest sequence similarity between SDH_Pp7 and the various SDRs is the area between Glu234 and Gly261, which spans most of the αG and βG region. Significantly, the level of sequence similarity among the various SDRs is very low in the region between residues Leu205 and Ala233, and this region also happens to contain a major insertion site in SDH_Pp7. This inserted area spans the αF including its N-; C-flanking loops, and is involved in an intermolecular helix-helix interaction with the αH of another molecule thereby forming the substrate-binding pocket. Additionally, the loop area connecting the βE and αE regions displays a low level of sequence similarity and some of the residues in this loop are in close contact with the substrate, such as Ile154, Ser155 and Ser156. Indeed, these observations are in agreement with a previous finding that some of the unique insertions or local heterogeneities for each SDR are to support the binding of their specific substrates (25).

In terms of the substrate binding pocket and the catalytic mechanism, all of these enzymes have a conserved catalytic triad containing Tyr(Y), Lys(K) and Ser(S) (26). In SDH_Pp7, the Tyr167 and Lys171 are located near the carboxyl end of the αE, whereas Ser153 is located in the loop connecting both the αE and βE regions. In its apo form, the

hydroxyl group of Ser153 is hydrogen bonded to the phenolic group of Tyr167, with the protonated amino group of Lys171 being located 4.0 Å from the Tyr167 phenolic group. (Fig. 6*A* and inset)

Lastly, in addition to the highly disordered N-terminal amino acids, three other areas show either partial disorder or higher temperature factors than the rest of the SDH_Pp7 molecule (Fig. 7); i.e., residues 48–61, which are around the αB region exposed to the surface, and residues 201–215 and 272–277. As clearly shown in Fig. 7, the area of residues 48–61 shows significantly reduced temperature factors upon complex formation reflecting its relationship with cofactor binding as discussed later. The other two areas of residues 201–215 and 272–277 are the least conserved regions among various SDRs and alcohol dehydrogenases (Fig. 5). Especially residues 200–218 are part of the above-mentioned inserted αF area that is in close proximity to the substrate binding pocket, and probably determine a unique specificity for (−)-secoisolariciresinol (**2a**); this local area also shows the largest conformational difference among the apo-form, binary and ternary complexes as discussed later.

Tetramer Formation—The individual subunits of SDH_Pp7 are arranged around two intersecting 2-fold axes forming a D_2 symmetry among the four subunits (Fig. 4*B*), thereby stabilizing the tetramer through tight intermolecular interactions among the four molecules. In the apo-form, the two monomers are in an asymmetric unit, related by the non-crystallographic 2-fold axis, and show a strong intermolecular interaction via a four-helix bundle motif formed by two αD and two αE regions (between violet and brown, yellow

and green in Fig. 4*B*). Accordingly, this monomer-monomer interface is mainly stabilized by a large number of hydrophobic interactions between the αD helices and the six side-chain hydrogen bonds between the αE helices. In both the binary and ternary complexes, however, this pseudo 2-fold axis becomes a crystallographic 2-fold axis. The second type of dimer that is related to the crystallographic two-fold axis in both apo- and complex forms of SDH_Pp7 is also stabilized via hydrogen bonding and hydrophobic interactions, especially between the αH of one subunit and αF of the other (between violet and green, yellow and brown in Fig. 4*B*). This exchange interaction of the αH between monomers is comparable to a domain swapping or an arm exchange. In particular, the ten C-terminal residues of one SDH_Pp7 monomer become inserted into a relatively hydrophobic cleft of the other thereby forming two intermolecular hydrogen bonds. Furthermore, near the point where the C-terminal arm leaves each monomer, there is a Pro267, i.e., a typical proline residue found in hinge regions of other domain swapping systems.

Overall, however, there are no major conformational differences in the backbone structures among the apo-form, and the binary and the ternary complexes (Fig. 8*A*), except for the area of residues 201–222 and C-terminal residues 273–277 (boxed areas in Fig. 8*A*). Furthermore, the Cα carbons of the binary and ternary complexes are superimposable with a r.m.s.d. of 0.71 Å; the apo-form can also be made superimposable with the binary and ternary complexes of 1.22 and 1.08 Å respectively, showing somewhat larger changes than between the binary and ternary complexes.

Cofactor Binding—The electron density corresponding to NAD^+ was clearly identifiable from the initial *Fo-Fc* map (Fig. 9*A*). In the apo-form, this cofactor binding site is filled with water molecules thereby forming a hydrogen bond network (Fig. 6*A*, inset) with the side-chains of the lined residues, particularly that of the highly conserved catalytic triad, Ser153, Tyr167 and Lys171 (Fig. 6*A*). As can be seen, the nicotinamide ring assumes a *syn*-conformation whereas the adenine ring has an *anti*-conformation (Fig. 6*B*). As in other NAD(P)H dependent oxidoreductases, the carboxy ends of βA, βB, βD, βE, βF, and the loop connecting βF and αF, also form a cleft for cofactor binding. In this position, the nicotinamide is stabilized by a stacking interaction with the Pro197, and its carbonyl oxygen is interacting with the backbone amide nitrogen of Val200 (Fig. 6*B*). The side chains of two residues, Tyr167 and Lys171, of the catalytic triad are within hydrogen bonding distance of the nicotine amide ribose, O2′ and O3′ respectively, as shown (see inset) thereby possibly fixing the position of the nicotinamide ring during catalysis and only permitting hydride transfer for the substrate to take up the pro-(*S*)-hydride form in the conversion to NADH. Additionally, the phenolic group of Tyr167 is located close (4.0 Å) to the C4 atom of the nicotine amide ring, whereas the side chain of Asp47 is within hydrogen bonding distance to the O3′ moiety of the adenine ribose and the N3 atom of the adenine ring (Fig. 6*B*). As mentioned above, the small loop following βB that contains this Asp47 residue and the adjacent αB thus become less flexible after NAD^+ is bound as reflected in the corresponding temperature factors (Fig. 7).

Like other typical SDRs, SDH_Pp7 has a glycine-rich motif, GxxGxxG, at the first β– α – β unit (^{23}GGAGGIG^{29}), which is known to participate in binding of the

pyrophosphate group of NAD^+ through a helical dipole of αA. The pyrophosphate group of the NAD^+ in particular is within hydrogen bonding distance to the backbone amide nitrogen of residue Ile28, thereby N-capping it and compensating for the helix macro dipole (27). All the amino acids in this tight turn between βA and αB thus show relatively high temperature factors probably facilitating their interaction with NAD^+ via conformational flexibility. A preference for NADH over NADPH is achieved accordingly by charge repulsion resulting from the highly conserved Asp47 in the pocket that is normally occupied by the 2′-phosphate group of NADPH (25). This acidic amino acid in βB and Asp47 is often found in SDR enzymes that preferentially bind NAD(H); i.e., to form a hydrogen bond to both hydroxyl groups of the adenine ribose of NAD(H) (25). Significantly, all of the residues involved in NAD^+ binding are highly conserved among SDH and SDRs from various species.

Substrate Binding of SDH and Catalytic Mechanism—The exact conformation of the bound (−)-matairesinol (**1a**) product and NAD^+ molecules were clearly defined in the experimental electron density map (Fig. 9*B*). In the observed arrangement of the tetrameric form, a binding pocket for both the substrate and cofactor of the neighboring molecule in the asymmetric unit was located at the similar position as observed in other SDR enzymes (Fig. 8*B*).

A deep groove containing the substrate-binding pocket is formed by three loops that connect βD and αD, βE and αE, βF and αF, respectively, as well as the C-terminal helix (αH) of the adjacent subunit. In particular, the residues corresponding to one of these loops

(200–213) and the αH (271–277) region have high temperature factors, these potentially serving as a gate for the substrate and helping it to bind in the proper orientation (Fig. 7). In this way, the substrate binding-pocket is lined with hydrophobic residues, reflecting the non-polar nature of the substrate. Both phenolic moieties of (–)-matairesinol (**1a**) and one of its methoxyl groups are surrounded by hydrophobic residues and anchored to the side chain hydroxyl groups of Ser104 and Ser164 and the backbone of Pro197 through hydrogen bonds (Fig. 6*C*). The B-face of the nicotine amide ring is open to the cleft with the C4 atom of being ~5 Å from the target hydroxyl group of (–)-secoisolariciresinol (**2a**) [shown with (–)-matairesinol (**1a**) in Fig. 6*C*]. Consequently, the nicotinamide and the substrate are in the proper orientation for the well-established B-face specific hydride transfer to C4 from the corresponding substrate reaction center. This substrate-binding pocket also contains the conserved catalytic triad, Ser153, Tyr167 and Lys171, and interestingly, upon NAD^+ and substrate binding, no change was detected in either the conformation or the position of these amino acid residues.

The Lys171 residue lowers the pKa of the phenolic hydroxyl group of the Tyr167 in the catalytic triad together with the positively charged NAD^+ (25,28). The Ser153 residue then shares its proton with the phenolic anionic group of Tyr167 (Fig. 6*C*, inset), and, in this way, the latter can serve as a general base in substrate deprotonation during catalysis (25,28). Concomitant deprotonation of the (–)-secoisolariciresinol (**2a**) is then presumed to occur via the phenolic anion of Tyr167 with hydride transfer to NAD^+, followed by nucleophilic attack to form the (–)-lactol intermediate (**12a**) from (–)-secoisolariciresinol (**2a**). Subsequent dehydrogenation of the (–)-lactol (**12a**) can then occur by the same

process involving Tyr167 as before and a newly bound NAD^+ molecule to afford the dibenzyl furanone, (–)-matairesinol (**1a**).

In this regard, only the one enantiomeric substrate, i.e., (–)-secoisolariciresinol (**2a**), was able to be modeled into this substrate-binding pocket using the same hydrogen bonds between the backbone and the two phenolic moieties observed in the quaternary (–)-matairesinol (**1a**) complex structure. In this modeled position, the target hydroxyl groups of the (–)-secoisolariciresinol (**2a**) are in the proper orientation and distance from the functional groups, but the corresponding hydroxyl groups of the stereoisomer, (+)-secoisolariciresinol (**2b**), cannot be fitted using the same constraints of the hydrogen bonds, this in turn perhaps explains the enantiospecificity of SDH_Pp7.

Concluding Remarks—In planta, matairesinol (**1**) is a central precursor in the biosynthesis of numerous lignans, including the important antiviral and anticancer agent, podophyllotoxin (6). In order to comprehensively understand the molecular basis of these multi-step enzymatic reactions and the structure/function relationships of the participating enzymes, we are systematically characterizing all of the enzymes in the biochemical pathway to this "phytoestrogenic" lignan. We previously employed X-ray crystallography to determine the structure of the enantiospecific enzyme in the same biosynthetic pathway; i.e., the bifunctional NADPH-dependent pinoresinol-lariciresinol reductases (PLR) (18), which converts pinoresinol (**10**) into lariciresinol (**11**) and then secoisolariciresinol (**2**). Based on detailed structural analyses and site-directed mutagenesis, the critical residues in

the enzyme active site of PLRs were determined, as well as that of the two related enzymes, phenylcoumaran benzylic ether reductase (PCBER) and isoflavone reductase (IFR) (18).

In this paper, we now report the crystal structure of the NAD(H)-dependent secoisolarici-resinol dehydrogenase (SDH), which catalyzes the next step, i.e., the enantiospecific conversion of (–)-secoisolariciresinol (**2a**) into (–)-matairesinol (**1a**). As noted above, the crystal structure of SDH_Pp7 has an overall structure quite similar to members of the SDR family, many of which, due to their relation to tumor growth, have been very well studied (26). The continuously increasing members of this SDR family perform diverse functions in both prokaryotes and eukaryotes and all appear to have the α/β single domain structure, i.e., including the dinucleotide-binding Rossmann fold and the homologous catalytic triad which contains a strictly conserved tyrosine residue. Interestingly, in spite of their similar folding (α/β single domain structure) and catalytic mechanisms, the substrates of the SDRs display very distinct substrate versatilities and can range from being rigid to flexible, from containing many polar groups to only a few, from being large to being small. Variation in the amino acid composition of the substrate-binding pocket is also well known for specificity of binding for various substrates, such as steroids, prostaglandins, sugars and alcohols (26). As for many of the other SDR enzymes, the SDH_Pp7 is a tetramer both in solution and in the crystal lattice state and possesses a highly conserved catalytic triad, i.e., in this case, Ser153, Tyr167 and Lys171. The triad is thus able to perform a concerted catalytic reaction in the conversion of (–)-secoisolariciresinol (**2a**) to (–)-matairesinol (**1a**).

Comprehensive understanding of the detailed reaction mechanisms of the participating enzymes in the phytoestrogenic lignan biosynthetic pathway thus provides crucial insight into how this complex family of plant natural products (the 8-8′ linked lignans) is formed. As we continue to add to our detailed knowledge of these enzymatic reactions, future work will be directed to metabolic engineering levels of both secoisolariciresinol (**2**) and matairesinol (**1**) in plant foodstuffs such as vegetables, grains, fruits or as supplements for processed food items; i.e., in order to provide a facile source of these beneficial lignans in staple dietary foodstuffs for the benefits of humanity, or to provide an increased supply of the antiviral/anticancer lignans such as podophyllotoxin (**5**).

REFERENCE

1. Lewis, N. G., and Davin, L. B. (1999) in *Comprehensive Natural Products Chemistry* (Barton, Sir D. H. R., Nakanishi, K., and Meth-Cohn, O., eds) Vol. 1, pp. 639-712, Elsevier, London
2. Davin, L. B., and Lewis, N. G. (2003) *Phytochemistry Reviews* **2**, 257-288
3. Adlercreutz, H., Höckerstedt, K., Bannwart, C., Hämäläinen, E., Forsis, T., and Bloigu, S. (1988) in *Progress in Cancer Research and Therapy* (Bresciani, F., King, R. J. B., Lippman, M. E., and Raynaud, J.-P., eds) Vol. 35 (Hormones and Cancer 3), pp. 409-412, Raven Press, Ltd., New York, NY
4. Adlercreutz, H., Mousavi, Y., Clark, J., Höckerstedt, K., Hämäläinen, E., Wähälä, K., Mäkelä, T., and Hase, T. (1992) *J. Steroid Biochem. Molec. Biol.* **41**, 331-337
5. Xia, Z.-Q., Costa, M. A., Proctor, J., Davin, L. B., and Lewis, N. G. (2000) *Phytochemistry* **55**, 537-549
6. Canel, C., Moraes, R. M., Dayan, F. E., and Ferreira, D. (2000) *Phytochemistry* **54**, 115-120
7. Gardner, J. A. F., Barton, G. M., and MacLean, H. (1959) *Can. J. Chem.* **37**, 1703-1709
8. Gardner, J. A. F., MacDonald, B. F., and MacLean, H. (1960) *Can . J. Chem.* **38**, 2387-2394
9. Gardner, J. A. F., Swan, E. P., Sutherland, S. A., and MacLean, H. (1966) *Can. J. Chem.* **44**, 52-58

10. Fujita, M., Gang, D. R., Davin, L. B., and Lewis, N. G. (1999) *J. Biol. Chem.* **274**, 618-627

11. Gang, D. R., Fujita, M., Davin, L. B., and Lewis, N. G. (1998) in *Lignin and Lignan Biosynthesis* (Lewis, N. G., and Sarkanen, S., eds) Vol. 697, pp. 389-421, ACS Symposium Series, Washington, DC

12. Davin, L. B., Wang, H.-B., Crowell, A. L., Bedgar, D. L., Martin, D. M., Sarkanen, S., and Lewis, N. G. (1997) *Science* **275**, 362-366

13. Halls, S. C., and Lewis, N. G. (2002) *Biochemistry* **41**, 9455-9461

14. Halls, S. C., Davin, L. B., Kramer, D. M., and Lewis, N. G. (2004) *Biochemistry* **43**, 2587-2595

15. Gang, D. R., Costa, M. A., Fujita, M., Dinkova-Kostova, A. T., Wang, H.-B., Burlat, V., Martin, W., Sarkanen, S., Davin, L. B., and Lewis, N. G. (1999) *Chem. Biol.* **6**, 143-151

16. Chu, A., Dinkova, A., Davin, L. B., Bedgar, D. L., and Lewis, N. G. (1993) *J. Biol. Chem.* **268**, 27026-27033

17. Dinkova-Kostova, A. T., Gang, D. R., Davin, L. B., Bedgar, D. L., Chu, A., and Lewis, N. G. (1996) *J. Biol. Chem.* **271**, 29473-29482

18. Min, T., Kasahara, H., Bedgar, D. L., Youn, B., Lawrence, P. K., Gang, D. R., Halls, S. C., Park, H., Hilsenbeck, J. L., Davin, L. B., Lewis, N. G., and Kang, C. (2003) *J. Biol. Chem.* **278**, 50714-50723

19. Xia, Z.-Q., Costa, M. A., Pélissier, H. C., Davin, L. B., and Lewis, N. G. (2001) *J. Biol. Chem.* **276**, 12614-12623

20. Umezawa, T., Davin, L. B., and Lewis, N. G. (1991) *J. Biol. Chem.* **266**, 10210-10217

21. Yang, J. K., Park, M. S., Waldo, G. S., and Suh, S. W. (2003) *Proc. Natl. Acad. Sci., USA* **100**, 455-460

22. Navaza, J. (1994) *Acta Crystallogr.* **A50**, 157-163

23. Holm, L., and Sander, C. (1993) *J. Mol. Biol.* **233**, 123-138.

24. Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) *Nucleic Acids Res.* **25**, 3389-3402

25. Powell, A. J., Read, J. A., Banfield, M. J., Gunn-Moore, F., Yan, S. D., Lustbader, J., Stern, A. R., Stern, D. M., and Brady, R. L. (2000) *J. Mol. Biol.* **303**, 311-327.

26. Duax, W. L., Griffin, J. F., and Ghosh, D. (1996) *Curr. Opin. Struct. Biol.* **6**, 813-823

27. Barbosa Pereira, P. J., Macedo-Ribeiro, S., Párraga, A., Pérez-Luque, R., Cunningham, O., Darcy, K., Mantle, T. J., and Coll, M. (2001) *Nat. Struct. Biol.* **8**, 215-220

28. Breton, R., Housset, D., Mazza, C., and Fontecilla-Camps, J. C. (1996) *Structure* **4**, 905-915

Table 1. Crystallographic data for the SDH-Pp7 apo, binary and ternary complex forms.

	Apo	Binary[a]	Ternary[b]
Data			
Wavelength (Å)	1.0332	1.54	1.54
Resolution (Å)	50 –1.6	50 – 2.8	50 – 2.0
Space group	$C222_1$	F222	F222
	a = 107.34	a = 58.51	a = 57.35
Cell dimensions (Å)	b = 133.56	b = 118.91	b = 118.74
	c = 69.35	c = 132.00	c = 131.25
Asymmetric unit	2 molecules	1 molecule	1 molecule
Total observations	259,713	20,597	70,809
Unique reflections	93,528	5,752	12,187
Completeness (%)	99.7	99.7	79.1
R_{sym}[c,d]	4.0 (13.6)	5.7 (11.4)	5.0 (12.6)
Refinement			
Resolution (Å)	10 – 1.6	10 – 2.8	10 –2.0
Number of reflections	63,119	4,585	11,716
R_{cryst}[e]	19.7	20.1	20.0
R_{free}[f]	22.6	22.1	22.9
rmsd bonds (Å)	0.04	0.03	0.05
rmsd angles (°)	2.9	3.2	3.3
Number of atoms			
Protein and ligand	3972	2030	2056
Water	288	28	127

[a] SDH_Pp + NAD^+

[b] SDH_Pp7 + NAD^+ + matairesinol

[c] Numbers in parentheses refer to the highest shell.

[d] $R_{sym}=\Sigma|I_h - \langle I_h \rangle|/\Sigma I_h$, where $\langle I_h \rangle$ is the average intensity over symmetry equivalent reflections.

[e] $R_{cryst} = \Sigma|F_{obs} - F_{calc}|/\Sigma F_{obs}$, where summation is over the data used for refinement.

[f] R_{free} was calculated as for R_{cryst} using 5% of the data that was excluded from refinement.

Figure legends

FIG. 1. Selected examples of various 8-8′ linked lignans with roles in human health protection, pharmaceuticals and plant defense.

FIG. 2. Biochemical pathway to (−)-matairesinol (**1a**) from *E*-coniferyl alcohol (**9**) in *Podophyllum* sp., where SDH catalyzes the enantiospecific conversion of (−)-secoisolariciresinol (**2a**) into (−)-matairesinol (**1a**) via intermediacy of (−)-lactol (**12a**). [Opposite enantiomers (structures not shown) are depicted in text as **1b**, **2b** and **10b**-**12b**.]

FIG. 3. Multi-angle laser light scattering elution profile of SDH_Pp7 revealing tetrameric status. Elution profiles are shown as molecular weight versus elution volume. The solid lines represent changes in refractive index on an arbitrary scale that is proportional to protein concentration. Dots indicate calculated molecular masses. [For elution details, see Experimental Procedures.]

FIG. 4. Crystal structures of SDH_Pp7. *A*, Structural element distribution of the monomeric form. Secondary structure elements have been numbered sequentially as αA–αH, βA–βG for the common helices and strands, respectively. N and C refer to N- and C-terminal regions, respectively. *B*, Arrangement of the tetrameric SDH_Pp7 in a D_2 symmetry. The pairs of yellow and brown molecules (or green and violet) are related by

the pseudo 2-fold axis, which becomes a crystallographic 2-fold axis in both binary and ternary complex crystals.

FIG. 5. Amino acid sequence alignment comparison of SDH_Pp7, Rv2002 from *Mycobacterium tuberculosis* (1NFF), 3α,20-β-hydroxysteroid dehydrogenase from *Streptomyces exfoliatus* (2HSD), R-specific alcohol dehydrogenase from *Lactobacillus brevis* (1NXQ), glucose dehydrogenase from *Bacillus megaterium* (1GCO) and 3-hydroxyacyl-CoA dehydrogenase from rat brain (1E6W). Secondary structural elements of SDH_Pp7 are highlighted in colored bars.

FIG. 6. Structures of the substrate-binding pocket of SDH_Pp7 in apo, binary and ternary complexes. *A*, Apo-form: the catalytic triad residues, Ser153 (brown), Tyr167 (violet) and Lys171 (pink) form a hydrogen bond network with each other and water molecules (shown as red circles). *B*, Binary complex with bound NAD^+ (blue), where the latter has its nicotinamide ring in a *syn*-conformation and the adenine ring in an *anti*-conformation. The Tyr167 phenolic and Lys171 protonated amino groups are hydrogen bonded to 2′ and 3′ OH respectively. ***C***, Ternary complex with NAD^+ (blue) and (–)-matairesinol (**1a**) (brown).

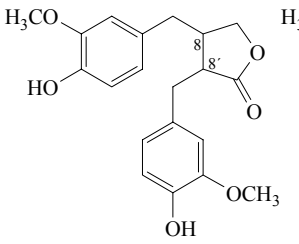
FIG. 7. Backbone temperature factor plots of SDH_Pp7 crystal structures; apo-form (red), binary (blue) and ternary (green) complexes. Besides N- and C-terminal residues, the residues 201–215 and 272–277 show the temperature factors of their corresponding Cα

atoms above 40 $Å^2$. In the two complex structures, the residues 48-61 have significantly lower temperature factors compared to that of the apo-form.

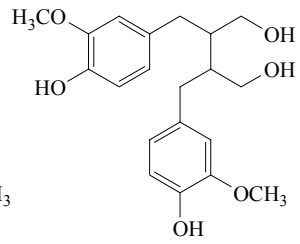
FIG. 8. SDH complexes. *A*, Superimposed view of Cα positions of apo-form (purple), as well as binary (yellow) and ternary (blue) complexes of SDH_Pp7. Two areas of residues 201–222 and 272–277 show differences among structures. *B*, The position of NAD^+ and product in the SDH_Pp7 tetramer. In each case, NAD^+ and (–)-matairesinol (**1a**) are depicted in violet and dark brown colors, respectively.

FIG. 9. Difference (*Fo-Fc*) Fourier maps of *A*, binary and *B*, ternary complexes of SDH_Pp7. NAD^+ and (–)-matairesinol (**1a**) have been modeled in. The figures were prepared using Quanta (Accelrys).

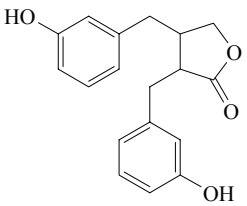
Fig. 1



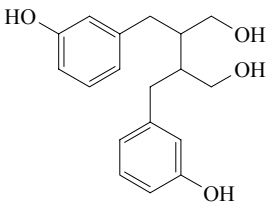
Matairesinol (**1**)



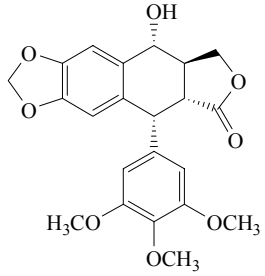
Secoisolariciresinol (**2**)



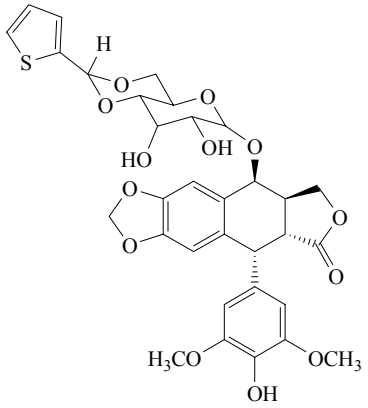
Enterolactone (**3**)



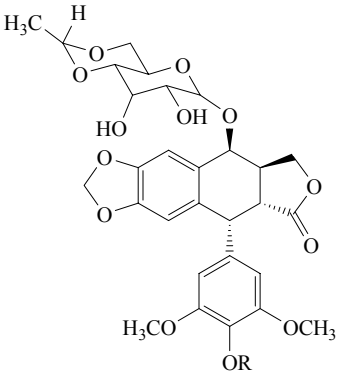
Enterodiol (**4**)



(−)-Podophyllotoxin (**5**)



Teniposide (**6**)



R = H, Etoposide (**7**)
R = $PO(OH)_2$, Etopophos® (**8**)

Fig. 2

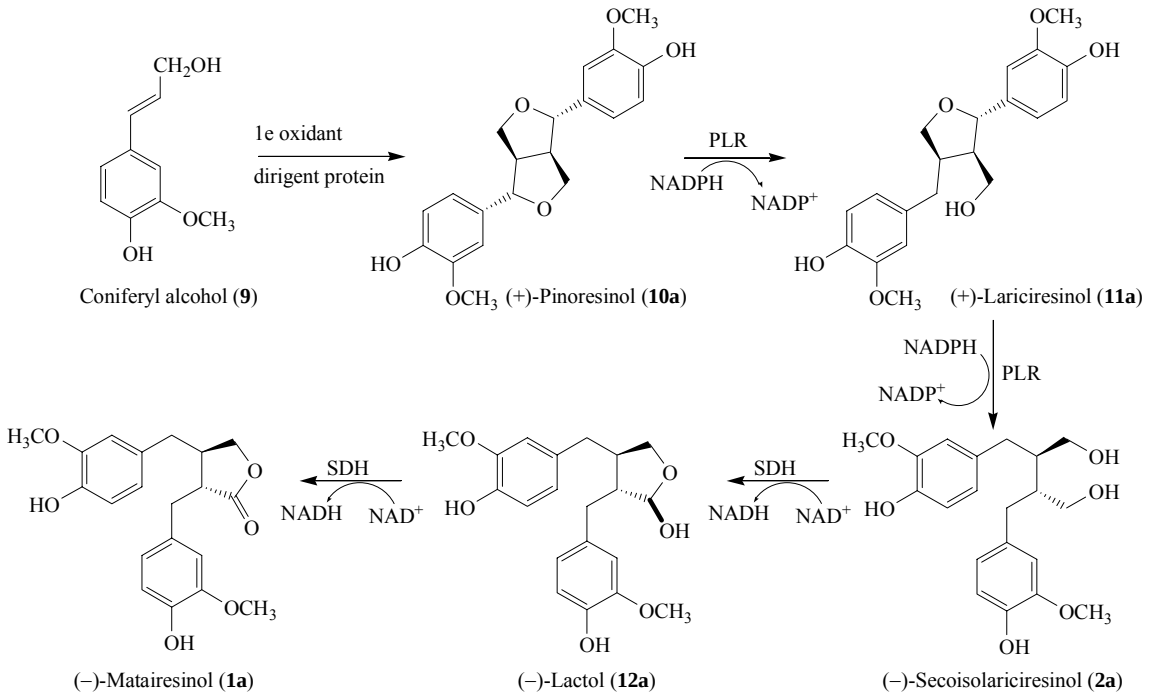

CH2OH
OCH3
OH
Coniferyl alcohol (9)
1e oxidant
dirigent protein
OCH3
OH
HO
O
O
OCH3
(+)-Pinoresinol (10a)
PLR
NADPH
NADP+
OCH3
OH
HO
O
HO
OCH3
(+)-Lariciresinol (11a)
NADPH
PLR
NADP+
H3CO
OH
OH
HO
OCH3
OH
(–)-Secoisolariciresinol (2a)
SDH
NADH
NAD+
H3CO
HO
O
OH
OCH3
OH
(–)-Lactol (12a)
SDH
NADH
NAD+
H3CO
HO
O
O
OCH3
OH
(–)-Matairesinol (1a)

Fig. 3

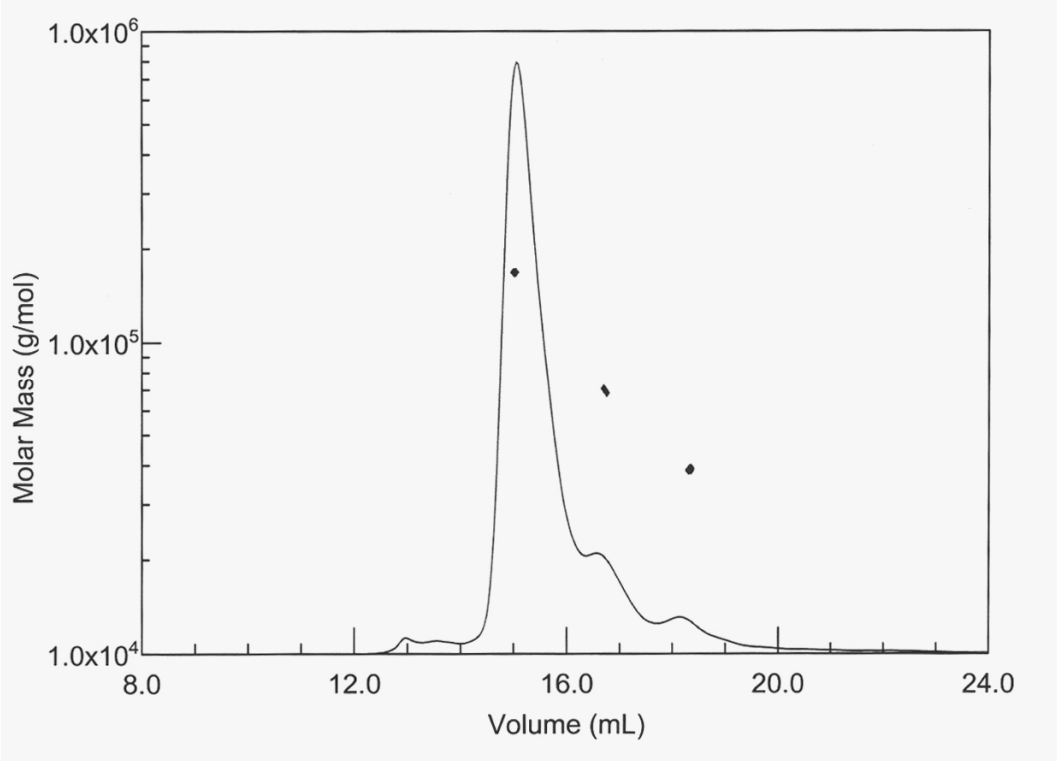

1.0x10^6
1.0x10^5
1.0x10^4
Molar Mass (g/mol)
8.0
12.0
16.0
20.0
24.0
Volume (mL)

Fig. 4 (A)

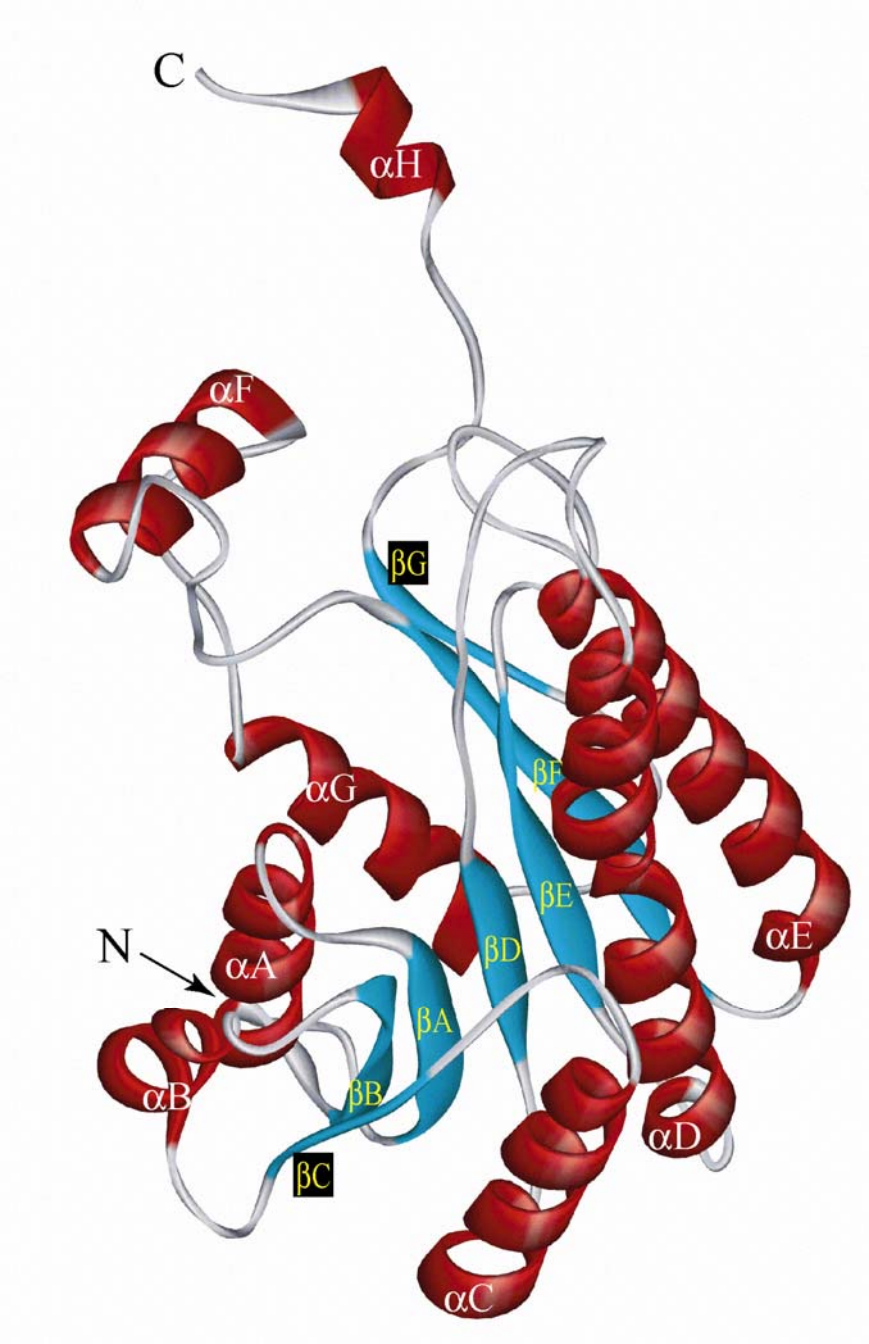


Fig. 4 (B)

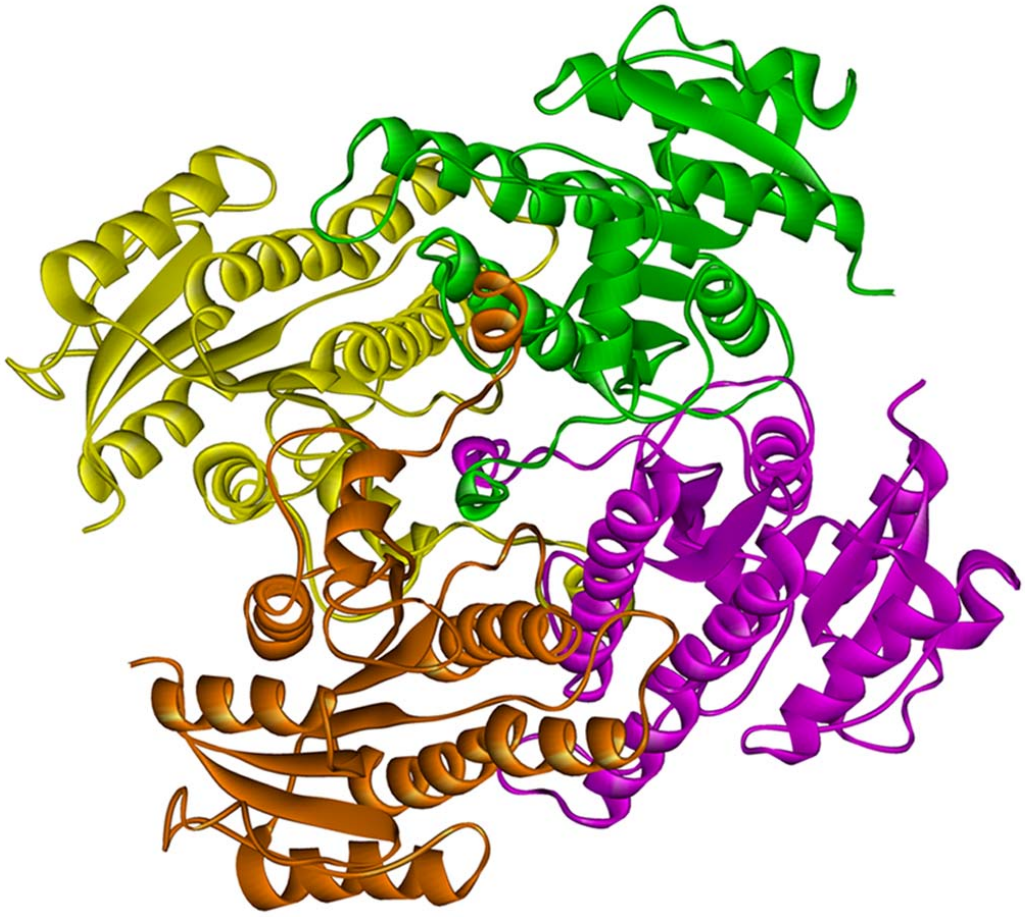


Fig. 5

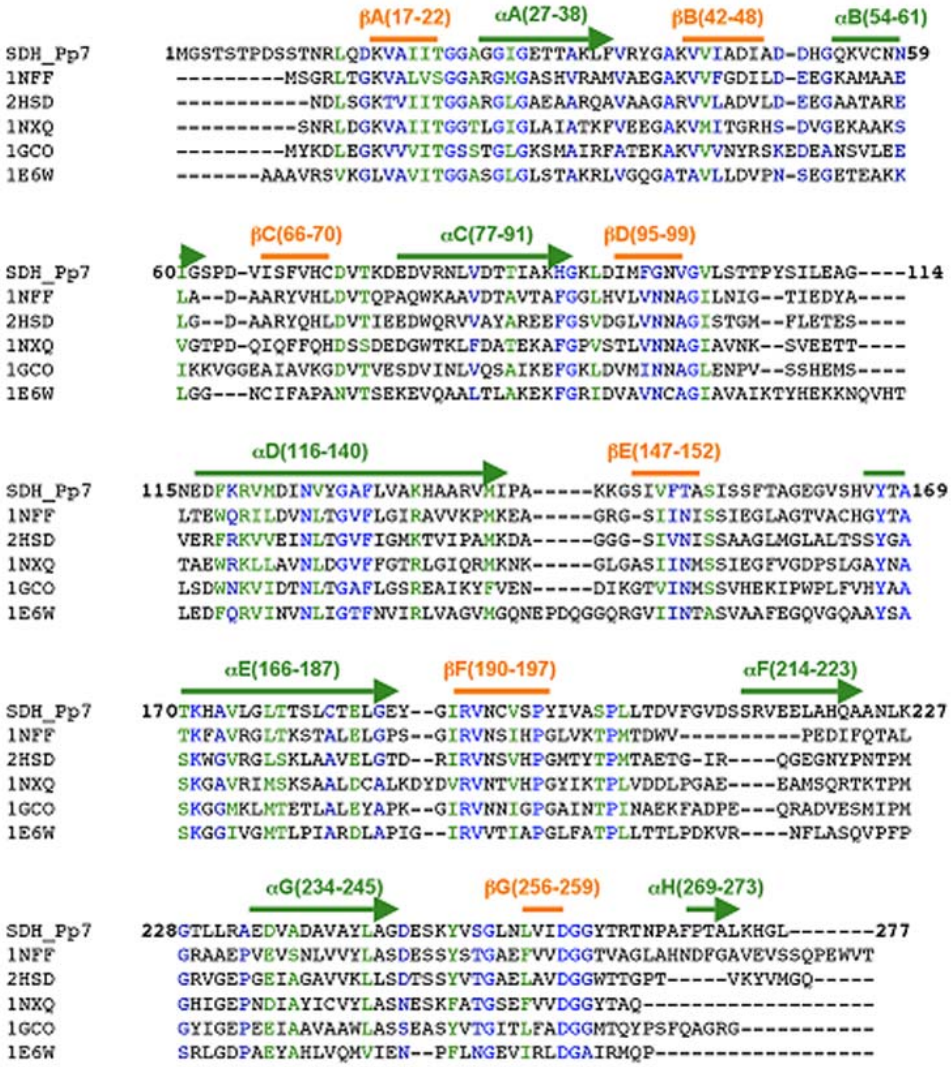

βA(17-22)
αA(27-38)
βB(42-48)
αB(54-61)
SDH_Pp7 1MGSTSTPDSSTNRLQDKVAIITGGAGGIGETTAKLFVRYGAKVVIADIAD-DHGQKVCNN59
1NFF ---------MSGRLTGKVALVSGGARGMGASHVRAMVAEGAKVVFGDILD-EEGKAMAAE
2HSD ----------NDLSGKTVIITGGARGLGAEAARQAVAAGARVVLADVLD-EEGAATARE
1NXQ ----------SNRLDGKVAIITGGTLGIGLAIATKFVEEGAKVMITGRHS-DVGEKAAKS
1GCO ---------MYKDLEGKVVVITGSSTGLGKSMAIRFATEKAKVVVNYRSKEDEANSVLEE
1E6W -------AAAVRSVKGLVAVITGGASGLGLSTAKRLVGQGATAVLLDVPN-SEGETEAKK
βC(66-70)
αC(77-91)
βD(95-99)
SDH_Pp7 60IGSPD-VISFVHCDVTKDEDVRNLVDTTIAKHGKLDIMFGNVGVLSTTPYSILEAG----114
1NFF LA--D-AARYVHLDVTQPAQWKAAVDTAVTAFGGLHVLVNNAGILNIG--TIEDYA----
2HSD LG--D-AARYQHLDVTIEEDWQRVVAYAREEFGSVDGLVNNAGISTGM--FLETES----
1NXQ VGTPD-QIQFFQHDSSDEDGWTKLFDATEKAFGPVSTLVNNAGIAVNK--SVEETT----
1GCO IKKVGGEAIAVKGDVTVESDVINLVQSAIKEFGKLDVMINNAGLENPV--SSHEMS----
1E6W LGG---NCIFAPANVTSEKEVQAALTLAKEKFGRIDVAVNCAGIAVAIKTYHEKKNQVHT
αD(116-140)
βE(147-152)
SDH_Pp7 115NEDFKRVMDINVYGAFLVAKHAARVMIPA-----KKGSIVFTASISSFTAGEGVSHVYTA169
1NFF LTEWQRILDVNLTGVFLGIRAVVKPMKEA-----GRG-SIINISSIEGLAGTVACHGYTA
2HSD VERFRKVVEINLTGVFIGMKTVIPAMKDA-----GGG-SIVNISSAAGLMGLALTSSYGA
1NXQ TAEWRKLLAVNLDGVFFGTRLGIQRMKNK-----GLGASIINMSSIEGFVGDPSLGAYNA
1GCO LSDWNKVIDTNLTGAFLGSREAIKYFVEN-----DIKGTVINMSSVHEKIPWPLFVHYAA
1E6W LEDFQRVINVNLIGTFNVIRLVAGVMGQNEPDQGGQRGVIINTASVAAFEGQVGQAAYSA
αE(166-187)
βF(190-197)
αF(214-223)
SDH_Pp7 170TKHAVLGLTTSLCTELGEY--GIRVNCVSPYIVASPLLTDVFGVDSSRVEELAHQAANLK227
1NFF TKFAVRGLTKSTALELGPS--GIRVNSIHPGLVKTPMTDWV----------PEDIFQTAL
2HSD SKWGVRGLSKLAAVELGTD--RIRVNSVHPGMTYTPMTAETG-IR----QGEGNYPNTPM
1NXQ SKGAVRIMSKSAALDCALKDYDVRVNTVHPGYIKTPLVDDLPGAE----EAMSQRTKTPM
1GCO SKGGMKLMTETLALEYAPK--GIRVNNIGPGAINTPINAEKFADPE---QRADVESMIPM
1E6W SKGGIVGMTLPIARDLAPIG--IRVVTIAPGLFATPLLTTLPDKVR----NFLASQVPFP
αG(234-245)
βG(256-259)
αH(269-273)
SDH_Pp7 228GTLLRAEDVADAVAYLAGDESKYVSGLNLVIDGGYTRTNPAFPTALKHGL-------277
1NFF GRAAEPVEVSNLVVYLASDESSYSTGAEFVVDGGTVAGLAHNDFGAVEVSSQPEWVT
2HSD GRVGEPGEIAGAVVKLLSDTSSYVTGAELAVDGGWTTGPT-----VKYVMGQ-----
1NXQ GHIGEPNDIAYICVYLASNESKFATGSEFVVDGGYTAQ-------------------
1GCO GYIGEPEEIAAVAAWLASSEASYVTGITLFADGGMTQYPSFQAGRG-----------
1E6W SRLGDPAEYAHLVQMVIEN--PFLNGEVIRLDGAIRMQP-----------------

Fig. 6 (A)

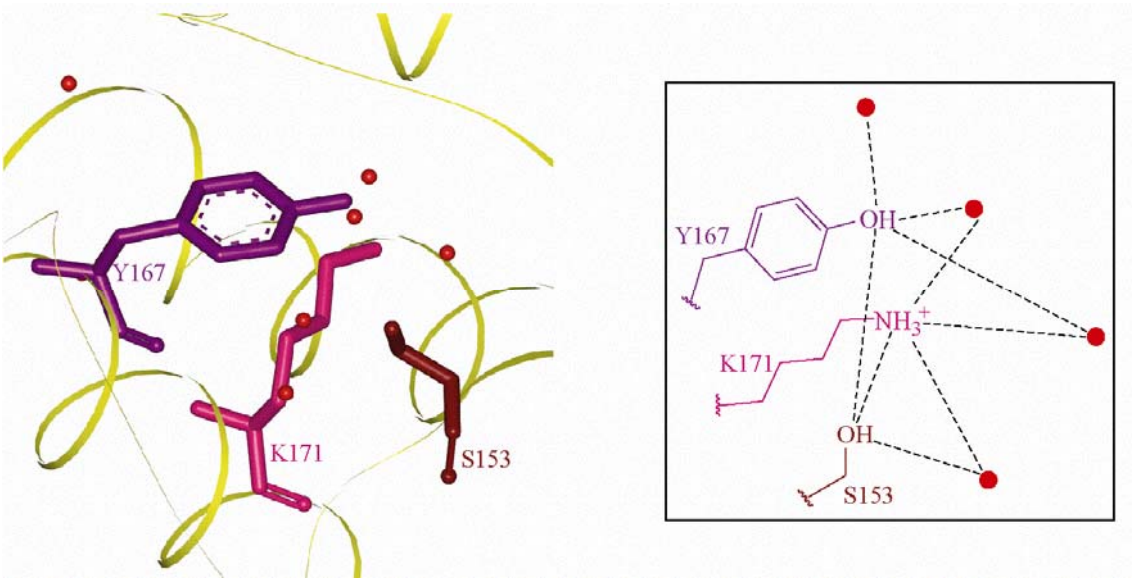


Fig. 6 (B)

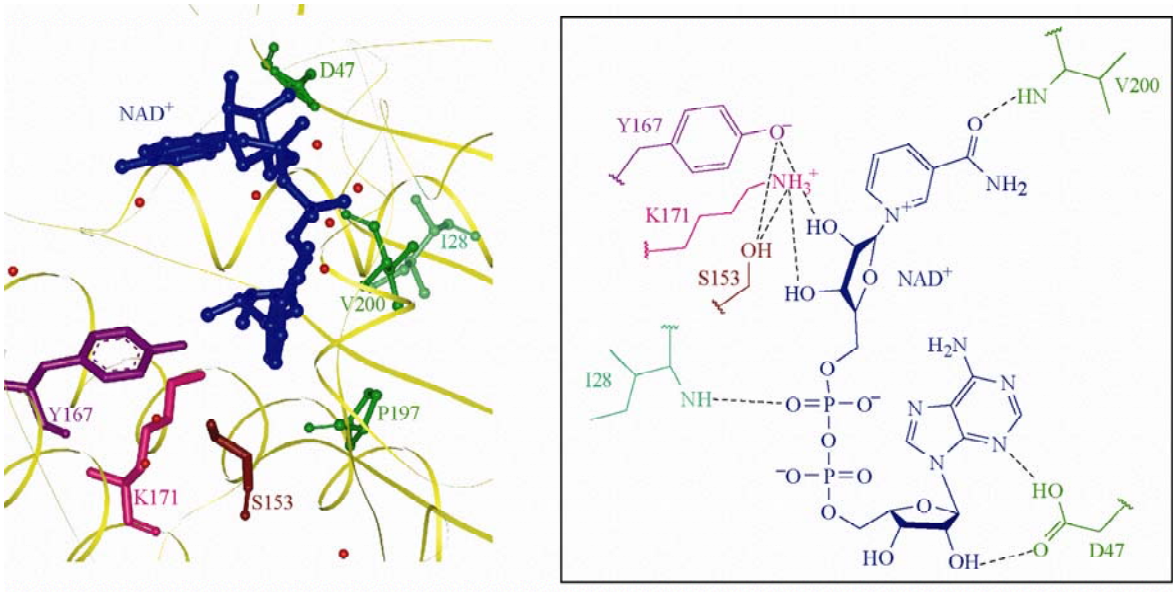

NAD+
D47
I28
V200
Y167
P197
K171
S153
Y167
K171
S153
NAD+
I28
V200
D47
P197

Fig. 6 (C)

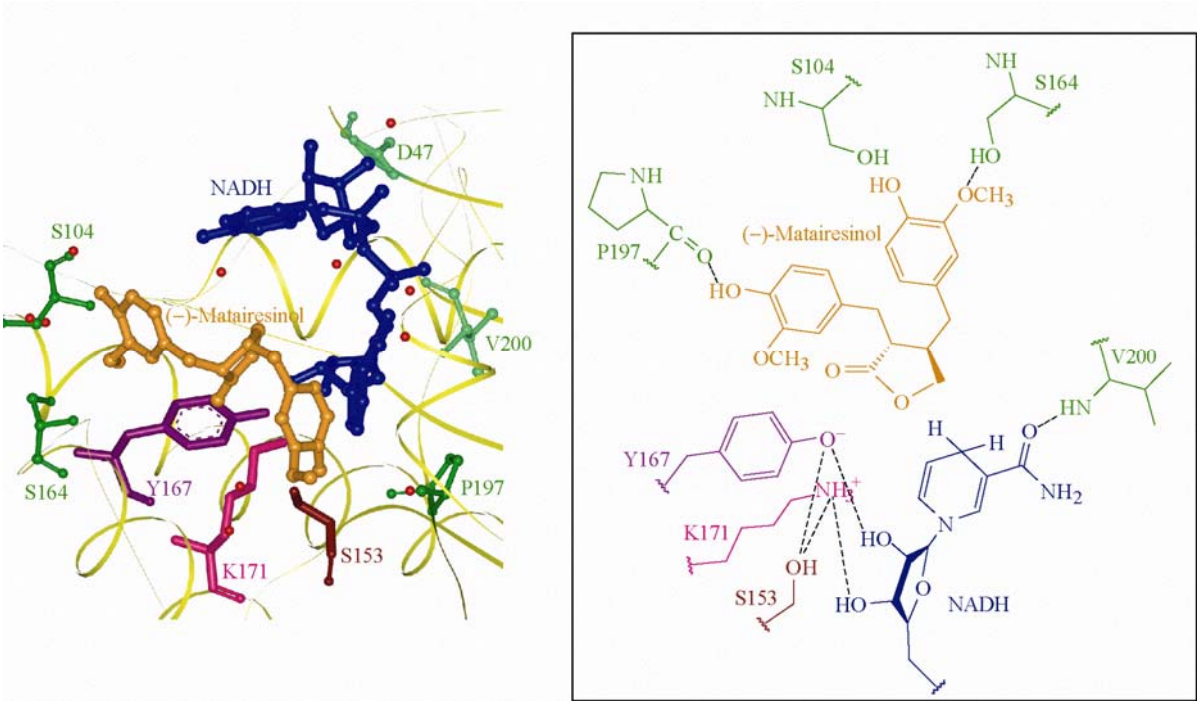

D47
NADH
S104
(−)-Matairesinol
V200
S164
Y167
P197
K171
S153
S104
NH
S164
OH
HO
NH
HO
OCH3
P197
C=O
(−)-Matairesinol
HO
OCH3
O
O
V200
HN
Y167
O−
H
H
O
NH3+
NH2
K171
N
HO
OH
S153
HO
O
NADH

Fig. 7

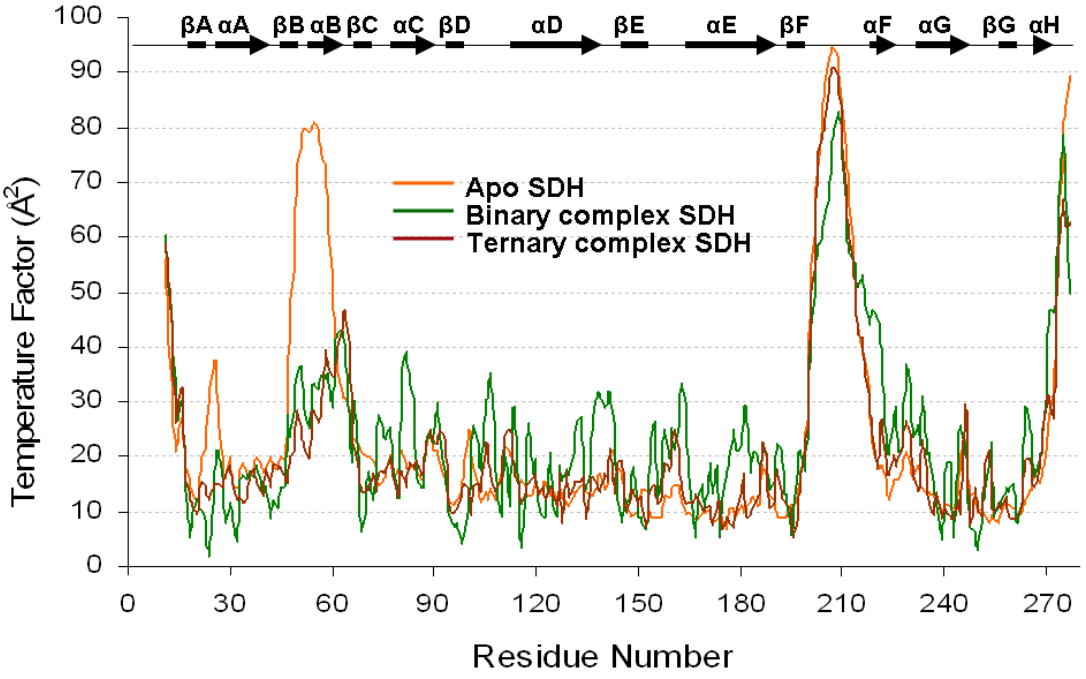

βA αA βB αB βC αC βD αD βE αE βF αF αG βG αH
Apo SDH
Binary complex SDH
Ternary complex SDH
Temperature Factor (Å²)
Residue Number
0
10
20
30
40
50
60
70
80
90
100
0
30
60
90
120
150
180
210
240
270

Fig. 8 (A)

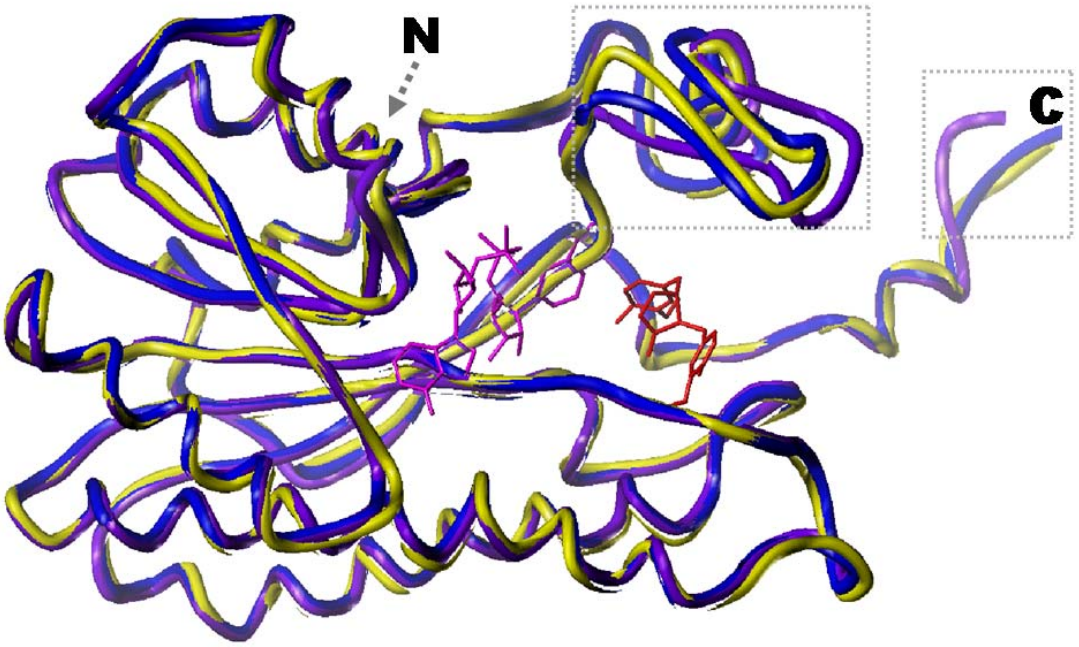


Fig. 8 (B)

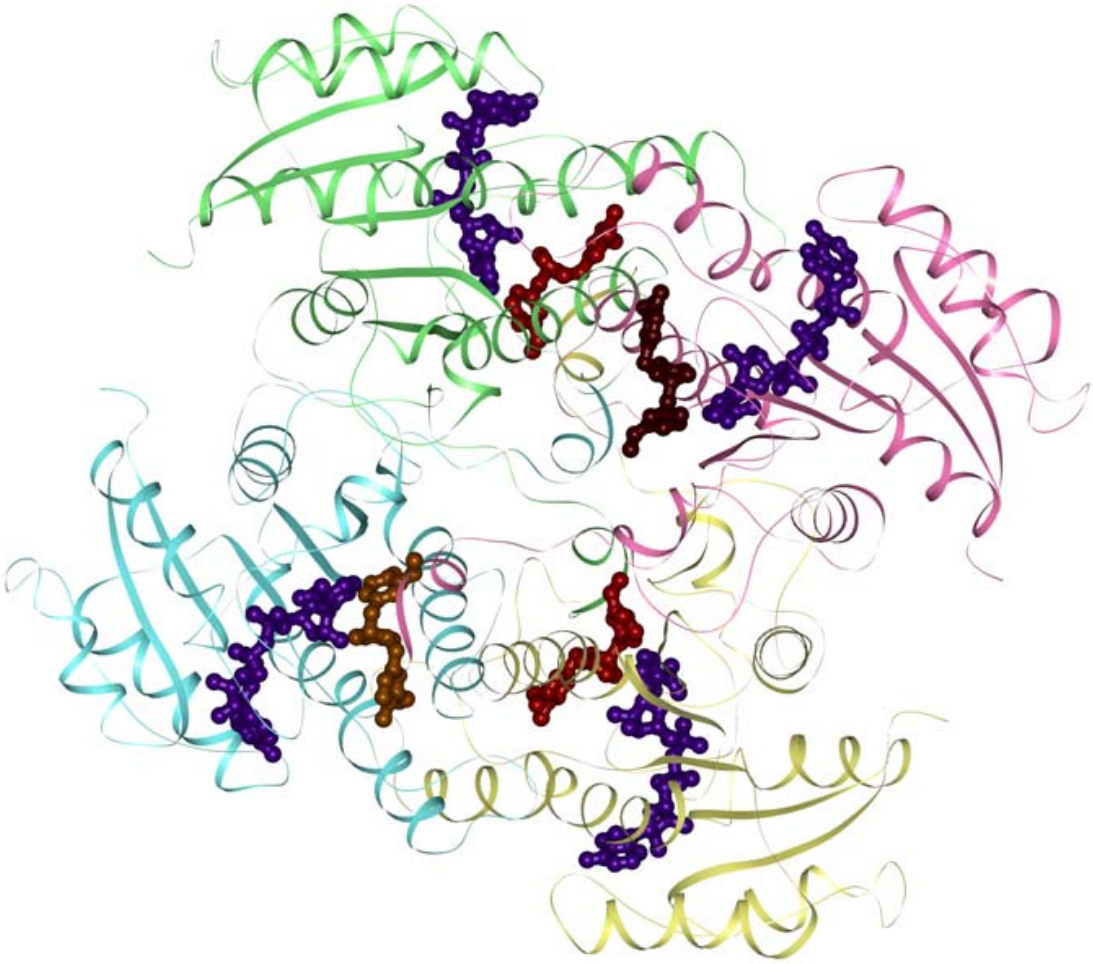


FIG. 9 (A)

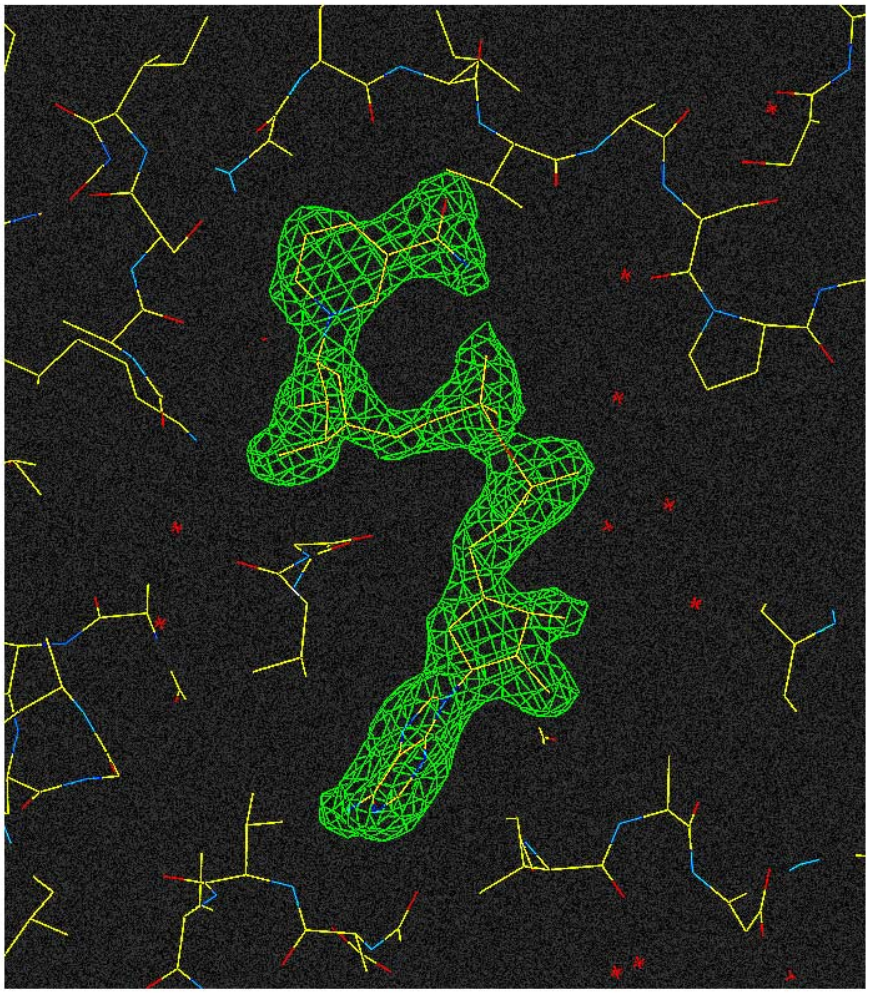
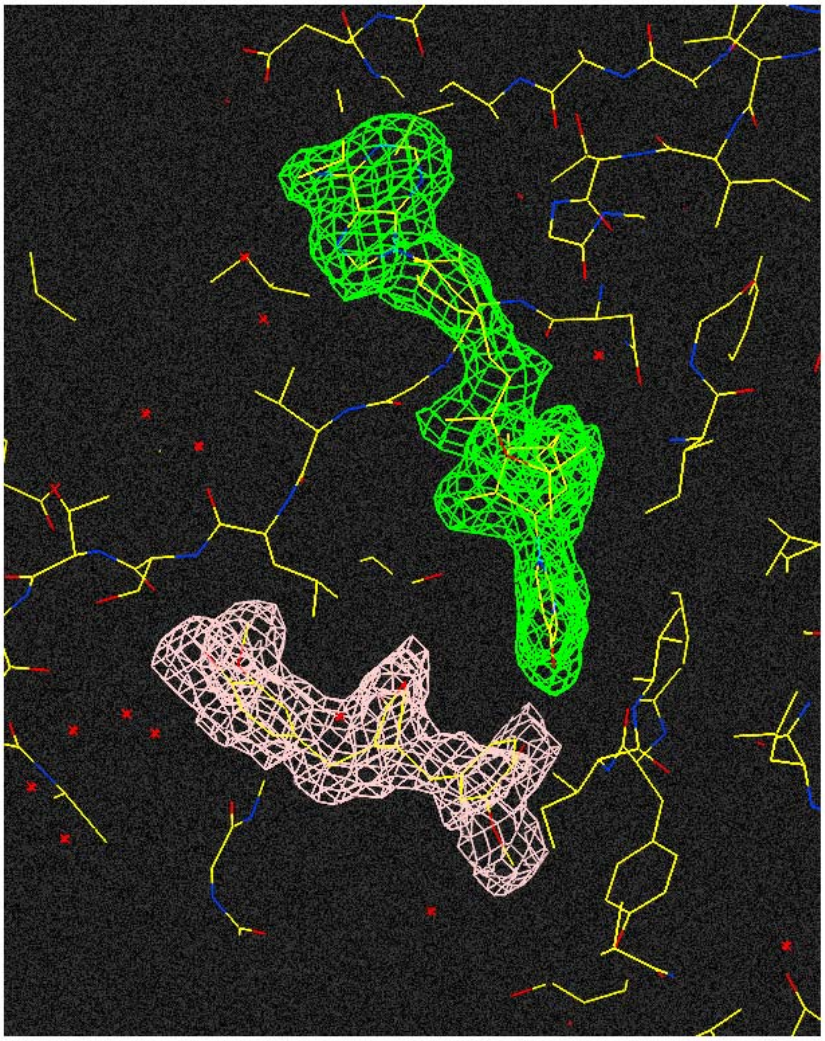


FIG. 9 (B)



SUPPLEMENT 1

MOLECULAR REPACEMENT (MR)

In the process of solving the crystal structures of SDH_Pp7 and mutant rubredoxin, molecular replacement (MR) was applied. MR was the choice for phase determination since there were trial models available in both cases. The basic principle of the MR is to determine the proper orientation and position of a molecule in the target unit cell. Following are the basic principle of MR.

CONCEPTS

The most serious problem or bottleneck of the crystal structure determination is assigning the phase of the corresponding intensity data. One way to solve this phase problem is to possess an atomic model, from which phase estimation can be tried. Such a model can be obtained from a related protein structure, or same protein structure in a different crystal form. However in order to solve a crystal structure, the correct orientation and position should be found in the new unit cell. Molecular replacement is the technique used to solve such a problem through allowing the proper alignment of the model coordinates.

As the public database of crystal structures gets larger and larger, this method is becoming more practical and popular for determining new structures. MR is also useful for

various ligand binding studies or studies of complexes from previously-determined structures (1, 2).

Approximately speaking, molecular replacement is applicable if a complete model is available and it has high amino acids identity (i.e. > 25 %) with the target protein. Of course, it is always helpful to have a good quality x-ray diffraction data of high completeness.

Two possible scenarios are considered when MR is applied for a phase determination. If a model is isomorphous to the target protein, the model phase can be used directly to calculate an electron density, ρ (x, y, z) from intensity data of the target protein by Eq.1, where $|F^{new}|$ is obtained from the crystal of the target protein and the phase α'^{model} is obtained from the model.

$$\rho(x,y,z) = \frac{1}{V}\sum_{hkl}\left|F_{hkl}^{new}\right| e^{\left[-2\pi i(hx+ky+lz)+i\alpha'^{model}(h,k,l)\right]} \qquad \text{Eq.1}$$

When the model used for phasing is not isomorphous with the given target structure, the problem becomes more complex. To find the correct position and orientation of the model in the target unit cell, two steps are required; rotational and translational searches. In the rotation search step, the relative spatial orientation of the model and target molecule is determined. In the following translation step, the correctly oriented model is superimposed onto the target molecule (2, 3).

Therefore, in this MR process, six parameters (three rotatonal and three translational) are required to express the relationship between the model and target molecules (Eq.2 and Fig.1).

$$\boldsymbol{x_1} = \boldsymbol{R}\boldsymbol{x_2} + \boldsymbol{T} \qquad \text{Eq.2}$$

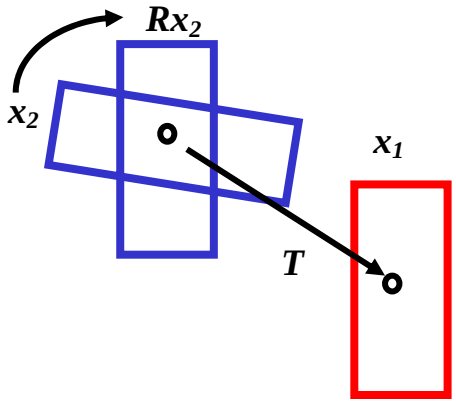


Fig. 1

($\boldsymbol{x_1}$: target atomic vector, $\boldsymbol{x_2}$: model atomic vector, $\boldsymbol{T}$: translation vector (tx, ty, tz) $\boldsymbol{R}$: rotation matrix expressed in eulerian angle (θ_1, θ_2, θ_3))

Because a Patterson function (Eq. 3, 4) is independent of the molecular position in the unit cell as long as molecule is not rotated, these processes can be simplified by Patterson overlapping function (4). A Patterson map is an interatomic vector map, so vectors between atoms in the unit cell are indicated as vectors from the origin to the maximum in the Patterson map. If a pair of atoms belongs to the same molecule, the corresponding vector is relatively short and their end-points are found nearby the origin in the map. This vector is called as a *self-vector* and the intramolecular vector is called as a *cross vector*.

The Patterson function $P(\mathbf{u})$ or $P(u, v, w)$ is a Fourier summation with intensities as coefficients and without phase angles, or rather with all phase angles equal to zero.

$$P(u,v,w)=\frac{1}{V}\sum_{hkl}\left|F(h,k,l)\right|^2 COS[2\pi(hu+kv+lw)] \qquad \text{Eq.3}$$

or shorter,

$$P(\mathbf{u}) = \frac{1}{V}\sum_{S}\left|F(S)\right|^2 COS[2\pi(\mathbf{u}\cdot\mathbf{S})] \qquad \text{Eq.4}$$

u, *v* and *w* are relative coordinates in the unit cell. To avoid confusion with *x*, *y* and *z* in the real cells, *u, v* and *w* are used in Patterson cell, which, however, has identical dimensions to the real cell. Because all phase angles are zero in the Patterson function it can be calculated without any previous knowledge of the structure.

Rotation Function

The principle of rotation function is based on the Rossman and Blow procedure (5). Self-vectors are used to determine the possible orientation of the model structure in a crystal lattice. As in Eq.2, the calculated Pattern function, $P(u)$, of the model structure is multiplied by a rotation matrix (**R**) to give a rotated function, $P_r(u_r)$. The $P^T(u)$ is the Patterson function of the crystal and the *u* is the coordinates in the Patterson map. The rotation function (*RF*) is defined as the overlap of two functions (Eq.5).

$$RF(\alpha,\beta,\gamma) = \frac{1}{v}\int_{\Omega} P^T(u) \times P_r(u_r)\, d^3u \qquad \text{Eq.5}$$

Where Ω is a region of integration with volume *v*. The maximum number of intramolecular vectors and fewest possible intermolecular vectors should be included in this integration region. Difficulty or complexity could happen if a large number of intermolecular vectors were included or unavoidable. Based on Eq.5, the better $P_r(u_r)$ and $P^T(u)$ overlap, the higher the RF value will be. The orientation (α, β, γ) corresponding to the highest RF value has a

highest possibility of being correct. The structure model is not always same as the target structure and the data quality and completeness are not always favorable. Therefore, it is not easy to guarantee that the highest RF value corresponds to the correct orientation (1, 2, 6, 7).

Translation Function

In order to obtain the final solution of the above mentioned MR method, the translation step is required to overlap the model molecule with target molecule in a real space (x, y, z). In a more advanced modern method than the previously used trial and error search, a translation function, $T(r)$, is used. This function gives the correlation coefficients between a set of cross-vectors of a model structure and the observed vectors.

Fist of all, the search model is rotated to proper orientation, which is determined in the previous rotational search process. Cross-vectors are calculated by using a crystallographic symmetry operation. Translation function, $T(r)$, is maximized by shifting the position of the search model. Finally, the position (x, y, z) corresponding to an orientation is fixed where $T(r)$ reaches the maximum value (Eq. 6).

$$T(\mathrm{r}) = \int_{v} P_{obs}(u) \times P_{calc}(r,u)\, du$$

Eq.6

$P_{obs}(u)$ is the observed Patterson vectors; $P_{calc}(r,u)$ is the calculated cross-Patterson vector of the search model at the position (r). Two other factors, R-factor (Eq. 7) and the

standard linear correlation coefficient, *C*, (Eq.8) are also used to aid the determination of the correct position for the target molecule.

$$R = \frac{\sum_{hkl} \left\| F_{obs} \right| - \left| F_{calc} \right\|}{\sum_{hkl} F_{obs}} \qquad \text{Eq.7}$$

$$C = \frac{\sum_{hkl} \left(\left| F\,obs \right|^2 - \overline{\left| F\,obs \right|^2} \right) \times \left(\left| F\,calc \right|^2 - \overline{\left| F\,calc \right|^2} \right)}{\sqrt{\left[\sum_{hkl} \left(\left| F\,obs \right|^2 - \overline{\left| F\,obs \right|^2} \right)^2 \times \sum_{hkl} \left(\left| F\,calc \right|^2 - \overline{\left| F\,calc \right|^2} \right)^2 \right]}} \qquad \text{Eq.8}$$

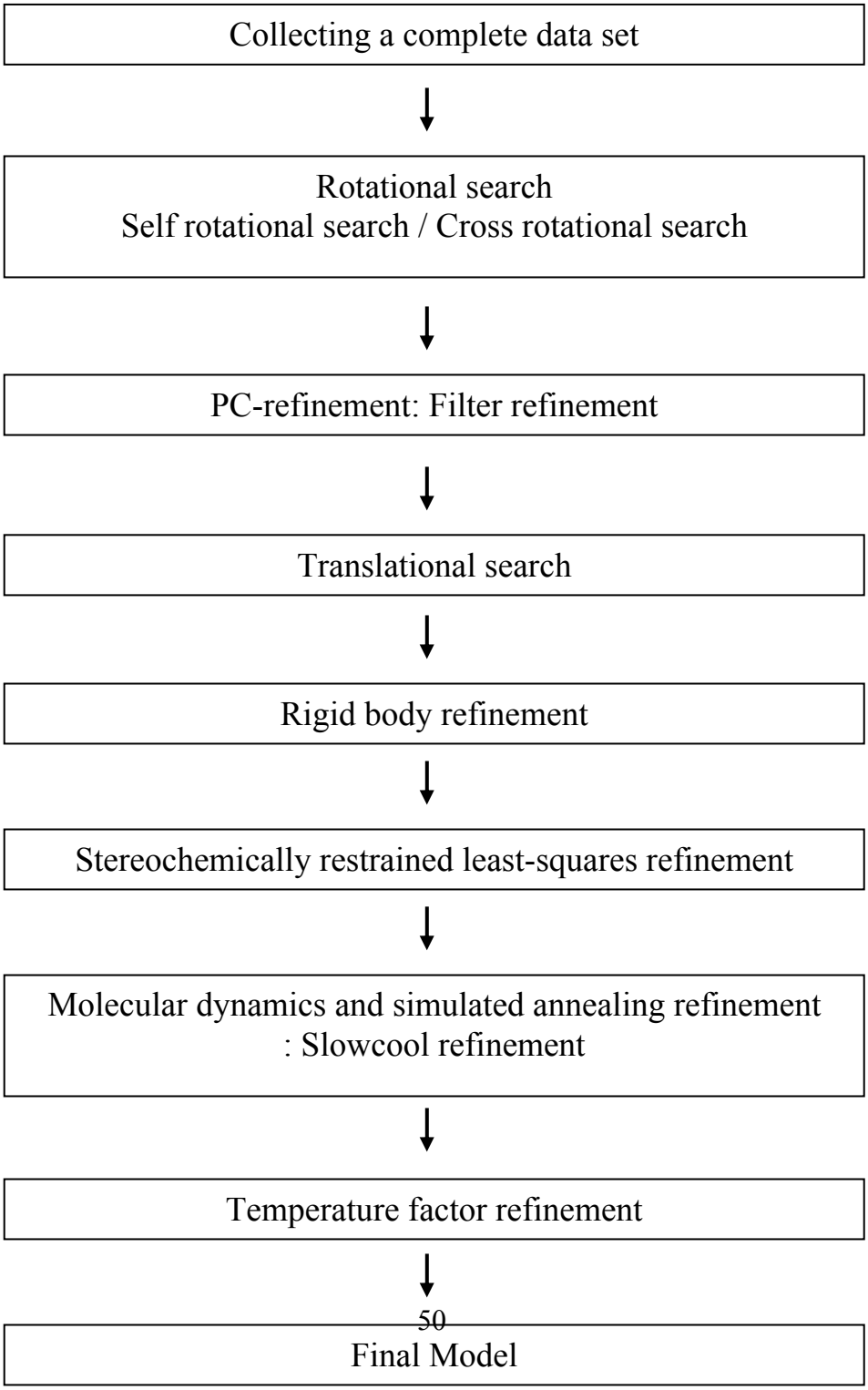
where $|F_{obs}|$ is the square root of the observed intensity of x-ray data, and $|F_{calc}|$ is the calculated structure factor from the oriented and positioned search model. The standard linear correlation coefficient has an advantage over the R-factor at some levels because it is not sensitive to the scaling error (1, 2, 3, 8, 9) .

Despite the tremendous advantage of MR over last several years, it often fails to find the correct position and orientation. Generally, the complete and high quality data, accurate model and suitable criteria for analysis can guarantee the success of this two-step processes.

In SDH-Pp7 experiment, X-PLOR software package was used for MR process. X-PLOR software package is based on the principle of Ramman and Blow (5, 10), and it uniquely uses the Patterson Correlation refinement (PC refinement) (11). PC-refinement uses the normalized structure factor (E), calculates the standard linear correlation coefficienct, between the square root of $(E_{obs})^2$ and $(E_{calc})^2$, and has capability of clustering the close peaks to a single peak. This technique is also extended to the translation search

and drastically increases the success rate in finding the correct orientation and position (1, 2, 9).

OVERALL PROCESS OF MOLECULAR REPLACEMENT METHOD



REFERENCE

1. Brunger, A.T. X-PLOR (1992) : a system for crystallography and NMR (version 3.1) Yale University, New Haven, CT 06511
2. Drenth, J. (1999) Principle of protein X-ray crystallography, 2nd Ed., Springer-Verlag, Inc.
3. Rhodes, G. (2000) Crystallography made crystal clear, 2nd Ed., Academic press.
4. Grosse-Kunstleve RW and Adams PD. (2001) *Acta Crystallogr D.* 57, 1390-1396
5. Rossman, M.G. and Blow M.G. (1962) *Acta Crystallogr A.* 15, 24-31
6. Stout, G. H. & Jensen, L. H. (1989) X-ray structure determination, 2nd Ed., John Wiley & Son, Inc.
7. Navaza, J and Saludjian, P. (1997) *Methods in enzymology* 276, 581-595
8. Hein J and Stovlbaek J. (1995) *J Mol Evol.* 40,181-189.
9. MacRee, D. (1999) Practical protein crystallography, 2nd Ed., Academic press
10. Rossmann MG. (1990) *Acta Crystallogr A.* 46, 73-82.
11. Badger J, Kumar RA, Yip P, Szalma S. (1999) Proteins 35, 25-33.

CHAPTER TWO

STRUCTURAL STUDY OF REDUCTION POTENTIAL OF *CLOSTRIDIIM PASTEURIANUM* RUBREDOXIN

Preface

Electron transfer occurs in many fundamental biological processes, such as photosynthesis, respiration, nitrogen fixation and many other biological reactions, transforming various types of energy into utilizable forms. The driving force for electron transfer is determined by the reduction potentials of participating donor and acceptor molecules. In most redox proteins, the reduction potential of the redox site is delicately determined and modulated by the corresponding local environment. Therefore, quite often differences of up to a few hundred mV occur in reduction potentials of homologous proteins with the same redox site. These differences become even larger among non-homologous proteins with the same redox site, and major determinants of these differences are believed to be the configuration of the local protein and solvents in the vicinity of the redox site. In spite of the rapidly growing amount of structural information and increasing numbers of reduction potentials determined for a given type of redox site, the structural origins of the apparent diversity in reduction potentials for these proteins remain unclear, and successful quantitative predictions are still difficult. Understanding the physical basis of the rates and redox potentials of electron transfer proteins, therefore, remains a subject of this chapter.

This following manuscript was published in *Journal of Biological Inorganic Chemistry* in June 2004.

The unique hydrogen bonded water in the reduced form of *Clostridium pasteurianum* rubredoxin and its possible role in electron transfer

Il Yeong Park[1,2], Buhyun Youn[1], Jill L. Harley[3], Marly K. Eidness[3], Eugene Smith[4], Toshiko Ichiye[5], and ChulHee Kang[1]

[1]*School of Molecular Biosciences, Washington State University, Pullman, WA 99164-4660*

[2]*College of Pharmacy, Chungbuk National University, Cheongju 361-763, Korea*

[3]*Department of Chemistry, University of Georgia, Athens, GA 30602-2556*

[4]*Wilkes Honors College, Jupiter, Fl 33458*

[5] *Department of Chemistry, Georgetown University, Washington DC 20057-1227*

ABSTRACT

Rubredoxin is a simple iron-sulfur (FeS_4) protein involved in oxidation-reduction reactions. The side chain of Leu41 near the iron-sulfur center has two conformations, which we suggested previously serves as a gate for a water molecule during the electron transfer process. To establish the role of residue 41 in electron transfer, an [L41A] mutant of *Clostridium pasteurianum* rubredoxin was constructed and crystallized in both oxidation states. Despite the lack of the gating side chain in this protein, the structure of the reduced [L41A] rubredoxin reveals a specific water molecule in the same position as observed in the reduced wild-type rubredoxin. In contrast, both the wild-type and [L41A] rubredoxins in the oxidized state do not have water molecules in this location. The reduction potential of the [L41A] variant was ~50mV more positive than wild-type. Based on these observations, it is proposed that the site around the Sγ of Cys9 serves as a port for an electron acceptor. Lastly, the Fe-S distances of the reduced rubredoxin are expanded, while the hydrogen bonds between Sγ of cysteines and backbone amide hydrogens are shortened compared to its oxidized counterpart. This small structural perturbation in the Fe(II)/Fe(III) transition is closely related to the small energy difference which is important in an effective electron transfer agent.

Keywords: Redox protein; rubredoxin; L41A mutant; electron transfer; crystal structure

INTRODUCTION

Oxidation-reduction is essential in bioenergetics or scavenging processes of life. Not only does it provide the means for transforming solar and chemical energy into an utilizable form for all living organisms, it also extends into a range of metabolic processes. The sign and relative magnitude of the reduction potential is the determining factor for transferring electrons from one component to another. In biological processes, such as photosynthesis, respiration and nitrogen fixation, electron transfer proteins determine the rate and the direction of the electron transfer reaction by modulating the reduction potential of their redox sites.

To date, however, the basic mechanisms of establishing and controlling the reduction potential of electron transfer proteins are largely unknown. Homologous redox proteins with the same redox sites have differences in reduction potentials that span a few hundred mV [1,2]. The origin that leads to the differences in reduction potential is crucial to understanding how the electron transfer protein can function. A single point mutation can drastically modify the reduction potential and subsequently influence the electron transfer rate. Despite the rapidly growing number of reduction potentials and structural information for a given type of redox site, the factors that contribute to differences in reduction potential are complex and quantitative predictions are still not possible. For example, factors such as subtle shifts in the local polar environment and solvent accessibility may be redox state dependent and dynamic. This complexity has made it impossible to determine basic mechanisms simply by examining a crystal structure of a single representative native

protein in one oxidation state. In fact, examining single structures has often led to very contradictory interpretations of the physical origins of reduction potential differences. To understand differences in reduction potential, it is clearly necessary to have highly accurate structures of both native and mutant proteins in both the oxidized and reduced states.

Efforts have been made to understand the determining factors for the reduction potential using rubredoxin [3, 4, 5, 6]. It is a small redox protein with a single iron atom redox site tetrahedrally coordinated by four cysteinyl thiolates. The reduction potentials of rubredoxins (Rds) cover the range from -100 to + 60 mV [7]. Unlike heme proteins, the redox sites of Rds are all relatively small so that the area of direct interactions with the protein is much smaller.

The redox center of rubredoxins, residing just beneath the surface of the protein, consists of a simple tetrahedral complex of four cysteinyl sulfurs coordinated to the iron atom in the center (Fig 1). Two cysteines (Cys6 and Cys39) are located inside the protein, while the other two cysteines (Cys9 and Cys42) are solvent accessible. The Sγ atom of Cys42 is partially covered by the side chains of two hydrophobic residues of Val8 and Val44, and the Sγ of Cys9 is also partially covered by the side chain of Leu41. Unlike the close Van der Waals contact between Val8 and Val44 over Cys42, Leu41 has freedom of movement in its side chain without perturbing the backbone or overall structure of the protein. A rigid backbone for this solvent-exposed side chain is provided by its neighbors-Pro40, Cys 39 and Cys42. The crystal structures of Rds deposited in the Protein Data Bank show the side chain of residue 41 in two conformations (1B20, 1C09, 1FHH, 1FHM, 1IRO, 4RXN; Fig 1) [8].

High-resolution crystal structures of both oxidation states of *Clostridium pasteurianum* (Cp) Rd were previously determined to elucidate their structural differences. A series of water molecules hydrogen-bonded to Sγ of Cys9 were only observed in its reduced state, and it was suggested that Leu41 near the iron-sulfur center serves as a gate for water entry during the redox reaction [5, 9, 10].

To test the role of the side chain of this Leu41 as a "flip-flop" gate, a variant of Leu41Ala (L41A) of *Clostridium pasteurianum* rubredoxin was constructed and crystallized. In addition to oxidized crystals, two kinds of reduced crystals were produced; one was obtained from the crystallization of the reduced protein solution under a nitrogen environment, the other was formed via direct reduction of the oxidized red crystal to the transparent colorless one with sodium dithionite.

MATERIALS AND METHODS

Cloning, over-expression, and purification of L41A Cp Rd variant.

The [L41A] Cp Rd clone was constructed by site-directed mutagenesis of the Cp Rd gene in plasmid pT7-7 [11] by the QuickChange procedure (Stratagene, La Jolla, CA), using complementary primers to introduce the ala codon, 5'-TGG GTT TGC CCG GCC TGC GGT GTT GG-3' and 5'-CCA ACA CCG CAG GCC GGG CAA ACC CA-3'. The QuickChange product was transformed into E. coli strain XL-1 Blue (Stratagene) according to manufacturer instructions. A few single colonies grew on Luria-Bertani (LB) medium and 1.5% agar (Difco Bacto) in the presence of ampicillin (100 μg/mL). One colony was selected and amplified by standard procedures [11] in a 100-mL growth and the plasmid purified by a QIAGEN midi-prep (QIAGEN, Valencia, CA). The plasmid sequence was determined by Sequetech, Inc. (Mountain View, CA) from the T7 primer site upstream of the Rd gene start. The L41A codon change was confirmed as well as the integrity of the remaining bases of the Cp Rd gene and the plasmid was named pL41A Cp Rd. It was then transformed into the E. coli expression strain, BL-21 Gold (DE3) (Stratagene). Cultures were grown at 250 rpm at 37°C in 4 L of LB medium supplemented with ampicillin (100 μg/mL) and induced by the addition of IPTG (0.4 mM final concentration), growth continued for 5 hours, then cells were harvested by centrifugation at 6000 x g, and lysed following published procedures [12].

[L41A] Cp Rd over-expressed protein was separated from other *E. coli* proteins by passage over a Hi-Trap Q 5-mL column (Amersham Biosciences) on an ÅKTA FPLC

system (Amersham Biosciences). The red-colored [L41A] Cp Rd protein eluted at ~50% B in a 20 column-volume gradient from 0-100%B where buffer A is 50 mM Tris-HCl, pH 7.6 and buffer B is Buffer A + 1.0 M NaCl. Fractions containing red color were pooled, diluted with Buffer A to reduce the NaCl concentration and loaded onto a Resource Q 6 mL column (Amersham Biosciences) and eluted under the same conditions as described above. A final round of purification was performed on a MonoQ HR10/10 column (Amersham Biosciences) under the same conditions. [L41A] Cp Rd protein was concentrated in YM3 Centricons (Millipore) to about 5 mM based on the visible absorbance extinction coefficient of the 490 nm peak [12].

Electrochemical Measurements

Cyclic voltammetry measurements were recorded via a BAS CV 50W potentiostat (Bioanalytical Systems, Lafayette, IN). Experiments were carried out at room temperature using a microscale electrochemical cell as previously described [13] using a pyrolytic graphite working electrode, platinum counter electrode and a saturated Ag/AgCl reference electrode. All potentials are reported versus the standard hydrogen electrode.

Crystallization of oxidized and reduced forms of Cp Rd variant L41A

The oxidized crystals of [L41A] Cp Rd were produced by vapor diffusion against 2.3 M ammonium sulfate, 0.1 M sodium acetate at pH 4.0 and 10 mM sodium chloride (oxidized). The reduced crystals were produced against 2.3 M ammonium sulfate and 0.1 M sodium acetate at pH 4.6 with 10 mM sodium dithionite as the reducing agent under a

nitrogen environment (drop-reduced). Each of 2 μ*L* of protein solution was mixed with 2 μ*L* of the reservoir solutions and placed over the reservoir as a hanging drop. The deep red colored crystals of the oxidized form, and the transparent, colorless crystals of the reduced form were grown to a suitable size for crystallographic study within 4 days. Another kind of reduced crystal was also produced by dipping the oxidized crystal into a cryo-solution containing sodium dithionite just before mounting on the goniometer head (soaked). The colorless (reduced) state did not change throughout the data collection procedure.

Data collection, molecular replacement and crystallographic refinement

All data for the three kinds of crystals (oxidized, drop-reduced, and soaked) of [L41A] Cp Rd were collected at the Advanced Light Source at Berkeley (Beam line 8.2.1) under a –160° C nitrogen stream with the crystal-to-detector distance of 110 mm. All three crystals belong to the same space group of R3 in the trigonal crystal system. The crystal data are listed in Table 1 together with the refinement statistics.

The structures of all three Cp Rds were solved by molecular replacement methods (supplement 1) using a coordinate of the previously solved oxidized Cp Rd (1FHH). The initial coordinates were obtained by the AMoRe software package [14]. The rigid-body refinement of the initial position was carried out using 15.0 Å to 3.0 Å resolution data and produced R-values around 30 %. Then the structures were refined by X-PLOR [15]. After several cycles of positional refinement and temperature factor refinement, we were able to fit all of the residues to the electron density.

The R-factor for the final models containing non-hydrogen atoms for three Rds are 17.9 (R_{free}=18.6), 19.0 (19.8) and 18.0 (19.4) % respectively. The root mean-square deviations (from standard geometry) of the three Rds are 0.006 Å for bonds and 1.65^{o} for angles (averages for the three Rds). The coordinates of the Rds structure have been deposited in the Protein Data Bank (1SMM for the oxidized form, 1SMU for the drop-reduced form, and 1SMW for the soaked form).

RESULTS

Reduction potential upon L41A mutation

A cyclic voltammogram of L41A Cp Rd is shown in Figure 2. The reduction potential was determined to be 50 mV more positive than wild-type Cp Rd.

Crystallization of oxidized and reduced L41A Cp Rd

Oxidized crystals were obtained under normal atmospheric conditions (~20 % oxygen). The colorless reduced crystals were obtained under two sets of conditions. In the first method, sodium dithionite as a reducing agent was added to both reservoir and crystal drop (drop-reduced). It was crucial to use the appropriate amount of reducing agent and to maintain the nitrogen atmosphere in the crystallization apparatus throughout the experiment. The protein reoxidized within 1-2 hours if insufficient reducing agent was used and if there were trace amounts of oxygen in the crystallization chamber. Likewise, the protein precipitated if too much sodium dithionite was used.

In the second method, reduced crystals were directly prepared by dipping them into a cryo-solution containing sodium dithionite just before data collection (soaked). Reduction of the crystal was confirmed by monitoring the color of the crystal changing from deep red to colorless transparent as the reducing agent soaked into the crystal. Some crystals were fractured and/or melted during the soaking, as there might be some structural rearrangement inside the crystalline lattice. However, by adjusting the concentration of sodium dithionite in the cryo-solution, some crystals survived during the reduction process,

and once reduced, the colorless state did not change throughout the whole data collection procedure.

Structural differences between oxidized and reduced forms of L41A Cp Rd

Though the volumes of the unit cells for the reduced Rds are slightly bigger than that for the oxidized form, the difference is small (0.4%). The overall conformations of the two oxidation states are very similar to wild-type Cp Rd with average rms distance of 0.72(8) Å (all atoms except for isopropyl group of 41 residue) and other Rds [10, 16].

The geometry around the [Fe-S] redox center is changed slightly upon reduction. The four Fe-S distances of the reduced form are slightly longer (Fe-S elongation; +0.012 Å on average) than those of the oxidized form, whereas the five hydrogen bonds between the Sγ of the cysteines and nearby peptide amide nitrogens are shorter (S-N shrinkage; -0.064 Å on average) than the oxidized form (Fig. 3). Although a magnitude of this observed rearrangement upon reduction is small, it is a general trend in Rds. Similar rearrangements were observed in native Cp Rd (1FHH-1FHM pair; Fe-S elongation 0.096 Å and S-N shrinkage –0.042 Å) as well as in *Pyrococcus furiosus* Rd (1CAA-1CAD pair; Fe-S elongation 0.033 Å and S-N shrinkage –0.094 Å) that both of the oxidized and reduced structures are available at the protein data bank. Interestingly, the elongation of inner shell Fe-S bonds are compensated by the shrinkage of outer shell hydrogen bonds, consequently the overall position of the backbone atoms around the [Fe-S] redox center remains unchanged upon a change in oxidation state.

The C-terminal amino acids, residues 53 and 54, are flexible as reflected in both their high temperature factors (B-factor 27.1 on average) and positional differences between the two oxidation states.

Unique water hydrogen bonded to Cys9-Sγ in the reduced form of [L41A] Cp Rd

The removal of the isopropyl group at position 41 (i.e. a substitution of leucine with alanine) does not perturb the local structure of [L41A] Cp Rd. The structure of the oxidized and reduced forms of [L41A] Cp Rd are very similar, with average rms distance between the oxidized and drop-reduced forms of 0.67 Å, and between the oxidized and soaked form of 0.64 Å, and between the drop-reduced and soaked form of 0.51 Å for all atoms excluding the two flexible terminal residues.

The reduced form of [L41A] Cp Rd (both the drop-reduced and soaked crystal) has a unique water molecule hydrogen bonded to Sγ-Cys9. Fig. 4 shows the well-localized electron density (B-factor 18.5 for drop-reduced, and 21.9 for soaked) of this water molecule residing closely to the Sγ-Cys9 (3.37 Å) and the Cβ-Ala41 (3.28 Å). In contrast, the oxidized form of [L41A] Cp Rd do not have a water molecule at this position, and it was not able to find a reasonable electron density within the hydrogen bonding distance from Sγ atoms in oxidized Rd. This water molecule of the reduced Rds was also found at the same position in our previous report of the reduced wild-type Cp Rd crystal.

DISCUSSION

Water molecule and electrochemistry.

Both the reduced forms of [L41A] Cp Rd and wild-type Cp Rd (1FHM) contain a hydrogen-bonded water molecule to Sγ of Cys9. In contrast, oxidized form of neither [L41A] Cp Rd nor wild-type Cp Rd (1FHH) has this specific hydrogen-bonded water, even though there is no steric hindrance that prevents a water molecule from hydrogen bonding at this position in the oxidized form of [L41A] Cp Rd. The electronic rearrangement around the iron center and/or the enrichment of negative charge in the Sγ of Cys9 upon reduction may induce some electrostatic attraction towards the partial positive charge of hydrogen in the bound water molecule. We previously proposed [10] that this hydrogen bonding mechanism of the Sγ atom to the solvent water molecules stabilizes the reduced state and is controlled (or gated) by the side chain of leucine at position 41. This nonpolar side chain allows transient penetration of water molecules when it is in reduced state [10]. The reduced state of the redox center is further stabilized by reducing the size of the sidechain at position 41 and consequently allowing an easier entry of water molecules.

Iron-sulfur cluster.

Upon reduction of Fe(III) to Fe(II), the attraction of Sγ electrons to the positive iron center decreases resulting in an expansion of the iron-sulfur cluster. However this expansion, or coordination-energy change, is effectively compensated by the contraction of hydrogen bonds between Sγ and backbone amide nitrogens, thus keeping the overall

environment of the [Fe-S] redox center unchanged over the Fe(II)/Fe(III) transition. The increased attraction between the electron–enriched Sγ and amide hydrogen of the reduced state and the overall peptide backbone around the [Fe-S] center play an important roles in keeping the redox center from expanding upon reduction.

This limited expansion observed in the [Fe-S] cluster is closely related to the relatively small energy difference associated with a change in oxidation state in Rds. The smaller inner-sphere reorganization energies associated with Fe(II)/Fe(III) couples of Rds, as compared to $[FeS_4]^{2-/1-}$ clusters [17], are important for efficient electron transfer reactions in biological systems. The structural relaxation of the [Fe-S] center associated with a change in oxidation state is typical in Rds whose high resolution structures of both redox states are available [10, 18].

Electron transfer pathway.

Electron transfer could happen through either direct tunneling between proteins or a solvent-mediated tunneling. Based on the crystal structures of both oxidation states, we have previously proposed [10] that the electron transfer in Rds proceeds via a solvent-mediated process using the surface cysteines of Cys9 and Cys42.

Recently, Tezcan et al. [19] suggested that van der Waals interactions and water-mediated hydrogen bonds are effective coupling elements for electron tunneling across a protein-protein interface. Based on our observation of the water molecule directly hydrogen bonded to the cysteinate in the reduced state and the increase of the reduction potential of the [L41A] Cp Rd variant, it appears that the transfer of electrons is most likely a solvent-

mediated process following the Fe(II)/Sγ(Cys9)/water/acceptor route. The umbrella of Val8 and Val44 residues over Cys42 makes the electron pathway more accessible through the Sγ atom of Cys9. The Sγ atom of Cys9 is exposed to the solvent by the "flip-flop" movement of the Leu41 side chain. This motion might serve as a gate or guide for the electron donor/acceptor protein during electron transfer.

If the reaction is a direct superexchange process (if not a solvent mediated one), the electron-acceptor atom of the acceptor molecule approaches directly to this position via a gate-like movement of the side chain in position 41 during the reaction. It is also plausible that Leu41 acts as a modulator or triggering mechanism for electrons to transduce through a Fe(II)/S□(Cys42)/acceptor protein route.

In summary, the environment around the Sγ atom of Cys9 is an important determinant of electron transfer reactivity. The site provides a facilitated mechanism for the electron transfer reaction that is influenced by the side chain at position 41. The electron can be transferred through the sulfur atom via a solvent-mediated manner or directly to the electron acceptor molecule, and interaction with the solvent is an important determinant for the reduction potential of the redox site.

REFERENCES

1. Cammack R (1992) Adv. Inorg. Chem. 38: 281-322

2. Beinert H, Holm RH, Munck E (1997) Science 277: 653-659

3. Sieker LC, Stenkamp RE, LeGall J (1994) In: Peck HD Jr, LeGall J (eds) methods in enzymology, vol 243. Academic Press, San Diego, pp 203-216.

4. Gray HB, Winkler JR (1996) An. Rev Biochem 65: 537-561

5. Swartz PD, Beck BW, Ichiye T (1996) Biophys J 71: 2958-2969

6. Johnson MK (1998) Curr Opin Chem Biol 2: 173-181

7. Meyer J, Moulis J-M (2001) In: Messerschmidt A, Huber R, Wieghardt K, Poulos T (eds) Handbook of metalloproteins, vol 1, Wiley, New York, pp 505-517

8. Maher MJ, Xiao Z, Wilce MC, Guss JM, Wedd AG (1999) Acta Crystallogr D55: 962-968

9. Yelle RB, Park N-S, Ichiye T (1995) Proteins 22: 154-167

10. Min T, Ergenacan C, Eidness M, Ichiye T, Kang C (2001) Protein Sci 10: 613-621

11. Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K (2000) Current Protocols in Molecular Biology. John Wiley & Sons, New York

12. Eidsness MK, Burden AE, Rchie KA, Kurtz Jr DM, Scott RA, Smith ET, Ichiye T, Beard B, Min T, Kang C (1999) Biochemistry 38: 14803-148

13. Smith ET, Bennett DW, Feinberg BA (1991) Anal Chim Acta 251: 27-33

14. Navaza J (1994) Acta Crystallogr A50: 157-163

15. Brunger AT (1992) X-PLOR: A system for crystallography and NMR (version 3.1). Yale University. New Haven

16. Dauter Z, Wilson KS, Sieker LC, Moulis JM, Meyer J (1996) Proc Natl Acad Sci USA 93: 8836-8840

17. Kennepohl P, Solomon EI (2003) Inorg Chem 42: 696-708

18. Day MW, Hsu BT, Joshua TL, Park JB, Zhou ZH, Adams MW, Rees DC (1992)

Protein Sci 1: 1494-1507

19. Tezcan FA, Crane BR, Winkler JR, Gray HB (2001) Proc Natl Acad Sci USA 98 : 5002-5006

Table 1 Summary of data collection and refinement statistics.

	Oxidized	Red (drop)	Red (soak)
Space group	R3	R3	R3
Cell constants			
$a = b$ (Å)	62.79	63.07	63.04
c (Å)	32.80	32.60	32.69
Molecule in an asymmetric unit	1	1	1
Resolution limits (Å)	1.36	1.43	1.38
Unique reflections ($F>2\sigma(F)$)	10231	8595	9923
Completeness (%)	99.0	98.2	99.3
$R(I)_{merge}$ overall (%)	2.9	3.0	3.8
$R(I)_{merge}$ at highest res. (%)	13.9	13.2	12.5
$I/\sigma(I)$ overall	17.1	15.7	17.7
$I/\sigma(I)$ at highest res.	1.9	1.4	2.5
R-value for all reflections (R_{free})	18.0 (18.9)	19.4 (21.3)	18.0 (20)
No. of residues	54	54	54
No. of protein atoms (inc H)	419 (787)	419 (787)	419 (787)
No. of solvent waters (sulfate)	37 (1)	46	48
B-factor$_{mean}$ overall	9.8	9.6	9.1
B-factor$_{mean}$ for solvents only	22.4	22.1	20.9
R.M.S. deviations from ideality			
Bond (Å)	0.021	0.021	0.022
Angle (°)	2.70	2.55	2.51

FIGURE LEGENDS

Fig. 1 The redox center of rubredoxin and the conformational variation of the side chain of Leu41 (6 structures overlapped-1B20, 1C09, 1FHH, 1FHM, 1IRO, 4RXN). The S□ of Cys42 is covered by the close van der Waals contact between Val8 and Val44.

Fig. 2 Cyclic voltammograms of [L41A] Cp Rd in 0.1M NaCl and 50 mM phosphate buffer at pH 7.6, with scan rate of 4 mV/s.

Fig. 3 The distances between Fe and Sγ, and between Sγ and amide N(Å) around the redox center of [L41A] Cp Rd. The values of the oxidized crystal are shown, and those of the reduced one (drop) are designated inside the parentheses.

Fig. 4 The electron density map showing the hydrogen bonded water molecule in [L41A] Cp Rd. (a) Oxidized form. (b) Reduced form (drop-reduced). (c) Reduced form (soaked). Note that there is no electron density for the water position in the oxidized form (a). The contour was drawn at 1.0 electrons/$Å^3$ level.

Fig. 1

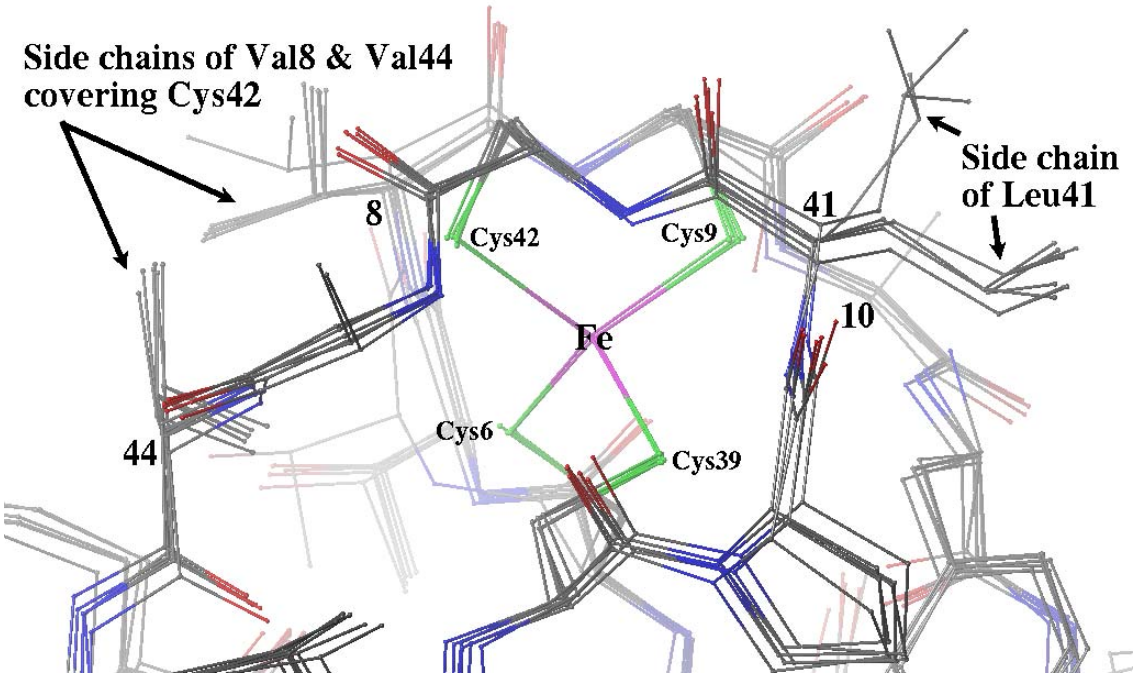

Side chains of Val8 & Val44
covering Cys42
Side chain
of Leu41
8
41
Cys42
Cys9
Fe
10
Cys6
44
Cys39

Fig. 2

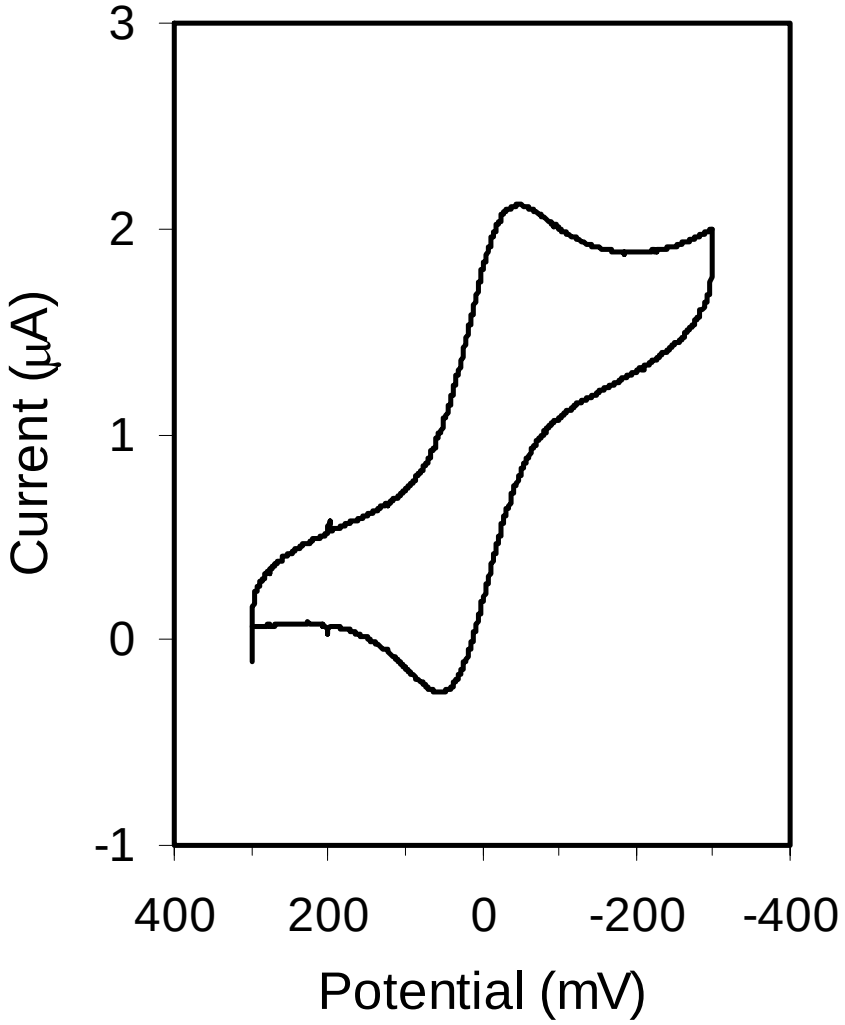


Fig. 3

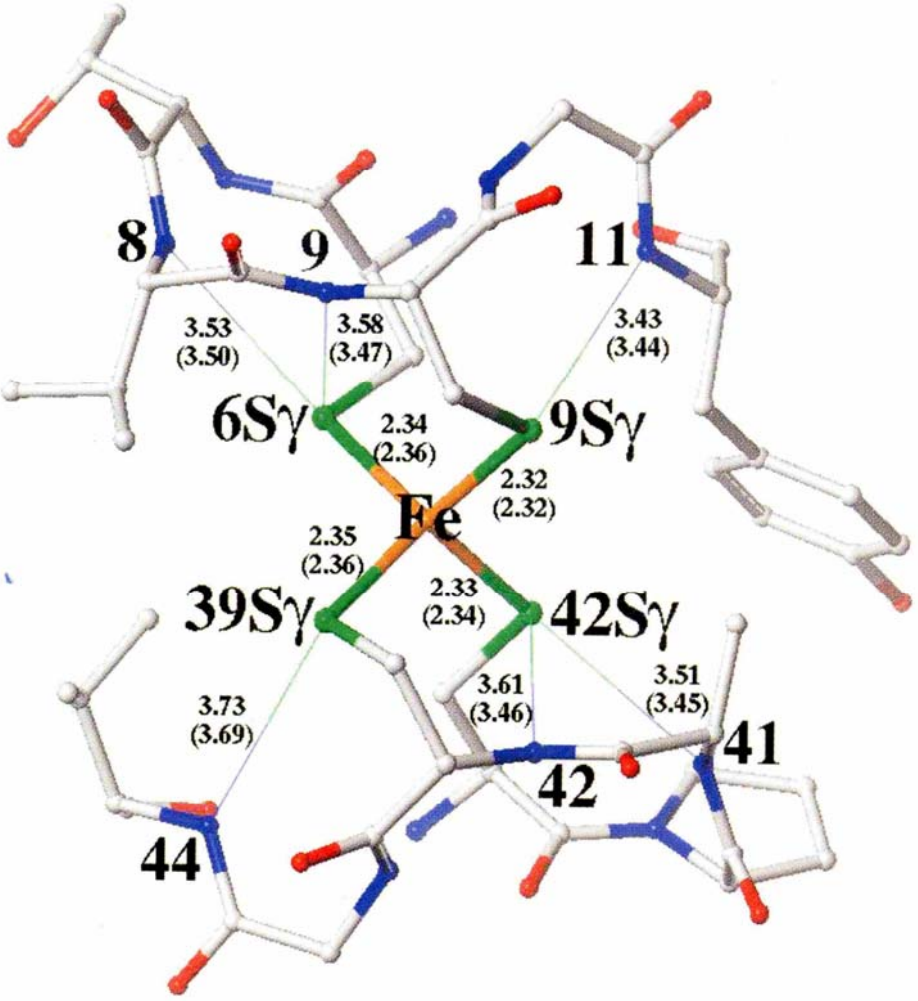

8
9
11
3.53
(3.50)
3.58
(3.47)
3.43
(3.44)
6Sγ
2.34
(2.36)
9Sγ
2.32
(2.32)
Fe
2.35
(2.36)
2.33
(2.34)
39Sγ
42Sγ
3.73
(3.69)
3.61
(3.46)
3.51
(3.45)
41
42
44

Fig. 4 (A)

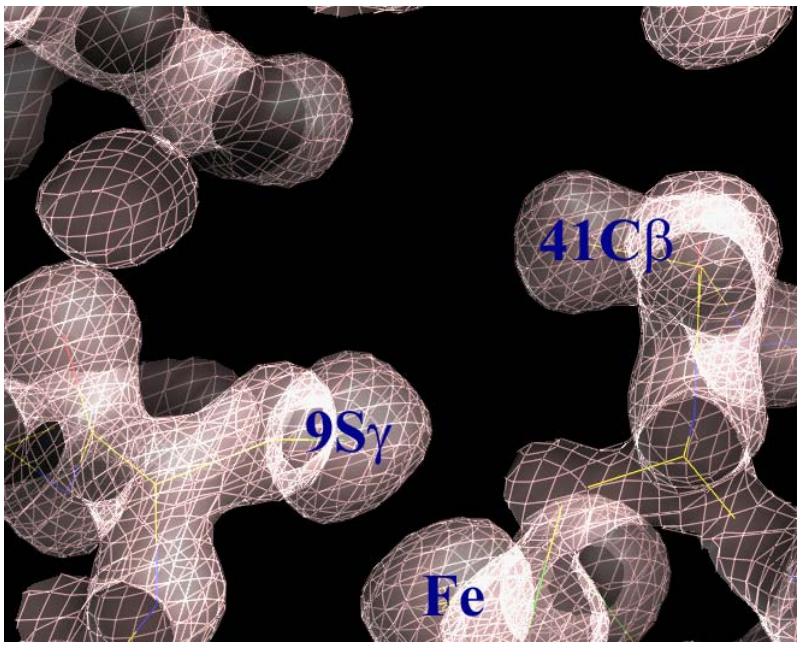


Fig. 4 (B)

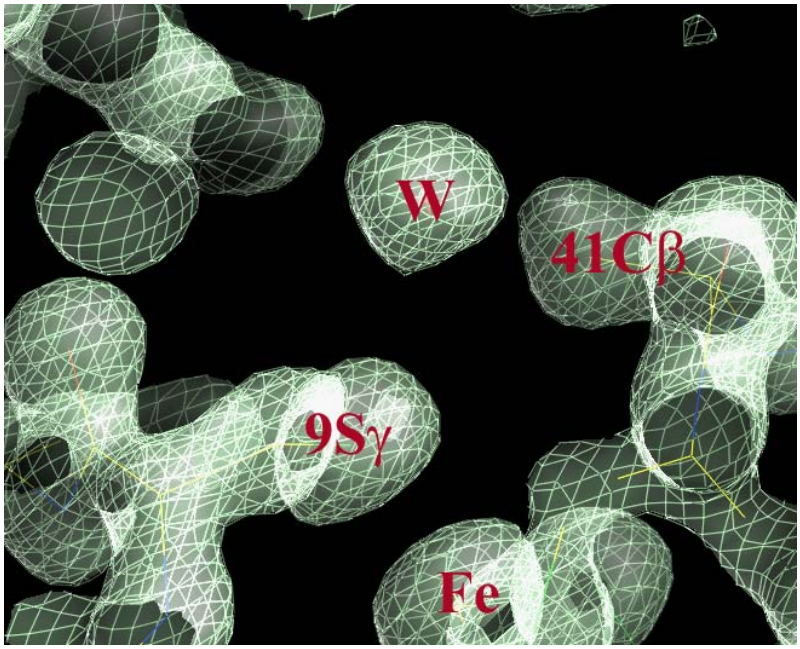
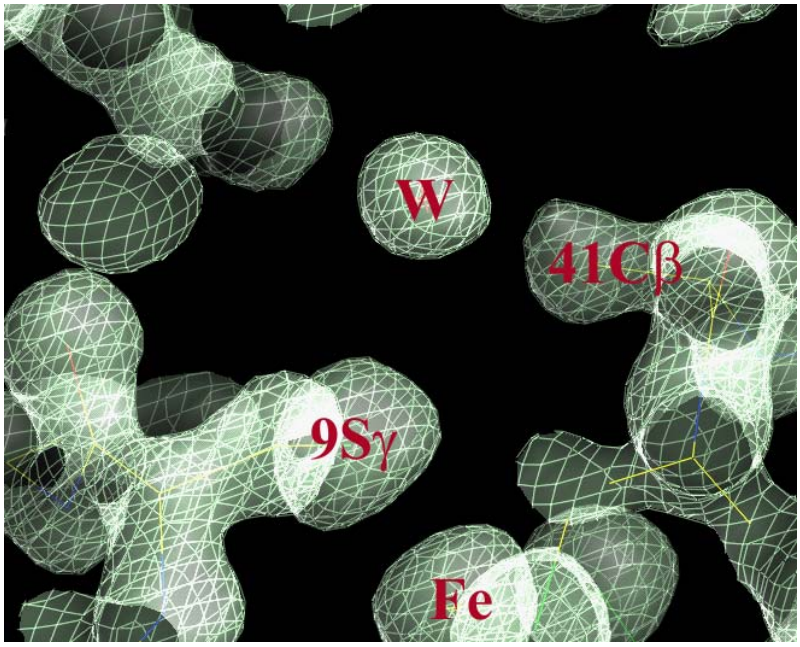


Fig. 4 (C)



CHAPTER THREE

CRYSTALLOGRAPHIC STUDY OF *ARABIDOPSIS* VACUOLAR SORTING RECEPTOR

Preface

Soluble proteins enter the secretory pathway in the endoplasminc reticulum and then move to the Golgi complex. If they have no specific sorting information on them, they exit the Golgi and are secreted to the cell exterior. Therefore, sorting away from this "bulk flow" requires both that the soluble protein carries a signal, and that an integral membrane receptor protein recognizes the signal and pulls the soluble protein into a vesicle that will carry it to a different destination. Somehow, within the plant Golgi apparatus, one receptor (BP-80) recognizes one type of signal and pulls a soluble protein into the lytic vacuole pathway, while a different receptor (RMR) recognizes a different signal and pulls a different soluble protein into the protein storage vacuole pathway. A similar process occurs in the mammalian Golgi complex, except the lytic pathway receptors are the mannose-6-phosphate receptor and sortilin. In order to understand this complex but decisive sorting mechanism, we have decided to solve the crystal structure of the receptor molecules. Here, we report crystallization and initial diffraction studies of tBP80 as a first step towards understanding its ability to bind ligands in a specific manner.

The following manuscript was published in *Acta Crystallographica* section D in November 2004. Two supplement sections of the crystallographic method follow the manuscript.

Purification, crystallization and preliminary crystallographic studies of the ligand-binding domain of a plant vacuolar sorting receptor

Buhyun Youn,[a+], Sally W. Rogers [b++], John C. Rogers[b] and ChulHee Kang [a]

[a]School of Molecular Biosciences, Washington State University, Pullman, WA 99164-4660, USA, [b]Institute of Biological Chemistry, Washington State University, Pullman, WA 99164-6340, USA.

Notes

[+ ++] These authors contributed equally to this article.

Keywords: vacuolar sorting receptor; BP80

Synopsis: Vacuolar sorting receptor (VSR) proteins target ligands to the lytic vacuole in a sequence-specific manner. Purification from *Drospohila* S2 cells, crystallization and preliminary X-ray diffraction studies of *Arabidopsis thalia* VSR, AtBP80b, have been carried out.

ABSTRACT

Vacuolar sorting receptor (VSR) proteins bind soluble protein ligands in a sequence-specific manner and target them to the lytic vacuole in plant cells. A VSR from *Arabidopsis thaliana*, AtBP80b, has been successfully purified after heterologous expression in *Drosophila* S2 cells. The AtBP80b protein (560 amino acids) was crystallized by the hanging-drop method with a PEG 400-based precipitant. Preliminary X-ray diffraction studies of an AtBP80b crystal showed that it belongs to the cubic space group $P2_13$ (or $P4_232$) and has unit-cell parameters $a = b = c = 145.9$ Å. Crystals of the VSR diffract beyond 2.5 Å resolution.

INTRODUCTION

In eukaryotic cells, soluble proteins enter the secretory pathway in the endoplasmic reticulum and then move to the Golgi complex. If they have no specific sorting information on them, they exit the Golgi and are secreted to the cell exterior. Therefore, sorting away from this `bulk flow' (Denecke *et al.*, 1990; Phillipson *et al.*, 2001) requires both that the soluble protein carries a signal and that an integral membrane-receptor protein recognizes the signal and directs the soluble protein into a vesicle that will carry it to a different destination. In plant cells, sorting mechanisms are complex because proteins are sorted into pathways to two separate vacuoles (Okita & Rogers, 1996; Vitale & Galili, 2001). Within the plant Golgi apparatus, one receptor, termed BP80 (Hadlington & Denecke, 2000; Kirsch *et al.*, 1994), recognizes one type of sorting determinant and targets a soluble protein to the lytic vacuole pathway (Humair *et al.*, 2001; Jiang & Rogers, 1998; Paris *et al.*, 1997), while a different receptor recognizes a different sorting determinant and pulls a different soluble protein into the protein storage vacuole pathway (Okita & Rogers, 1996; Vitale & Galili, 2001).

Both types of sorting determinants are encoded by the polypeptide sequence, but differ in composition and position within the polypeptide (Matsuoka & Neuhaus, 1999). The structural details by which each receptor interacts with the correct sorting determinant is a central question in plant cell biology. Sorting determinants for the lytic vacuole pathway are sequence-specific and the best studied such determinants share a central Asn-Pro-Ile-Arg (NPIR) motif (Matsuoka & Neuhaus, 1999). This knowledge led to the

biochemical purification of an ~80 kDa protein, termed BP80 (Fig.1 (A)), from lysates of pea clathrin-coated vesicle membranes by means of an affinity column carrying the peptide SSSFADS**NPIR**PVTDRAASTYC (Kirsch *et al.*, 1994). This peptide, termed the proaleurain peptide, contains the vacuolar sorting determinants for a plant cysteine protease (Holwerda *et al.*, 1992). BP80 binding to the proaleurain peptide occurred with a K_d of approximately 40 nm and was pH-dependent, with optimal binding between pH 6.0 and 6.5, while binding was abolished at pH ≤4.5. Additionally, binding was sequence-specific: the peptide SRF**NPIR**LPT efficiently competed for binding, while the peptide SRFNPGRLPT, with an Ile-to-Gly mutation in the NPIR motif, did not compete (Kirsch *et al.*, 1994).

Further studies of BP80-proaleurain peptide interactions utilized a recombinant form of the BP80 lumenal portion (termed tBP80 (Fig. 1(B)), lacking the transmembrane domain and cytoplasmic tail) that was expressed in *Drosophila* S2 cells, from which it was secreted and could be purified from the medium (Cao *et al.*, 2000). This tBP80 is comprised of three distinct domains: an N-terminal `PA domain' (Cao *et al.*, 2000; Luo & Hofmann, 2001; Mahon & Bateman, 2000), a central region without defined homology to other proteins and a C-terminal domain comprised of three epidermal growth factor (EGF) repeats. Expression in a eukaryotic cell was required because the tBP80 requires chaperone-mediated folding in the endoplasmic reticulum to acquire a proper three-dimensional structure and to ensure correct placement of disulfide bonds among its 34 Cys residues. These results (Cao *et al.*, 2000) indicated that tBP80 had two separate proaleurain peptide ligand-binding sites, one NPIR-sequence-specific and the other not specific for

NPIR. Results from binding studies using a recombinant protein lacking the EGF-repeat domain indicated that NPIR-specific binding was defined by interactions of the PA domain and the central unique domain. The EGF repeats appeared not to participate in NPIR-specific binding, but rather affected the conformation of the other two domains in a favorable manner. Additionally, the EGF repeats participated in Ca^{2+}-dependent non-NPIR-specific binding. Limited proteolysis experiments demonstrated that a site between the PA and unique domains and a site between the unique and EGF-repeat domains were very protease-sensitive. Once one of the two sites was cut, the other became relatively protease resistant.

We wanted to determine the crystal structure of tBP80 as a first step towards understanding its ability to bind ligands in a specific manner. Because chromatographic purification in a previous study resulted in large losses and a low yield (Cao *et al.*, 2000), we chose to express a form with six His residues present at the C-terminus in order to facilitate its purification. Probably for technical reasons, expression of a tBP80 homolog from *Arabidopsis* termed truncated AtBP80b (Fig.1(C))(Hadlington & Denecke, 2000) with a C-terminal His tag was successful, while we were not able to obtain a His-tagged tBP80. We therefore have expressed truncated AtBP80b and purified it to homogeneity in two steps. Here, we report its crystallization and initial diffraction studies.

EXPERIMENTAL PROCEDURE

Expression of truncated AtBP80b in Drosophila S2 cells

The lumenal domain of plasmid Z38123 (Paris *et al.*, 1997) encoding the BP80 homolog AtBP80b was amplified by PCR using primers 5'-CCCTCCGGAGCCAACTTTGCCTGA AC and 5'-GGGGGTACCGGAATCATGAAGCTTGG. This resulted in the coding sequence for the lumenal domain, comprised of 560 amino acids bracketed by *Kpn*I and *Bsp*EI restriction sites. The DNA sequence of the PCR product was determined to ensure its fidelity. The PCR product was then digested with *Kpn*I and *Bsp*EI and inserted into that interval in plasmid pMT/V5 (Invitrogen, Carlsbad, CA, USA) to give an in-frame protein fusion at its C-terminus with the plasmid-encoded sequence LESRGPFEGKPIPNPLLGLD STGTGHHHHHH. Methods for the culture of *Drosophila* S2 cells, DNA transfection, selection of transformants and expression of recombinant protein have been described elsewhere (Cao *et al.*, 2000).

Purification of truncated AtBP80b

When cells reached a density of 1×10^7 ml^{-1}, recombinant protein expression was induced by the addition of copper sulfate to a final concentration of 500 μM. 3 d after induction, the medium was separated from the cells by centrifugation for 10 min at 1000*g*. The medium was dialyzed extensively at 277 K against 0.1 *M* Na_3PO_4, 0.01 *M* Tris pH 8 and then passed over a column of Ni^{2+}-nitriloacetic acid agarose (Qiagen, Valencia, CA, USA) to select protein with the C-terminal His tag. After washing the column with the

same buffer until $A_{280} \leq 0.01$, protein was eluted in 8-10 1 ml fractions with 0.1 *M* imidazole pH 7.7. Truncated AtBP80b-containing fractions were pooled and dialyzed overnight at 277 K against repeated changes of 20 m*M* MES, 50 m*M* NaCl pH 6.3. This material was then passed over a proaleurain peptide-affinity column (Kirsch *et al.*, 1994). The column was washed with the same buffer until $A_{280} \leq 0.01$. Bound protein was eluted in five 1 ml fractions of 0.1 *M* sodium acetate pH 4.0, which were then brought to neutrality with 1 *M* Tris base. Confirmation of the presence and purity of the protein was made by SDS-PAGE and immunoblot analysis using monoclonal antibody 17F9 (Cao *et al.*, 2000). The pooled fractions were then dialyzed against 20 m*M* Tris-HCl pH 7.4 at 277 K overnight.

Crystallization

The purified protein was concentrated to 12.0 mg ml^{-1} in 20 m*M* Tris buffer pH 7.5 using a YM30 membrane (Amicon). Crystallization trials were performed using the hanging-drop vapor-diffusion method at 277 K using 24-well VDX plates (Hampton Research). Each hanging drop was prepared by mixing equal volumes (2 µl each) of protein solution and reservoir solution, followed by placement over 0.5 ml reservoir solution. Initial crystallization conditions were established using the screening kits from Hampton Research Inc. (Crystal Screens I and II and PEG/Ion Screen) and Emerald BioStructures Inc. (Wizard I and II).

X-ray data collection and processing

The VSR crystal was flash-vitrified using the same reservoir solution as cryoprotectant. A crystal was picked up in a CryoLoop (0.5 mm diameter; Hampton Research) and immediately frozen in a nitrogen stream at 100 K. Diffraction data consisting of 200 images with 1° oscillation and extending to 2.5 Å were collected at the Stanford Synchrotron Radiation Laboratory (SSRL) beamline 7-1 (MAR 345 image-plate detector) using the BLUE-ICE interface. Indexing, integration of images and scaling of the intensity data were performed using the programs *DENZO* and *SCALEPACK* from the *HKL* package (Otwinowski & Minor, 1997). The native crystal data extends to 2.6 Å with a completeness of 82%. The R_{merge} and $\langle I/\sigma(I)\rangle$ values are 0.04 and 12.8, respectively, as summarized in Table 1 together with the data statistics.

RESULTS AND DISCUSSION

The *Drosophila* S2 expression system and our purification protocol allowed us to obtain pure AtBP80b protein with yields of approximately 1-2 mg of homogeneous AtBP80 protein per litre of S2 cell-culture medium. Initially, small cube-shaped crystals were obtained from two different screening solutions: (i) 0.2 *M* trisodium citrate dehydrate, 0.1 *M* Tris-HCl pH 8.5, 30%(*v*/*v*) PEG 400 and (ii) 0.05 *M* cesium chloride, 0.1 *M* MES pH 6.5, 30%(*v*/*v*) Jeffamine M-600. These initial conditions were refined to optimize the crystallization by varying the concentration of the ingredients and the pH. The best crystals were obtained by mixing 2.0 µl protein solution (6 mg ml^{-1} in 20 m*M* Tris-HCl pH 7.5) with an equal volume of reservoir solution containing 0.15 *M* tricitrate dehydrate, 0.1 *M* Tris-HCl pH 8.0, 30%(*v*/*v*) PEG 400. Under these optimized conditions, the crystals grew to dimensions of 0.3 × 0.3 × 0.3 mm after two weeks (Fig. 2) and the crystals belong to the cubic space group $P2_13$ (or $P4_232$), with unit-cell parameters $a = b = c = 145.9$ Å. The calculated V_M (Matthews coefficient) is 2.13 $Å^3$ Da^{-1}, with a solvent content of 42.14% (Matthews, 1968; Kantardjieff & Rupp, 2003), which corresponds to one or two molecules in the asymmetric unit with respect to the choice of space group.

In order to overcome the crystallographic phase problem, multiple-wavelength anomalous dispersion (supplement 2) and multiple isomorphous replacement (supplement 3) methods are being applied, since the molecular-replacement approach is not currently possible.

REFERENCE

Cao, X., Rogers, S. W., Butler, J., Beevers, L. & Rogers, J. C. (2000). *Plant Cell*, **12**, 493-506.

Denecke, J., Botterman, J. & Deblaere, R. (1990). *Plant Cell*, **2**, 51-59.

Hadlington, J. L. & Denecke, J. (2000). *Curr. Opin. Plant Biol.* **3**, 461-469.

Holwerda, B. C., Padgett, H. S. & Rogers, J. C. (1992). *Plant Cell*, **4**, 307-318.

Humair, D., Hernández Felipe, D., Neuhaus, J.-M. & Paris, N. (2001). *Plant Cell*, **13**, 781-792.

Jiang, L. & Rogers, J. C. (1998). *J. Cell Biol.* **143**, 1183-1199. Kantardjieff, K. A. & Rupp, B. (2003). *Protein Sci.* **12**, 1865-1871.

Kirsch, T., Paris, N., Butler, J. M., Beevers, L. & Rogers, J. C. (1994). *Proc. Natl Acad. Sci. USA*, **91**, 3403-3407.

Luo, X. & Hofmann, K. (2001). *Trends Biochem. Sci.* **26**, 147-148.

Mahon, P. & Bateman, A. (2000). *Protein Sci.* **9**, 1930-1934.

Matsuoka, K. & Neuhaus, J.-M. (1999). *J. Exp. Bot.* **50**, 165-174.

Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491-497.

Okita, T. W. & Rogers, J. C. (1996). *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **47**, 327-350.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307-326.

Paris, N., Rogers, S. W., Jiang, L., Kirsch, T., Beevers, L., Phillips, T. E. & Rogers, J. C. (1997). *Plant Physiol.* **115**, 29-39.

Phillipson, B. A., Pimpl, P., daSilva, L. L., Crofts, A. J., Taylor, J. P., Movafeghi, A., Robinson, D. G. & Denecke, J. (2001). *Plant Cell*, **13**, 2005-2020.
Vitale, A. & Galili, G. (2001). *Plant Physiol.* **125**, 115-118.

FIGURE REGEND

Fig. 1. Schematic diagrams of BP80 (A), tBP80 (B) and AtBP80b(C) structure.

Fig. 2. Picture of an AtBP80b crystal. This crystal was grown at 277 K. After two weeks of growth, the crystal grew to dimensions of 0.3 × 0.3 × 0.3 mm.

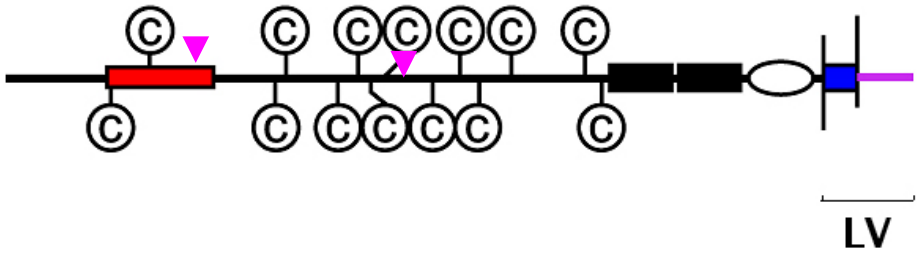
Table 1. Data-collection statistics

X-ray wavelength (Å)	0.97 (SSRL BL 7-1)
Space group	$P2_13$ (or $P4_23_2$)
Unit-cell parameters (Å)	$a = b = c = 145.9$
Resolution (Å)	50.00 – 2.66
No. of measured reflections	186,569
No. of unique reflections	24,159
R_{merge}[#] (%)	4.3 (66.1)
Average redundancy	7.87 (9.21)
Data completeness (%)	81.9 (24.1)
$\langle I/\delta(I)\rangle$	12.8 (2.3)

[#]$R_{merge} = \sum_h \sum_i |I(h)_i - \langle I \rangle| / \sum_h \sum_i I(h)_i$, where $I(h)$ is the intensity of reflection h, $\sum_h$ is the sum over all reflections and $\sum_i$ is the sum over i measurements of reflection h.

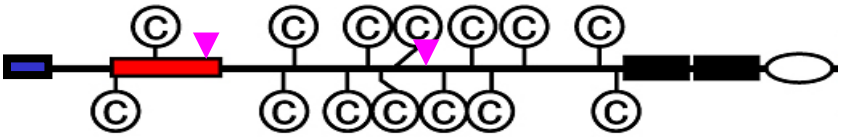
*Values in parentheses refer to the highest resolution shell (2.71-2.66 Å).

Fig. 1 (A)



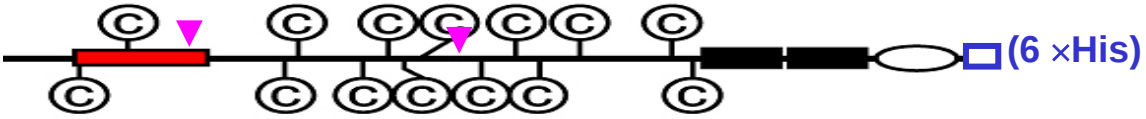
© :cystein residue, :Protease associated (PA) domain, :Transmembrane domain (TMD) , : Cytoplasmic tail (CT), : EGF repeats, : Calcium binding EGF repeat, : Glycosylation site

Fig. 1 (B)



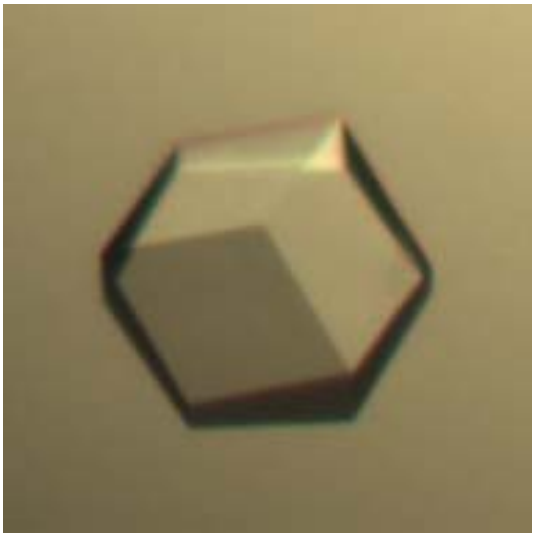
: Signal Peptide

Fig. 1 (C)



(6×His): Extra sequence by expression vector

Fig. 2



SUPPLEMENT 2

MULTIPLE WAVELENGTH ANOMALOUS DISPERSION (MAD) METHOD

The x-ray diffraction data for the AtBP80b crystal has been collected. In order to overcome the previously mentioned crystallographic phase problem of AtBP80b, multiple-wavelength anomalous dispersion method is going to be applied, which is an almost unique choice in this case due to its lack of sequence similarity with any known structure.

CONCEPTS

Generally, electrons in an atom are regarded as free electrons, which scatter the beam at a phase difference of 180 ° with respect to the incident beam. This is the case for the most of the light atoms in biomolecules such as H, C, O, N, S. When the x-ray wavelength approaches an absorption edge, scattered x-ray maximizes an imaginary component to its phase (f″ scattering coefficient becomes non-zero) because of retardation compared to a normally scattered x-ray. This effect is called *anomalous scattering*. The scattering factor *f* of anomalous scattering can be written as

$$f_{anom} = f + f' + if'' \qquad \text{Eq.1}$$

where f' and f'' are real and imaginary contributions of anomalous signal, respectively. Anomalous scattering does not follow typical Friede's Law (1) that states, "*members of a Friedel pair have equal amplitude and opposite phase* ($\mathbf{F}hkl = \mathbf{F}\overline{hkl}$)". The real scattering

component f' is related to f'' via the Kramers-Kronig relationship (2, 3). This anormalous behavior can be used for the structure determination.

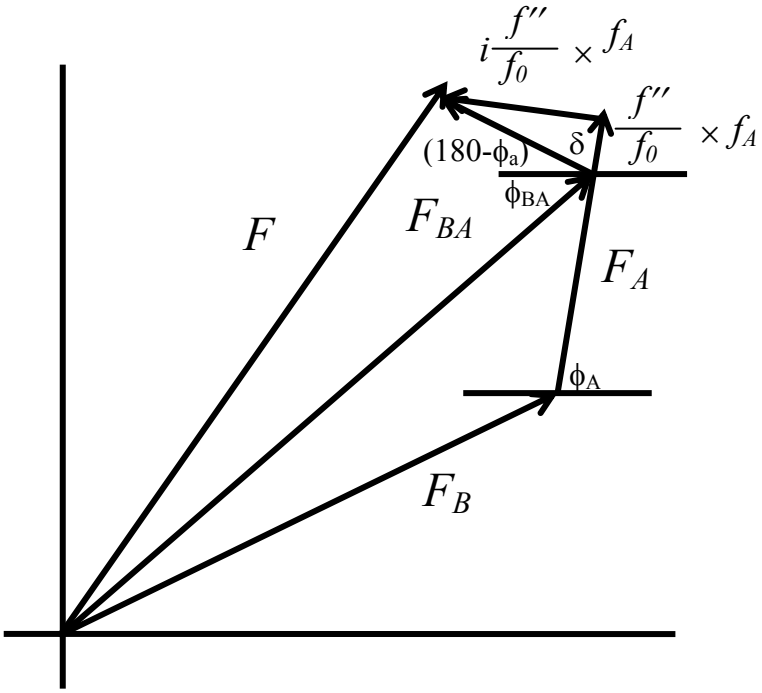
Changes in the structure factor are induced by tuning the X-ray wavelength to different points in the absorption spectrum of the incorporated heavy atom, thus changing the values of the anomalous scattering factors f' and f'', which in turn affect the structure factor. In MAD method, this may be achieved by measuring diffraction data of at least two different X-ray energies. Usually diffraction data sets are collected at three different X-ray energies; the edge, where f' is minimum; the peak, where f'' is maximum and the Fridel pair has the largest difference; the remote, where f' and f'' are relatively small (3, 4).

Algebraic process of MAD data

This part is largely based on the Chapter 9 in the "*Principle of Protein X-ray Crystallography*", by Jan Drenth (2). Each anomalously scattered atom has an atomic scattering factor of $f = f_0 + f' + if''$ (Eq.1). In the complex vector diagram, F_B is the contribution to the structure factor by the nonamomalously scattered atoms, F_A is the nonamomalous contribution of the anomalously scatted atom, and the complete nonamomalous part is $F_{BA} = F_B + F_A$.

The anomalous scattering contribution, a, is

$$a = \frac{f'}{f_0} \times F_A + i\frac{f''}{f_0} \times F_A \qquad \text{Eq.2}$$



Φ_{BA} is the phase angle of F_{BA}, Φ_A of vector F_A, and Φa of vector a. $\Delta\Phi = \Phi_{BA} + (180° - \Phi_a) = 180° + (\Phi_{BA} - \Phi_a)$.

Amplitude of the structure factor *F* is

$$|F|^2 = |F_{BA}|^2 + |a|^2 - 2\,|F_{BA}| \times |a| \times \cos\Delta\Phi \qquad \text{Eq.3}$$

with

$$|a|^2 = \frac{(f')^2 + (f'')^2}{f_0^{\,2}} \times |F_A|^2,$$

$$|F|^2 = |F_{BA}|^2 + \frac{(f')^2 + (f'')^2}{f_0^{\,2}} \times |F_A|^2 + 2\frac{f'}{f_0}\cos\delta\,|F_{BA}| \times |F_A|\ \cos(\Phi_{BA} - \Phi_a)$$

$$+2\frac{f''}{f_0}\ \sin\delta\,|F_{BA}| \times |F_A|\ \cos(\Phi_{BA} - \Phi_a)$$

and this can be rewritten as

$$\left|F\right|^2 = \left|F_{BA}\right|^2 + p\left|F_A\right|^2 + \left|F_{BA}\right| \times \left|F_A\right| \times [\, q\cos(\Phi_{BA}-\Phi_A) + r\sin(\Phi_{BA}-\Phi_A)] \qquad \text{Eq.4}$$

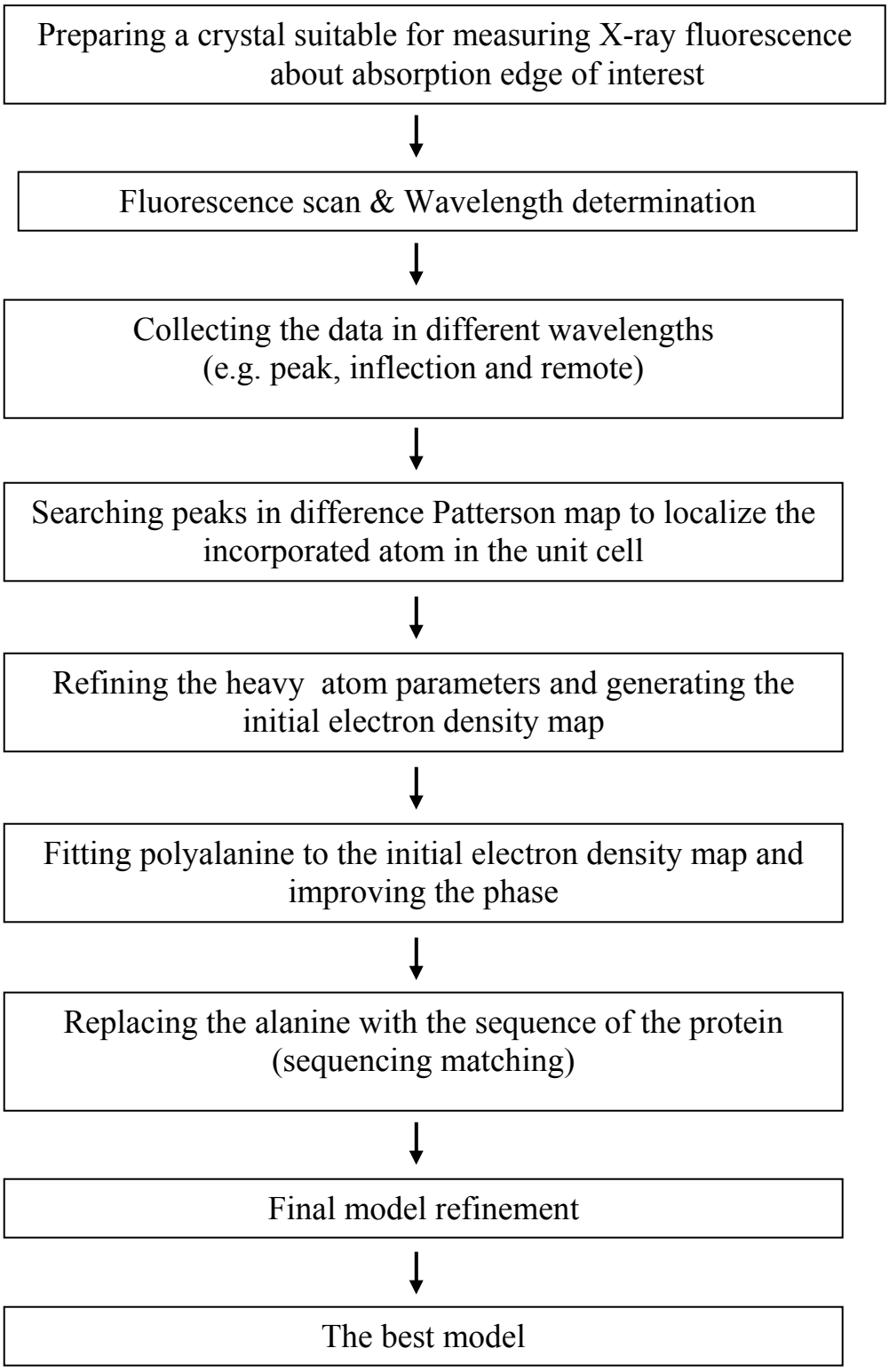
$$\text{with } p = \frac{\left(f'\right)^2 + \left(f''\right)^2}{\left(f_0\right)^2},\ q = 2\frac{f'}{f_0} \text{ and } r = 2\frac{f''}{f_0}.$$

p, *q* and *r* are functions of wavelength and can be derived from the atomic absorption coefficient. $|F|^2$ can be determined experimentally, which is intensity data of the crystal. Unknown quantities $|F_{BA}|$, $|F_A|$ and (Φ_{BA} - Φ_A) are independent of wavelength and equal for the *Friedel* pairs, except for the sign of (Φ_{BA} - Φ_A). Therefore, a minimum of two different wavelengths is sufficient to find $|F_{BA}|$, $|F_A|$ and (Φ_{BA} - Φ_A) for each reflection. Φ_{BA} can obtained by calculating Φ_A, which can be derived from a Patterson map with coefficient $|F_A|^2$ or by direct methods.

Because no anomalous scattering is taken into account for the calculation of the A-structure, the real phase or its enantiomorph is obtained for F_A. The solution for this is to calculate Φ_A for both structures. This gives two solutions for Φ_{BA} and best electron density map will be the final solution (2, 5).

Typically, the selenomethionine containing minimal media is being used in order to incorporate the selenium in places of sulfer. By collecting the anormalous data in a tunable wavelength in synchrotron beam, the phasing can be done.

OVERALL PROCESS OF MULTIWAVELENGTH ANOMALOUS DISPERSION



REFERENCE

1. Hauptman H. (1993) *Acta Crystallogr D.* 49, 3-8.
2. Drenth, J. (1999) Principle of protein X-ray crystallography, 2nd Ed., Springer-Verlag, Inc.
3. MacRee, D. (1999) Practical protein crystallography, 2nd Ed., Academic press
4. Rhodes, G. (2000) Crystallography made crystal clear, 2nd Ed., Academic press
5. Read RJ. (1996) *Structure*. 4, 11-14.

SUPPLEMENT 3

MULTIPLE ISOMORPHOUS REPLACEMENT (MIR) METHOD

There is no structure available in PDB that has an adequate sequence similarity with AtBP80b. Therefore molecular replacement can not be used for the structure determination. We have been having a trouble to use the ^{Se}Met-containing media for culturing AtBP80b clone, therefore, as a third method, we have to depend on MIR.

CONCEPTS

Isomorphous replacement method, developed by Green, Ingram and Perutz (1), is still one of the most popular phasing methods and it is, except MAD, the only method available for solving the protein structures with very limited structural information.

Application of the isomorphous replacement method requires the X-ray diffraction data of the native protein crystal as well as that of the heavy atom derivative crystal (2). The single isomorphous replacement method (SIR) and multiple isomorphous replacement (MIR) are based on the introduction of one or more heavy atoms as new scatterers of high atomic number without disturbing the crystal packing or molecular structure. For perfect isomorphism, the conformation of the protein, the position and orientation of its molecules as well as the unit cell parameters in the native and in the derivative crystals must be exactly same. Under this restricted conditions, the intensity differences between the native and the other patterns are then exclusively due to the attached heavy atom.

$\mathbf{F}_P$ is the structure factor of protein, $|\mathbf{F}_P|$ with its magnitude and ϕ_P with its phase. $\mathbf{F}_{PH}$ is contribution from the heavy atom plus protein. If perfect isomorphism is assumed, structure factor of the heavy-atom isomorphous derivatives are simply defined as Eq.1.

$$\mathbf{F}_{PH} = \mathbf{F}_P + \mathbf{F}_H \qquad \text{Eq.1}$$

where $\mathbf{F}_H$ is the structure-factor contribution from the heavy atom only. The final goal is to derive a value of ϕ_H, but only $|\mathbf{F}_P|$ and $|\mathbf{F}_{PH}|$ can be measured. $|\mathbf{F}_H|$ can not be measured, but can be calculated by means of *difference-Patterson*-techniques (3). Once the position of the heavy atom is more or less accurately determined, the calculated value of ϕ_H can be used to estimate angle of structure factor of the native protein.

$$\phi_P = \phi_H + \cos^{-1}[(|\mathbf{F}_{PH}|^2 - |\mathbf{F}_P|^2 - |\mathbf{F}_H|^2) / 2|\mathbf{F}_P||\mathbf{F}_H|] \qquad \text{Eq.2}$$

Since this equation contain an inverse cosine term, two solutions can exist for ϕ_P. This ambiguity is illustrated in Fig. 1. Actually, the MIR was the first method to solve the phase ambiguity.

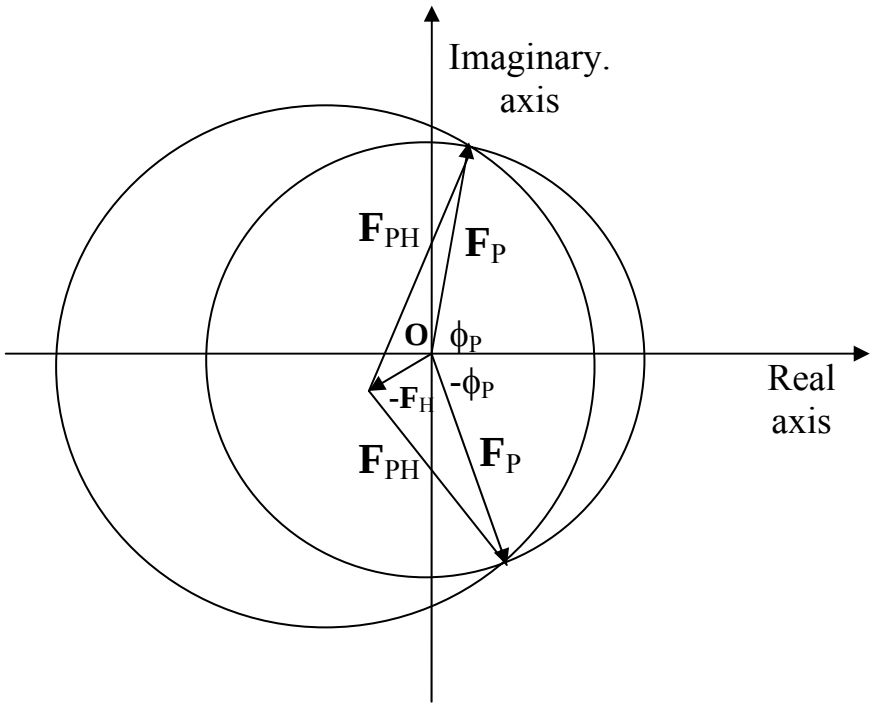


Fig .1

It is based on second heavy atom bound to the protein molecule in a position different from the first one. Therefore, both heavy atom positions must be derived (2, 4, 5, 6). The graphical construction in Fig.2 is equivalent to solving the pair of simultaneous equations occurring in Eq.3.

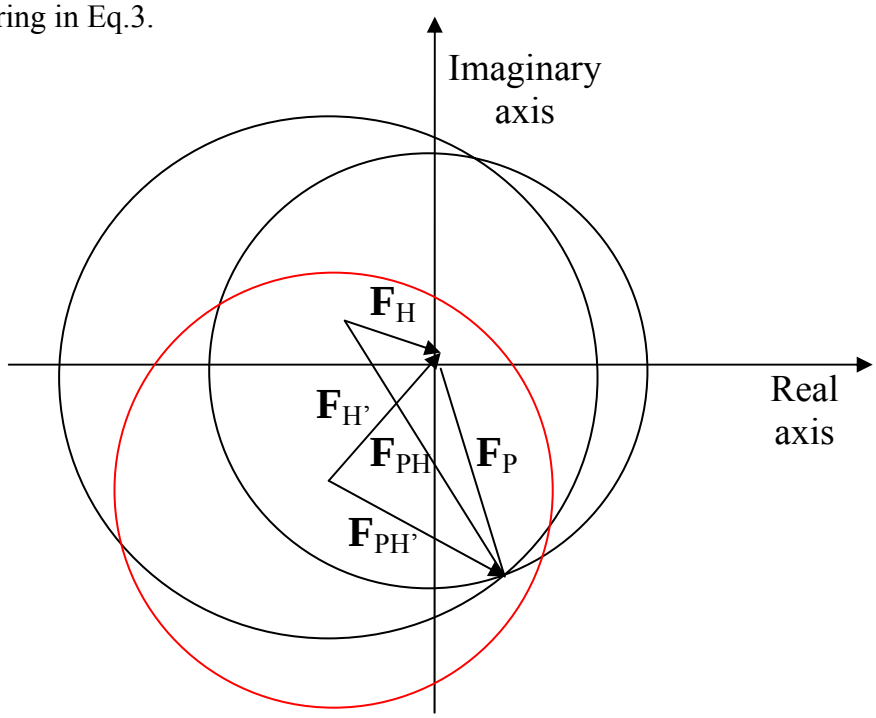


Fig. 2

$$\phi_P = \phi_H + \cos^{-1}[(|\mathbf{F}_{PH}|^2 - |\mathbf{F}_P|^2 - |\mathbf{F}_H|^2) / 2 |\mathbf{F}_P| |\mathbf{F}_H|]$$

$$\phi_P = \phi_{H'} + \cos^{-1}[(|\mathbf{F}_{PH'}|^2 - |\mathbf{F}_P|^2 - |\mathbf{F}_{H'}|^2) / 2 |\mathbf{F}_P| |\mathbf{F}_{H'}|] \quad \text{Eq.3}$$

Theoretically, two different heavy-atom derivatives yield an unambiguous value for the phase angle of the native protein, but even small experimental errors in the amplitudes will lead to rather uncertain phases. Therefore, in practice, we need more than two derivative data and a sophisticated weighting algorithm to sort this ambiguity.

<u>The phasing power of Heavy atom</u>

The phasing power and figure of merit are the most popular way to judge the quality of phasing by MIR. Following are the mathematical basis for these concepts. This part is based on Appendix 2 in the "*Principle of Protein X-ray Crystallography*", by Jan Drenth (2).

Isomorphous phasing power:

$$\sqrt{\frac{\sum_{hkl}\left|F_H(calc)\right|^2}{\sum_{hkl}\left|E\right|^2}}$$

with $$\sum_{hkl}\left|E\right|^2 = \sum_{hkl}\left[\left|F_{PH}(obs)\right| - \left|F_{PH}(calc)\right|\right]^2$$

alternative expression is: $$\frac{\sum_{hkl}\left|F_H(calc)\right|}{\sum_{hkl}\left|E\right|}$$

The figure of merit for a given reflection (*h k l*) is defined as

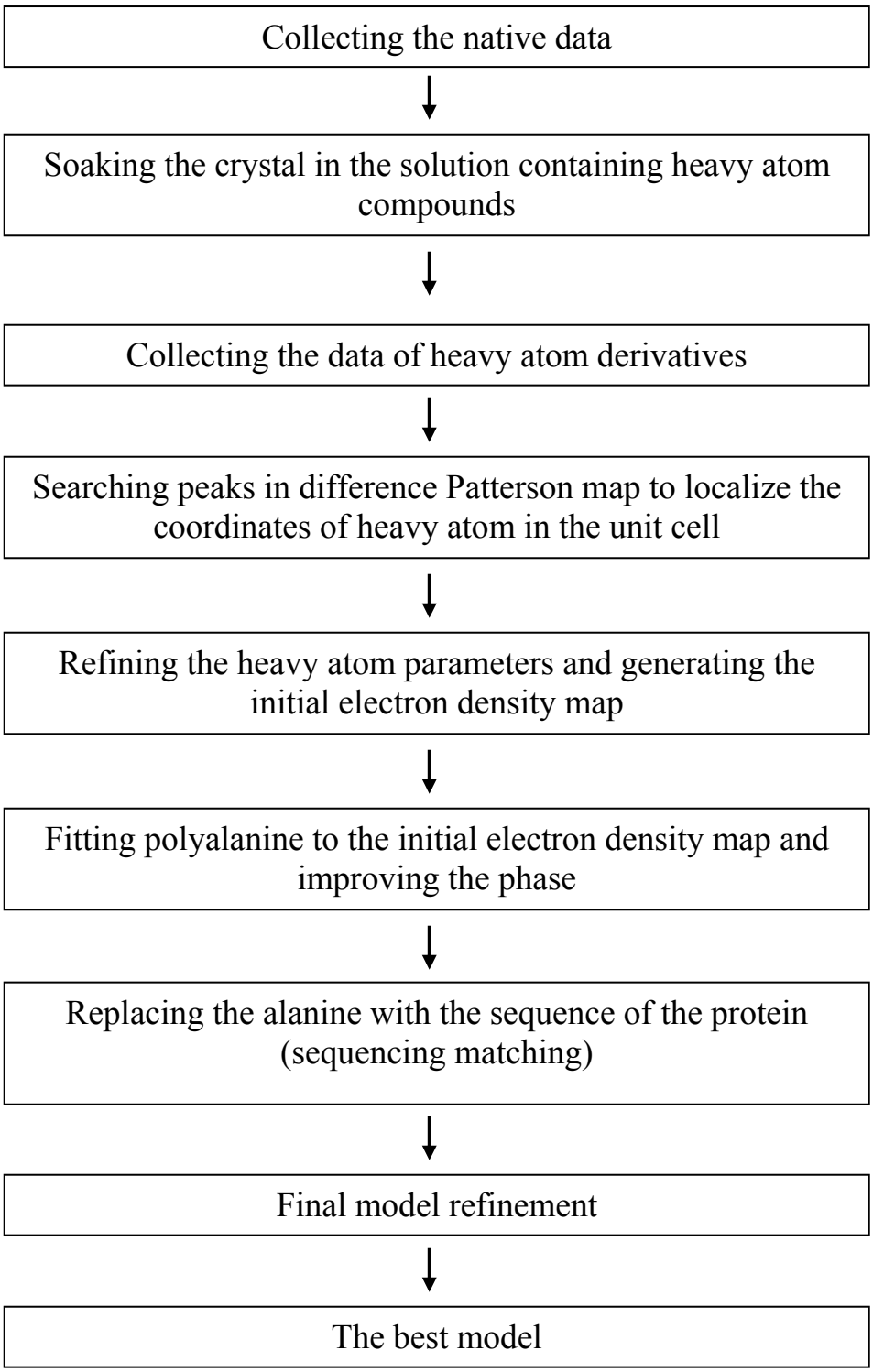
$$m = \frac{\left|F(hkl)_{\text{best}}\right|}{\left|F(hkl)\right|}$$

Where

$$\mathrm{F}(hkl)_{\text{best}} = \frac{\sum_{\alpha} P(\alpha)\mathrm{F}_{hkl}(\alpha)}{\sum_{\alpha} P(\alpha)}$$

It can be shown that the figure of merit is the weighted mean of the cosine of the deviation of the phase angle from $\alpha_{best:}$ $m = \overline{\cos[\alpha - \alpha(\text{best})]}$.

OVERALL PROCESS OF MULTIPLE ISOMORPHOUS REPLACEMENT



REFERENCE

1. Green, D.W., Ingram, V.M. and Perutz M.F. (1954) *Proc. Roy. Soc.* 225, 287-307
2. Drenth, J. (1999) Principle of protein X-ray crystallography, 2nd Ed., Springer-Verlag, Inc.
3. Reed RJ.(2003) *Acta Crystallogr D.* 59, 1881-1890
4. MacRee, D. (1999) Practical protein crystallography, 2nd Ed., Academic press
5. Karle J. (1989) *Acta Crystallogr A.* 45, 765-781.
6. Rhodes, G. (2000) Crystallography made crystal clear, 2nd Ed., Academic press