This is to certify that the
dissertation entitled

X-RAY CRYSTALLOGRAPHIC STUDIES OF RNA POLYMERASE III TRANSCRIPTION FACTOR TFIIIB AND 1L-*MYO*-INOSITOL 1-PHOSPHATE SYNTHASE

presented by

Xiangshu Jin

has been accepted towards fulfillment
of the requirements for the

Ph.D. degree in Chemistry

James J. Stezowski
Major Professor's Signature

12/9/02
Date

X-RAY CRYSTALLOGRAPHIC STUDIES OF RNA POLYMERASE III TRANSCRIPTION FACTOR TFIIIB AND 1L-*MYO*-INOSITOL 1-PHOSPHATE SYNTHASE

By

Xiangshu Jin

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Chemistry

2002

ABSTRACT

X-RAY CRYSTALLOGRAPHIC STUDIES OF RNA POLYMERASE III TRANSCRIPTION FACTOR TFIIIB AND 1L-*MYO*-INOSITOL 1-PHOSPHATE SYNTHASE

By

Xiangshu Jin

TFIIIB is the central initiation factor in the RNA polymerase III transcription system. Other transcription factors of RNA polymerase III such as TFIIIA, TFIIIC, and SNAPc act as assembly factors to recruit TFIIIB to promoters. TFIIIB then allows the recruitment of RNA polymerase III for transcription. In order to investigate the transcription initiation mechanism of RNA polymerase III, three-dimensional structures of TFIIIB-DNA complexes are essential. Toward this goal, recombinant mutants of individual members of TFIIIB, namely BRF, B″, and TBP, containing functional domains were designed, expressed in *E.coli* cells, and purified to homogeneity for crystallization. Various promoter DNAs with different lengths were also designed and purified. BRF functional domains, *S. cereviaise* BRF435-531, *S. cereviaise* BRF435-596 (Δ 538-550), *K. lactis* BRF395-501 have been crystallized. Crystallization of BRF/TBP/DNA complexes was carried out extensively and crystals of 11 different BRF/TBP/DNA complexes were obtained.

1L-*myo*-inositol 1-phosphate (MIP) synthase catalyzes the isomerization of D-glucose 6-phosphate to 1L-*myo*-inositol 1-phosphate, the first committed and rate-limiting step during the *de novo* biosynthesis of inositol-containing compounds important

in signal transduction, cell wall biogenesis, etc. NAD^+ serves as a cocatalyst during the reaction. In order to investigate the mechanism of MIP synthase, structures of MIP synthase in its *apo* form, NAD^+-bound form, NADH-bound form, and in complex with NAD^+ and a high affinity inhibitor, 2-deoxy-D-glucitol 6-*(E)*-vinylhomophosphonate were studied by X-ray crystallography. While the active site residues 351-375 were missing in the *apo* structure, the fully occupied NAD^+ folds a short helix, α13, that encompasses 351-361, making interactions with it. When the enzyme is coupled with NADH, several dramatic structural changes were observed: all active site residues that were disordered in the *apo* and NAD^+-bound structures are now completely ordered; the conformation of the NADH molecule changed significantly from its position in the NAD^+-bound structure; two small molecules, phosphate and glycerol, are bound in the enzyme active site mimicking the substrate binding. The unambiguous position of the phosphate in the NADH-bound structure differs significantly from the position of the phosphate in the previously reported structure of MIP synthase/NAD^+/2-deoxy-D-glucitol 6-phosphate. In the structure of MIP synthase in complex with NAD^+ and an inhibitor, 2-deoxy-D-glucitol 6-(*E*)-vinylhomophosphonate, the inhibitor is fully occupied in the active site of one of two molecules in the asymmetric unit. Based on the new structural data, a mechanism of MIP synthase was proposed.

To my parents

ACKNOWLEDGMENTS

First of all, I would like to acknowledge my advisor, Prof. James H. Geiger for his guidance, encouragement, and support throughout the entire course of my graduate study. I sincerely thank him for all the invaluable experiences that I have had in his lab, they helped me become a better researcher and a stronger person. I would like to thank our collaborator, Dr. Steven Hahn, for providing the plasmids of BRF constructs and B″ and carrying out the Gel Shift assays. I also want to thank Prof. John W. Frost and his group for providing the inhibitor of MIP synthase that was used in this study.

I would like to acknowledge former and current Geiger group members, Dr. Stacy Hovde, Dr. Marta Abad, Dr. Adam Stein, Paul Booth, Christopher Geiger, Mike Wolf, Keith Freel, Erika Mathes, Sara Weaver, and Aimee Brooks for their help, support, and friendship. Ever since I joined the Geiger group, it has been like a family to me; my life would never have been the same without having such a nice group of people around. I want to acknowledge Stacy for all of her help and encouragement, her kind care helped me to go on with courage and strength during my difficult times. Thanks a lot, Stacy! I also want to thank Marta for patiently helping me out with the basics and giving me helpful advice for all the days that we shared in the Geiger lab. I will always remember your cheerful character. I thank Adam for his grounding work on the MIP synthase project. I want to thank Paul for his great patience in communicating with me and teaching me the basics in the early days. I will never forget your encouraging smiles. A big thanks goes to Mike for his help working on the TFIIIB project as well as the Valproate inhibition assay. Christopher, Sara, and Keith have also worked on the TFIIIB project, I am very appreciative of their help! I also want to mention Erika. I remember

vividly those days when you started working in the lab, and I am so happy for what you have accomplished. All members of the Geiger group from past to present that I had the chance to meet and work with have my sincere gratitude for their friendship, I wish all of you good luck on your adventures in Crystallography.

Finally, I want to thank my parents for their unconditional love, support, and understanding. Without the love from my family, which is inseparable by distance, it would not have been possible for me to overcome a variety of challenges that I have encountered in my pursuit of academic goals.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER II: CRYSTALLIZATION AND PRELIMINARY X-RAY DIFFRACTION ANALYSIS OF THE TFIIIB COMPLEX

CHAPTER III: X-RAY CRYSTALLOGRAPHIC STRUCTURAL STUDIES OF 1L-*MYO*-INOSITOL 1-PHOSPHATE SYNTHASE

LIST OF FIGURES

Images in this dissertation are presented in color.

CHAPTER I: INTRODUCTION

CHAPTER II: CRYSTALLIZATION AND PRELIMINARY X-RAY DIFFRACTION ANALYSIS OF THE TFIIIB COMPLEX

CHAPTER III: X-RAY CRYSTALLOGRAPHIC STRUCTURAL STUDIES OF 1L-*MYO*-INOSITOL 1-PHOSPHATE SYNTHASE

LIST OF ABBREVIATIONS

A – alanine
AC – adenylate cyclase
ADA2 – adenosine deaminase 2
A. fulgidus – Archaeglobus fulgidus
AMP - adenosine monophosphate
ATP – adenosine triphosphate

BM – bending magnet
BME - β-mercaptoethanol
bp – base pair
BRF – TFIIB-related factor

C – cysteine
C/CYT – cytosine
Cα - the alpha carbon
CDP – cytodine diphosphate
CNS – Crystallography & NMR System
C terminal - carboxy terminal

D - aspartic acid
DAG - diacylglycerol
DNA – deoxyribonucleic acid
DPE – downstream promoter element
DSE –distal sequence element
DTT - dithiothreitol

E - glutamic acid
E. coli - Escherichia coli
EDTA – ethylenediamine tetraacetic acid
EMSA – electrophoretic mobility shift assay
EPS - electrostatic potential surface

F - phenylalanine
F_{cal} - calculated structure factors
F_{obs} - observed structure factors
FPLC – fast pressure liquid chromatography

G - glycine
G/GUA – guanine
GDP – guanine diphosphate
GTP - guanine triphosphate
G-protein – guanine nucleotide binding protein
Gu - guanidine

H/His - histidine
HEPES - N-[2-hydroxyethyl] piperazine-N'-[ethane sulfonic acid]
HPLC – high pressure liquid chromatography

I - isoleucine
I – intensity
ID – insertion device
IMCA-CAT – Industrial Macromolecular Crystallography Association Collaborative Access Team
1,4,5-IP_3 – *myo*-inositol-1,4,5-triphosphate
IPTG – isopropyl-β-D-thiogalactoside

K - lysine
K. lactis – *Kluyveromyces lactis*

L – leucine
LB – Luria-Bertani

M - methionine
MI – *myo*-inositol
MIP – *myo*-inositol 1-phosphate
MPD - 2-Methyl-2,4-pentanediol
M.W. – Molecular weight
M. tuberculosis – *Mycobacterium tuberculosis*

N – asparagine
NAD^+ - nicotinamide adenine dinucleotide
NADH - nicotinamide adenine dinucleotide, reduced form
N-cor – nuclear receptor co-repressor

O.D. – Optical density

P – proline
PEG - polyethylene glycol
PDB – protein data bank
PKA – protein kinase A
PLC – phospholipase C
PMSF – phenyl methyl sulfonyl fluoride
POL – RNA polymerase
PSE – proximal sequence element
PtdIns – phosphatidyl inositol
PtdIns-(4,5)-P_2 - phosphatidyl inositol-4,5-bisphosphate

Q – glutamine

R - arginine
RNA – ribonucleic acid
rmsd - root mean square deviation
rRNA – ribosomal RNA

S – serine
SANT – SWI3, ADA2, N-cor, TFIIIB
S. cerevisiae – Sacchromyces cerevisiae
SDS-PAGE – sodium dodecyl sulfate–polyacrylamide gel electrophoresis
SNAPc – small nuclear RNA activating protein complex
snRNA – small nuclear RNA
SWI3 – SWItch gene product 3

T - threonine
T/THY – thymine
TAF – TBP associated factor
TBP – TATA binding protein
TFB – transcription factor B
TFIIA – transcription factor IIA
TFIIB – transcription factor IIB
TFIIIA – transcriptionm factor IIIA
TFIIIB – transcriptionm factor IIIB
TFIIIC – transcriptionm factor IIIC
TRF1 – TBP related factor 1
Tris – 2-amino-2-(hydroxymethyl)-1,3-propanediol
tRNA – transfer RNA

U – uracil
U6-MLP – U6-major late promoter
UBF – upstream binding factor

V - valine

W - tryptophan
WT - wild type

Y - tyrosine

CHAPTER I

INTRODUCTION

1.1 TFIIIB

1.1.1 TRANSCRIPTION

Transcription is the process by which the genetic message in DNA is transcribed into RNA. Eukaryotic cells contain three distinct transcription systems, each consisting of one of three RNA polymerases and a unique set of transcription factors. Accurate and efficient transcription requires RNA polymerase in association with several transcription factors to be properly located at a specific start sequence of DNA called a promoter.

RNA polymerase I (Pol I) transcribes ribosomal RNA (rRNA) genes. Accurate initiation of transcription by RNA polymerase I requires the upstream binding factor (UBF) and the selectivity factor SL-1 complex, which includes the ubiquitous TATA binding protein (TBP) in addition to several other novel factors unique to this complex (1-3).

RNA polymerase II (Pol II) transcribes the protein-coding messenger RNA (mRNA) genes and some small nuclear RNA (snRNA) genes. RNA polymerase II transcription is the most complex among the three transcription systems, and is by far the most studied system. All of the RNA polymerase II transcription initiation factors have been identified, and much is known about how these factors associate at the core promoter to recruit RNA polymerase II to initiate transcription (4,5). TFIID is the only basal initiation factor capable of specifically recognizing a core promoter element and it can do so by binding to one or more of the core promoter DNA sequences (6). TFIID is a large multi-protein complex consisting of TBP and several ancillary TBP associated

factors (TAFs) (7-10). While TBP specifically recognizes the TATA box core promoter element, other TAFs can recognize the initiation core element or the downstream promoter element (DPE). TBP is the central factor in this complex and its recognition of the TATA element is thought to be the most general and critical of core promoter recognition events for most RNA polymerase II promoters (11,12).

Though each of the three eukaryotic RNA polymerases uses distinct sets of transcription initiation factors, all three RNA polymerases share some similarities. First the RNA polymerases have five shared subunits and several other homologous subunits(13). Second, all three polymerases require TBP both *in vitro* and *in vivo* for initiation, even though TBP is contained in distinct complexes specific to each polymerase(14). Third, both RNA polymerase II and RNA polymerase III transcription systems have a TFIIB-like factor required for transcription initiation (15).

1.1.2 Transcription by RNA polymerase III

1.1.2.1. Promoters of RNA polymerase III-transcribed genes

The genes transcribed by RNA polymerase III (Pol III) encode a variety of small RNA molecules. Many of these have essential functions in cellular metabolism: tRNA and 5S rRNA are required for protein synthesis; 7SL RNA is involved in intracellular protein transport; U6, H1, and MRP RNAs are involved in posttranscriptional processing; VA RNAs of adenovirus serve to divert the translational machinery of an infected cell towards the more effective production of viral proteins (16). RNA polymerase III also transcribes genes with no known function, for example, the 7SK genes and the short interspersed repeat (SINE) (17,18). The distinguishing feature of RNA polymerase III

promoters as a class is the preponderance of gene-internal (transcribed) promoter elements. The promoters of RNA polymerase III genes include discontinuous intragenic structures that are composed of essential sequence blocks separated by nonessential nucleotides. It is convenient to distinguish three classes of these promoters (Figure 1.1): type I genes (5S rRNA genes), type II genes (tRNA and 7SL RNA genes), and type III genes (U6 small nuclear RNA and 7SK small cytoplasmic RNA genes)(16).

Type I gene promoters:

The principal feature of type I genes is a binding site for TFIIIA. The mode of DNA binding by TFIIIA is best understood for the *Xenopus* protein, which interacts sequence-specifically with three intragenic sites, encompassing a span of approximately 50 base pairs: the transcription start-proximal box A, the intermediate element (ie), and box C.

Type II gene promoters:

Type II gene promoters have highly conserved internal promoter elements, box A and box B, that constitute the noncontiguous binding site of TFIIIC. The box A-box B separation is variable over a wide range, but separation of 30-60 bp is optimal for concurrent occupancy of both sites by TFIIIC.

Type III gene promoters:

The promoters of type III genes have exclusively external elements, such as the entirely upstream-located promoters of the vertebrate U6 snRNA genes. The upstream promoter elements TATA box, Proximal Sequence Element (PSE), and Distal Sequence Element (DSE) play a dominant role in type III genes. The organization of these promoters is particularly important because they contain dual function-Pol II/III elements

and they determine distinctive pathways for assembling transcription initiation complexes.

The intragenic sequences of type III genes are highly conserved between different genes and different species. This reflects a strong conservation of the general transcription factors employed by RNA polymerase III. The flanking sequences of type I and II promoters show little or no conservation, this suggests that these sequences are more likely to be recognized by gene- or species-specific factors or that their cognate factors have very flexible DNA binding specificities (16).

1.1.2.2 Transcription factors utilized by RNA polymerase III

Purified RNA polymerase III initiates transcription randomly. Accurate and specific initiation by RNA polymerase III requires the assistance of transcription factors in order to recruit the polymerase to the appropriate start sites of genes. RNA polymerase III transcription factors include both general factors and gene-specific factors. General factors include TFIIIB, TFIIIC, SNAPc etc., and gene-specific factors include TFIIIA, PSE-binding factor, OCT-1, etc. (19,20). Among all RNA polymerase III transcription factors, TFIIIB is the central initiation factor since it alone can recruit the polymerase and specifies the transcription start site. TFIIIC and TFIIIA, the other components of the core transcription apparatus, bind DNA and serve as assembly factors for TFIIIB; the six-subunit TFIIIC complex interacts directly with TFIIIB, and TFIIIA serves as a 5S rRNA gene-specific platform for TFIIIC. TFIIIB is composed of three subunits: TBP, BRF, and B″. TFIIIB alone usually does not show sequence-specific DNA binding, it must be recruited onto DNA by interaction with other transcription factors. For instance yeast TFIIIB can bind to tRNA genes only if TFIIIC is bound (Figure 1.2).

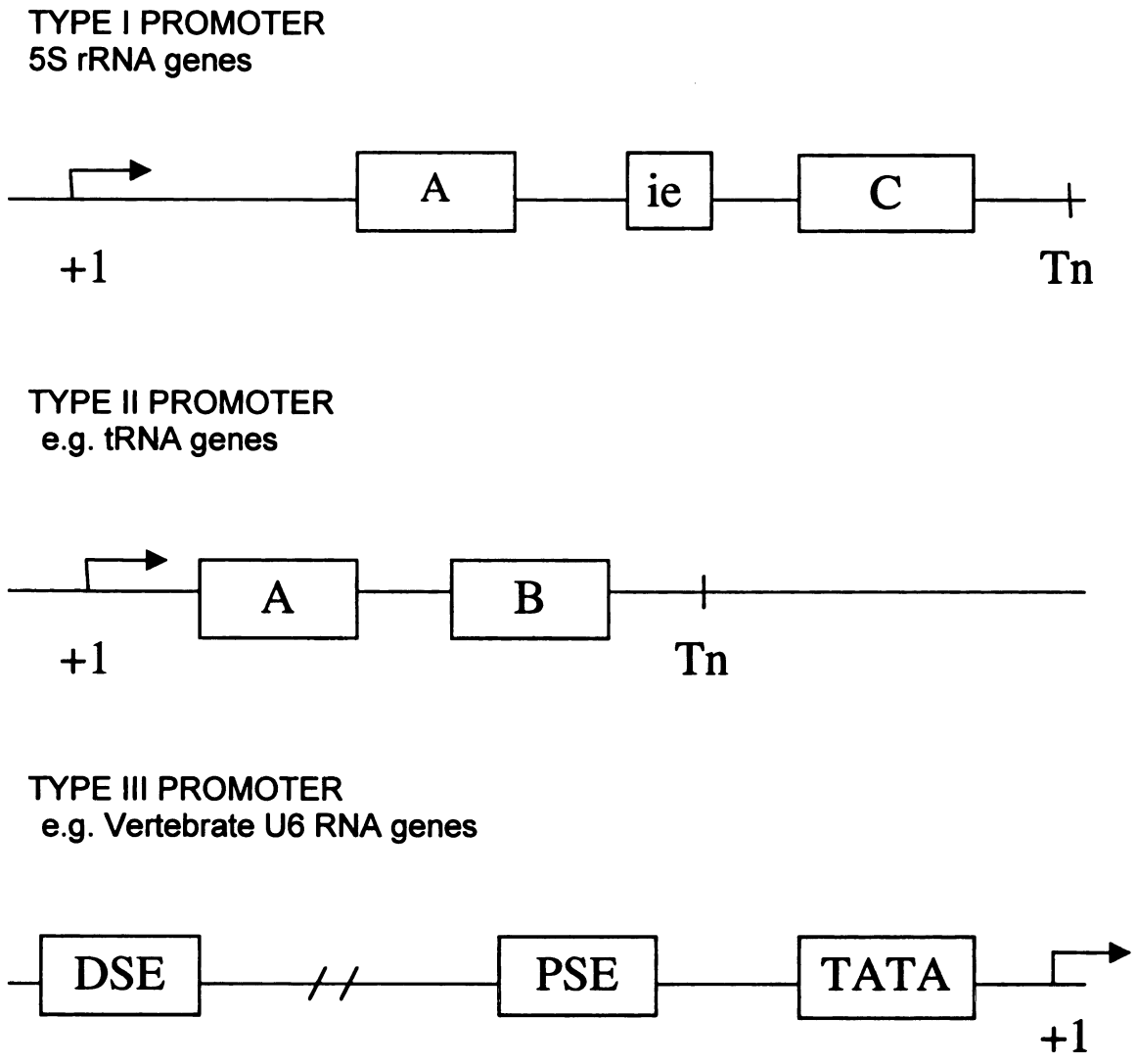


Figure 1.1 Diagram of the three types of RNA polymerase III genes. +1 indicates the start site of transcription; Tn indicates the termination site of transcription.

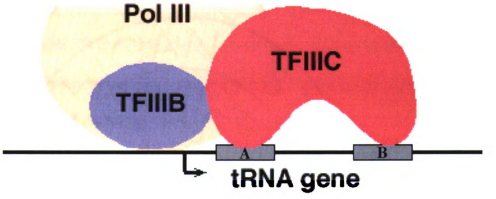


Figure 1.2 A model of TFIIIC-dependent TFIIIB binding to a tRNA promoter.

TBP

It is now well established that TATA-binding protein (TBP) is an essential component of the transcriptional machinery of all three RNA polymerases, and therefore is required for the expression of all nuclear genes (21,22). The N-terminal region of TBP is variable in both size and sequence, but the C-terminal domain is highly conserved. Mutations affecting basal transcription by RNA polymerases I, II and III all map to the C-terminal domain of TBP. This domain alone has been shown to support transcription by RNA polymerase III (23-26). While it has been supposed that the constitution of TFIIIB is similar in most other eukaryotes, it appears that TBP is replaced by TRF1 (TBP-related factor 1) in *Drosophila melanogaster* (27).

BRF

BRF is the TFIIB-related factor of the TFIIIB complex. Sequence alignment between *S. cerevisiae, K. lactis, C. albican, Arabidopsis* and human BRF shows its strong homology (Figure 1.3). Although the homology is most pronounced in the amino-terminal half, conserved regions also exist in the carboxy-terminal half that are unique to BRF. In *S. cerevisiae* TFIIIB, BRF is a protein of 596 amino acids, with a molecular mass of ~67 kD and an isoelectric point (pI) of ~6.9. The amino-terminal half of BRF is homologous to the Pol II transcription factor IIB (TFIIB) (Figure 1.4). The N-terminal 320 amino acids of BRF are 19% identical to TFIIB, with an overall similarity of 44%. Three regions are conserved between BRF and the various TFIIB proteins. One consists of the putative Zn finger near the N-terminus, and the other two imperfect direct repeats (Figure 1.5). The repeat regions of BRF are 25% and 28% identical to the corresponding regions of human TFIIB.

Figure 1.3 Sequence alignment of BRF from *S. cerevisiae, K. lactis, C. albican, Arabidopsis thaliana* and human. Identical residues are highlighted in yellow.

```
TFIIB  M M T R E S I D K R A G R R G P N L N I V L T C P E C K V - - Y P P K I V E R F  38
BRF    M P V - - - - - - - - - - - - - - - - - - - - C K N C H G T E F E R D L S N - -  18

TFIIB  S E G D V V C A L C G L V L S D K L V D T R S E W R T F S N D D H N G D D P S R  78
BRF    A N N D L V C K A C G V V S E D N P I V S E V - - - T F G E T S - - - - - - - -  47

TFIIB  V G E A S N P L L D G N N L S T R I G K G E T T D M R F T K E L N K A O G K N V  118
BRF    - - - A G A A V V O G S F - - - - I G A G O S H A A - F G - - - - - - - - G S S A  72

TFIIB  M D K K D N E V O A A F A K I T M L C D A A E L P K I V K D C A K E A Y K L C H  158
BRF    L E S R E A T L N N A R R K L R A V S Y A L H I P E Y I T D A A F O W Y K L A L  112

TFIIB  D E K T L K G K S M E S I M A A S I L I G C R R A E V A R T F K E I O S L I H V  198
BRF    A N N F V O G R R S O N V I A S C L Y V A C R K E K T H H M L I D F S S R L O V  152

TFIIB  K T K E F G K T - L N I M K N I L R G K S E D G F L K I D T D N M S G A O N L T  237
BRF    S V Y S I G A T F L K M V K - - - - - - - - - - - - K L H I T E L P L A D P S L  180

TFIIB  Y I P R F C S H L G L P M O - - - - V T T S A E Y T A K K C K E I K E I A G K S  273
BRF    F I O H F A E K L D L A D K K I K V V K D A V K L A O R M S K D - W M F E G R R  219

TFIIB  P I T I A V V S I Y L N I L L F O I P I T A A K V G O T L O V T E G T I K S G Y  313
BRF    P A G I A G A C I L L A C R M N N L R R T H T E I V A V S H V A E E T L O O R L  259

TFIIB  K I L Y E H R D - K L V D P O L I A N G V V S L D N L P - - - - - - - - - - - -  340
BRF    N E F K N T K A A K L S V O K F R E N D V E D G E A R P P S F V K N R K K E R K  299

TFIIB  - - - - - - - - - - - - - - - - G V E K K                                    345
BRF    I K D S L D K E E M F O T S E E A L N K N                                    320
```

Figure1.4 Sequence alignment between *S. cerevisiae* TFIIB and the N-terminal 320 amino acid residues of BRF, identical residues are highlighted in yellow.

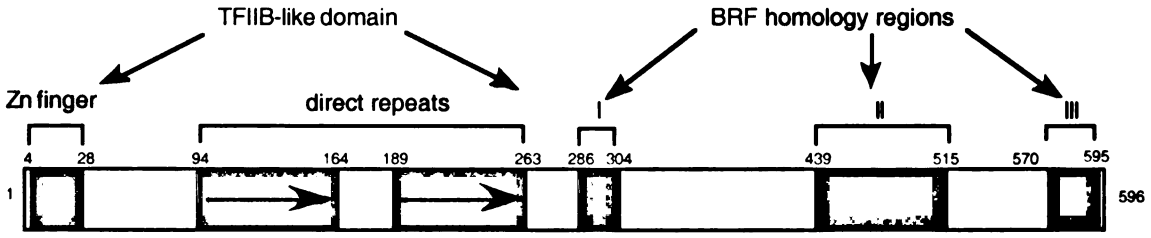


Figure 1.5 Schematic representation of *S. cerevisiae* BRF. Rose, TFIIB homology regions; blue, regions that are conserved among other species.

BRF is highly homologous to TFIIB and *Archae* TFB over the Zn binding domain and the core domain (two repeats). However, BRF also contains a ~30 kD domain at its C-terminus that is conserved only among other BRFs (Figure 1.3) and appears to play a major role in interaction with TBP/DNA. Both TFIIB and *Archea* TFB bind the TBP/DNA complex tightly (28-31). It was quite plausible to anticipate that TBP would interact with the homologous domain of BRF and TFIIB, but in fact BRF appears to act differently from these homologous factors. The region of BRF conserved with the TFIIB family members does not detectably bind TBP (32,33). On the other hand, the C-terminal half of BRF, which is not conserved with the TFIIB family members, strongly interacts with TBP and DNA, and it alone can form a complex with TBP, B″, and DNA (34,35). Deleting homology regions II and III in the BRF C-terminal region abrogates this interaction (34,35). TBP mutations that selectively inhibit RNA polymerase III transcription *in vivo* impair interactions between TBP and the BRF C-terminal domain (36-38).

An important unanswered question in the Pol III system is the role of the N-terminal TFIIB-homologous region of BRF. This region is essential for transcription and appears to act differently from its relatives in the *Archae* and RNA polymerase II systems. It has been shown that the N-terminal domain of BRF interacts with τ131, a subunit of TFIIIC in two-hybrid assays (32), and with C34, a subunit of RNA polymerase III in affinity chromatography (33). Previous studies have shown that BRF does not interact on the same face of the TBP/DNA complex as TFIIB. On the other hand, BRF may overlap in its position with TFIIA for binding to TBP, DNA, or both (37). The N-

terminal domain of BRF is also necessary for TFIIIC-dependent transcription *in vitro* (34,35).

Extensive deletion mutagenesis of the C-terminal domain of BRF has shown that a small domain encompassing 435-545 is sufficient for the formation of both the BRF/TBP/DNA complex and the BRF/B″/TBP/DNA complex. When this domain is reconstituted with the N-terminal domain of BRF, it recovers almost full wild-type BRF activity in TFIIIC-independent and TFIIIC-dependent transcription initiation complex formation, as well as TFIIIC-mediated TATA-less transcription and TFIIIC-independent TATA-dependent transcription *in vitro* (34,35).

B″

B″ is the third subunit of TFIIIB. Since TFIIB and the N-terminal half of BRF sit in corresponding locations in their respective DNA complexes (30,34), and exercise similar functions, it is surprising that BRF and TBP are not by themselves competent to direct transcriptional initiation by RNA polymerase III (37). In fact, B″ is absolutely required for transcription by RNA polymerase III *in vivo* as well as *in vitro* in duplex DNA or chromatin, at TATA-containing and TATA-less promoters. It is also B″ that makes the TFIIIB-DNA complex extraordinarily stable (20). In *S. cerevisiae*, B″ consists of 594 amino acids (Figure 1.6) and has almost the same molecular weight as BRF (~67kD), although it migrates anomalously in SDS-PAGE (as a ~90kD protein). The ready renaturability and anomalous electrophoretic mobility of B″ suggests that it may contain a highly stable core structure in the TFIIIB complex (39). *S. cerevisiae* B″ contains a SANT domain (Figure 1.7). The SANT domain is related to a Myb repeat and was originally identified in a number of proteins, including the SWI3, ADA2, N-Cor, and

yeast TFIIIB B″ proteins (40). In *S. cerevisiae* B″, C-terminal deletions of B″ that lack most of the SANT domain are inactive for *in vitro* transcription of a TATA-less tRNA gene, although they are still active for transcription of the TATA-containing U6 snRNA gene (41).

B″ can associate with a preassembled TBP/BRF/DNA but not with a complex lacking BRF. B″ contacts not only BRF but also TBP. At least one mutation in *S. cerevisiae* TBP prevents association of B″ without affecting association of BRF (37). B″ also makes interactions with DNA. It has been shown that B″ interacts upstream of the TATA box. B″ binding to the TBP/BRF/DNA complex extends the DNA footprint ~10 bp upstream of the TATA box, and is stabilized by an additional 15-20 bp DNA stretch upstream of the TATA box (37). Thus, the DNA segment that is additionally covered when B″ adds to the TBP/BRF/DNA complex harbors a site of efficient B″ cross-linking and contributes to the stability of the TFIIIB-DNA complex (42). The binding of B″ to the TBP/BRF/DNA complex induces a bend in the DNA between the TATA box and the transcription start site, which is in phase with the bend imposed by TBP on the TATA box (23,25,26). This bending of DNA has been postulated to contribute to the stabilization of the TFIIIB-DNA complex by helping impede sliding of the DNA out of the complex (23), a hypothesis consistent with thermodynamic and kinetic data indicating B″-dependent kinetic trapping of the DNA (43). It has been suggested that BRF may also undergo a conformational change upon B″ binding to the TBP/BRF/DNA complex.

```
  1 MSSIVNKSGTRFAPKVRQRRAATGGTPTPKPRTPQLFIPESKEIEEDNSD
 51 NDKGVDENETAIVEKPSLVGERSLEGFTLTGTNGHDNEIGDEGPIDASTQ
101 NPKADVIEDNVTLKPAPLQTHRDQKVPRSSRLASLSKDNESRPSFKPSFL
151 DSSSNSNGTARRLSTISNKLPKKIRLGSITENDMNLKTFKRHRVLGKPSS
201 AKKPAGAHRISIVSKISPPTAMTDSLDRNEFSSETSTSREADENENYVIS
251 KVKDIPKKVRDGESAKYFIDEENFTMAELCKPNFPIGQISENFEKSKMAK
301 KAKLEKRRHLRELRMRARQEFKPLHSLT[redacted]
351 [redacted]KAHTAIQLKLNPDGTMAIDEETMVVDRHKNASIENEYKEKVDEN
401 PFANLYNYGSYGRGSYTDPWTVEEMIKFYKAL[redacted]
451 [redacted]ALRSKLPPNFDEYCCEIKKNIGTVADFNE
501 KLIELQNEHKHHMKEIEEAKNTAKEEDQTAQRLNDANLNKKGSGGIMTND
551 LKVYRKTEVVLGTIDDLKRKKLKERNNDDNEDNEGSEEEPEIDQ
```

Figure 1.6 The amino acid sequence of *S. cerevisiae* B″. Blue, 329-357, glutamate rich domain; rose, 415-472, SANT domain.

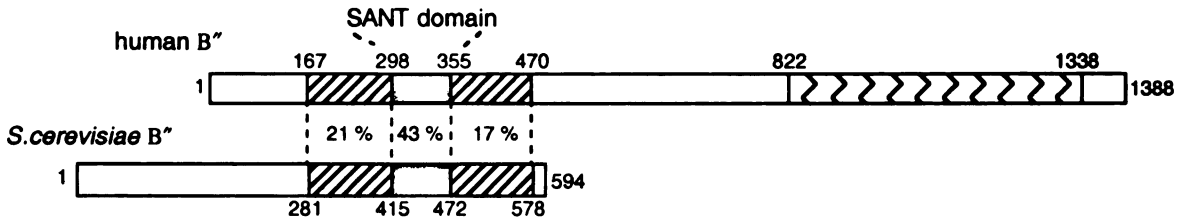


Figure 1.7 Regions of similarity in the human and *S. cerevisiae* B″ sequences. The rose boxes correspond to the SANT domain; the hatched boxes indicate regions of lower but still significant similarity on either side of the SANT domain. The percentages indicate identities between the two proteins in the regions delimited by the dotted lines. The small arrows indicate the repeats in human B″.

A central ~225 amino acid segment of B″ appears to encompass its functional core. Two domains, one at each end of this core region (in *S. cerevisiae* B″, amino acids 272-292 and 424-449), are required for TFIIIC-dependent transcription *in vitro*, and one or the other of these segments is required for TFIIIC-independent transcription (15). An additional and partly overlapping ~65 amino acid segment (amino acids 355-421) is required for transcription of linear DNA (35). The principal target for DNA cross-linking of B″ lies between amino acids 277 and 315, *i.e.*, it encompasses the N-proximal "essential" amino acids 272-292 segment. This segment of B″ faces the upstream part of the DNA site occupied by TFIIIB. The N-terminal 223 amino acids of B″ are not essential for transcription of bare DNA (41). This segment of B″ diminishes DNA-protein cross-linking upstream of the TATA box at least 2-fold, as though it formed an obstruction of the corresponding DNA-interacting segments of B″. Deleting B″ amino acids 1-185 has the same effect on cross-linking. A segment of B″ extending approximately from amino acid 190 to amino acid 210 becomes more accessible to cleavage by hydroxyl radical cleavage upon entry into the TFIIIB-DNA complex (41). This is consistent with the notion that DNA access of B″ might be blocked by its internal folding, and uncovered upon formation of the TFIIIB-DNA complex.

1.2 1L-*myo*-inositol 1-phosphate synthase

1.2.1 Cell communication and the importance of inositol

Cells of multi-cellular organisms communicate with each other using substances such as ions, hormones and neurotransmitters. Lipophilic agents such as steroid hormones pass through the lipid bilayer of the cell membrane and bind to specific intracellular receptors triggering the appropriate cellular responses. Hydrophilic substances, which cannot cross the cellular lipid bilayer, deliver their message by binding to specific receptors located on the cell surface. This type of signal transduction, also called transmembrane signaling, depends on surface receptors. Several classes of receptors are involved in signal transduction. The first class of cell surface receptor is linked to an ion channel; stimulation of the receptor can trigger the ion channel to pump ions into or out of the cell. A change in the given ionic species inside the cell is perceived as an internal signal, evoking an ultimate response to the external signal. The second class of cell surface receptors are related to tyrosine kinases (44). Binding of the extracellular receptor by an agonist activates an intracellular tyrosine kinase and leads to the phosphorylation of tyrosine residues on the target proteins inside the cell and causes them to respond. The third class of receptor is coupled via a class of guanine nucleotide binding proteins (G-protein) to the intracellular enzymes or ion channels, through which the receptors evoke their responses. The family of G-proteins includes several members that regulate different intracellular pathways. Upon agonist binding to the receptor, the activated specific G-protein stimulates or inhibits other membrane bound enzymes that act as amplifiers. This, in turn, generates second messengers inside the cell. Only a few second messengers have been identified so far. They include adenosine 3′, 5′ cyclic

monophosphate (cAMP), guanosine 3′, 5′ cyclic monophosphate (cGMP), diacylglycerol (DAG), Ca^{2+}, and *myo*-inositol 1,4,5-triphosphate (1,4,5-IP_3). Most important signal transduction pathways include a cAMP cascade and phosphoinositide cascade.

cAMP cascade

The binding of hormones such as adrenaline, calcitonin, glucagons, thyroid-stimulating hormone, etc. to a seven-helix receptor in the plasma membrane triggers the exchange of GTP for GDP bound to the stimulated G protein (Gs). The α subunit of the G protein (Gsα-GTP) then dissociates from the βγ subunits (Gβγ). Gsα-GTP activates adenylate cyclase, an integral membrane protein. Adenylate cyclase (AC) catalyzes the cyclization of ATP to produce cAMP. cAMP then activates protein kinase A (PKA) by binding its regulatory subunit, thus unleashing its catalytic subunits. PKA alters the activity of many target proteins by phosphorylating serine and threonine residues (Figure1.8).

Phosphoinositide (PI) cascade

Phosphatidyl inositol (PtdIns) is the most abundant inositol lipid in nature, making up 5 % of the total membrane phospholipid content in eukaryotes. Subsequent phosphorylation on the inositol moiety of PtdIns produce a variety of phosphorylated inositides including phosphatidylinositol-4, 5-bisphosphate (PtdIns (4,5) P_2). Agonist binding to a heterotrimeric G-protein-coupled receptor causes the exchange of GDP for GTP and subsequent activation of phospholipase C (PLC). Hydrolysis of PtdIns (4,5) P_2 by activated PLC produces *myo*-inositol 1,4,5-triphosphate (1,4,5-IP_3), which binds to its

receptor buried in the membrane of the endoplasmic reticulum (ER) and stimulates the release of Ca^{2+}. Ca^{2+} acts by binding to calmodulin and other calcium sensors. Ca^{2+}-bound calmodulin activates target proteins by binding to positively charged amphipathic helices, causing a diverse set of cellular responses including secretion, excitation, contraction, growth and proliferation (45-47).

Cell communication in brain

In brain, communication between neurons makes use of the signal transduction system involving 1,4,5-IP_3 and the phosphoinositide cascade and therefore requires a suitable amount of *myo*-inositol. In most mammalian cells, *myo*-inositol is obtained from diet, however in brain cells, the major portion of *myo*-inositol is biosynthesized from D-glucose 6-phosphate since plasma *myo*-inositol does not effectively cross the blood-brain barrier (48,49).

The *de novo* biosynthesis of *myo*-inositol begins with the conversion of D-glucose-6-phosphate to 1L-*myo*-inositol-1-phosphate by MIP synthase, followed by dephosphorylation of this product by *myo*-inositol monophosphatase to produce *myo*-inositol. *Myo*-inositol is then combined with cytodine diphosphate-diacylglycerol (CDP-DAG) in a reaction catalyzed by phosphatidyl inositol (PtdIns) synthase to form phosphatidyl inositol.

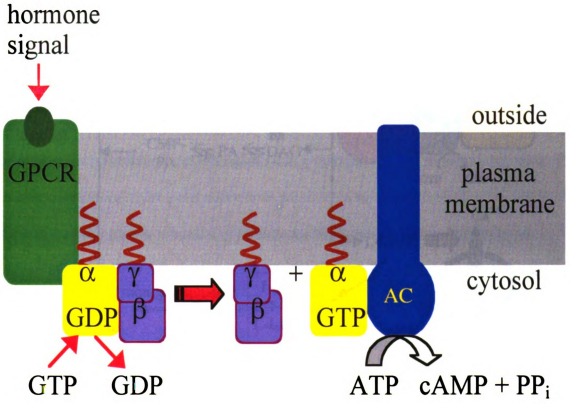


Figure1.8 cAMP cascade.

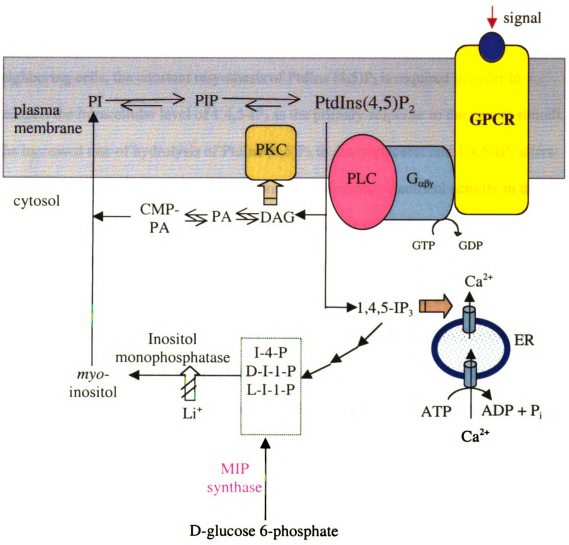


Figure1.9 Phosphoinositide cascade.

Cellular signaling between hyperactive neurons and neighboring cells causes an increase in the turnover of the Phosphoinositide cascade. In the over stimulated neighboring cells, the constant resynthesis of PtdIns $(4,5)P_2$ is required in order to maintain the intracellular level of 1,4,5-IP_3 as the primary response to the excess stimuli. The increased rate of hydrolysis of PtdIns $(4,5)P_2$ to diacylglycerol and 1,4,5-IP_3 alters the electrical activity of the affected neurons. The abnormal electrical activity in the brain tissue surrounding the pathological cells is thought to result in manic disorders. *In vivo*, the inhibition of the phosphoinositide cascade phosphatases by lithium results in the depletion of cellular *myo*-inositol. The decrease in the *myo*-inositol availability to regenerate the PtdIns (4,5) P_2 pool results in signal termination and provides control over manic disorders. Effects of inositol phosphatase inhibition are only observed in brain cells since no dietary extra cellular *myo*-inositol can be pumped into the phosphoinositide cascade (Figure 1.9) (45,47).

Lithium carbonate is used with variable success to treat manic disorder because of its ability to inhibit various enzymes of the phosphoinositide cascade, particularly *myo*-inositol monophosphatase (50,51). However, the established success of lithium as the therapeutic treatment for manic disorder is not complete. Suppressing the hydrolysis of inositol monophosphatases leads to the depletion of *myo*-inositol and the scarcity of PtdIns (4,5) P_2, the precursor of 1,4,5-IP_3. Consequently the hyperactive neuron would be dampened since it loses a critical pool of PtdIns $(4,5)P_2$ for signal transduction. In addition, a large amount of lithium carbonate (>2000 mg/day) is required to maintain an effective drug concentration at the cellular level (0.5-1.2 mM), which is relatively close to the toxic level (3-5 mM). In addition to *myo*-inositol monophosphatase, lithium also

inhibits other enzymes and causes accumulation of various metabolites. Along with the side effects of lithium, insensitivity to this drug and a decrease in effectiveness due to its required continual uptake have also been reported (52).

Inhibition of MIP synthase in brain tissues represents an alternative approach for treating manic depression. This inhibition is expected to reduce *myo*-inositol as well as MIP levels and may present different neurological outcomes than that observed with lithium treatment and *myo*-inositol depletion. The accumulated D-glucose 6-phosphate upon inhibition of MIP synthase could be used as a substrate for other intracellular enzymes.

Recent *in vivo* experiments in yeast and *Dictyostelium* showed that valproate, a drug used in the treatment of depression, bipolar disorder, and seizure disorder, may act by inhibition of MIP synthase, thus lowering neuronal inositol pools similar to the action of lithium (53,54).

1.2.2 1L-*myo*-inositol 1-phosphate synthase

1L- *myo*-inositol 1-phosphate (MIP) synthase (EC 5.5.1.4) is found in all eukaryotes investigated thus far including protozoa, fungi, algae, plants and animals. The amino acid sequence of MIP synthase has been remarkably well conserved throughout evolution over virtually its entire sequence (Figure 1.10).

MIP synthase catalyzes the isomerization of D-glucose 6-phosphate into 1L-*myo*-inositol 1-phosphate (Figure 1.11), β-nicotinamide adenine dinucleotide (NAD^+) is required to catalyze this reaction (55). MIP synthase employs NAD^+ as a prosthetic group, which is neither consumed nor generated during the catalytic cycle.

Loewus and Kelly were the first to propose that a cyclization mechanism involving the generation of a C5 ketose prior to the formation of the C-C bond was the most likely pathway (56,57). The reaction is proposed to begin with the binding of D-glucose 6-phosphate to the enzyme-NAD^+ complex. The resulting open form of D-glucose 6-phosphate is oxidized at the C5 position with the simultaneous reduction of NAD^+ to NADH and forms the intermediate B, 5-keto-D-glucose 6-phosphate, in which the acidity of the α protons at C6 is greatly increased. The subsequent enolization of intermediate B to intermediate C is followed by the aldol condensation reaction between C6 and C1. Intermediate D, *myo*-2-inosiose 1-phosphate, is then reduced by the enzyme-bound NADH forming NAD^+ and MIP, which is subsequently released from the enzyme active site.

Figure 1.10 Sequence alignment of MIP synthase from *Saccharomyces cerevisiae*, *Arabidopsis thaliana*, human, *Mycobacterium tuberculosis*, and *Archaeoglobus fulgidus*. Identical residues are highlighted in yellow.

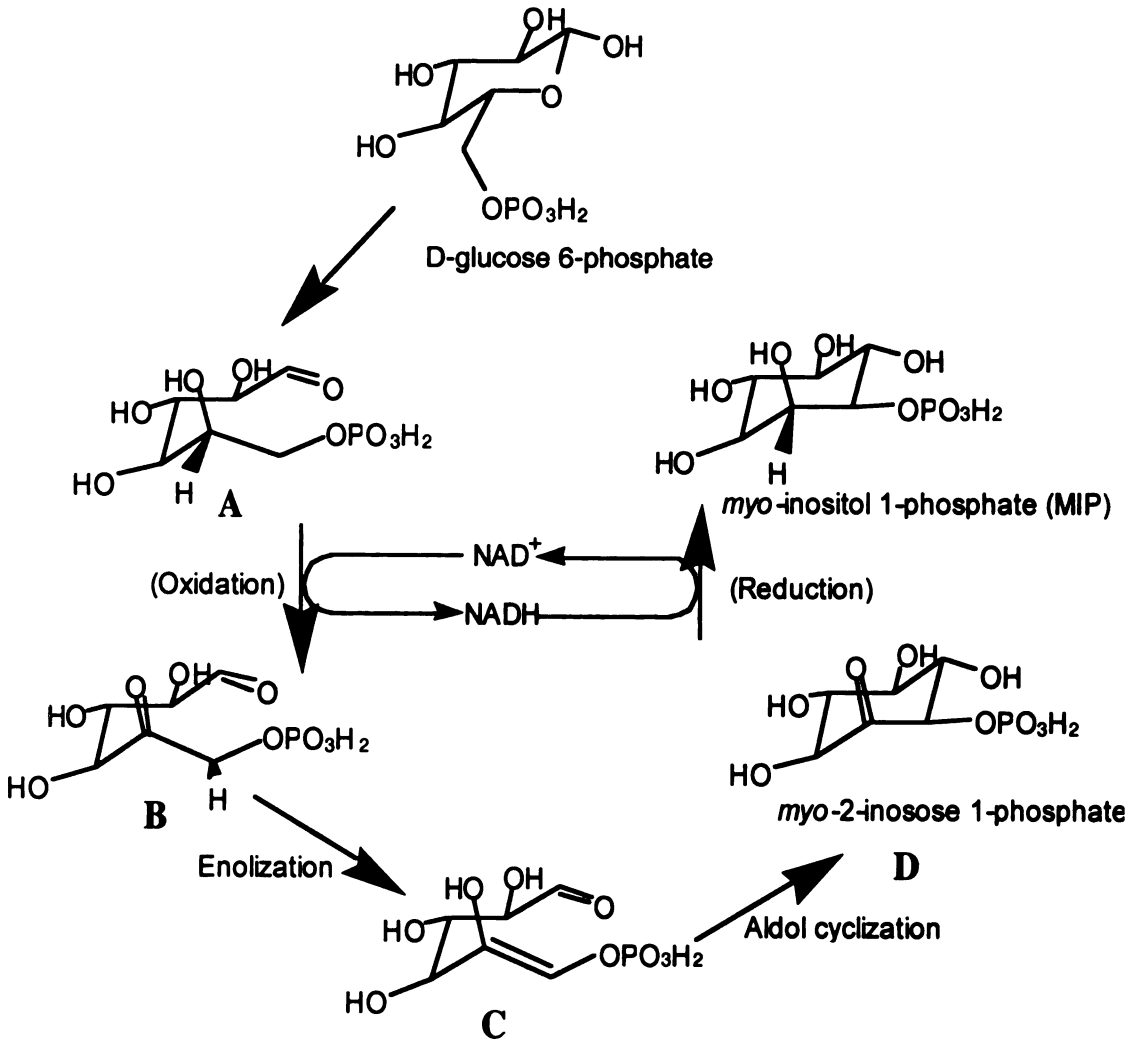


Figure 1.11 Proposed mechanism of MIP synthase.

Various experiments support the mechanism of Figure1.11: (a) MIP synthase requires NAD^+ for its activity and NADH is tightly associated with the enzyme. Incubation of independently synthesized intermediate B with *apo* MIP synthase reconstituted with [4-^{3}H] NADH resulted in the formation of D- [5-^{3}H] glucose 6-phosphate and [^{3}H]-*myo*-inositol 1-phosphate (58,59). *In situ* formation of NADH during substrate turnover has been observed by UV-vis spectroscopy (60). (b) A trapping experiment using tritiated sodium borohydride ([^{3}H]-$NaBH_4$) indicated the transient existence of intermediate D (61). (c) Isotope effects were observed at both the C5 and C6 positions of D-glucose 6-phosphate. This observation is consistent with the presence of intermediate B, which indicates that the enolization is rate limiting (62,63). (d) The *pro-R* hydrogen of the C6 methylene group is preferentially removed during enolization (64). Sherman, Eisenberg, and Barnett found that the reduction of the MIP synthase reaction mixture with [^{3}H]-$NaBH_4$ did not furnish any lysyl derivative of glucose, regardless of the source of MIP synthase used (59,65,66). Another experiment using $H_2{}^{18}O$ showed that ^{18}O incorporation into the product was not the result of the hydrolysis of a Schiff base but came from the nonenzymatic exchange of the glucose aldehyde carbonyl with the medium. These various observations established that MIP synthase could not be classified as a type I aldolase, which involves the formation of a Schiff base (Figure 1.12). On the other hand, MIP synthase was not inhibited by EDTA, therefore MIP synthase does not utilize a type II aldolase mechanism, in which a divalent metal ion such as Zn^{2+}, Mn^{2+} acts as the Lewis acid (Figure 1.13) (67-69). It has been reported recently that MIP synthase from *Archaeoglobus fulgidus* requires divalent cations and therefore is thought to be a type II aldolase (70).

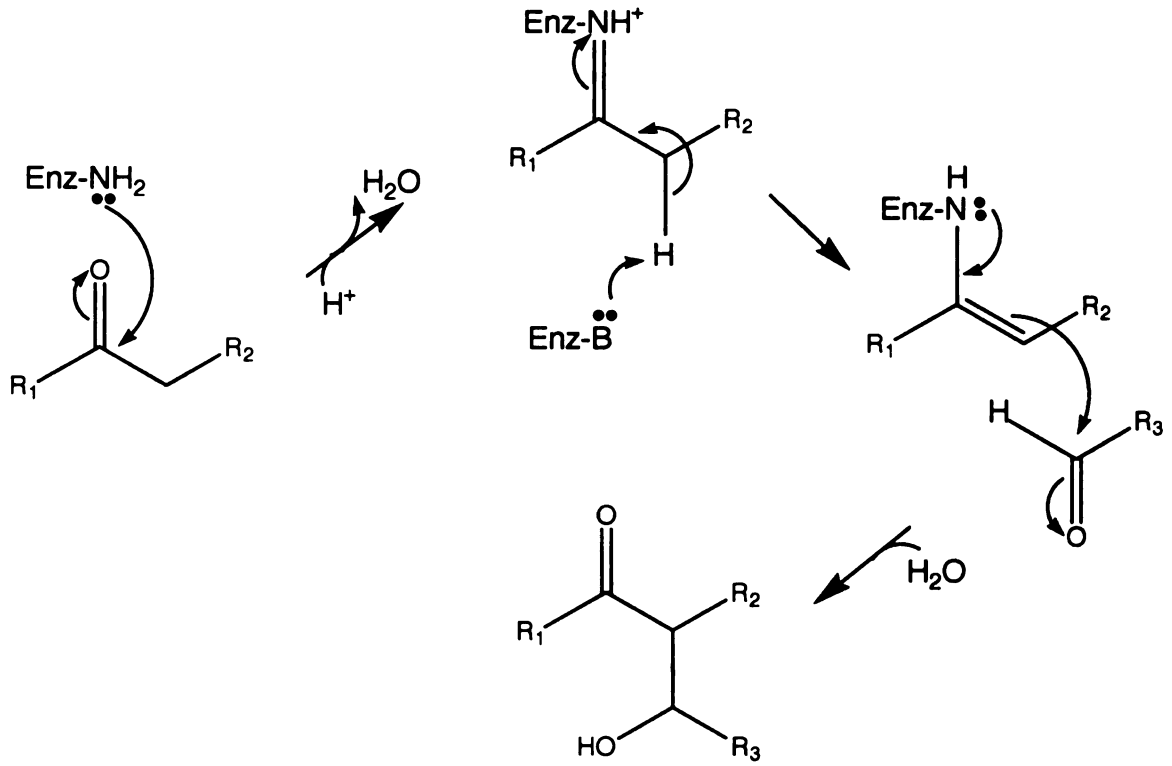


Figure 1.12 The mechanism of a type I aldolase.

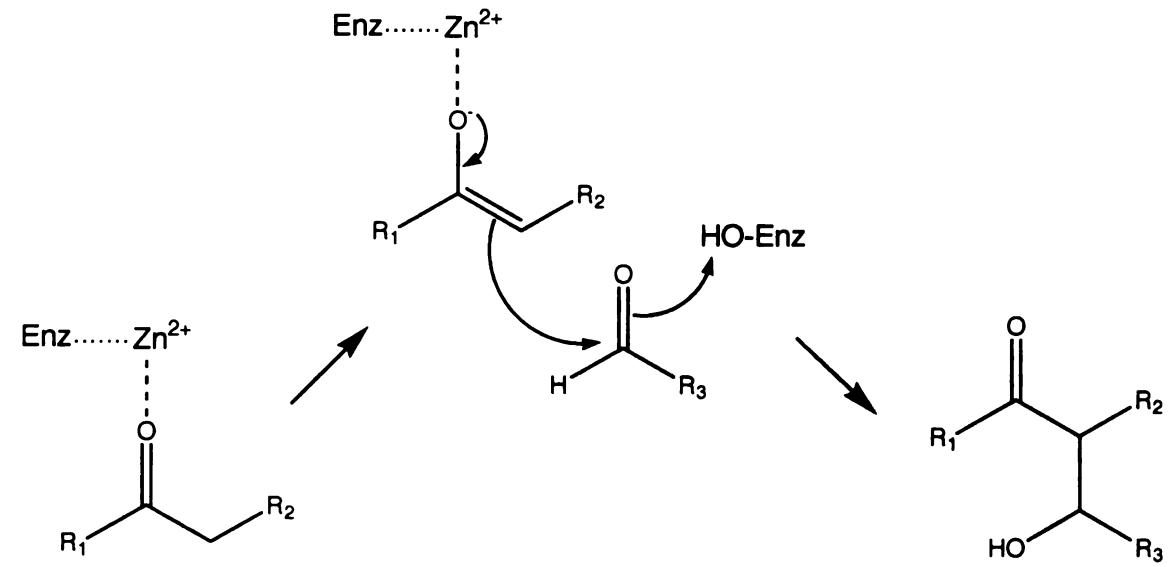


Figure 1.13 The mechanism of a type II aldolase.

MIP synthase might be responsible for the type of base-catalyzed cyclization where the appropriate activation of the substrate is carried out by the enzyme oxido-reductase function. Loewus showed that only the C6 *pro-R* hydrogen was lost during turnover and that the *cis*-enol C resulted directly from this selective removal (56). Sherman demonstrated that intermediate B was reduced by a base to produce two major products identified as *epi*-inositol 3-phosphate and D-*myo*-inositol 3-phosphate (71). The difference in stereo selectivity observed between a base-catalyzed cyclization and MIP synthase catalyzed cyclization indicated the strong involvement of the enzyme active site during aldol condensation. Enzymatic base-catalyzed enolization must be stereo controlled and the orientation of the newly formed *myo*-inositol 1-phosphate must be under strict enzymatic control.

Although the mechanism proposed above accounts for the available data, a number of mechanistic questions still remain, especially the nature of the aldol condensation.

1.2.3 Previous structural investigations on MIP synthase

Crystal structures of MIP synthase from *S. cerevisiae* with partially occupied NAD^+, and in complex with NAD^+ and an inhibitor, 2-deoxy-D-glucitol 6-phosphate have recently been determined (72). The results showed an example of induced fit, where nearly 60 residues, residues 351-409, in the active site that were disordered in the structure with a partially occupied NAD^+ become ordered upon the binding of the inhibitor (Figure 1.14). It appeared that the complete folding of the enzyme active site required either complete NAD^+ occupancy or inhibitor/substrate binding, or both. The

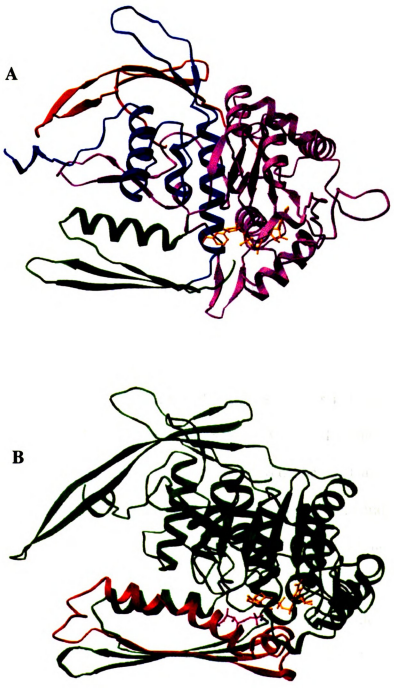


Figure 1.14 (A) Ribbon model of the MIP synthase monomer. Red, the N-terminal region; purple, the NAD^+-binding region; green, the tetramerization region; blue, the C-terminal domain. (B) Ribbon model of the MIP synthase/NAD^+/2-deoxy-D-glucitol 6-phosphate complex. Green, the residues that were ordered in the structure with low occupancy NAD^+; red, the newly ordered residues; yellow, NAD^+; magenta, 2-deoxy-D-glucitol 6-phosphate.

inhibitor, 2-deoxy-D-glucitol 6-phosphate was bound in its extended conformation inconsistent with the substrate during the reaction. Based on the structural data, the substrate modeling in the conformation necessary for cyclization was performed and a mechanism was proposed (Figure 1.15). It was proposed that the first step involves an oxidation at C5. Subsequently the substrate is reoriented to a conformation where the phosphate can act as a base at the enolization step. The developing negative charge is stabilized by a nearby NH_4^+. Nucleophilic attack by C6 on C1 is promoted by K369 stabilization of the developing negative charge on O1. Subsequent protonation at O1 by K369 and reduction of C5 by NADH yields the product MIP.

The crystal structure of MIP synthase from *Mycobacterium tuberculosis* with NAD^+ has also been published recently (73). In this structure, residues 241-267, which are conserved as 362-391 in *S. cerevisiae* MIP synthase, are missing (Figure 1.16). The structure has an NAD^+ moiety and a zinc ion in the active site. The nicotinamide moiety of the NAD^+ is in the *syn* conformation about the N-glycosidic bond and is held there by coordination with the Zn^{2+}. The Zn^{2+} ion was identified by its tetrahedral coordination and from bond lengths of close to 2.2 Å for each of its four ligands. It lies adjacent to the NAD^+ nicotinamide moiety, bridging the nicotinamide oxygen atom and phosphodiester oxygen NO2 and coordinating the S311 OG atom and a water molecule. The presence of Zn^{2+} may influence the pK_a of the acidic side chains of D235 and D310, which coordinate this water molecule. The Zn^{2+} ion is more buried than that present in similar enzymes that employ Zn^{2+} directly in the mechanism, e.g., horse liver alcohol dehydrogenase (PDB code 3bto). A detailed analysis of the catalytic mechanism of MIP synthase was

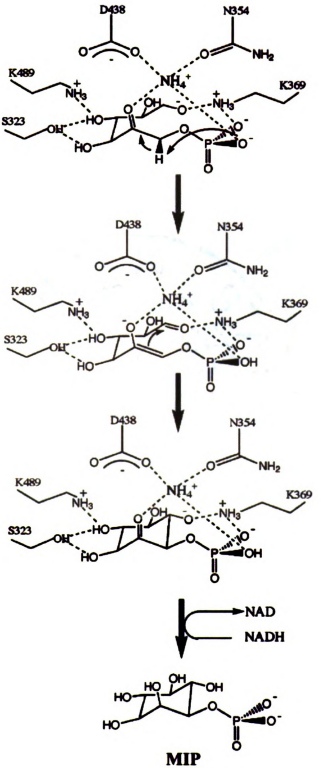


Figure 1.15 Proposed mechanism for the transformation catalyzed by MIP synthase (72).

not performed owing to the lack of a structure complexed with the substrate or suitable inhibitors.

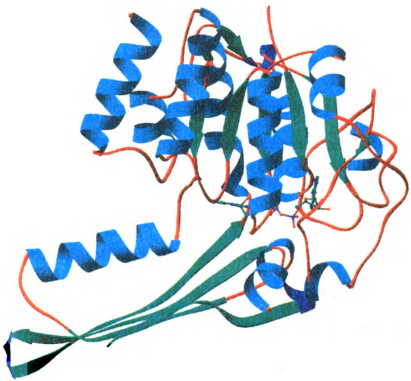


Figure 1.16. The structure of MIP synthase from *Mycobacterium tuberculosis* (73).

1.2.4 Remaining challenges of MIP synthase

1) The effect of NAD^+

NAD^+ is one of the most commonly used cofactors in living cells. As first noted by Rossmann *et al.* (74), most NAD^+ binding proteins are similar in tertiary structure in the region where NAD^+ binds. This core topology region consists of a βαβαβ unit with at least one additional parallel β strand. Within the βαβαβ unit is a sequence of 30-35 amino acids forming the finger print region, which can be used to identify the location of NAD^+ binding. This finger print region includes a glycine-rich phosphate binding consensus sequence (GXGXXG), where the first strictly conserved glycine allows for a tight turn of the main chain, which is important for positioning the second glycine. The second glycine, because of its missing side chain, allows for close contact of the main chain to the pyrophosphate of NAD^+. It is thought that any side chain in this position would protrude into the binding site of NAD^+ and disrupt binding. The third glycine is important for the close packing of the secondary structure elements of the first β strand and the first α helix. Instead of the common GXGXXG motif, MIP synthase from eukaryotes has a GXGGXXG motif. Interestingly, it appears that MIP synthase from *Archaeglobus fulgidus* does not have a similar GXG(G)XXG motif. In addition, several other NAD^+-interacting residues are not conserved in *A. fulgidus* enzyme (Figure 1.10). Therefore, it can be presumed that the NAD^+ binding of *A. fulgidus* MIP synthase could be quite different from that of MIP synthase from eukaryotes and *Mycobateria.*

Enzymes that bind NAD^+ catalyze reactions central to energy production, storage, and transfer by exploiting the ability of the nicotinamide group to transfer hydride ions or electrons. In most NAD^+-dependent catalysis, NAD^+ is a cosubstrate. However, there is

also a family of NAD^+-requiring enzymes where NAD^+ plays the role of a catalytic prosthetic group (75). These enzymes catalyze reactions such as epimerization, aldol condensation, cyclization, α,β-elimination, and decarboxylation, in which strong bonds with no obvious lability are cleaved. A transient oxidation of the appropriate substrate carbon along with the simultaneous reduction of NAD^+ labilizes the scissile bond by introducing a carbonyl or an imine. NADH then reduces the oxidized center at a later step subsequent to the enzymatic transformation of the oxidized activated substrate. MIP synthase uses NAD^+ as a prosthetic group, which is neither consumed nor generated at the end of each catalytic cycle as shown in Figure 1.11. The three-dimensional structure of the MIP synthase/NAD^+ complex must represent the enzyme at its state when the substrate is not yet bound or the product is already produced and released from the enzyme active site. This structure should also provide insights into any conformational changes that might occur in both the Rossmann fold domain and the active site of MIP synthase due to the binding of NAD^+. One big question to be answered is whether the active site disorder observed previously in the structure of MIP synthase with partially occupied NAD^+ (72) could be due to not fully occupied NAD^+.

2) The structure of the NADH-bound MIP synthase

Radio labeling experiments have shown that the hydride transferred from the C5 position of the substrate to NAD^+ during the oxidation step was returned to *myo*-2-inosose 1-phosphate by NADH at the reduction step (76). The nicotinamide ring of NAD^+ must be located adjacent to the pyranose ring since there is a direct hydride transfer, and the same face of the nicotinamide ring must participate in the oxidation and

reduction steps. After the substrate is oxidized with the formation of NADH from NAD^+, the substrate or nicotinamide ring may move in order to reposition the carbonyl oxygen at C5, and consequently the side chains of active site residues may need to experience conformational changes as well. Furthermore, all non-redox steps of the reaction catalyzed by MIP synthase occur when the enzyme is coupled with NADH after the oxidation of the substrate, implicating that the three-dimensional structure of the MIP synthase/NADH complex may provide more information on the enzyme in its active state.

3) The structure of MIP synthase in complex with structural analogues of the substrate and reaction intermediates

Even with the previously published structural data on the MIP synthase/NAD^+/2-deoxy-D-glucitol 6-phosphate complex, some questions still remain unanswered. More structures complexed with various structural analogues of the substrate and reaction intermediates should produce a more complete picture of the MIP synthase mechanism. Previous inhibition studies have provided clues on the substrate binding to MIP synthase. First, in order to investigate whether the cyclic or acyclic form of the substrate binds MIP synthase, a series of cyclic and acyclic substrate analogues were synthesized. None of the molecules that are covalently locked in the cyclic form or that undergo ring opening slowly led to any detectable inhibition of MIP synthase (Figure 1.17) (77).
Several cyclic analogues of the substrate D-glucose 6-phosphate that can undergo ring opening such as 2-deoxy-D-glucose 6-phosphate and 2-deoxy-D-glucose 6-homophosphonate are inhibitors of MIP synthase (Figure 1.18) (78). One of the

intermediates in the proposed reaction pathway, *myo*-2-inosose 1-phosphate has been synthesized and found to be a potent inhibitor (K_i= 3.6 x 10^{-6} M), and several analogues of this intermediate were also found to be inhibitors of MIP synthase (Figure 1.19) (79). On the other hand, several acyclic analogues of the substrate were found to be very potent inhibitors of MIP synthase (Figure 1.20) (60). Structures of MIP synthase complexed with these acyclic and cyclic inhibitors will present the conformation of MIP synthase at its different steps of the catalytic cycle.

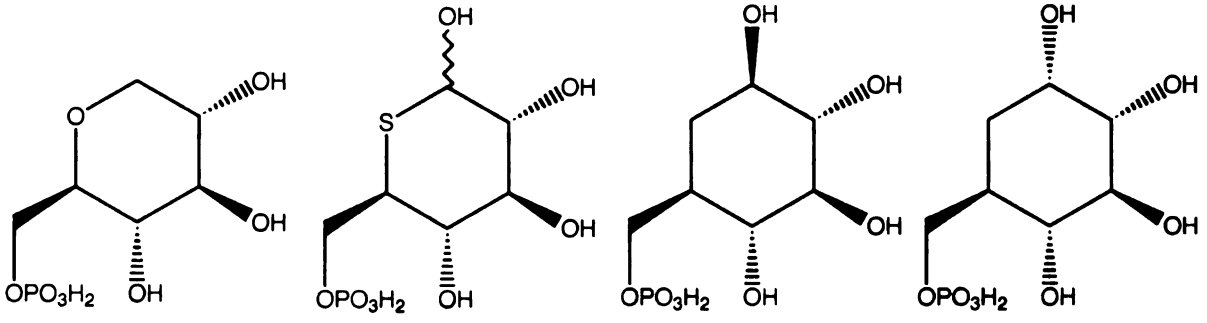
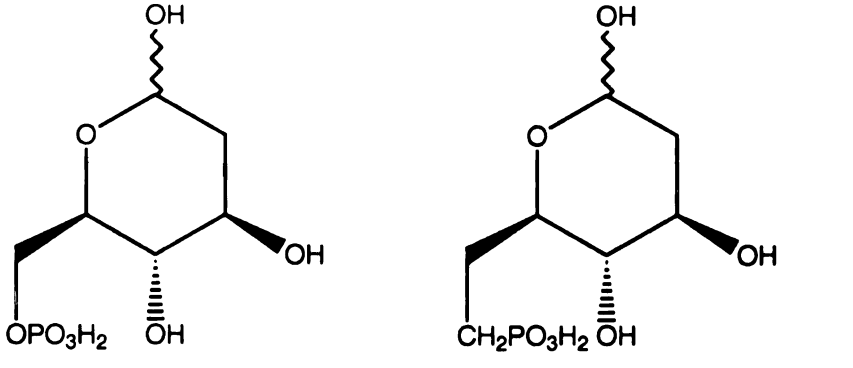


Figure 1.17 Cyclic analogues of D-glucose 6-phosphate with no inhibition of MIP synthase (77).



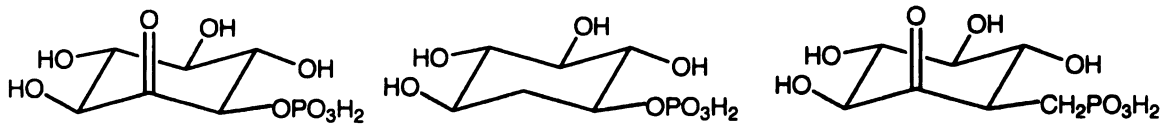
2-deoxy-D-glucose 6-phosphate

$K_i = 9.1 \times 10^{-6}$ M

2-deoxy-D-glucose 6-homophosphonate

$K_i = 7.1 \times 10^{-5}$ M

Figure 1.18 Cyclic analogues of D-glucose 6-phosphate that can undergo ring opening with inhibition of MIP synthase (78).



myo-2-inosose 1-phosphate

$K_i = 3.6 \times 10^{-6}$ M

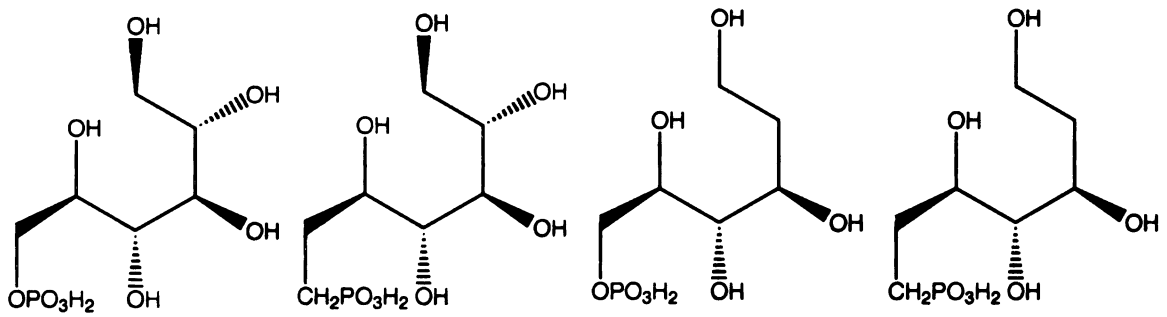
2-deoxy-*myo*-inositol 1-phosphate

$K_i = 170 \times 10^{-6}$ M

1-deoxy-1-(phosphonomethyl)-*myo*-2-inosose

$K_i = 37 \times 10^{-6}$ M

Figure 1.19 The intermediate D, *myo*-2-inosose 1-phosphate, and its analogues are inhibitors of MIP synthase (79).



D-glucitol 6-phosphate

$K_i = 1.5 \times 10^{-4}$ M

D-glucitol 6-homophosphonate

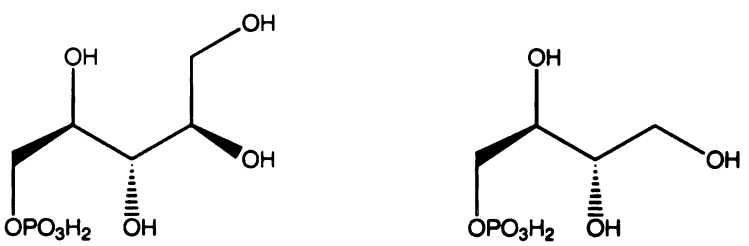
$K_i = 1.1 \times 10^{-4}$ M

2-deoxy-D-glucitol 6-phosphate

$K_i = 2.3 \times 10^{-6}$ M

2-deoxy-D-glucitol 6-homophosphonate

$K_i = 5.8 \times 10^{-6}$ M



D-arabinitol 5-phosphate

$K_i = 1.7 \times 10^{-5}$ M

D-erythritol 4-phosphate

$K_i = 4.7 \times 10^{-5}$ M

Figure 1.20 Acyclic analogues of D-glucose 6-phosphate are inhibitors of MIP synthase (60).

Second, in order to investigate the mechanism proposed by Floss (80) that a single, active site base and its conjugate acid mediate all of the non-redox steps (Figure 1.21), conformationally restricted *(Z)*- and (*E*)- vinylhomophosphonate analogues of the substrate were synthesized (77). While none of the (*Z*)-vinylhomophosphonates were inhibitors of MIP synthase, all (*E*)-vinylhomophosphonates were competitive inhibitors. These results are consistent with the reaction pathway where transoid conformations of the substrate D-glucose 6-phosphate and the intermediate 5-keto-D-glucose 6-phosphate may be found. A transoid conformation would position the dibasic phosphate monoester of 5-keto-D-glucose 6-phosphate for the removal of the *pro-R* hydrogen of C6 (Figure 1.22). Stereo selective proton removal would result in the monobasic phosphate monoester, which could then deliver the abstracted proton back to the C1 carbonyl oxygen to complete its catalytic role in the formation of the bond between C1 and C6. Intramolecular removal of a proton by a phosphate monoester is precedented for 3-dehydroquinate synthase, an enzyme mechanistically related to MIP synthase in its catalytic use of NAD^+ and aldol cyclization (81). Though these inhibition studies are consistent with the catalytic involvement of the phosphate monoester, they do not prove it. The structures of MIP synthase complexed with (*E*)-vinylhomophosphonate analogues of the substrate will answer this question.

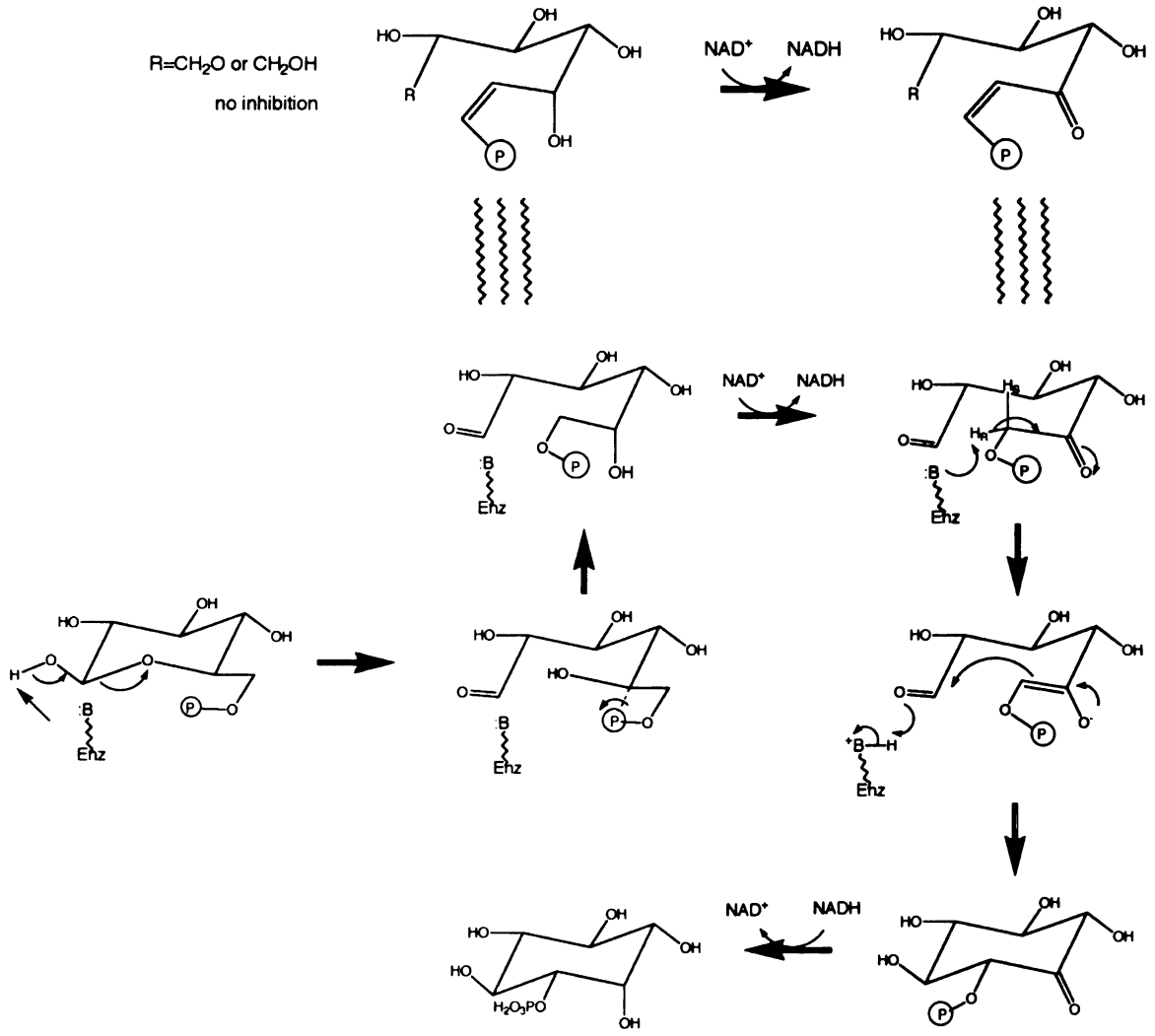
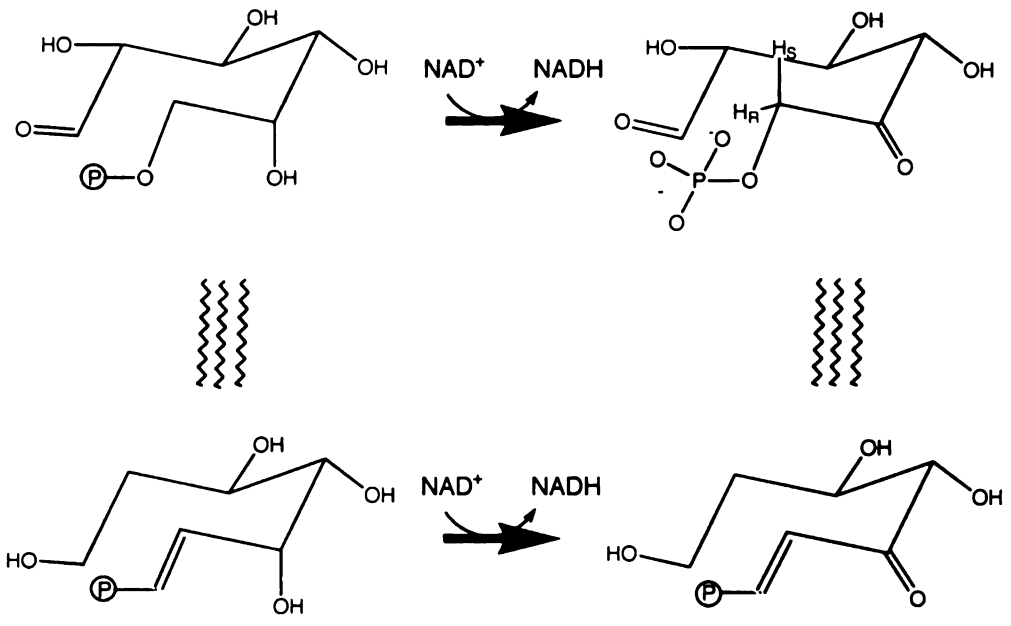


Figure 1.21 The mechanism of MIP synthase proposed by Floss (80).



2-deoxy-D-glucitol 6-(*E*)-vinylhomophosphonate

$K_i=0.67\times10^{-6}$ M

Figure 1.22 The mechanism consistent with the substrate binding in a transoid conformation and the phosphate monoester acting as the base in the enolization step (77).

Third, non-charged species such as D-glucose or 2-deoxy-D-glucose are not inhibitors of MIP synthase (82), indicating the importance of the phosphate in the substrate. In order to establish whether the phosphate moiety of the substrate is interacting with one or two binding pockets during each turnover, a diphosphate analogue of the substrate was synthesized and found to be an inhibitor with K_i = 0.06 mM (Figure 1.23) (78). The structure of MIP synthase complexed with this inhibitor will put to rest the question of whether the phosphate would move during the enzyme catalysis.

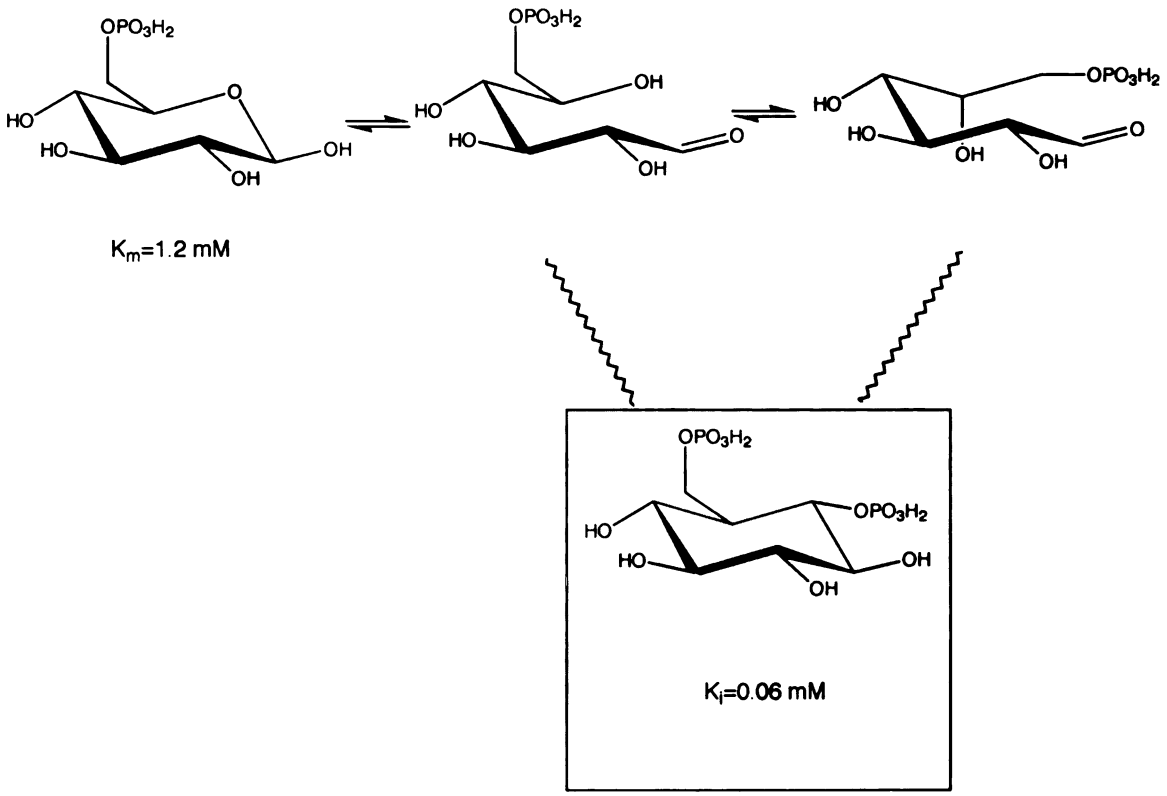


Figure 1.23 Analogues of reaction intermediates of the early steps during the catalysis by MIP synthase(78).

1.3 Literature cited

1. Jacob, S. T., and Ghosh, A. K. (1999) *J.Cell.Biochem.Suppl*, 41-50
2. Paule, M. R. (1990) *Nature* **344**, 819-820
3. Paule, M. R., and White, R. J. (2000) *Nucleic Acids Res* **28**, 1283-1298
4. Woychik, N. A., and Hampsey, M. (2002) *Cell* **108**, 453-463
5. Orphanides, G., Lagrange, T., and Reinberg, D. (1996) *Genes Dev* **10**, 2657-2683
6. Kaufmann, J., and Smale, S. T. (1994) *Genes Dev* **8**, 821-829
7. Verrijzer, C. P., Chen, J. L., Yokomori, K., and Tjian, R. (1995) *Cell* **81**, 1115-1125
8. Gill, G., and Tjian, R. (1992) *Curr Opin Genet Dev* **2**, 236-242
9. Dynlacht, B. D., Hoey, T., and Tjian, R. (1991) *Cell* **66**, 563-576
10. Tanese, N., Pugh, B. F., and Tjian, R. (1991) *Genes Dev* **5**, 2212-2224
11. Zawel, L., and Reinberg, D. (1995) *Annu Rev Biochem* **64**, 533-561
12. Carey, M., and Smale, S. T. (2000) *Transcriptional regulation in eukaryotes. Concepts, strategies, and techniques*, 1st Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor
13. Chedin, S., Ferri, M. L., Peyroche, G., Andrau, J. C., Jourdain, S., Lefebvre, O., Werner, M., Carles, C., and Sentenac, A. (1998) *Cold Spring Harbor Symp. Quant. Biol.* **63**, 381-389
14. Hernandez, N. (1993) *Genes Dev* **7B**, 1291-1308
15. Kumar, A., Grove, A., Kassavetis, G. A., and Geiduschek, E. P. (1998) *Cold Spring Harb Symp Quant Biol* **63**, 121-129

16. White, R. J. (1998) *RNA Polymerase III Transcription*, Springer-Verlag, New York

17. Geiduschek, E. P., and Kassavetis, G. A. (2001) *J Mol Biol* **310**, 1-26

18. Huang, Y., and Maraia, R. J. (2001) *Nucleic Acids Res* **29**, 2675-2690

19. White, R. J. (1998) *Int J Oncol* **12**, 741-748

20. Kassavetis, G. A., Braun, B. R., Nguyen, L. H., and Geiduschek, E. P. (1990) *Cell* **60**, 235-245

21. Apone, L. M., and Green, M. R. (1998) *Nature* **393**, 114-115

22. Geiduschek, E. P., and Kassavetis, G. A. (1995) *Current Opinion in Cell Biology* **7**, 344-351

23. Grove, A., Kassavetis, G. A., Johnson, T. E., and Geiduschek, E. P. (1999) *J Mol Biol* **285**, 1429-1440

24. Grove, A., Galeone, A., Yu, E., Mayol, L., and Geiduschek, E. P. (1998) *J Mol Biol* **282**, 731-739

25. Braun, B. R., Kassavetis, G. A., and Geiduschek, E. P. (1992) *J Biol Chem* **267**, 22562-22569

26. Leveillard, T., Kassavetis, G. A., and Geiduschek, E. P. (1991) *J Biol Chem* **266**, 5162-5168

27. Takada, S., Lis, J. T., Zhou, S., and Tjian, R. (2000) *Cell* **101**, 459-469

28. Kuras, L., Kosa, P., Mencia, M., and Struhl, K. (2000) *Science* **288**, 1244-1248

29. Kosa, P. F., Ghosh, G., DeDecker, B. S., and Sigler, P. B. (1997) *Proc Natl Acad Sci U S A* **94**, 6042-6047

30. Nikolov, D. B., Chen, H., Halay, E. D., Usheva, A. A., Hisatake, K., Lee, D. K., Roeder, R. G., and Burley, S. K. (1995) *Nature* **377**, 119-128

31. Bagby, S., Kim, S., Maldonado, E., Tong, K. I., Reinberg, D., and Ikura, M. (1995) *Cell* **82**, 857-867

32. Chaussivert, N., Conesa, C., Shaaban, S., and Sentenac, A. (1995) *J Biol Chem* **270**, 15353-15358

33. Khoo, B., Brophy, B., and Jackson, S. P. (1994) *Genes Dev* **8**, 2879-2890

34. Kassavetis, G. A., Kumar, A., Ramirez, E., and Geiduschek, E. P. (1998) *Mol Cell Biol* **18**, 5587-5599

35. Kassavetis, G. A., Bardeleben, C., Kumar, A., Ramirez, E., and Geiduschek, E. P. (1997) *Mol Cell Biol* **17**, 5299-5306

36. Buratowski, S., and Zhou, H. (1992) *Cell* **71**, 221-230

37. Colbert, T., Lee, S., Schimmack, G., and Hahn, S. (1998) *Mol Cell Biol* **18**, 1682-1691

38. Cormack, B. P., and Struhl, K. (1993) *Science* **262**, 244-248

39. Huet, J., Conesa, C., Manaud, N., Chaussivert, N., and Sentenac, A. (1994) *Nucleic Acids Res* **22**, 3433-3439

40. Aasland, R., Stewart, A. F., and Gibson, T. (1996) *Trends Biochem Sci* **21**, 87-88

41. Kumar, A., Kassavetis, G. A., Geiduschek, E. P., Hambalko, M., and Brent, C. J. (1997) *Mol Cell Biol* **17**, 1868-1880

42. Shah, S. M., Kumar, A., Geiduschek, E. P., and Kassavetis, G. A. (1999) *J Biol Chem* **274**, 28736-28744

43. Cloutier, T. E., Librizzi, M. D., Mollah, A. K., Brenowitz, M., and Willis, I. M. (2001) *Proc Natl Acad Sci U S A* **98**, 9581-9586

44. Goutsouliak, V., and Rabkin, S. W. (1997) *Cell Signal* **9**, 505-512

45. Lenox, R. H., and Watson, D. G. (1994) *Clin Chem* **40**, 309-314

46. Rana, R. S., and Hokin, L. E. (1990) *Physiol Rev* **70**, 115-164

47. Duman, R. S., Heninger, G. R., and Nesteler, E. J. (1997) *Arch. Gen. Psychiatry* **54**, 597

48. Margolis, R. U., Press, R., Altszuler, N., and Stewart, M. A. (1971) *Brain Res* **28**, 535

49. Sherman, W. R. (1989) *Inositol Lipids in Cells Signaling* (Downes, C. P., Ed.), Academic, New York

50. Berridge, M. J., Downes, C. P., and Hanley, M. R. (1982) *Biochem J* **206**, 587-595

51. Sherman, W. R., Gish, B. G., Honchar, M. P., and Munsell, L. Y. (1986) *Fed Proc* **45**, 2639-2646

52. Peet, M., and Pratt, J. P. (1993) *Drugs* **46**, 7

53. Vaden, D. L., Ding, D., Peterson, B., and Greenberg, M. L. (2001) *J Biol Chem* **276**, 15466-15471

54. Williams, R., S. B., Cheng, L., Mudge, A. W., and Harwood, A., J. (2002) *Nature* **417**, 292-295

55. Barnett, J. E., and Corina, D. L. (1968) *Biochem J* **108**, 125-129

56. Loewus, M. W., and Loewus, F. A. (1980) *Carbohydr Res* **82**, 333-342

57. Loewus, F. A., and Kelly, S. (1962) *Biochem Biophys Res Commun* **7**, 204-208

58. Barnett, J. E., Rasheed, A., and Corina, D. L. (1973) *Biochem Soc Trans* **1**, 1267-1269

59. Barnett, J. E., Rasheed, A., and Corina, D. L. (1973) *Biochem J* **131**, 21-30

60. Peng, J. (1996) Ph.D. Dissertation, Michigan State University

61. Chen, C. H., and Eisenberg, F., Jr. (1975) *J Biol Chem* **250**, 2963-2967

62. Chen, I. W., and Charalampous, F. C. (1967) *Biochim Biophys Acta* **136**, 568-570

63. Loewus, M. W. (1977) *J Biol Chem* **252**, 7221-7223

64. Loewus, M. W., Loewus, F. A., Brillinger, G. U., Otsuka, H., and Floss, H. G. (1980) *J Biol Chem* **255**, 11710-11712

65. Sherman, W. R., Hipps, P. P., Mauck, L. A., and Rasheed, A. (1978) *In Cyclitols and phosphoinositides* (Eisenberg, F., Ed.), Academic Press, New York

66. Maeda, T., and Eisenberg, F., Jr. (1980) *J Biol Chem* **255**, 8458-8464

67. Mauck, L. A., Wong, Y. H., and Sherman, W. R. (1980) *Biochemistry* **19**, 3623-3629

68. Pittner, F., and Hoffmann-Ostenhof, O. (1976) *Hoppe Seylers Z Physiol Chem* **357**, 1667-1671

69. Pittner, F., and Hoffmann-Ostenhof, O. (1978) *Hoppe Seylers Z Physiol Chem* **359**, 1395-1400

70. Chen, L., Zhou, C., Yang, H., and Roberts, M. F. (2000) *Biochemistry* **39**, 12415-12423

71. Kiely, D. E., and Sherman, W. R. (1975) *J Am Chem Soc* **97**, 6810-6814

72. Stein, A. J., and Geiger, J. H. (2002) *J Biol Chem* **277**, 9484-9491

73. Norman, R. A., McAlister, M. S., Murray-Rust, J., Movahedzadeh, F., Stoker, N. G., and McDonald, N. Q. (2002) *Structure (Camb)* **10**, 393-402

74. Rossmann, M. G. (1976) *Fed Proc* **35**, 2112-2114

75. Frey, P. A. (1987) *In Pyridine Nucleotide Coenzyes: Chemical, Biochemical and Medicinal Aspects. Part B.* (Avramovic, O., Ed.), Vol.2, John Wiley & Sons, Inc., New York

76. Byun, S. M., and Jenness, R. (1981) *Biochemistry* **20**, 5174-5177

77. Tian, F., Migaud, M. E., and Frost, J. W. (1999) *J. Am. Chem. Soc.* **121**, 5795-5796

78. Migaud, M. E. (1996) Ph. D. Dissertation, Michigan State University

79. Migaud, M. E., and Frost, J. W. (1995) *J. Am. Chem. Soc.* **117**

80. Floss, H. G. (1997) *Nat Prod Rep* **14**, 433-452

81. Carpenter, E. P., Hawkins, A. R., Frost, J. W., and Brown, K. A. (1998) *Nature* **394**, 299-302

82. Gumber, S. C., Loewus, M. W., and Loewus, F. A. (1984) *Arch Biochem Biophys* **231**, 372-377

CHAPTER II

CRYSTALLIZATION AND PRELIMINARY X-RAY DIFFRACTION ANALYSIS OF THE TFIIIB COMPLEX

In order to determine the structure of TFIIIB, the central initiation factor of Pol III transcription, recombinant mutants of individual members of TFIIIB, namely BRF, B″, TBP, and different promoter DNAs were designed based on published and unpublished data (1-4). This chapter will describe the experimental procedures of purification of BRF, B″, TBP, and DNAs, crystallization, as well as preliminary X-ray diffraction analysis.

2.1 BRF

Crystals of a full-length member of a TFIIB-like protein have not been obtained. In fact, it has been necessary to remove the Zn ribbon domain in order to obtain crystals in all crystal structures of TFIIB-like proteins published so far (5-8). An NMR structure of the C-terminal repeat domains of TFIIB indicates that there are only weak interactions between the two cyclin domains in the absence of TBP binding (9). Since the TFIIB-homology N-terminal region of BRF does not detectably interact with TBP/DNA, it may be difficult to elucidate the structure of this region. Therefore, we decided to focus on the C-terminal region of BRF. Figure 2.1 shows the deletion mutants of BRF that were constructed for the structural investigations; these include six *S. cerevisiae* BRF mutants and four *K. lactis* BRF mutants. All BRF mutant plasmids were generated by subcloning the coding sequence for BRF with a six-His tag at the C-termini into pET vectors. Our collaborator, Dr. Steven Hahn, prepared and provided all plasmids.

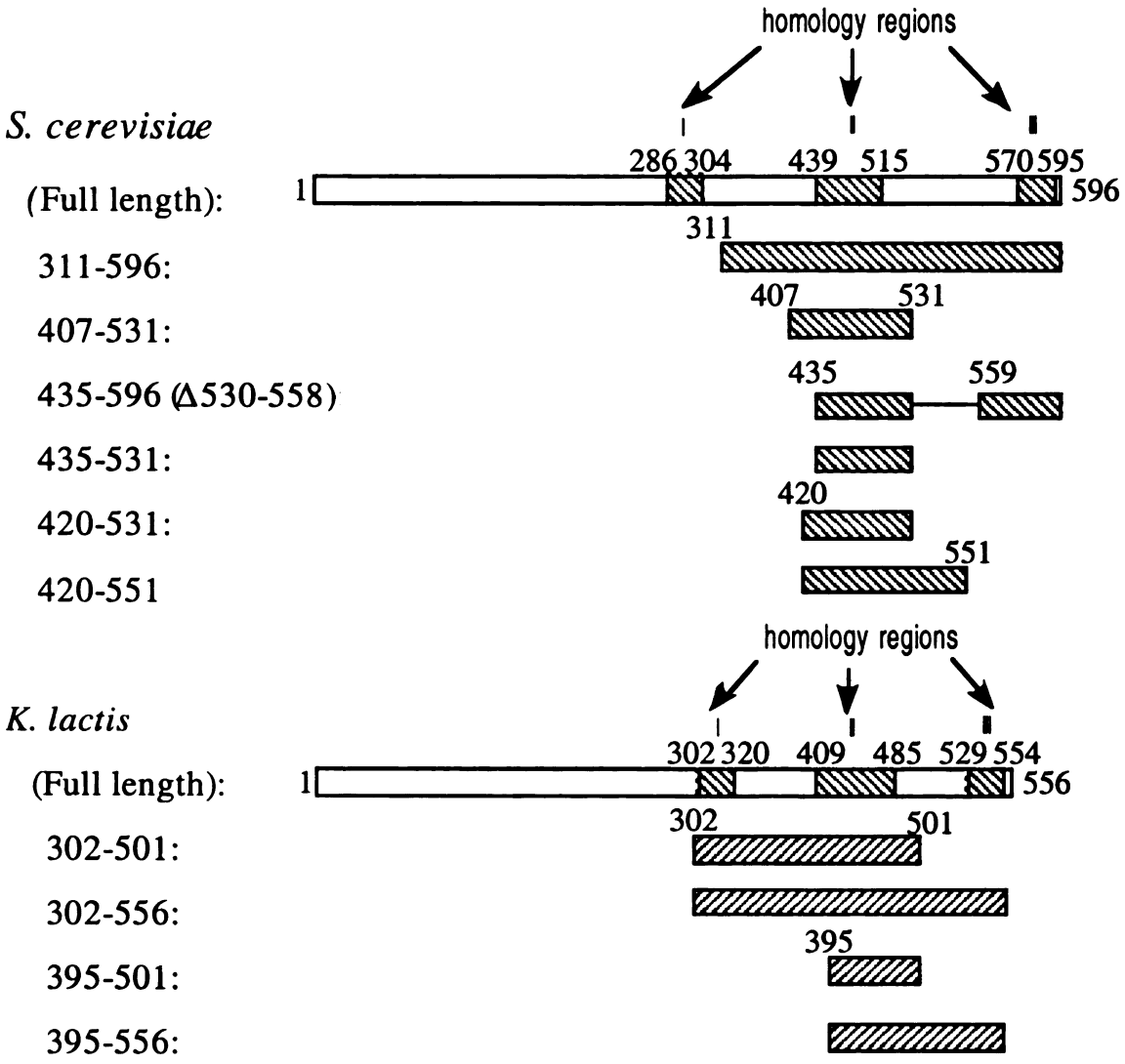


Figure 2.1 BRF constructs used in this study.

2.1.1 Over-expression and purification of BRF

2.1.1.1 Over-expression of BRF

All *S. cerevisiae* BRF plasmids were transformed into *E. coli* BL21 (DE3) strains. The LB medium containing 100mg/L ampicillin was inoculated by a single colony of transformed *E. coli* BL21 (DE3) cells. Growth conditions of 37 °C and rapid shaking at 250 rpm were maintained for 18 hours or until the medium was saturated with cells. One liter of LB/ampicillin was then inoculated with the saturated culture and growth continued until an O.D. at 600 nm of 0.6. The inducer, isothiopropyl-β-galactoside (IPTG), was then added to a final concentration of 0.5 mM, and shaking continued for another 3 hours. Cells were then harvested by centrifugation at 5000 rpm.
All *K. lactis* BRF plasmids were transformed into *E. coli* BL21 (DE3) codon plus cells, all growth media contained 34 mg/L chloramphenicol in addition to 100 mg/L ampicillin, but the rest of the protocol remained the same.

2.1.1.2 Purification of BRF

Cells were re-suspended in the lysis buffer (6M Guanidine·HCl, 0.1M NaH_2PO_4, 0.01M Tris·Cl, pH 8.0), and sonicated using 2 or 3 one-minute pulses on ice. The lysate was then centrifuged for 20-30 minutes at 15000 rpm to spin the debris down; the proteins were in the supernatant.

Ni-NTA affinity chromatography

Ni-NTA agarose (Qiagen) was used for 6x His-tagged purification. Buffers used for purification of BRF proteins under denaturing conditions were:

- Lysis buffers

Buffer A: 6 M Gu·HCl; 0.1M NaH_2PO_4; 0.01M Tris·Cl, pH 8.0

Buffer B: 6 M Urea; 0.1M NaH_2PO_4; 0.01M Tris ·Cl, pH 8.0

- Wash buffer

Buffer C: 6 M urea; 0.1M NaH_2PO_4; 0.01M Tris ·Cl, pH 6.8

- Elution buffers

Buffer D: 6 M urea; 0.1M NaH_2PO_4; 0.01M Tris ·Cl, pH 4.5

Buffer E: 6M Gu·HCl; 0.2M Acetic acid

Ni-NTA superflow slurry was resuspended completely and poured into the column to allow the resin to settle. The column was then equilibrated with 5 bed volumes of the lysis buffer. The above lysate was loaded onto the column and washed with the lysis buffers and wash buffer, and the BRF was eluted with the elution buffers. Fraction samples were analyzed by the Bradford method and SDS-PAGE. The purity of the protein can be seen in the SDS-PAGE shown in Figure 2.2. Fractions containing BRF of similar purity were pooled together for further preparation.

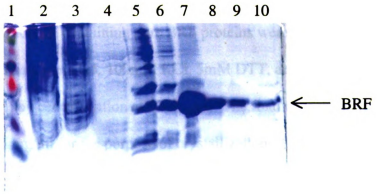


Figure 2.2 SDS-PAGE of Ni-NTA affinity purified *K. lactis* BRF 395-501.
Marker; 2. Flow through; 3. Wash A; 4. Wash B; 5. Wash C; 6-10. Elutions in 5 fractions in order. M.W. marker: Purple 42,000, Orange 32,000, Red 17,900, and Blue 7,200.

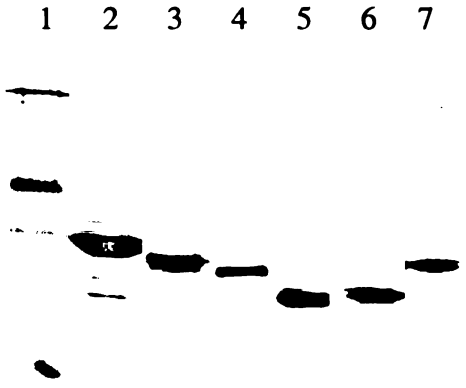
Protein refolding

Proteins were diluted in a dilution buffer (6M Urea, 10% Glycerol, 20mM Tris·Cl, 500mM KCl, 5mM DTT) to about 0.5mg/ml. They were then dialyzed against the refolding buffer (10% Glycerol, 20mM Tris·Cl, 500mM KCl, 2mM EDTA, 5mM DTT, 1mM PMSF) for 6 hours. The refolding buffer was changed and the dialysis was repeated three more times.

FPLC ion exchange chromatography

Pharmacia FPLC equipment was used to purify refolded proteins by ion exchange chromatography. Source 15Q matrix (anion exchanger with an average particle size of 15μm) was used as the stationary phase. Refolded BRF proteins were diluted 3 fold with a buffer containing 10% Glycerol, 20mM Tris·Cl, pH 8.5, 5mM DTT, then loaded onto a Source-Q column, and a gradient was run from 50 mM KCl (buffer A: 10% Glycerol, 20mM Tris.Cl, pH 8.5, 50mM KCl, 5mM DTT) to 1M KCl (buffer B: 10% Glycerol, 20mM Tris.Cl, pH 8.5, 1M KCl, 5mM DTT). BRF proteins elute at the salt concentration of about 200mM. The fractions were analyzed using the Bradford assay and SDS-PAGE. The fractions containing pure BRF proteins were buffer-exchanged in 10% Glycerol, 20mM Tris·Cl, pH 8.5, 100mM KCl, 5mM DTT, and concentrated using Centriprep-10 (Amicon) by centrifugation at 6000 rpm. The final concentration of BRF was from 5mg/ml to 35 mg/ml for the purpose of crystallization. Figure 2.3 shows the purified BRF constructs on SDS-PAGE.

A. Over-expressed and purified *S. cerevisiae* BRF constructs



B. Over-expressed and purified *K. lactis* BRF constructs

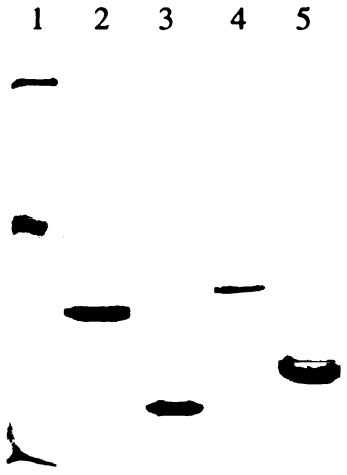


Figure 2.3 Over-expressed and purified BRF constructs used in crystallization.
(A) 1. M.W. marker 2. *S. cerevisiae* BRF311-596 3. *S. cerevisiae* BRF407-531 4. *S. cerevisiae* BRF435-596 (Δ530-558) 5. *S. cerevisiae* BRF435-531 6. *S. cerevisiae* BRF435-551 7. *S. cerevisiae* BRF 420-551.
(B) 1. M.W. marker 2. *K. lactis* BRF302-501 3. *K. lactis* BRF395-501 4. *K. lactis* BRF302-556 5. *K. lactis* BRF395-556.
M.W. marker: Purple 42,000, Orange 32,000, Red 17,900, and Blue 7,200.

2.1.2 Crystallization and data collection of BRF

In order to obtain well diffracting crystals of BRF functional domains, extensive screening of conditions was performed on all 10 constructs of BRF. Both the hanging drop vapor diffusion method (Figure 2.4) and the batch crystallization method (Figure 2.5) were used. In the hanging drop vapor diffusion method, a drop containing a mixture of protein and reservoir solution is equilibrated against the reservoir. In the batch crystallization method, the protein and precipitant are mixed directly under oil to reach super saturation. Nearly 3000 conditions were screened to produce the first crystal of BRF.

Crystals of *S. cerevisiae* BRF435-531 grew over 10 months in a 1:1 mixture of protein solution (10mg/ml) and reservoir solution (8 % PEG 8000, 0.1 M Tris, pH 8.5) at room temperature. The crystals of BRF are exquisitely sensitive to small changes in protein concentration, precipitant concentration, and pH. Despite the difficulty of obtaining crystals, crystals of *S. cerevisiae* BRF435-531 were reproduced and improved. Crystals of *S. cerevisiae* BRF435-596 (Δ530-558) and *K. lactis* BRF395-501 were also grown with similar difficulty (Table 2.1). An example of a *K. lactis* BRF395-501 crystal is depicted in Figure 2.6. The crystals were cryoprotected using a cryoprotectant solution containing 30% glycerol, 8 % PEG 8000, 0.1 M Tris, pH 8.5 for data collection. Despite the difficulty of obtaining crystals and the small crystal sizes, data were collected on an *S. cerevisiae* BRF435-531 crystal. Data collection statistics are listed in Table 2.2.

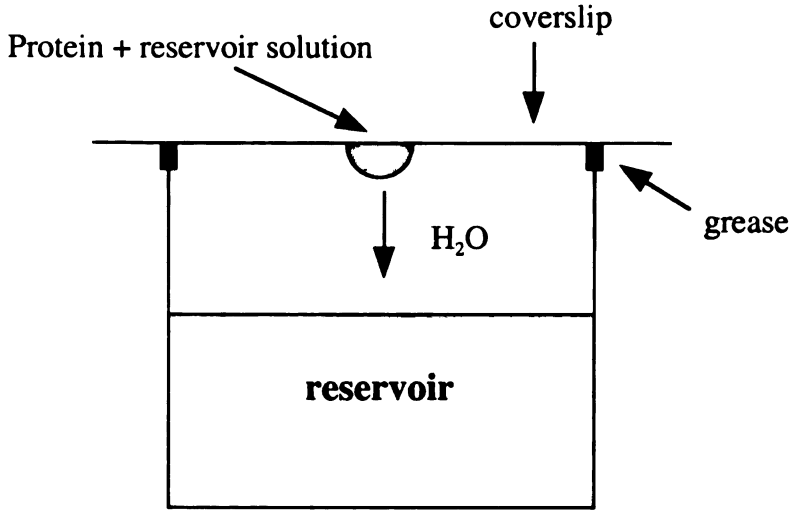


Figure 2.4 Hanging drop vapor diffusion crystallization method.

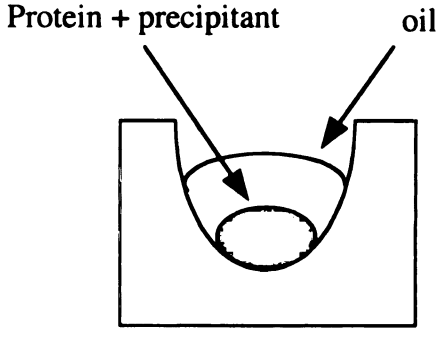


Figure 2.5 Batch crystallization method.

Table 2.1 BRF crystals

	Protein concentration	Buffer	Reservoir conditions	Resolution
S. cerevisiae BRF 435-531	8.0 mg/ml	20 mM Tris, pH 8.0, 100 mM KCl, 10 % Glycerol, 5 mM DTT	8 % PEG 8000, 0.1 M Tris, pH 8.5	3.5 Å
S. cerevisiae BRF 435-596 (Δ530-558)	6.0 mg/ml	20 mM Tris, pH 8.0, 100 mM KCl, 10 % Glycerol, 5 mM DTT	8 % PEG 8000, 0.1 M Tris, pH 8.5	4.0 Å
K. lactis BRF 395-501	20.0 mg/ml	20 mM phosphate, pH 8.0	9.5 ~10.5 % PEG 8000, 0.1M Tris, pH 8.5	3.0 Å

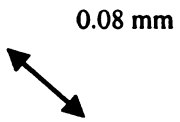


Figure 2.6 A crystal of *K. lactis* BRF 395-501, with dimensions of 0.08 × 0.08 × 0.005 mm^3.

Table 2.2 Data collection statistics of *S. cerevisiae* BRF 435-531 crystal*‡

Wavelength (Å)	0.938
Resolution range (Å)	30.0-3.5 (3.62-3.5)
Space group	P2
Unit cell parameters	a=131.087 Å, b=90.773 Å, c=143.59 Å, β=103.314°
Matthew's coefficient (Å^3/Dalton)	3.94
Solvent content (%)	69
Molecules per asymmetric unit	~ 20
Number of observations	139555
Unique reflections	37517
Completeness (%)	88.4 (90.1)
I/σ	4.47 (2.73)
R_{sym}(I)# (%)	32.4 (62.0)

* Data were collected at the Advanced Photon Source, BIOCARS BM14 beamline.

‡ Values in parenthesis refer to the last resolution shell.

$R_{sym} = \frac{\sum \left| I - \langle I \rangle \right|}{\sum I}$, where I is an individual intensity measurement and <I> is the average intensity for this reflection, with summation over all data.

The crystals of *S. cerevisiae* BRF 435-531 belong to the space group P2 with the unit cell parameters a=131.087 Å, b=90.773Å, c=143.59 Å, β=103.314°. Given the relatively small size of the protein, this unit cell is huge with about 20 molecules in the asymmetric unit, and a very high solvent content of 69 %. The length of time for crystal growth, difficulty of reproduction, low diffraction quality, and complex asymmetric unit have hindered progress towards structure determination of this functional domain of BRF.

2.2 B″

For structural investigation, a deletion mutant of *S. cerevisiae* B″ was generated by subcloning the coding sequence for B″ 240-520 with a six-His tag at the C-termini into a pET vector by our collaborator, Dr. Steven Hahn. This mutant B″ contains the SANT domain and was found to have full activity in binding TBP-DNA.

2.2.1 Over-expression and purification of B″

2.2.1.1 Over-expression of B″

The plasmid encoding B″240-520 was transformed into *E. coli* BL21 (DE3) cells. The concentrations of ampicillin and kanamycin in all growth media were 50 mg/L and 30 mg/L respectively. An overnight culture was grown from a single colony of transformed *E. coli* BL21 (DE3) cells in the LB media containing ampicillin and kanamycin, and then diluted 250 fold into fresh LB media containing ampicillin and kanamycin. The culture was incubated at 37°C with shaking at 250 rpm until the O.D. at 600 nm reached 0.5-0.6. The inducer IPTG was then added to a final concentration of

0.5mM, and shaking continued at 30°C for 3 more hours. The cells were then spun down by centrifugation at 5000 rpm and stored at –80°C.

2.2.1.2 Purification of B″

Ni NTA affinity purification

Ni-NTA agarose (Qiagen) was used for 6x His-tagged purification, buffers used for purification of B″ under native conditions are:

- Lysis Buffer

20mM HEPES, 300mM NaCl, 10% Glycerol (pH 8.0), add 5mM β- mercaptoethanol (BME) and protease inhibitors (PMSF, Benzamidine, Leupeptin, Pepstatin, Chymostatin) before use.

- Binding/Wash Buffer

Lysis Buffer + 20mM imidazole (pH 8.0), add BME and PMSF before use.

- Elution Buffer

Lysis Buffer + 100mM imidazole (pH 8.0), add BME and PMSF before use.

Ni-NTA agarose (Qiagen) was spun at 4°C to remove the storage buffer and equilibrated in 30ml binding buffer, again spun and the buffer was removed. Cells were resuspended in the lysis buffer (50ml per 2L of cells) on ice and sonicated. The crude extract was spun down by centrifugation at 15000 rpm. The supernatant containing B″ was added to the pre-equilibrated agarose and incubated at 4°C for 60 minutes. Protein-bound agarose was spun for the supernatant to be removed and washed twice with 40ml of the wash buffer. Protein-bound agarose was then packed into an HR 10/10 column, and washed with 5 bed volumes of the wash buffer at a flow rate 2ml/min, protein was eluted with 7 bed volumes of the elution buffer at a flow rate1ml/min. Upon analysis by

the Bradford method and SDS-PAGE (Figure 2.7), fractions containing B″ were pooled together for further purification.

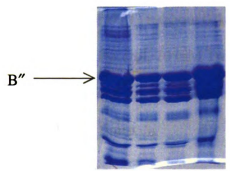


Figure 2.7 SDS-PAGE of *S. cerevisiae* B″ 240-520 after Ni-NTA affinity purification; the presence of degradation products and impurities requires further purification.

FPLC ion exchange chromatography

Pharmacia FPLC equipment was used to purify Ni-NTA purified B″ by ion exchange chromatography using Source 15S matrix (cation exchanger with an average particle size of 15μm) as the stationary phase. Ni-NTA affinity chromatography-purified proteins were buffer-exchanged into buffer A (10% Glycerol, 20mM HEPES, pH 8.0, 75 mM NaCl, 5mM BME, 1 mM PMSF). A gradient was run against buffer B (10% Glycerol, 20mM HEPES, pH 8.0, 500 mM NaCl, 5mM BME, 1 mM PMSF). B″ elutes over most of the gradient. This indicates the unusual conformational and/or aggregation heterogeneity of B″. Therefore, it was thought that refolding BRF and B″ together might lead to the production of a more homogeneous material.

2.2.2 Refolding and purification of the BRF/B″ complex

Refolding of BRF and B″

Ni-NTA purified BRF and B″ 240-520 were combined together in 1:1 molar ratio, and this mixture was diluted with a denaturing buffer (6M Urea, 10% Glycerol, 20 mM Tris·Cl, pH 8.0, 500mM KCl, 5mM DTT) to about 0.4mg/ml. The mixture was dialyzed against the refolding buffer (10% Glycerol, 20 mM Tris·Cl, pH 8.0, 500mM KCl, 2mM EDTA, 5mM DTT, 1mM PMSF) for 6 hours. The refolding buffer was replaced fresh and the dialysis was repeated three times.

Purification of the BRF/ B″ complex

Refolded BRF/ B″ complex was buffer-exchanged into buffer A (10 % Glycerol, 20 mM Tris·Cl, pH 8.0, 50 mM KCl, 2mM EDTA, 5mM DTT, 1mM PMSF), and loaded onto an FPLC Source-Q column. A salt gradient was run against buffer B (10 % Glycerol, 20 mM Tris·Cl, pH 8.0, 2mM EDTA, 5mM DTT, 1mM PMSF). The BRF/B″ complex eluted at about 130 mM KCl in a nice sharp peak, in contrast to the broad range of peaks of B″ alone. SDS-PAGE (Figure 2.8) confirmed that the peak fractions indeed contain the purified BRF/B″ complex.

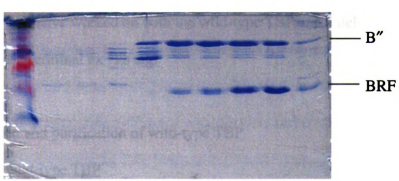


Figure 2.8 SDS-PAGE of the BRF/B″ complex purified by FPLC.

M.W. marker: Purple 42,000, Orange 32,000, Red 17,900, and Blue 7,200.

2.2.3 Crystallization of the BRF/B″ complex

Two BRF/B″ complexes listed in Table 2.3 were extensively screened for crystallization using the hanging drop vapor diffusion method (Figure 2.4). Unfortunately, neither of them produced crystals. Efforts are being made towards crystallization of the BRF/B″/TBP/DNA complexes as will be described in section 2.4.

Table 2.3 The BRF/B″ complexes used in crystallization attempts.

BRF	B″
S. cerevisiae BRF 435-531	*S. cerevisiae* B″ 240-520
K. lactis BRF 395-501	*S. cerevisiae* B″ 240-520

2.3 TBP

The C-terminal conserved domain of *S. cerevisiae* TBP was used in attempts to produce crystals of BRF/TBP/DNA complexes. This domain was chosen based on the previous success on structure determination of the TBP/TATA and the TFIIA/TBP/DNA complexes (10,11). We have worked on both the wild-type TBP and a deletion mutant, TBP 56-240 with an N-terminal 6x His tag.

2.3.1 Over-expression and purification of wild-type TBP

<u>Over-expression of wild-type TBP</u>

In the process of optimizing the over-expression level, the plasmid encoding wild-type *S. cerevisiae* TBP was transformed into several different strains of *E. coli* cells,

including BL21 (DE3), BL21 (DE3) codon plus, BL21 (DE3) pLys S, Tuner, and Tuner pLys S. It has been found that wild-type *S. cerevisiae* TBP is over-expressed in BL21 (DE3) codon plus and Tuner pLys S cells equally well and better than all the rest of the *E. coli* strains so far tested.

An LB medium containing carbenicillin and chloramphenicol was inoculated with a single colony of transformed *E. coli* BL21 (DE3) codon plus or Tuner pLys S cells. The concentrations of carbenicillin and chloramphenicol in all growth media were 100 mg/L and 34 mg/L respectively. Growth conditions of 37 °C and rapid shaking at 250 rpm were maintained for 18 hours or until saturated. The saturated culture was then diluted 40 fold into fresh LB media containing carbenicillin and chloramphenicol, and growth continued until an O.D. at 600 nm of 1.0 was reached. The inducer IPTG was then added to a final concentration of 0.4 mM, and shaking continued for another 3 hours at 30 °C. Cells were then harvested by centrifugation at 5000 rpm.

Purification of wild-type TBP

Cells were resuspended in the lysis buffer (30 mM Tris, pH 7.5, 10 % Glycerol, 50 mM KCl, 1 mM DTT, 1 mM PMSF, 2 mM EDTA), and sonicated in a dry ice/Ethanol bath. The crude extract was spun down by centrifugation at 15000 rpm. The supernatant containing TBP was loaded onto a Phast-Q (Pharmacia) in tandem with a Phast-S (Pharmacia) column that were pre-equilibrated with the lysis buffer. While most *E. coli* proteins bind to the Phast-Q column, TBP binds to the Phast-S column. The columns were washed with the lysis buffer until the O.D. at 280 nm of the flow through reached lower than 0.1. The Phast-Q column was then disconnected, and a gradient was run on

the Phast-S column from 0.2 M KCl to 0.6 M KCl in the buffers containing 30 mM Tris, pH 7.5, 10 % Glycerol, 1 mM DTT, 1 mM PMSF, 2 mM EDTA. Fractions containing TBP detected by UV-vis and evaluated by SDS-PAGE were pooled together and concentrated to about 1mg/ml for proteolysis as will be described in section 2.3.2.

Source 15S matrix (cation exchanger with an average particle size of 15μm) was used for ion-exchange chromatography. Proteolyzed TBP was buffer-exchanged into buffer A (10% Glycerol, 30mM Tris, pH 7.5, 50 mM KCl, 1 mM DTT, 1 mM PMSF, 2 mM EDTA), and loaded onto an FPLC Source-S column. A salt gradient was run against buffer B (10% Glycerol, 30mM Tris, pH 7.5, 1 M KCl, 1 mM DTT, 1 mM PMSF, 2 mM EDTA). The eluted fractions were analyzed by the Bradford assay and SDS-PAGE. The fractions containing pure TBP were concentrated for BRF/TBP/DNA complex formation and crystallization.

2.3.2 Proteolysis of wild-type TBP

In order to remove the unconserved N-terminal segment that is dispensable for basal transcription and yeast cell viability (12-14), proteolysis of wild-type *S. cerevisiae* TBP was performed using two enzymes, endoproteinase Lys-C and Trypsin.

Proteolysis with endoproteinase Lys-C

The enzyme endoproteinase Lys-C purchased from Roche was dissolved in the storage buffer (50 mM Tricine, pH 8.0, 10 mM EDTA) in 15 units/ml. Phast-S purified and concentrated TBP was introduced into the proteolysis buffer (50 mM Tricine, pH 8.0, 20 % Glycerol, 200 mM KCl, 5 mM DTT); one unit of the endoproteinase Lys-C was

added to each 20 mg of TBP to carry out the proteolysis. Reaction was done at room temperature for 4 hours followed by additional 12 hours at 4 °C.

Many trials of proteolysis with endoproteinase Lys-C have been unsuccessful; a different approach to obtain proteolyzed TBP became necessary.

Proteolysis with Trypsin

Immobilized Trypsin purchased from PIERCE was used in proteolysis attempts. The immobilized Trypsin gel was washed with 3 bed volumes of the digestion buffer (0.1M NH_4HCO_3, pH 8.0) for three times, then added to TBP in the proteolysis buffer (30 mM Tris, pH 7.5, 10 % Glycerol, 250 mM KCl). The reaction mixture was incubated at room temperature with gentle shaking at 100 rpm for 2 hours. The reaction then continued at 4 °C for additional 12 hours. The immobilized Trypsin was then separated from the reaction mixture by centrifugation. This method successfully proteolyzed TBP at the desired site. Figure 2.9 shows an example of SDS-PAGE of the proteolysis result.

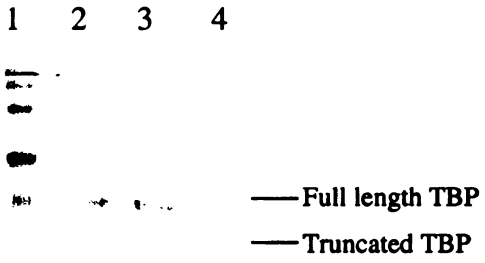


Figure 2.9 SDS-PAGE of the proteolysis of TBP.

1. M.W. marker; 2. Before the proteolysis (at the time point zero); 3. After 2 hours of reaction at room temperature; 4. After 12 hours at 4°C (the stop point).

M.W. marker: Purple 42,000, Orange 32,000, Red 17,900, and Blue 7,200.

2.3.3 Over-expression and purification of TBP 56-240

Over-expression of TBP 56-240

The plasmid encoding *S. cerevisiae* TBP 56-240 with an N-terminal 6x His tag was transformed into one of several different strains of *E. coli* cells, including BL21 (DE3), BL21 (DE3) codon plus, BL21 (DE3) pLys S, Tuner, and Tuner pLys S with Tuner pLys S cells producing the optimal over-expression. The cell growth condition was the same as that of wild-type TBP.

Purification of TBP 56-240

Ni-NTA affinity purification under native conditions was performed. Cells were re-suspended in the lysis buffer (50mM Tris, pH 7.5, 20% Glycerol, 500 mM KCl, 1mM PMSF, 5mM BME, and 3 tablets/L protease inhibitor cocktail, EDTA free). Cells were lysed by sonication and centrifuged. The supernatant containing TBP was loaded on a Ni-NTA column that was pre-equilibrated with the lysis buffer, and the column was washed with the lysis buffer, wash buffers (lysis buffer plus 20mM imidazole and lysis buffer plus 100mM imidazole). Subsequently, TBP was eluted with elution buffers (lysis buffer plus 250mM imidazole and lysis buffer plus 400mM imidazole). Fractions containing TBP identified by the Bradford method and SDS-PAGE were pooled together for further purification. The 6x His tag at its N-terminus may affect the formation of protein-TBP-DNA complex and prevent the formation of crystals, so the Ni-NTA purified TBP was proteolyzed using Thrombin before purification by ion exchange chromatography as will be described in section 2.3.4.

For FPLC ion exchange purification, protein was diluted with no salt buffer (30 mM Tris, pH 7.5, 10 % Glycerol, 1 mM DTT, 1 mM PMSF, 3 tablets/L protease inhibitor cocktail, EDTA-free), and loaded on a Source-S (cation exchange) column. A salt gradient was run from 30mM KCl to 1000mM KCl. The fractions containing TBP were identified, pooled together, and concentrated for the BRF/TBP/DNA complex formation and crystallization setups.

2.3.4 Proteolysis of TBP 56—240

Purified TBP with its N-terminal 6x His tag was successfully complexed with BRF and DNAs, and these complexes were crystallized. However, compared to the crystals of complexes formed of TBP with its His tag removed, they were evaluated to be of lower quality based on their appearances and small sizes. Thrombin was used for the removal of the N-terminal 6x His tag, Biotinylated Thrombin was much more effective than free Thrombin. Biotinylated Thrombin was buffer-exchanged to the proteolysis buffer (30 mM Tris, pH 7.5, 10 % Glycerol, 250 mM KCl) before the reaction. For each milligram of TBP, 1.5 units of Biotinylated Thrombin were sufficient for a successful proteolysis. Figure 2.10 shows an example of the proteolysis being monitored at different time points and with varied amount of the enzyme used. After the proteolysis is completed, Biotinylated Thrombin was removed by adding streptavadin into reaction mixture followed by filter centrifugation, only TBP but not thrombin remained in the filtrate by this method.

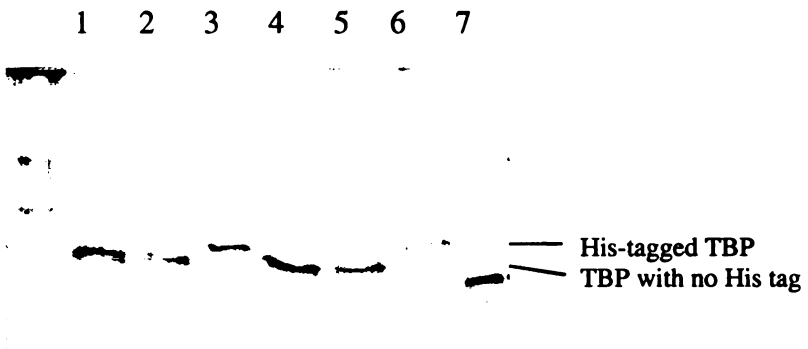


Figure 2.10 Monitored proteolysis of TBP by Biotinylated Thrombin.

1. M.W. Marker. Purple 42,000, Orange 32,000, Red 17,900, and Blue 7,200;

2. After 12 hours of reaction when 1.5 units thrombin/mg of TBP was used;

3. After 12 hours of reaction when 0.75 unit thrombin /mg of TBP was used;

4. After 12 hours when no thrombin was used;

5. After 18 hours of reaction when 1.5 units thrombin/mg of TBP was used;

6. After 18 hours of reaction when 0.75 unit thrombin/mg of TBP was used;

7. After 18 hours when no thrombin was used.

2.4 The BRF/TBP/DNA and BRF/B″/TBP/DNA complexes

The primary goal of this research is to understand the transcription initiation mechanism of the Pol III transcription system; this goal can be achieved through elucidation of the structural details of the TFIIIB complex bound to a promoter DNA sequence. All purified BRF constructs and B″ described above were evaluated for TBP/DNA binding activities by electrophoretic mobility shift assay (EMSA) by our collaborator, Dr. Steven Hahn. From the EMSA results shown in Figures 2.11-2.14, all BRF constructs and the B″ construct that we have over-expressed and purified are active in BRF/TBP/DNA and BRF/B″/TBP/DNA complex formation, thus all can be used for the crystallization of BRF/TBP/DNA and BRF/B″/TBP/DNA complexes.

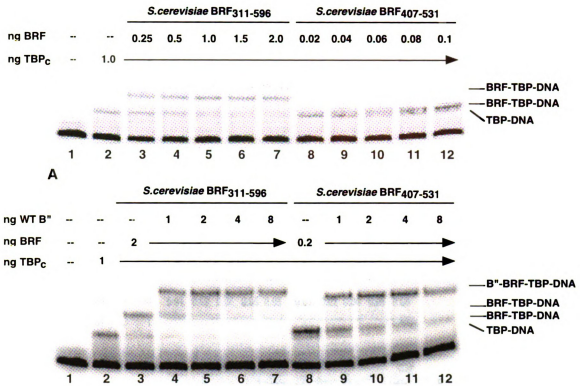


Figure 2.11 Gel mobility shift assays using1 ng *S. cerevisiae* TBP56-240, and indicated amounts of BRF and B". The probe is the U6 promoter with its TATA box modified to that of the AdMLP TATA box (U6-MLP promoter). Binding reactions for 45 min. at room temperature run on 6% TGOE (25 mM Tris, 190 mM glycine, pH 8.3) with no MgOAc gel for 60 min. at 4 °C.

(A) Titrations of *S. cerevisiae* BRF311-596 and *S. cerevisiae* BRF407-531.

(B) Titrations of WT *S. cerevisiae* B" with either 2ng *S. cerevisiae* BRF311-596 or 0.2 ng *S. cerevisiae* BRF407-531.

Experiment done by Dr. Steven Hahn.

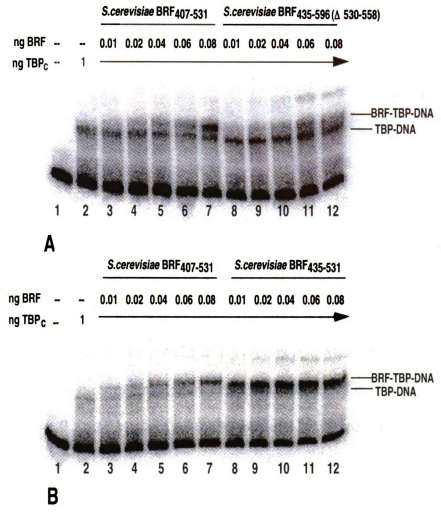


Figure 2.12 Gel mobility shift assays of *S. cerevisiae* BRF constructs using1 ng *S. cerevisiae* TBP56-240, and indicated amounts of BRF. The probe is the U6 promoter with its TATA box modified to that of the AdMLP TATA box (U6-MLP promoter). Binding reactions for 45 min. at room temperature run on 6% TGOE (25 mM Tris, 190 mM glycine, pH 8.3) with no MgOAc gel for 90 min. at 4 °C.

(A) Titrations of *S. cerevisiae* BRF407-531 and *S. cerevisiae* BRF435-596 (Δ 530-558).

(B) Titrations of *S. cerevisiae* BRF407-531 and *S. cerevisiae* BRF435-531.

Experiment done by Dr. Steven Hahn.

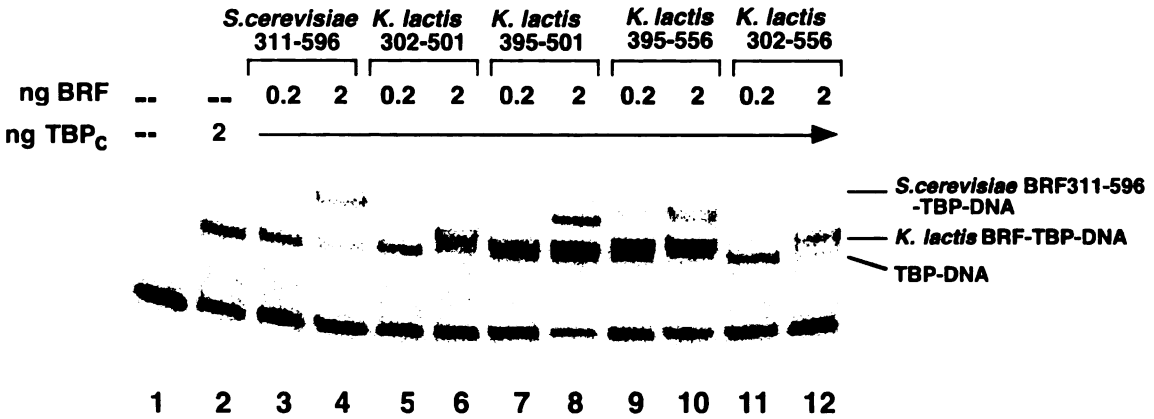


Figure 2.13 Gel mobility shift assays of *K. lactis* BRF constructs using 2 ng *S. cerevisiae* TBP 56-240, and indicated amounts of BRF. The probe is the U6 promoter with its TATA box modified to that of the AdMLP TATA box (U6-MLP promoter). Binding reactions for 45 min. at room temperature run on 6% TGOE (25 mM Tris, 190 mM glycine, pH 8.3) + 0.2 mM MgOAc gel for 90 min. at 4 °C. Experiment done by Dr. Steven Hahn.

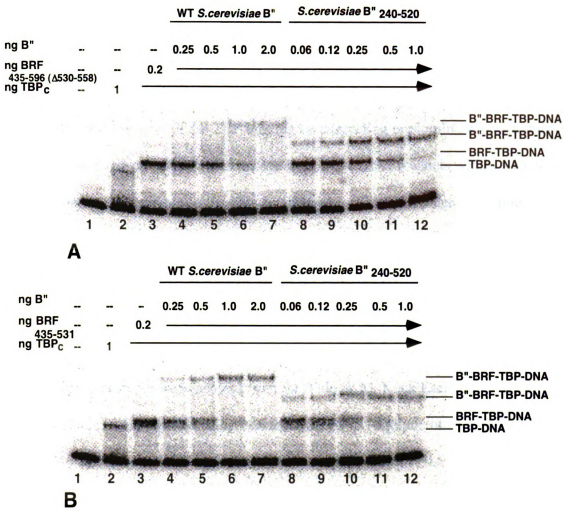


Figure 2.14 Gel mobility shift assays of *S. cerevisiae* B″ using1ng TBP 56-240, 0.2ng BRF, and indicated amounts of B". The probe is the U6 promoter with its TATA box modified to that of the AdMLP TATA box (U6-MLP promoter). Binding reactions for 45 min. at room temperature run on 6% TGOE (25 mM Tris, 190 mM glycine, pH 8.3) with no MgOAc gel for 90 min. at 4 °C. (A) Titrations of WT *S. cerevisiae* B" and *S. cerevisiae* B"240-520 with 0.2ng *S. cerevisiae* BRF435-596 (Δ 530-558). (B) Titrations of WT *S. cerevisiae* B" and *S. cerevisiae* B"240-520 with 0.2 ng *S. cerevisiae* BRF435-531. Experiment done by Dr. Steven Hahn.

2.4.1 DNAs

Previous mutagenesis, DNA footprinting, and photocrosslinking results have provided information on the architecture of TFIIIB on the promoter DNA. In addition to being responsible for most of the TBP binding by interacting predominantly with the N-terminal top and stirrup of TBP, BRF makes most of its DNA contacts within a 15 bp region immediately upstream of the TBP binding site. B″, on the other hand, interacts with DNA within 10 bp immediately downstream of the TBP binding site(2,3,15-17). According to these data, DNAs were designed for complex formation and crystallization.

Purification of DNAs

Oligonucleotides were ordered from the Yale W.M. Keck Facility. Oligonucleotides were purified using Perkin Elmer HPLC equipment on a Source-Q column; UV absorbance at 260 nm was monitored. A strand of DNA was loaded onto the Source-Q column with buffer A (10 mM NaOH, 0.2 M NaCl), and eluted with a gradient against buffer B (10 mM NaOH, 1.0 M NaCl). Collected fractions were neutralized with 1 M Tris, pH 7.5, then diluted with10 mM Tris, pH 7.5 and loaded on a DEAE column. DNAs were then eluted with a buffer of 10 mM Tris, pH 7.5, 1 M NaCl. Eluted DNAs were concentrated and annealed in equimolar amounts. Table 2.4 and Table 2.5 list all of the oligonucleotides used in TFIIIB complex crystallization attempts. For DNAs (Table 2.5) that were used for the BRF/B″/TBP/DNA complex formation, four strands with overhangs were annealed together in equimolar amounts.

Table 2.4 DNAs utilized in crystallization of the BRF/TBP/DNA complexes

	DNA sequence
DNA1	TACTATAAAAGAATGTTTTTTTCGCATA ATGATATTTTCTTACAAAAAAAGCGTAT
DNA2	AATACTATAAAAGAATGTTTTTTTCGCA TTATGATATTTTCTTACAAAAAAAGCGT
DNA3	TACTATAAAAGAATGTTTTTTTCGCAT ATGATATTTTCTTACAAAAAAAGCGTA
DNA4	ATACTATAAAAGAATGTTTTTTTCGCA TATGATATTTTCTTACAAAAAAAGCGT
DNA5	TACTATAAAAGAATGTTAATTTCGCA ATGATATTTTCTTACAATTAAAGCGT
DNA6	TACTATAAAAGAATGTTTTTTTCGC ATGATATTTTCTTACAAAAAAAGCG
DNA7	TACTATAAAAGAATGTTTTTTTCG ATGATATTTTCTTACAAAAAAAGC
DNA8	CTATAAAAGAATGTTTTTTTCGCA GATATTTTCTTACAAAAAAAGCGT
DNA9	TTACTATAAAAGAATGTTTTTTTCGC AATGATATTTTCTTACAAAAAAAGCG
DNA10	TTACTATAAAAGAATGTTTTTTTCGC ATGATATTTTCTTACAAAAAAAGCGT
DNA11	ATACTATAAAAGAATGTTTTTTTCGC ATGATATTTTCTTACAAAAAAAGCGA
DNA12	GTACTATAAAAGAATGTTTTTTTCGC ATGATATTTTCTTACAAAAAAAGCGC
DNA13	TTACTATAAAAGAATGTTTTTTTCGC ATGATATTTTCTTACAAAAAAAGCGA
DNA14	ATACTATAAAAGAATGTTTTTTTCGC ATGATATTTTCTTACAAAAAAAGCGT

Table 2.5 DNAs utilized in crystallization of the BRF/B″/TBP/DNA complexes

	DNA sequence
DNA 15	ACTATTTTCGGCTA CTATAAAAGAATGTTTTTTTCGCA TGATAAAAGC CGATGATATTTTCTTACAAAAAAAGCGT
DNA 16	CGTCCACTATTTT CGGCTACTATAAAAGAATGTTTTTTTCGCA GCAGGTGATAAAAGCCGAT GATATTTTCTTACAAAAAAAGCGT
DNA 17	CGTCCACTATTTTC GGCTACTATAAAAGAATGTTTTTTTCGCAAC GCAGGTGATAAAAGCCGAT GATATTTTCTTACAAAAAAAGCGTTG

2.4.2 Crystallization of the BRF/TBP/DNA and BRF/ B″/TBP/DNA complexes

BRF, TBP, and DNA were diluted and combined in the ratio of 1:1:1.5 in a buffer containing 50 mM Tris, pH 8.0, 10 % Glycerol, 300 mM KCl, 1 mM DTT, then left on ice for approximately 20 minutes to get complete binding. Very often, aggregation of the formed complexes was observed even with a concentration as low as 0.08 mg/ml. In these cases, a buffer containing higher concentration of salt was added to the mixture. Any precipitates were removed by centrifugation before crystallization setups. So far 26 different BRF/TBP/DNA complexes have been made and used in crystallization as listed in Table 2.6, of which 11 comlexes crystallized. Table 2.7 lists the conditions where the BRF/TBP/DNA complex crystals were grown. In order to confirm the presence of individual members of BRF/TBP/DNA complexes in the grown crystals, SDS-PAGE was performed on dissolved crystals. The results shown in Figure 2.15 verified the presence of BRF and TBP. Since BRF was known to interact with TBP poorly in the absence of DNA, although the presence of DNA was not tested we are sure that the DNAs are present in those complexes. A few examples of BRF/TBP/DNA complex crystals are

shown in Figure 2.16. Currently, the attempts to crystallize the BRF/B″/TBP/DNA complexes are being made.

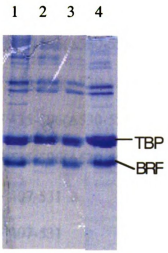


Figure 2.15 SDS-PAGE of the BRF/TBP/DNA complexes. (1) *S. cerevisiae* BRF 435-531/TBP/DNA1 (2) *S. cerevisiae* BRF 435-531/TBP/DNA2. (3) *S. cerevisiae* BRF 435-531/TBP/DNA3. (4) *S. cerevisiae* BRF 435-531/TBP/DNA6.

Table 2.6 The BRF/TBP/DNA complexes tested for crystallization so far

Complexes	BRF	DNA*	Crystals
1	*S. cerevisiae* BRF 311-596	DNA1	Yes
2	*S. cerevisiae* BRF 435-596(Δ530-558)	DNA1	No
3	*S. cerevisiae* BRF 435-596(Δ530-558)	DNA2	No
4	*S. cerevisiae* BRF 435-596(Δ530-558)	DNA3	Yes
5	*S. cerevisiae* BRF 435-596(Δ530-558)	DNA6	No
6	*S. cerevisiae* BRF 407-531	DNA1	No
7	*S. cerevisiae* BRF 407-531	DNA2	No
8	*S. cerevisiae* BRF 407-531	DNA3	No
9	*S. cerevisiae* BRF 407-531	DNA4	No
10	*S. cerevisiae* BRF 407-531	DNA5	No
11	*S. cerevisiae* BRF 407-531	DNA6	No
12	*S. cerevisiae* BRF 407-531	DNA7	No
13	*S. cerevisiae* BRF 407-531	DNA8	No
14	*S. cerevisiae* BRF 435-531	DNA1	Yes
15	*S. cerevisiae* BRF 435-531	DNA2	Yes
16	*S. cerevisiae* BRF 435-531	DNA3	Yes
17	*S. cerevisiae* BRF 435-531	DNA4	No
18	*S. cerevisiae* BRF 435-531	DNA5	No
19	*S. cerevisiae* BRF 435-531	DNA6	Yes
20	*S. cerevisiae* BRF 435-531	DNA7	Yes

Table 2.6 (cont'd)

	BRF	DNA	Crystals
21	*S. cerevisiae* BRF 420-531	DNA9	No
22	*K. lactis* BRF 395-501	DNA1	Yes
23	*K. lactis* BRF 395-501	DNA2	Yes
24	*K. lactis* BRF 395-501	DNA5	Yes
25	*K. lactis* BRF 395-501	DNA6	Yes
26	*K. lactis* BRF 395-501	DNA7	No

* DNA numberings are consistent with Table 2.5.

Table 2.7 Crystals of the BRF/TBP/DNA complexes grown so far.

	Complexes	Size (at the longest direction)	Conditions
1	*S. cerevisiae* BRF 311-596/TBP/DNA1	0.05 mm	30 % PEG 8000, 0.2 M NaAc, 0.1 M Na Cacodylate , pH 6.5
2	*S. cerevisiae* BRF 435-596 (Δ530-558)/TBP/DNA4	0.05 mm	30 % PEG 4000, 0.2 M $MgCl_2$, 0.1 M Tris, pH 7.5
		0.2 mm	2.0 M NH_4H_2PO4 2.0 M, 0.1 M Tris, pH 8.5
3	*S. cerevisiae* BRF 435-531/TBP/DNA1	0.1 mm	30 % PEG400, 0.2 M $CaCl_2$, 0.1 M Na-HEPES, pH 7.5
		0.08 mm	28 % PEG 4000, 0.2 M $MgCl_2$, 0.1 M Tris, pH 7.5
4	*S. cerevisiae* BRF 435-531/TBP/DNA2	0.15 mm	30 % PEG 4000, 0.2 M $MgCl_2$, 0.1 M Tris, pH 7.5
5	*S. cerevisiae* BRF 435-531/TBP/DNA3	0.08 mm	30 % PEG 4000, 0.2 M $MgCl_2$, 0.1 M Tris, pH 7.5
6	*S. cerevisiae* BRF 435-531/TBP/DNA6	0.05 mm	35 % PEG 4000, 0.2M $MgCl_2$, 0.1 M Tris, pH 7.5
7	*S. cerevisiae* BRF 435-531/TBP/DNA7	0.05 mm	30 % PEG 4000, 0.2M $MgCl_2$, 0.1 M Tris, pH 7.5
		0.05 mm	18 % PEG 8000, 0.2 M $Ca(CH_3COO)_2$, 0.1 M Na-Cacodylate, pH 6.5
8	*K. lactis* BRF 395-501/TBP/DNA1	0.2 mm	35 % PEG400, 0.2 M $CaCl_2$, 0.1 M Na-HEPES, pH 7.5

Table 2.7 (cont'd)

	Complexes	Size (at the longest direction)	Conditions
9	*K. lactis* BRF 395-501/TBP/DNA2	0.1 mm	30 % PEG 4000, 0.2 M $MgCl_2$, 0.1M Tris, pH 9.5
10	*K. lactis* BRF 395-501/TBP/DNA5	0.05 mm	28 % PEG 4000, 0.2 M $MgCl_2$, 0.1M Tris, pH 7.5
11	*K. lactis* BRF 395-501/TBP/DNA6	0.05 mm	30 % PEG 400, 0.2 M $CaCl_2$, 0.1 M Tris, pH 7.5

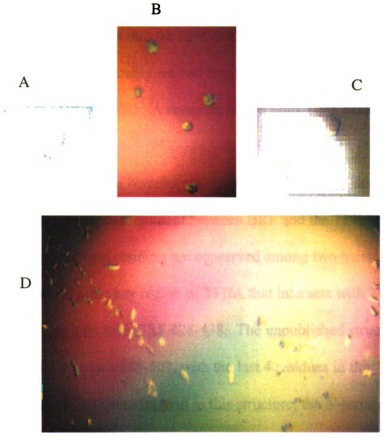


Figure 2.16 (A) Crystals of the *S. cerevisiae* BRF 435-531/TBP/DNA1 complex grown in 30 % PEG400, 0.2 M $CaCl_2$, 0.1 M Na-HEPES, pH 7.5. (B) Crystals of the *S. cerevisiae* BRF 435-531/TBP/DNA6 complex grown in 30 % PEG 4000, 0.2 M $MgCl_2$, 0.1 M Tris, pH 7.5. (C) Crystals of the *S. cerevisiae* BRF 435-531/TBP/DNA2 complex grown in 30 % PEG 4000, 0.2 M $MgCl_2$, 0.1 M Tris, pH 7.5. (D) Crystals of the *S. cerevisiae* BRF 435-531/TBP/DNA7 complex grown in 30 % PEG 4000, 0.2M $MgCl_2$, 0.1 M Tris, pH 7.5.

2.5 Discussion

Regardless of all the efforts that we have made on this project and the above-mentioned preliminary data, regrettably we had to give up our pursuit of this project, since the structure of the BRF/TBP/DNA complex has recently been determined by another research group (18).

Our final thoughts go to the homology between the C-terminal domain and TFIIA. As shown in the sequence alignment between BRF and the TFIIA small subunit (Figure 2.17), few but significant residues are conserved among two transcription initiation factors. Especially, the key region of TFIIA that interacts with TBP, residues 67-72, shares a good homology with BRF 428-438. The unpublished structure (18) of BRF/TBP/DNA contains residues 435-507, with the last 4 residues in this homology region at the N-terminal. It is interesting that in this structure, the N-terminus including these 4 residues begins with a helix, while in the TFIIA structure this region is a strand. However, It can still be assumed that interactions between this region of BRF and TBP/DNA would be similar to those of TFIIA. It was also noted that the C-termini of two proteins, BRF 564-585 within the homology region III, and TFIIA 99-120, are remarkably well conserved (Figure 2.17).

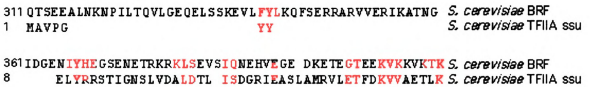


Figure 2.17 Sequence alignment of *S. cerevisiae* BRF C-terminal domain and TFIIA small subunit. Identical residues are highlighted in yellow. Boxed in blue are BRF428-438 and TFIIA 67-72; this region of TFIIA makes interactions with the TBP/DNA.

TFIIA interacts with TBP near the N-terminal stirrup and interacts with the DNA upstream of the TATA box (10,19). TFIIB interacts both with the C-terminal stirrup and DNA on one face of TBP/DNA complex (5,7). It was suggested from the Gel Mobility Shift Assay that BRF does not interact on the same face of the TBP/DNA complex as TFIIB, however, BRF likely overlaps in its position with TFIIA for binding to TBP, DNA, or both (2). The folding of some domains of BRF can be roughly predicted, based on the structures of the TFIIA/TBP/DNA complex and the TFIIB/TBP/DNA complex that have already been determined. Figure 2.18 shows the modeled TFIIA/TFIIB/TBP/DNA quaternary complex, this modeling was done by overlaying the structure of TBP in the human TFIIBc/TBPc/MLP complex (7) onto the structure of TBP

in the yeast TFIIA/TBP/DNA complex (20). The first β strand of the TFIIA small subunit, β1, contains the important residues that interact with TBP and shares a good homology with BRF. Therefore, this region of BRF, BRF 428-438, is likely to make similar interactions with TBP. The primary sequence of the C-terminal β strand of the TFIIA small subunit, β3, is quite similar to that of BRF, therefore it can be assumed that the C-terminus of BRF makes similar interactions to that of TFIIA. It is very interesting that although the N-terminal half of BRF shares a high homology with TFIIB, it does not compete with TFIIB for TBP/DNA binding (2). Yet, given the high sequence homology with TFIIB, the three-dimensional folding of the N-terminal half of BRF is very likely to be very similar to that of TFIIB (Figure 2.18).

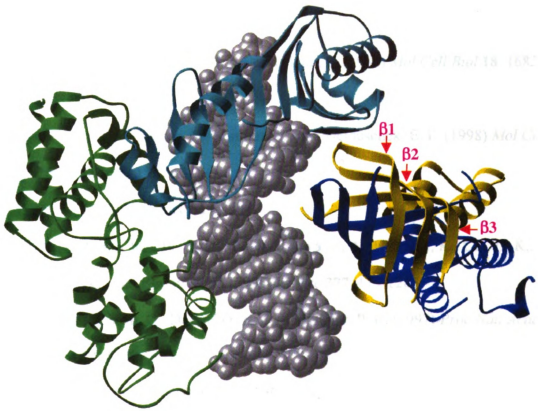


Figure 2.18 Modeled structure of the TFIIA/TFIIB/TBP/DNA complex. TFIIA small subunit is in yellow, TFIIA large subunit is blue, TBP is in cyan, TFIIB is in green, and DNA is in silver. The presumed folding of BRF 1-286 would be somewhat similar to TFIIB; the presumed protein-protein and protein-DNA contacts of BRF 428-438 would be similar to those of β1 of TFIIA small subunit; the presumed folding of BRF 564-585 would mimic β3 of TFIIA small subunit.

2.6 Literature cited

1. Kassavetis, G. A., Bardeleben, C., Kumar, A., Ramirez, E., and Geiduschek, E. P. (1997) *Mol Cell Biol* **17**, 5299-5306
2. Colbert, T., Lee, S., Schimmack, G., and Hahn, S. (1998) *Mol Cell Biol* **18**, 1682-1691
3. Kassavetis, G. A., Kumar, A., Ramirez, E., and Geiduschek, E. P. (1998) *Mol Cell Biol* **18**, 5587-5599
4. Hahn, S. (2000) Unpublished results
5. Nikolov, D. B., Chen, H., Halay, E. D., Usheva, A. A., Hisatake, K., Lee, D. K., Roeder, R. G., and Burley, S. K. (1995) *Nature* **377**, 119-128
6. Kosa, P. F., Ghosh, G., DeDecker, B. S., and Sigler, P. B. (1997) *Proc Natl Acad Sci U S A* **94**, 6042-6047
7. Tsai, F. T., and Sigler, P. B. (2000) *Embo J* **19**, 25-36
8. Littlefield, O., Korkhin, Y., and Sigler, P. B. (1999) *Proc Natl Acad Sci U S A* **96**, 13668-13673
9. Bagby, S., Kim, S., Maldonado, E., Tong, K. I., Reinberg, D., and Ikura, M. (1995) *Cell* **82**, 857-867
10. Geiger, J. H., Hahn, S., Lee, S., and Sigler, P. B. (1996) *Science* **272**, 830-836
11. Kim, J. L., Nikolov, D. B., and Burley, S. K. (1993) *Nature* **365**, 520-527
12. Reddy, P., and Hahn, S. (1991) *Cell* **61**, 349-357
13. Gill, G., and Tjian, R. (1991) *Cell* **65**, 333-340
14. Cormack, C., Strubin, M., Ponticelli, A. S., and Struhl, K. (1991) *Cell* **65**, 341-348

15. Persinger, J., Sengupta, S. M., and Bartholomew, B. (1999) *Mol Cell Biol* **19**, 5218-5234

16. Shen, Y., Kassavetis, G. A., Bryant, G. O., and Berk, A. J. (1998) *Mol Cell Biol* **18**, 1692-1700

17. Shah, S. M., Kumar, A., Geiduschek, E. P., and Kassavetis, G. A. (1999) *J Biol Chem* **274**, 28736-28744

18. Personal communication with Juo, S.

19. Tan, S., Hunziker, Y., Sargent, D. F., and Richmond, T. J. (1996) *Nature* **381**, 127-151

20. Unpublished 2.5 Å structure of yeast TFIIA/TBP/DNA complex.

CHAPTER III

X-RAY CRYSTALLOGRAPHIC STRUCTURAL STUDIES OF 1L-*MYO*-INOSITOL 1-PHOSPHATE SYNTHASE

This chapter will describe the structural investigation of 1L-*myo*-inositol 1-phosphate (MIP) synthase followed by a proposal of the mechanism. In this chapter, "MIP synthase" refers to MIP synthase from *S. cerevisiae* unless specified otherwise.

3.1 Theory

3.1.1 Structure determination from X-ray diffraction data

When a crystal is exposed to X-rays, constructive interference between rays scattered from successive planes in the crystal will only take place if the path difference between the rays is equivalent to an integral number of wavelengths. This is known as Bragg's Law:

$$2d\sin\theta = n\lambda \tag{3.1}$$

The structure factor for a particular reflection from a crystal is a complex quantity, which can be represented by its amplitude and phase:

$$\mathbf{F}(hkl) = \sum f_j e^{2\pi i(hx_j + ky_j + lz_j)} = F(hkl)e^{i\alpha(hkl)} \tag{3.2}$$

Where *F(hkl)* is the amplitude and $\alpha(hkl)$ is the phase. When the diffracted X-ray is recorded, all information on the phase is lost and only a measurement of the intensity of the diffracted beam is recorded. The intensity is given by

$$I(hkl) = \mathbf{F}(hkl) \bullet \mathbf{F}^*(hkl) = [\mathbf{F}(hkl)]^2 \tag{3.3}$$

The electron density can be computed by Fourier transform theory. The structure factor equation 3.2 can be rewritten in terms of a continuous summation over the volume of the unit cell.

$$F(hkl) = \sum f_j e^{2\pi i(hx_j + ky_j + lz_j)} = \sum f_j e^{2\pi i \mathbf{r_j} \bullet \mathbf{S}} = \int_{\text{unit cell}} \rho(\mathbf{r}) \mathbf{e}^{2\pi i \mathbf{r} \bullet \mathbf{S}} dv \tag{3.4}$$

where **S** is used to denote the position in diffraction space. By multiplying both sides by exp (-2πi**r'**•**S**) and integrating over the volume of diffraction space, it can be shown that

$$\rho(\mathbf{r}) = \int_{\text{diffraction space}} \mathbf{F(S)} e^{-2\pi i \mathbf{r} \bullet \mathbf{S}} dv_s \tag{3.5}$$

where dv_s is a small unit of volume in diffraction space. The integration can be replaced by a summation since **F(S)** is not continuous and is non-zero only at the reciprocal lattice points. Therefore,

$$\rho(xyz) = \frac{1}{V} \sum_h \sum_k \sum_l \mathbf{F}(hkl) e^{-2\pi i(hx+ky+lz)}$$

$$= \frac{1}{V} \sum_h \sum_k \sum_l F(hkl) e^{i\alpha(hkl)} e^{-2\pi i(hx+ky+lz)} \tag{3.6}$$

Since only intensities but not phases are measured in the recorded diffraction pattern, it is impossible to determine a structure directly from a recorded diffraction pattern; this is known as the phase problem in Crystallography.

There are four methods by which the phase problem can be solved. 1) The Patterson summation. This is a Fourier summation based on the experimentally observable $[F(hkl)]^2$. It is basically a vector map of the structure and is applied for molecules containing relatively few atoms. 2) Direct methods. In this method

mathematical relationships between the reflections can be used to provide phase information. 3) Heavy atom isomorphous replacement. In this method a heavy atom is introduced to a structure to provide phase information. 4) Anomalous scattering. In this method phase information is obtained from the information contained in the scattering by an atom whose natural absorption frequency is close to the wavelength of the incident radiation.

3.1.2 Molecular replacement

The molecular replacement method makes use of a known protein as an initial phasing model for the structure to be solved. The search and target molecules must have reasonable sequence identity (> 25 %) for there to be a good chance of success. Likewise, having a dataset with good completeness can be crucial. Generally there are two steps in molecular replacement and these are known as the rotation and translation functions. If successful, a preliminary model of the target structure will be obtained by correctly orienting and positioning the search molecule in the target cell. This solution can be optimized by rigid body refinement. Finally, the target structure can be put through cycles of map calculation, model fitting, and refinement, which help to reduce the bias introduced by the starting model. Even in the absence of a suitable search molecule, the self-rotation function can be used to determine the direction and nature of non-crystallographic symmetry elements. The problem is six dimensional because it involves three rotational and three translational parameters. This is simplified by separating the search into two stages namely the rotation and translation searches.

Rotation function

The rotation function should allow the orientation of the search molecule, which produces a maximal overlap with the target structure to be determined in the absence of any phases for the unknown structure. To do this, it compares the Patterson self-vectors of the known and unknown structures at different orientations of the search model. It should be noted that Patterson functions can be calculated from the amplitudes only and using Patterson space means that the translation vector is irrelevant since all intra-molecular vectors are shifted to the origin. The rotation function is usually calculated as a function of Eulerian angles: α, β, and γ. The molecule is placed in an orthogonal co-ordinate system with the axis of highest symmetry along Z (about α) to reduce the amount of computation.

Translation function

Having determined the angles α, β, and γ from the rotation search, the rotation matrix can be calculated and applied to the coordinates of the search molecule. The shift vector, which is required to position the search molecule correctly relative to the symmetry elements of the target molecule, can be determined by one of a number of translation searches. Patterson methods can be used to measure the overlap of the target Patterson cross-vectors with those calculated for the oriented search molecule as it ranges through the target cell. The simpler way to solve the translation problem is R-factor search. It involves the calculation of an R-factor as the search molecule and its symmetry mates are moved through the unit cell of the target crystal. The correct position should give the lowest R-factor. Other parameters, such as the correlation coefficient, can be

used to measure the agreement between the F_{obs}s and F_{cal}s as the search model is moved around. The R-factor is defined as follows.

$$R = \frac{\sum \left| |F_{obs}| - |F_{cal}| \right|}{\sum |F_{obs}|} \tag{3.7}$$

The molecular replacement method was used to determine the structure of the MIP synthase/NAD^+ complex from the diffraction data on a $P2_1$ crystal that will be discussed in this chapter.

3.1.3 Structure refinement

Once a structure is determined and the electron density map calculated, the model needs to be refined in order to find a best agreement of the model with both observed diffraction data and chemical restraints. The refinement in macromolecular crystallography is complicated not only because of the high dimensionality of the parameter space (typically at least three times the number of atoms in the model), but also because of the crystallographic phase problem described in section 3.1.1. Electron density maps computed from a combination of native crystal amplitudes and multiple isomorphous replacement (MIR) or multiwavelength anomalous diffraction (MAD) phases sometimes are insufficiently accurate to allow a complete and unambiguous tracing of the macromolecule. When structures are solved by molecular replacement described in section 3.1.2, the resulting electron density maps can be severely "model-biased" in that they seem to confirm the existence of the search model without providing clear evidence of actual differences between it and the true crystal structure. Therefore, initial models require extensive refinement.

A crystallographic refinement can be formulated as a search for the global minimum of the target function E (1),

$$E = E_{chem} + w_{xray} E_{xray} \tag{3.8}$$

E_{chem} is a function of all atomic positions and an empirical force field, describing both covalent and noncovalent interactions. E_{xray} describes the difference between observed and calculated diffraction data, and w_{xray} is a weight chosen to balance the forces arising from each term. In a macromolecular structure, the many local minima of the target function tend to defeat gradient descent optimization techniques such as conjugate gradient or least-squares methods (2). These methods are not capable of shifting the atomic coordinates enough to correct errors in the initial model. Simulated annealing is an optimization technique well suited to the multiple-minima characteristic of crystallographic refinement.

Simulated annealing refinement

Simulated annealing refinement can overcome barriers between local minima, and thus can explore a greater volume of the parameter space to find "deeper" minima. Annealing refers to a physical process wherein a solid is heated until all particles randomly arrange themselves in a viscous liquid phase, and then cooled slowly so that all particles arrange themselves in the lowest energy state. By formally defining the target E in Eq. 3.9 to be the equivalent of the potential energy of the system, one can simulate the annealing process (3). Simulated annealing is an approximation algorithm; there is no guarantee that it will find the global minimum except in the asymptotic limit of an infinite search. Compared to gradient descent methods where search directions must

follow the gradient, simulated annealing achieves more optimal solutions by allowing motion against the gradient. The likelihood of counter gradient motion is determined by a control parameter "temperature", the higher the temperature the more likely the optimization can overcome the barriers.

Simulated annealing requires a generation mechanism to create a Boltzmann distribution at a given temperature T,

$$B(q_1,.....,q_i) = e^{-\frac{E(q_1,.....q_i)}{k_b T}} \tag{3.9}$$

where E is the target function given by equation 3.8, k_b is the Boltzmann constant, and q_1, $.....,q_i$ are adjustable parameters such as atomic coordinates. Simulated annealing also requires an annealing schedule, which is a sequence of temperatures $T_1 \geq T_2 \geq \ldots \geq T_l$ at which the Boltzmann distribution is computed. Monte Carlo (4) and molecular dynamics (5) simulations are the two most commonly used generation mechanisms.

Simulated annealing has improved the efficiency of crystallographic refinement significantly; however, simulated annealing alone is still insufficient to refine a crystal structure without human intervention by manual refitting of the model to electron density maps using interactive graphics programs (6).

For all structures included in this chapter, structure refinements were done by simulated annealing methods in combination with manual refitting of the models.

3.2 Experimental procedures

3.2.1 Protein over-expression and purification

Over-expression

S. cerevisiae MIP synthase was over-expressed in *E.coli* BL21 (DE3) cells. The LB medium containing 100 mg/L carbenicillin was inoculated with a single colony of *E.coli* strain BL21 (DE3) bacteria containing MIP synthase expression plasmid. Cells were grown at 37 °C with rapid shaking for 18 hours or until saturated. One liter of LB/carbenicillin was inoculated with 25 ml of saturated culture and growth was maintained at 37 °C with rapid shaking until the O.D. $_{600\ nm}$ reached 0.6. The inducer, IPTG, was then added to a final concentration of 0.25 mM, and growth continued for another 3 hours. The cells were harvested by centrifugation at 5000 rpm.

Purification

- Cells were re-suspended in the lysis buffer (Tris 20 mM, NH_4Cl 10 mM, DTT 10 mM, pH 7.7, protease inhibitor cocktail, complete EDTA-free) and lysed by sonication. Cell debris was spun down by centrifugation at 15000 rpm for 40 minutes.
- The supernatant was loaded onto a Phast-Q ion exchange column and washed with the lysis buffer until the $OD_{280\ nm}$ of the eluent is lower than 0.5. A gradient of NH_4Cl (0.01-0.75 M) in the same buffer (20 mM Tris, pH 7.7, 10 mM DTT, protease inhibitor cocktail, complete EDTA-free) was then applied onto the column.

- Eluted fractions containing MIP synthase were then pooled together and passed directly through a Blue A affinity column to remove glucose 6-phosphate dehydrogenase.
- Activated charcoal was then added to the sample in a ratio of 1g charcoal to 5mg protein. This treatment was performed at 4 °C for 30 minutes. The cofactor-bound charcoal was removed by centrifugation and filtration.
- Protein samples were loaded onto another ion exchange column, an FPLC Source-Q column, a gradient of NH_4Cl (0.01-0.3 M) in the same buffer (20 mM Tris, pH 7.7, 10 mM DTT, protease inhibitor cocktail, complete EDTA-free) was applied. Fractions containing pure MIP synthase elutes at 0.18 M NH_4Cl. Fractions were further analyzed by SDS-PAGE for homogeneity, and were concentrated to 8mg/ml for crystallization setups.

3.2.2 Crystallization

Crystallization of *apo* MIP synthase

Crystals of *apo* MIP synthase were grown by the hanging drop vapor diffusion method (Figure 2.4). If good nucleations were not observed, micro or macro-seeding methods were applied in order to get diffraction quality crystals. Crystals with the dimensions 0.5mm x 0.5 mm x 0.5 mm grew in a condition previously reported (7).

Crystallization of MIP synthase in complex with cofactors and various ligands

Purified MIP synthase was incubated with cofactors and various concentrations of different inhibitors at 37 °C for 10 minutes before co-crystallization setups.

In cases where co-crystallization setups of MIP synthase with various cofactors and ligands did not produce suitable crystals for data collection, a soaking approach was applied. Pre-grown *apo* MIP synthase crystals were soaked into stabilizers containing various concentrations of cofactors and ligands. Table 3.1 lists all crystals of MIP synthase with cofactors and ligands obtained for data collection by both co-crystallization and soaking methods. For each category, multiple data sets were collected on crystals containing various concentrations of cofactors and ligands at various pH's, details will be described in later sections.

Table 3.1 Crystals of MIP synthase with cofactors and ligands

	Co-factor	Ligands	Co-crystallized	Soaked	Space group	Resolution*
1	None	None	N/A	N/A	C2	2.6 Å
2	NAD^+	None		Yes	$P2_1$	1.9 Å
3	NAD^+	None		Yes	C2	2.4 Å
4	NADH	None		Yes	C2	1.7 Å
5	NAD^+	2-deoxy-D-glucitol 6-(*E*)-vinylhomophosphonate	Yes		C2	2.1 Å
6	NAD^+	2-deoxy-D-glucitol 6-(*E*)-vinylhomophosphonate	Yes		$C222_1$	2.3 Å
7	NAD^+	2-deoxy-D-glucitol 6-(*E*)-vinylhomophosphonate		Yes	C2	1.9 Å
8	NADH	2-deoxy-D-glucitol 6-(*E*)-vinylhomophosphonate	Yes		C2	2.2 Å
9	NADH	2-deoxy-D-glucitol 6-(*E*)-vinylhomophosphonate		Yes	C2	1.8 Å
10	None	2-deoxy-D-glucitol 6-(*E*)-vinylhomophosphonate		Yes	C2	2.0 Å
11	NAD^+	2-deoxy-D-glucose 6-phosphate	Yes		C2	2.0 Å
12	NAD^+	2-deoxy-D-glucose 6-phosphate		Yes	C2	2.1 Å

Table 3.1 (cont'd)

	Co-factor	Ligands	Co-crystallized	Soaked	Space group	Resolution*
13	NADH	2-deoxy-D-glucose 6-phosphate	Yes		C2	2.0 Å
14	NADH	2-deoxy-D-glucose 6-phosphate		Yes	C2	1.9 Å
15	NAD^+	D-glucitol 6-phosphate	Yes		C2	2.0 Å
16	NADH	D-glucitol 6-phosphate		Yes	C2	2.1 Å
17	NAD^+	2-deoxy-D-glucitol 6-phosphate		Yes	C2	1.9 Å
18	NADH	2-deoxy-D-glucitol 6-phosphate	Yes	Yes	C2	2.0 Å
19	NAD^+	Valproate	Yes		C2	2.0 Å
20	NAD^+	Valproate	Yes		R32	2.4 Å
21	NAD^+	Valproate			P2	3.9 Å
22	NAD^+	Valproate	Yes	Yes	C2	2.0 Å
23	NADH	Valproate			C2	2.6 Å
24	NADH	Valproate	Yes	Yes	C2	2.4 Å
25	NAD^+	DL-*myo*-inositol 1-phosphate			C2	2.0 Å

Table 3.1 (cont'd)

	Co-factor	Ligands	Co-crystallized	Soaked	Space group	Resolution*
26	NADH	D-glucose 6-phosphate		Yes	C2	2.0 Å
27	NAD^+	NH_4^+		Yes	C2	3.5 Å
28	NAD^+	Rb^+		Yes	C2	1.8 Å
29	NADH	EDTA		Yes	C2	2.1 Å

* For each category, multiple datasets were collected, listed is the highest resolution data set among each category.

3.3 Data collection and refinement

For simplicity and clarity, only the datasets that produced the structures to be discussed in this chapter are described below.

3.3.1 The *apo* MIP synthase

Crystals were transferred to a cryoprotecting stabilizer (5 % PEG 8000, 0.1 M CH_3COONa, pH 4.5, 30 % glycerol) before flash-frozen for data collection; this method of cryoprotection was applied to all the data collection to be described in this chapter. Data were collected at Michigan State University Macromolecular X-ray diffraction facility. Data reduction and scaling were performed using Denzo and SCALEPACK respectively (8). The structure of MIP synthase/NAD^+/2-deoxy-D-glucitol 6-phosphate(9) was used as the phasing model, the map was traced and refinement was done using TURBO-FRODO (6,10) and CNS (11,12) respectively. The data collection and refinement statistics are listed in Table 3.2. The final refinement model consisted of residues 10-350, 376-533 in both molecules in the asymmetric unit. The structure also included 133 water molecules. Figure 3.1 represents the Ramachandran plot of the *apo* MIP synthase structure; only 3 residues out of 997 are in disallowed regions.

Table 3.2 Data collection and refinement statistics of the *apo* MIP synthase[a]

Space group	C2
Resolution range (Å)	40-2.6 (2.69-2.6)
Unit cell	a=153.543 Å, b=97.062 Å, c=122.064 Å, β=125.716°
Number of total reflections	44160
Completeness (%)	98.3 (96.8)
R_{sym} (%)[b]	12.0 (60.5)
I/σ	11.6 (1.9)
R factor (%)[c]	19.0
R_{free} (%)[d]	27.8
Average B-factor (Å^2)	47
RMSD of bond lengths (Å)	0.0072
RMSD of bond angles (°)	1.39

[a] The parentheses denote those values for the last resolution shell.

[b] $R_{sym} = \frac{\sum \| I | - \langle I \rangle |}{\sum |I|}$ where I is the observed intensity and <I> is the average intensity obtained from multiple observations of symmetry-related reflections.

[c] $R = \frac{\sum \| F_{obs} | - | F_{cal} \|}{\sum |F_{obs}|}$

[d] $R_{free} = \frac{\sum \| F_{obs} | - | F_{cal} \|}{\sum |F_{obs}|}$, where reflections belong to a test set of 10 % randomly selected data.

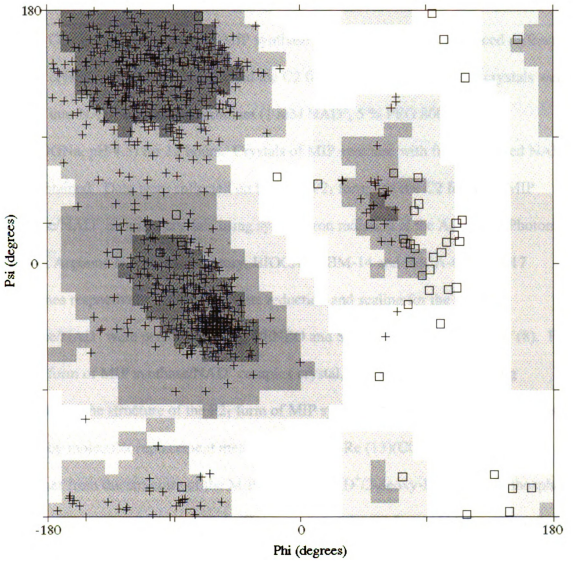


Figure 3.1 Ramachandran plot of the *apo* MIP synthase structure.

3.3.2 The MIP synthase/NAD^+ complex

Co-crystallization trials of MIP synthase and NAD^+ have not produced diffraction quality crystals. Both the $P2_1$ form and the C2 forms of *apo* MIP synthase crystals were soaked into NAD^+-containing stabilizer (1 mM NAD^+, 5 % PEG 8000, 0.1 M CH_3COONa, pH 4.5) for 12 hours. Crystals of MIP synthase with fully occupied NAD^+ were obtained. Data were collected on both the $P2_1$ form and the C2 forms of MIP synthase/NAD^+ complex crystals using synchrotron radiation at the Advanced Photon Source, Argonne National Laboratory, BIOCARS BM-14 and IMCA-CAT ID-17 beamlines respectively. Diffraction data reduction and scaling for the $P2_1$ MIP synthase/NAD^+ were performed using DENZO and SCALEPACK respectively (8). For the C2 form of MIP synthase/NAD^+ complex crystal, data were processed using HKL2000. The structure of the $P2_1$ form of MIP synthase/NAD^+ complex crystal was solved by molecular replacement method using AMoRe (13)(CCP4 package). A monomer from the structure of the MIP synthase/NAD^+/2-deoxy-D-glucitol 6-phosphate complex (9) was used as the model. The structure of the C2 form of MIP synthase/NAD^+ crystal was solved using the same structure (9) as the initial phasing model. Maps were traced using TURBO-FRODO (6,10) and refinements were performed using CNS(11). Data collection and refinement statistics are tabulated in Table 3.3.

The contents of each molecule in the asymmetric unit of both crystal forms are tabulated in Table 3.4. Figure 3.2 represents the Ramachandran plot of the $P2_1$ MIP synthase/NAD^+ structure, only 6 out of 2001 residues are in the disallowed region; Figure 3.3 represents the Ramachandran plot of the C2 MIP synthase/NAD^+ structure, only 3 out of 1005 residues are in disallowed regions.

Table 3.3 Data collection and refinement statistics of the MIP synthase/NAD^+ complexes[a]

Space group	$P2_1$	C2
Resolution range (Å)	40-1.9 (1.97-1.9)	50-2.4 (2.49-2.4)
Unit cell	a=90.794 Å, b=185.581 Å, c=94 Å, β=114.77°	a=152.384 Å, b=97.344 Å, c=122.887 Å, β=126.526°
Number of total reflections	214444	56444
Completeness (%)	99.7 (100)	96.6 (99.2)
R_{sym} (%)[b]	6.4 (42.2)	9.9 (45.3)
I/σ	23.5 (2.5)	28.6 (3.0)
R factor (%)[c]	20.9	20.9
R_{free} (%)[d]	27.8	28.8
Average B-factor ($Å^2$)	49.6	58.8
RMSD of bond lengths (Å)	0.007	0.0078
RMSD of bond angles (°)	1.26	1.36

[a] The parentheses denote those values for the last resolution shell.

[b] $R_{sym} = \frac{\sum \|I| - \langle I \rangle|}{\sum |I|}$ where I is the observed intensity and <I> is the average intensity obtained from multiple observations of symmetry-related reflections.

[c] $R = \frac{\sum \|F_{obs}| - |F_{cal}\|}{\sum |F_{obs}|}$

[d] $R_{free} = \frac{\sum \|F_{obs}| - |F_{cal}\|}{\sum |F_{obs}|}$, where reflections belong to a test set of 10 % randomly selected data.

Table 3.4 Contents of the MIP synthase/NAD^+ complex structures

Space group	Molecule	Residues	Number of water molecules
$P2_1$	A	10-361, 380-533	1529
	B	10-361, 380-464, 472-533	
	C	10-464, 472-533	
	D	10-361, 380-390, 410-533	
C2	A	10-362, 376-533	160
	B	10-362, 376-464, 472-533	

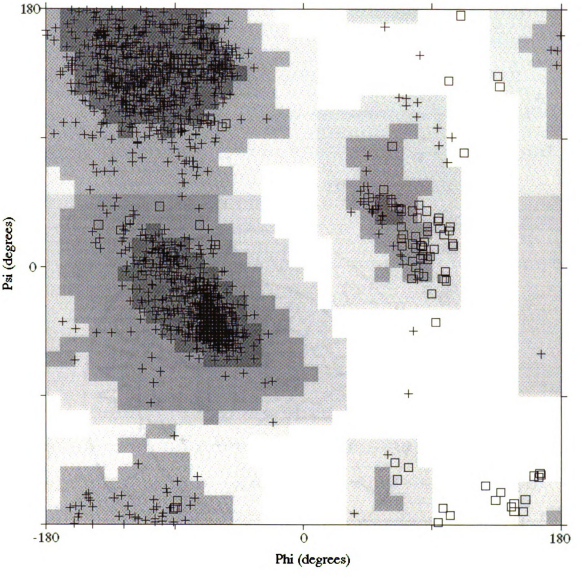


Figure 3.2 Ramachandran plot of the $P2_1$crystal form of the MIP synthase/NAD^+ complex structure.

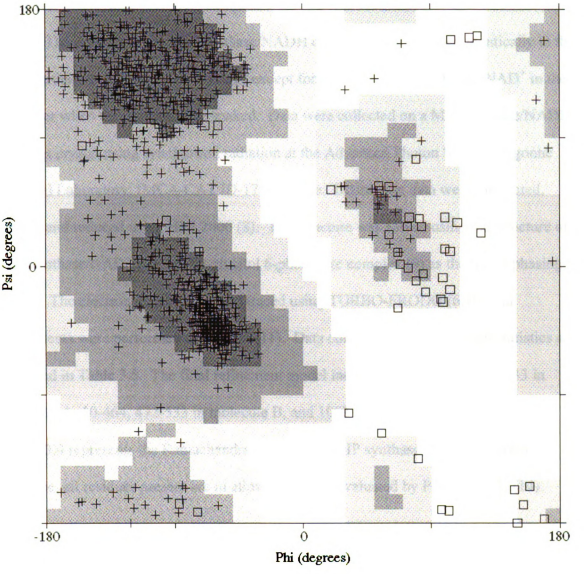


Figure 3.3 Ramachandran plot of the C2 crystal form of the MIP synthase/NAD^+ complex structure.

3.3.3 The MIP synthase/NADH complex

The crystals of the MIP synthase/NADH complex were obtained identically to the MIP synthase/NAD^+ complex crystals, except for the change of NADH for NAD^+ in the stabilizer when the crystals were soaked. Data were collected on a MIP synthase/NADH complex crystal using synchrotron radiation at the Advanced Photon Source, Argonne National Laboratory, IMCA-CAT ID-17 beamline. Diffraction data were integrated, scaled, and reduced using HKL2000 (8). The structure was solved using the structure of MIP synthase/NAD^+/2-deoxy-D-glucitol 6-phosphate complex (9) as the initial phasing model. The electron density map was traced using TURBO-FRODO (6,10) and refinements were performed using CNS(11). Data collection and refinement statistics are tabulated in Table 3.5. The final refinement model included residues from 10-533 in molecule A, 10-464, 472-533 in molecule B, and 1068 water molecules.

Figure 3.4 represents the Ramachandran plot of the MIP synthase/NADH complex structure, all residues present are in allowed regions evaluated by PROCHECK (14).

Table 3.5 Data collection and refinement statistics of the MIP synthase/NADH complex[a]

Space group	C2
Resolution range (Å)	50-1.7 (1.8-1.7)
Unit cell	a=149.86 Å, b=99.851 Å, c=122.57 Å, β=126.64°
Number of total reflections	160868
Completeness (%)	84.4 (69.5)
R_{sym} (%)[b]	7.4 (25.6)
I/σ	36.8 (4.4)
R factor (%)[c]	16.5
R_{free} (%)[d]	19
Average B-factor (Å^2)	22
RMSD of bond lengths (Å)	0.0053
RMSD of bond angles (°)	1.35

[a] The parentheses denote those values for the last resolution shell.

[b] $R_{sym} = \frac{\sum \| I \| - \langle I \rangle \|}{\sum |I|}$ where I is the observed intensity and <I> is the average intensity obtained from multiple observations of symmetry-related reflections.

[c] $R = \frac{\sum \| F_{obs} | - | F_{cal} \|}{\sum |F_{obs}|}$

[d] $R_{free} = \frac{\sum \| F_{obs} | - | F_{cal} \|}{\sum |F_{obs}|}$, where reflections belong to a test set of 10 % randomly selected data.

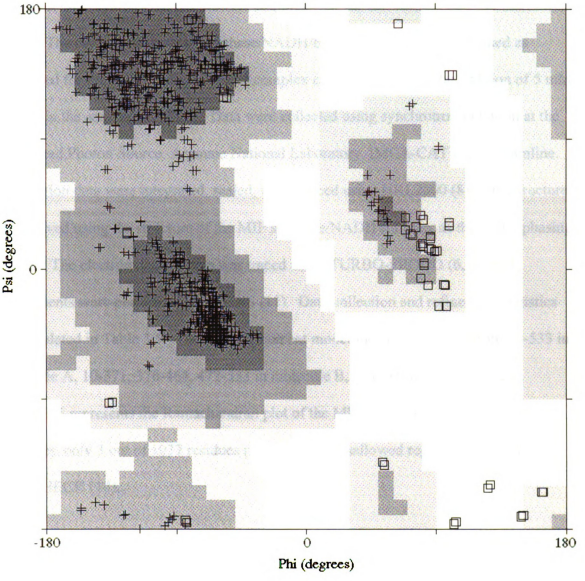


Figure 3.4 Ramachandran plot of the MIP synthase/NADH complex structure.

3.3.4 The MIP synthase/NADH/EDTA complex

The crystals of the MIP synthase/NADH/EDTA complex were obtained as described for the MIP synthase/NADH complex crystals except for the addition of 5 mM EDTA in the soaking stabilizer. Data were collected using synchrotron radiation at the Advanced Photon Source, Argonne National Laboratory, IMCA-CAT ID-17 beamline. Diffraction data were integrated, scaled, and reduced using HKL2000 (8). The structure was solved using the structure of the MIP synthase/NADH complex as the initial phasing model. The electron density map was traced using TURBO-FRODO (6,10) and refinements were performed using CNS (11). Data collection and refinement statistics are tabulated in Table 3.5. The final refinement model included residues from 10-533 in molecule A, 10-371, 376-464, 472-533 in molecule B, and 180 water molecules. Figure 3.5 represents the Ramachandran plot of the MIP synthase/NADH/EDTA structure, only 3 out of 1037 residues present are in disallowed regions evaluated by PROCHECK (14).

Table 3.6 Data collection and refinement statistics of the MIP synthase/NADH/EDTA complex[a]

Space group	C2
Resolution range (Å)	50-2.1 (2.18-2.1)
Unit cell	a=151.98 Å, b=97.61 Å, c=121.72 Å, β=126.15°
Number of total reflections	161339
Completeness (%)	99.8(99.6)
R_{sym} (%)[b]	6.7 (42.2)
I/σ	31.8 (3.2)
R factor (%)[c]	20.9
R_{free} (%)[d]	26.1
Average B-factor (Å^2)	43.4
RMSD of bond lengths (Å)	0.0072
RMSD of bond angles (°)	1.30

[a] The parentheses denote those values for the last resolution shell.

[b] $R_{sym} = \frac{\sum \| I \| - \langle I \rangle \|}{\sum |I|}$ where I is the observed intensity and <I> is the average intensity obtained from multiple observations of symmetry-related reflections.

[c] $R = \frac{\sum \| F_{obs} | - | F_{cal} \|}{\sum |F_{obs}|}$

[d] $R_{free} = \frac{\sum \| F_{obs} | - | F_{cal} \|}{\sum |F_{obs}|}$, where reflections belong to a test set of 10 % randomly selected data.

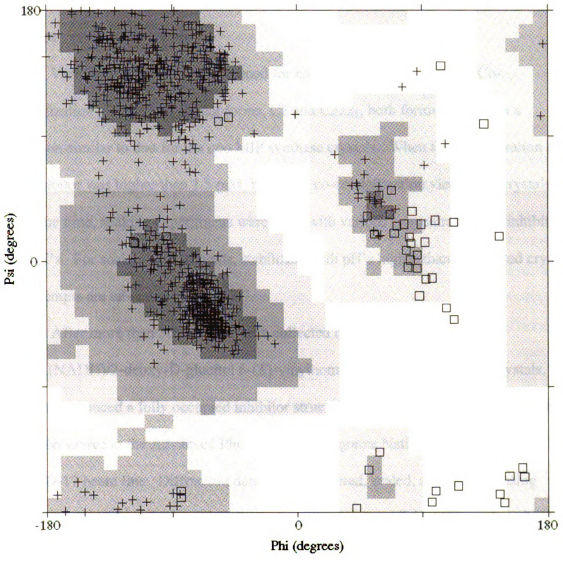


Figure 3.5 Ramachandran plot of the MIP synthase/NADH/EDTA complex structure.

3.3.5 The MIP synthase/NAD^+(NADH)/2-deoxy-D-glucitol 6-(*E*)-vinylhomophosphonate complex

Various conditions were screened for co-crystallization attempts. Co-crystallization yielded two crystal forms, C2 and $C222_1$, both forms grew from a condition similar to that for the *apo* MIP synthase crystals. When the concentration of the inhibitor was higher than 1.5 mM, however, co-crystallization yielded no crystals. At the same time, soaking experiments were done with various concentrations of inhibitors and pH's. For soaking experiments, stabilizers with pH's higher than 5.5 melted crystals. All attempts are tabulated in Table 3.7.

After more than 30 data sets were collected on the MIP synthase /NAD^+(NADH)/2-deoxy-D-glucitol 6-(*E*)-vinylhomophosphonate complex crystals, the data that produced a fully occupied inhibitor structure were collected using synchrotron radiation source at the Advanced Photon Source, Argonne National Laboratory, IMCA-CAT ID-17 beam line. Diffraction data were integrated, scaled, and reduced using HKL2000 (8). The structure was solved using the structure of the MIP synthase/NADH complex as the initial phasing model. The electron density map was traced using TURBO-FRODO (6,10) and refinements were performed using CNS(11). Data collection and refinement statistics are tabulated in Table 3.8. The final refinement model included residues from 10-533 in molecule A, 10-464, 472-533 in molecule B, and 400 water molecules. Figure 3.6 represents the Ramachandran plot of the MIP synthase/NADH structure/2-deoxy-D-glucitol 6-*(E)*-vinylhomophosphonate structure, only 3 out of 1041 residues present are in disallowed region evaluated by PROCHECK (14).

Table 3.7 Data collected on the MIP synthase/NAD^+(NADH)/2-deoxy-D-glucitol 6-(*E*)-vinylhomophosphonate complex crystals

Co-factor	Concentration of the inhibitor (mM)	pH	Method	Space group	Resolution[a](Å)	Observation
NAD^+	0.17	4.5	Co-crystallized	C2	3.1	Not occupied
NAD^+	0.34	4.5	Co-crystallized	C2	2.7	Not occupied
NAD^+	0.68	4.5	Co-crystallized	C2	3.2	Not occupied
NAD^+	1	4.5	Co-crystallized	C2	2.1	Not occupied
NADH	1	4.5	Co-crystallized	C2	3.0	Not occupied
NAD^+	1.5	4.5	Co-crystallized	$C222_1$	2.3	Not occupied
NAD^+	2	4.5	Soaked for 12 hrs	C2	2.6	Not occupied
NADH	2	4.5	Soaked for 24 hrs	C2	1.8	Occupied[b]
NAD^+	5	4.5	Soaked for 12 hrs	C2	2.7	Not occupied
NADH	5	4.5	Soaked for 7 hrs	C2	2.6	Not occupied
NADH	6	4.5	Soaked for 12 hrs	C2	2.6	Not occupied
NAD^+	7.5	4.5	Soaked for 12 hrs	C2	2.1	Not occupied
NAD^+	10	4.5	Soaked for 6 hrs	C2	2.5	Not occupied
NAD^+	13.5	4.5	Soaked for 24 hrs	C2	1.9	Partially occupied
NADH	13.5	4.5	Soaked for 24 hrs	C2	2.0	Partially occupied
NAD^+	13.5	**5.5**	Soaked for 24 hrs	C2	2.0	Fully occupied

[a]The highest resolution data set among multiple data sets in each experiment.

[b]Occupied with two small molecules that were present in the NADH-bound structure but not the inhibitor.

Table 3.8 Data collection and refinement statistics of the MIP synthase/NAD^+/2-deoxy-D-glucitol 6-(*E*)-vinylhomophosphonate complex[a]

Space group	C2
Resolution range (Å)	50-2.0 (2.1-2.0)
Unit cell	a=151.81 Å, b=97.782 Å, c=122.293 Å, β=126.3°
Number of total reflections	110451
Completeness (%)	98.2(98.6)
R_{sym} (%)[b]	7.4 (50)
I/σ	17.7 (2.46)
R factor (%)[c]	18.8
R_{free} (%)[d]	24.4
Average B-factor ($Å^2$)	37.3
RMSD of bond lengths (Å)	0.0067
RMSD of bond angles (°)	1.29

[a] The parentheses denote those values for the last resolution shell.

[b] $R_{sym} = \frac{\sum \| I \| - \langle I \rangle \|}{\sum |I|}$ where I is the observed intensity and <I> is the average intensity obtained from multiple observations of symmetry-related reflections.

[c] $R = \frac{\sum \| F_{obs} | - | F_{cal} \|}{\sum |F_{obs}|}$

[d] $R_{free} = \frac{\sum \| F_{obs} | - | F_{cal} \|}{\sum |F_{obs}|}$, where reflections belong to a test set of 10 % randomly selected data.

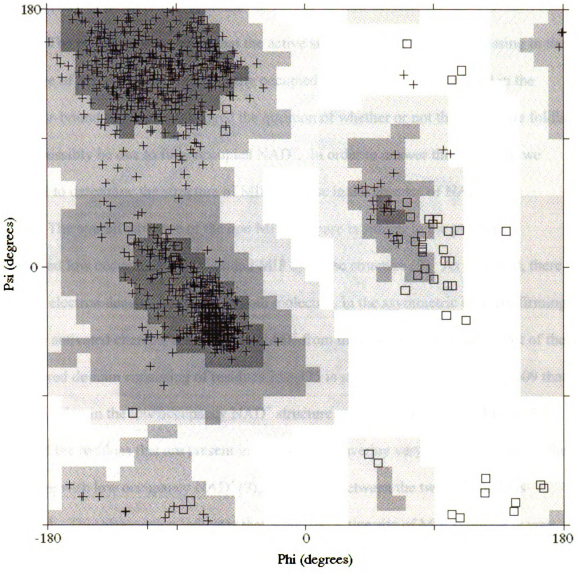


Figure 3.6 Ramachandran plot of the MIP synthase/NAD^+/2-deoxy-D-glucitol 6-(*E*)-vinylhomophosphonate structure.

3.4 Structure of the *apo* MIP synthase

The previous observation that the active site residues 351-409 are missing in the structure of MIP synthase with partially occupied NAD^+ but became ordered in the inhibitor-bound structure (9) begged the question of whether or not the active site folding could possibly be due to fully occupied NAD^+. In order to answer this question, we decided to determine the structure of MIP synthase in the absence of NAD^+.

The overall structure of the *apo* MIP synthase is similar to the previously published low occupancy NAD^+-bound MIP synthase structure (9). As expected, there was no electron density for NAD^+ in both molecules in the asymmetric unit, confirming that the activated charcoal removed the NAD^+ from the enzyme. Although a part of the disordered domain consisting of residues 351-375 is still missing, residues 376-409 that were missing in the low occupancy NAD^+ structure are now well ordered (Figure 3.7). Most of the residues that are present in this structure overlay very well with those of the structure with low occupancy NAD^+ (9), the RMSD between the two structures is 0.8648Å. This observation indicates that while the active site of MIP synthase is very flexible, the NAD^+ binding domain is very rigid.

However, the loop encompassing residues 191-198 differs significantly between the *apo* MIP synthase and the previously published low occupancy NAD^+-bound structure (Figure 3.8). The location of this loop in the *apo* structure overlays well with the structure of the MIP synthase/NAD^+/2-deoxy-D-glucitol 6-phosphate complex rather than the low occupancy NAD^+-bound structure, inconsistent with the previous conclusion that the folding of the active site forces this loop to flip out (9).

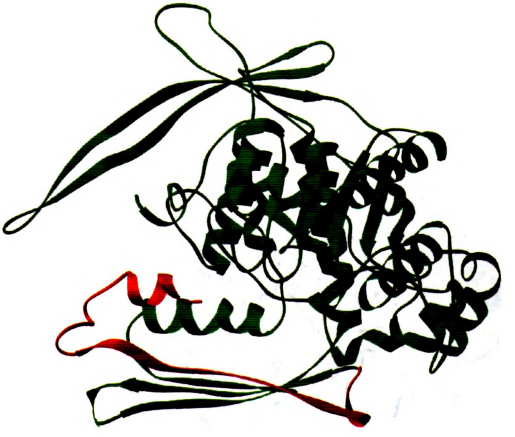


Figure 3.7 Structure of the *apo* MIP synthase. The region that was disordered in the previous low occupancy NAD^+- bound structure (9) is colored red.

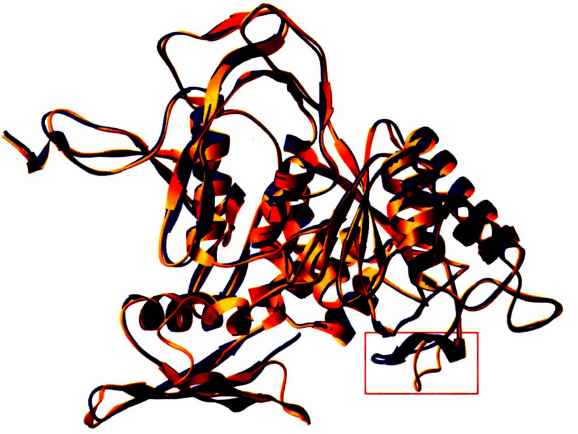


Figure 3.8 Superposition of the structures of *apo* MIP synthase in gold and the low occupancy NAD^+-bound MIP synthase (9) in blue. Boxed in red are residues 191-198.

3.5 Structure of the MIP synthase/NAD^+ complex

In order to investigate any possible conformational changes in the MIP synthase active site and the Rossmann fold domain in the presence of fully occupied NAD^+, and more specifically, in order to answer the question of whether or not the active site folding that was observed in the structure of the MIP synthase/NAD^+/2-deoxy-D-glucitol 6-phosphate complex could be due to the fully occupied NAD^+, the three-dimensional structure of the MIP synthase/NAD^+ complex was elucidated. This structure represents the enzyme at its state when the substrate is not yet bound or the product is already produced and released from the enzyme active site.

3.5.1 Structure of the MIP synthase/NAD^+ complex tetramer

The $P2_1$ form contains an entire tetramer of MIP synthase in the asymmetric unit as it is in solution (Figure 3.9). Molecule A and B, C and D are related by a noncrystallograhic two-fold rotation axis. The AB and CD dimers are related by another noncrystallographic two-fold axis that is perpendicular to the first two-fold axis. The interface between two monomers buries a surface area of 9539.66 $Å^2$, while the tetramerization interface between the AB and CD dimers buries an additional 6096.4 $Å^2$ of surface area.

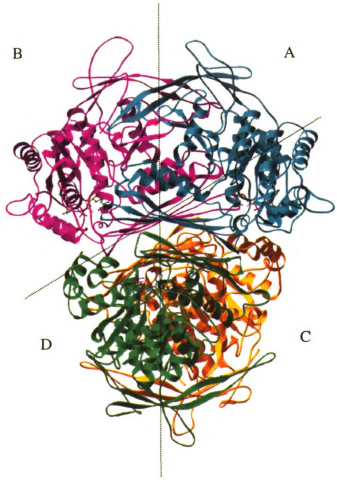


Figure 3.9 The entire tetramer of the $P2_1$ structure of the MIP synthase/NAD^+ complex. Molecule A is colored in cyan, B is in magenta, C is in gold, and D is in green.

3.5.2 Structure of the MIP synthase/NAD^+ complex monomer

In the $P2_1$ crystal structure of the MIP synthase/NAD^+ complex, molecules A, B, and D are similar in that they are all missing residues 362-380 as shown in Figure 3.10. Molecule C in the $P2_1$ structure is an exception in that the entire active site is folded. In the C2 structure, both of the molecules in the asymmetric unit are missing residues 363-375. In addition, molecules B and C in the $P2_1$ structure and molecule B in the C2 structure are missing residues 465-471. The ordered part of the structure in each monomer is overall similar to that of *apo* MIP synthase, the RMSD between the two structures is 1.07 Å. On the other hand, α13, the helix encompassing residues 352-361, which was missing in the *apo* structure became ordered upon NAD^+ binding. Residues N355 and D356 of this helix make direct interactions with NAD^+. In all molecules except molecule C in the $P2_1$ structure, the strand that was missing in the previously published structure of the low occupancy NAD^+-bound structure is less well ordered than in the *apo* structure.

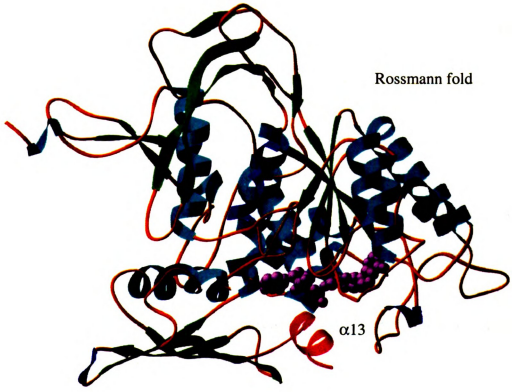


Figure 3.10 A monomer of the MIP synthase/NAD^+ complex. The helix in red is the newly ordered region upon NAD^+ binding; NAD^+ is in lavender. The rest of the α helices are cyan; all β strands are green; the loops are brown.

3.5.3 NAD^+ binding

In both the $P2_1$ and the C2 crystal forms of the MIP synthase/NAD^+ complex structures, the NAD^+ molecules are fully occupied. The nicotinamide portion of the molecule has higher B factor values indicating that this part is less well ordered overall than the rest of NAD^+.

MIP synthase binds NAD^+ in a way similar to that of other Rossmann fold containing NAD^+ binding proteins. The NAD^+ molecule runs across the bottom of the Rossmann fold domain. The Rossmann fold domain of MIP synthase contains the core topology region, $\beta_3\alpha_1\beta_8\alpha_8\beta_{11}\alpha_{10}\beta_{12}\alpha_{11}\beta_{13}$, the structural motif common in many NAD^+ binding proteins. Numerous interactions were observed between NAD^+ and MIP synthase as delineated in Table 3.9. Overall 1054 $Å^2$ of surface area was buried by the interaction of MIP synthase with NAD^+. The consensus phosphate binding sequence GXGGXXG motif is also present in MIP synthase as the loop connecting β_3 and α_1, and the N-terminus of α_1 (residues 72-78, Figure 3.11). The interaction of this motif and NAD^+ is similar to that of other NAD^+ binding proteins. The first conserved glycine, G72, allows for a tight turn of the main chain from β_3 to α_1, and this is important for positioning the second and the third glycines, G73 and G74 respectively. Because of its missing side chain, G74 enables close contact of the main chain to the pyrophosphate of NAD^+. The last glycine in this motif, G78, is important for the close packing of β_3 and α_1. The main chain nitrogen atoms as well as the side chain nitrogen atoms of N76 and N77 within this motif make hydrogen bond interactions with the pyrophosphate oxygen atoms as shown in Figure 3.11. It is important to note that since this GXGGXXG motif

is not conserved in *A. fulgiudus* MIP synthase, NAD^+-binding could be significantly different from that of *S. cerevisiae* MIP synthase.

Unique in MIP synthase are three insertions. The first insertion encompassing residues 93-140 contributes to the dimerization via van der Waals interactions and several tight hydrogen bonds (Table 3.10). The second insertion encompasses residues 149-215, this insertion completely surrounds the adenine portion of NAD^+ making strong hydrophobic interactions as well as hydrogen bond interactions with NAD^+ (Figure 3.12). These interactions seem to contribute strongly to the high binding affinity of NAD^+ to MIP synthase (K_m=0.017 mM). The third insertion encompassing residues 247-276 is involved in crystal packing as shown in Figure 3.13. This region interacts with the loop of a neighboring tetramer encompassing residues 464-472.

With the active site residues 362-379 being disordered, the nicotinamide portion of NAD^+ is completely exposed to the surface. In fact, there is a very tight hydrogen bond with an interatomic distance of 2.5 Å that stabilizes this conformation as shown in Figure 3.14.

Table 3.9 Interactions between the MIP synthase and $NAD^{+\#}$

Portion of NAD^+	van der Waals contacts	Hydrogen bonds (Distances in Å)	Portion of NAD^+	van der Waals contacts	Hydrogen bonds (Distances in Å)
Adenine	I71 G72 W147 D148 I149 S184 I185 W243 A245 N246 P277	AN_1:S184 OG (2.71) AN_6:I185 O (2.96)	Phospho diester	G75 N76 N77 N246 N355 D356	AO_2:N76 N(3.22) NO_2:N77 N(2.95) AO_1:N355ND_2(2.7)
			Nicotinamide Ribose	N77 T244 A245 N246 T247 G295 S296 D356	NO_3*:A245 O(2.9) NO_2*:T247OG_1(3.2) NO_3*:T247 N(3.01)
Adenine Ribose	G72 G74 G75 N76 D148 I149 R198 T244 A245 N246	AO_3*:D148 OD_1(2.82) AO_2*:D148 OD_2(2.88) AO_2*:R198 NE(3.21)	Nicotinamide	N77 G295 D320 L321 D438 A442	NN_7:N77 OD_1(3.16)

Listed are residues making van der Waals interactions within a 4.0 Å cutoff.

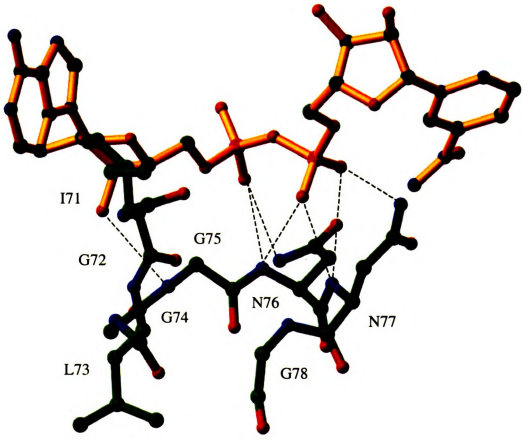


Figure 3.11 The GXGGXXG phosphate-binding motif within the MIP synthase Rossmann fold domain.

Table 3.10 Dimerization interactions* by the first insertion (93-140)

A	B	Hydrogen bonds (Distances in Å)	A	B	Hydrogen bonds (Distances in Å)
F94	L424		C112	L387	
E98	H427		S113	D338	
K101	G425		S113	N384	
N104	M423		S113	I386	
N104	L424		T114	S383	T114 OG_1: S383 O
F106	G340		T114	N384	(2.45)
F106	K342		T114	I386	
F106	L387		L117	V37	
F106	L392		L117	F45	
F106	E421		G118	A35	
F106	M423		G118	V37	
G107	A339	G107 N:G340 O (2.76)	I119	N34	
G107	G340		I119	A35	
G107	I341		I119	V36	I119 N : A35 O (3.03)
S108	A339		I119	V37	
M109	D338		D120	V37	I119 O : V37 N (2.77)
Q111	I386		E122	G496	
C112	G340		N124	H498	
C112	N384		V126	V37	
C112	I386		V126	F45	
			Y127	S383	

- van der Waals interactions are within a 4.0 Å cutoff.

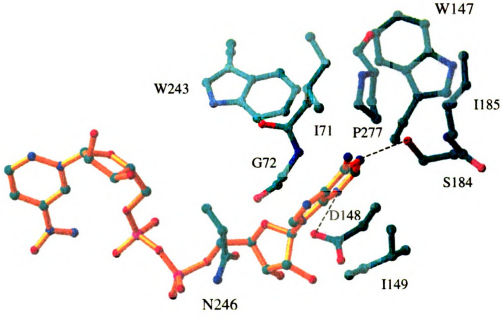


Figure 3.12 Interactions between the adenine portion of NAD^+ and MIP synthase.

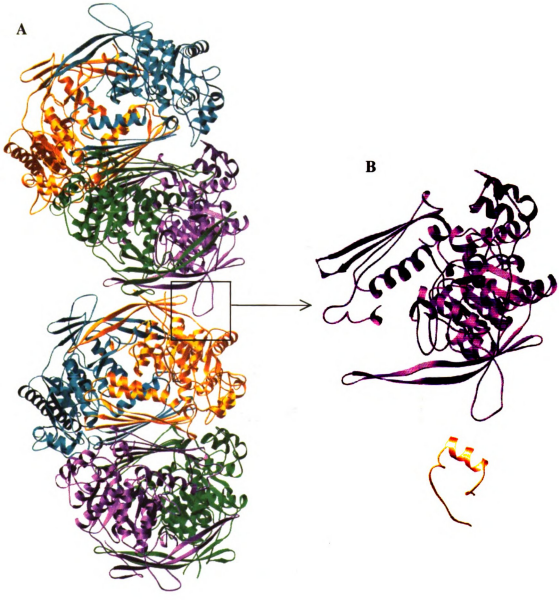


Figure 3.13 The insertion encompassing residues 247-276 is involved in the interaction with a neighboring tetramer.

(A) Two neighboring tetramers. Lavender, molecule A; green, molecule B; gold, molecule C; cyan, molecule D.

(B) The insertion in molecule C makes interactions with the loop 464-472 of molecule A in the neighboring tetramer.

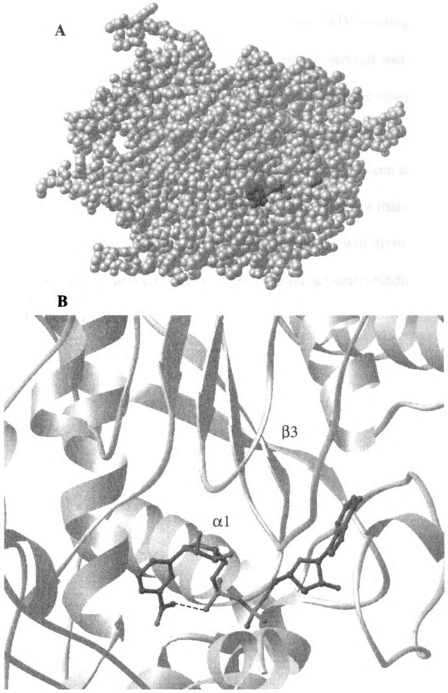


Figure 3.14 (A) Space-filling model of the NAD^+-bound MIP synthase, the nicotinamide is exposed as shown in atom-color. (B) The tight hydrogen bond between the nicotinamide nitrogen and the phosphodiester oxygen stabilizes this conformation.

3.5.4 Electrostatic charge distribution

An analysis of surface charge distribution (15) of the NAD^+-binding site was performed using the program GRASP (16). The electrostatic potential surface of the NAD^+-binding region is shown in Figure 3.15. The apparent negative charge on the NAD^+-binding surface within the dotted circle defines the binding site of the positively charged nicotinamide, indicating that there are polar interactions between this site and the nicotinamide portion of NAD^+. Furthermore, this observation implies that as NAD^+ oxidizes the substrate to become NADH, these polar interactions will diminish and a motion of the nicotinamide ring might occur. The potential substrate-binding site is denoted within the dotted square.

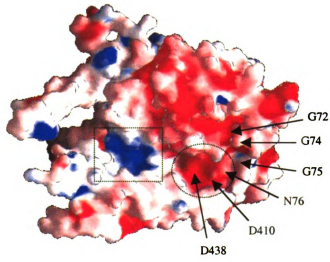


Figure 3.15 The electrostatic potential surface of MIP synthase with a view looking into the NAD^+ binding surface. Blue, EPS> 6kcal/mol; red, EPS<-6 kcal/mol; white, EPS~0. The dotted circle denotes the nicotinamide-binding site; the dotted square denotes the potential substrate-binding site. For clarity, residues 181-201, 351-361 were removed.

3.5.5 The exception of molecule C in the $P2_1$ structure of the MIP synthase/NAD^+ complex

In molecule C in the $P2_1$ structure of the MIP synthase/NAD^+ complex, the entire active site was ordered, although the electron density for some regions was not complete. A significant electron density feature was present in the center of the 2-deoxy-D-glucitol 6-phosphate binding site, indicating that a small molecule is bound in the enzyme active site, consistent with the previous observation that the enzyme active site folds only when occupied with a small molecule (9). Most of the active site residues including S323, N354, K369, K412, D438, and K489, occupy similar positions to those seen in the previously reported structure of the MIP synthase/NAD^+/2-deoxy-D-gluctiol 6-phosphate complex. However, D356, which was flipped out of the enzyme active site in the structure of the MIP synthase/NAD^+/2-deoxy-D-glucitol 6-phosphate complex (9), is now swung into the phosphate position of the inhibitor 2-deoxy-D-glucitol 6-phosphate. Q325, which was implicated in substrate binding when the substrate was modeled in a pseudocyclic conformation consistent with aldol cyclization (9), has also moved and makes a hydrogen bond with K412, buttressing its orientation in the enzyme active site. These changes compared to the structure of the MIP synthase/NAD^+/2-deoxy-D-glucitol 6-phosphate complex (9) brought about questions regarding some details of the model reported previously (9).

3.6 Structure of the MIP synthase/NADH complex

All non-redox steps during the reaction catalyzed by MIP synthase occur when the enzyme is coupled with NADH after NAD^+ has oxidized the substrate, therefore, the three-dimensional structure of the MIP synthase/NADH structure must represent the enzyme in its active catalytic state. In addition, after the hydroxyl group of the substrate C5 is oxidized by NAD^+, either the nicotinamide ring or the substrate may move in order to reposition the carbonyl oxygen of the substrate C5 since it would collide with the hydrophobic nicotinamide ring if there were no motion. Consequently, some of the active site residues may move as well to accommodate any repositioning of the substrate and/or the nicotinamide ring. In order to answer these questions, it was decided to elucidate the structure of the MIP synthase/NADH complex. As expected and even beyond that, the structure of MIP synthase/NADH presented a picture rather different from any structures of MIP synthase seen thus far. Several dramatic structural changes were observed in the enzyme active site as well as in the cofactor.

3.6.1 Overall structure of the MIP synthase/NADH complex

When MIP synthase is bound with NADH instead of NAD^+, the diffraction of the crystals improved significantly to 1.6 Å though the data is only complete to 1.7 Å. Figure 3.16A shows an example of the 1.6 Å electron density map for the NADH-bound MIP synthase structure.

The entire active site became ordered when the enzyme is bound to NADH. Although quite surprising, this observation was consistent with the speculation that the enzyme is at its active stage when coupled with NADH instead of NAD^+, since all of the

non-redox steps catalyzed by the enzyme occur after NAD^+ is reduced to NADH. There were no significant conformational changes in the Rossmann fold domain of MIP synthase compared with the *apo* structure. The two structures overlay well with an RMSD of 0.87 Å. Figure 3.16B depicts the superimposed structures of the *apo* and the NADH-bound MIP synthase.

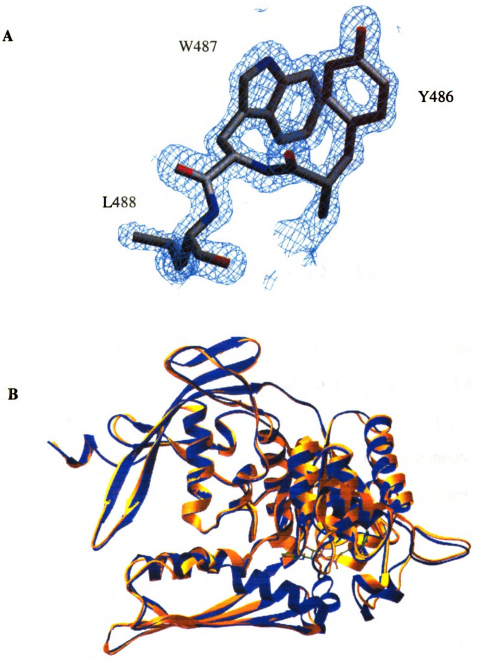


Figure 3.16 (A) An example of the $2F_o$-F_c electron density map of the MIP synthase/NADH structure contoured at 2.4 σ. (B) Superposition of the structures of *apo* MIP synthase and the NADH-bound MIP synthase. The *apo* MIP synthase is in gold and the NADH-bound MIP synthase is in blue.

3.6.2 Conformation of NADH

When compared with the structure of the MIP synthase/NAD^+ complex, a significant conformational change of the cofactor was observed. The nicotinamide ring deviates from its position in the NAD^+-bound structure by about 1.3 Å, and the tight hydrogen bond that pinches the nicotinamide and phosphodiester together was eliminated (Figure 3.17). In fact, there is a 13σ electron density peak between the nicotinamide and phosphodiester as shown in Figure 3.18 A at position 0. This feature coordinates the nicotinamide oxygen and phosphodiester oxygen with interatomic distances of 2.34 Å and 2.36 Å respectively. A conserved water molecule 2.32 Å away and another 10 σ feature 2.29 Å away (Figure 3.18A at position 4) make the third and fourth coordination sites of the tetrahedrally coordinated feature. These distances are short for hydrogen bonds; instead this feature is more likely a metal ion. A number of different divalent and monovalent cations were modeled in this position including Zn^{2+}, Mn^{2+}, Ca^{2+}, Mg^{2+}, Na^+, and K^+. Modeling of Zn^{2+} and Mn^{2+} produced negative density around the atom in the F_o-F_c map, indicating that they are too electron rich to account for this feature. On the other hand, Mg^{2+} accounts for the electron density the best with only a bit of positive density in the F_o-F_c map.

When the *apo* MIP synthase crystals were soaked in EDTA in addition to NADH, all the active site residues are ordered as in the MIP synthase/NADH structure, with the exception of residues 372-375 in molecule B. However, the conformation of the nicotinamide is completely different from that of the NADH-bound structure, in fact it is similar to that of the NAD^+-bound structure. Importantly, the electron density feature between the nicotinamide and phosphodiester has disappeared (Figure 3.18 B), indicating

that this is indeed a divalent cation. However, it is still unclear which specific divalent cation this is. The density of the nicotinamide is also much worse, indicating the increased mobility of the nicotinamide ring in the absence of a putative divalent cation.

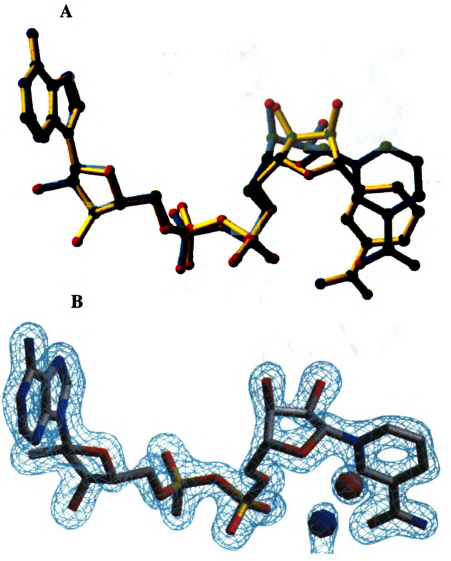


Figure 3.17 (A) Overlay of cofactors in the NAD^+-bound structure in yellow and the NADH-bound structure in cyan. (B) The $2F_o$-F_c electron density map of the MIP synthase/NADH complex structure contoured at 2.4σ around the NADH. Shown in blue is the putative divalent cation; red is a water molecule that coordinates with the divalent cation.

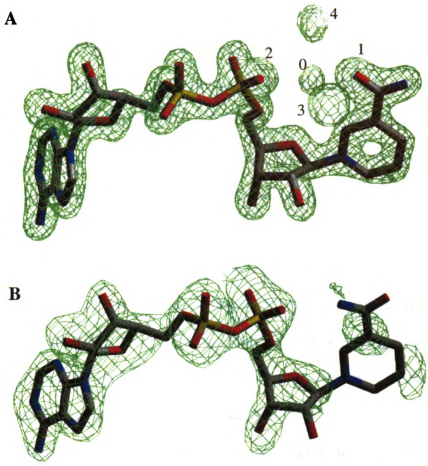


Figure 3.18 (A) Simulated annealing omit electron density map of the MIP synthase/NADH complex structure contoured at 5σ.

(B) Simulated annealing omit electron density map of the MIP synthase/NADH/EDTA complex structure contoured at 1.8 σ.

The conformation of nicotinamide and the position of the putative divalent cation are similar to those of MIP synthase from *Mycobacterium tuberculosis* (Figure 3.19) (17). In the *M. tuberculosis* MIP synthase structure, Zn^{2+} was modeled at this position, however, its coordination is different from that of the putative divalent cation in the structure of the *S.cerevisiae* MIP synthase/NADH complex. Distances from Zn^{2+} to the nicotinamide oxygen and phosphodiester oxygen are 2.16 Å, 2.08 Å respectively, the third ligand 2.21 Å away is a conserved water molecule residing at the same position in the *S. cerevisiae* MIP synthase/NADH structure. The fourth ligand is S311, and is conserved as S439 in *S. cerevisiae* MIP synthase. In the structure of the MIP synthase/NADH complex, S439 is 2 Å away from this position making a tight hydrogen bond with the fourth ligand of the putative divalent cation. S439 is in the same position in all structures of *S. cerevisiae* MIP synthase determined so far.

The mechanistic role of the divalent cation is not clear at this point. Previous experiments done on *S. cerevisiae* MIP synthase have ruled out the direct involvement of a divalent cation during the catalysis (18). The structural role of this putative divalent cation is critical in that it bridges the cofactor and thus might bring the nicotinamide C4 close to the substrate oxidation center C5. This divalent cation also coordinates an important water molecule that makes hydrogen bonds with two absolutely conserved aspartate residues, D356 and D438. It might influence the pK_a of the acidic side chains of D356 and D438 by chelating the conserved water molecule, which makes hydrogen bonds with D356 and D438. Further biochemical experiments are required to clearly identify this cation and define its precise role in catalysis.

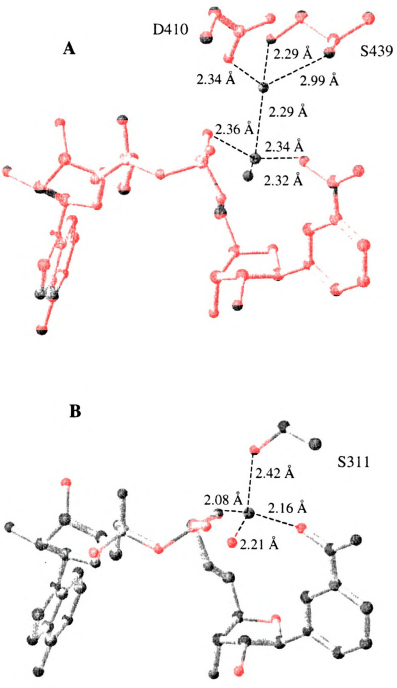


Figure 3.19 (A) The NADH-bound MIP synthase from *S. cerevisiae.* Cyan, the modeled putative divalent cation; Orange, the fourth ligand of the putative divalent; Red, the conserved water molecule. (B) The NAD^+-bound MIP synthase from *M. tuberculosis.* Zn^{2+} is in cyan.

3.6.3 The MIP synthase active site

While all of the active site residues are ordered, significant structural changes compared to the structures of the *apo* MIP synthase and the MIP synthase/NAD^+ complex are also evident on the opposite side of the active site. There are two small molecules bound in the enzyme active site. The peak heights of both of these molecules in the F_o-F_c electron density map are higher than 7σ. One of these molecules is clearly a tetrahedron with a heavy atom as the central atom consistent with either a phosphate or sulfate ion, since the peak height at the center is 22 σ in the F_o-F_c electron density map. Given the fact that there are often micro-molar quantities of phosphates in the reagents that were used in protein purification and crystallization, a phosphate anion was modeled in this position. The other feature overlaps with the position of the inhibitor in the previously published structure of the MIP synthase/NAD^+/2-deoxy-D-glucitol 6-phosphate complex. A glycerol molecule was modeled because of the shape and size of the electron density and also the fact that a 30 % glycerol solution was used as the cryoprotectant when freezing these crystals. Figure 3.20 shows the 7σ simulated annealing omit map with these two molecules modeled in.

A significant conformational change also occurred in a loop connecting β13 and α12, in order to accommodate the phosphate. The Cα of S323 moved 2.1 Å and the side chain oxygen moved 4 Å away from their positions in all structures of MIP synthase determined previously including the *apo*, NAD^+-bound, low occupancy NAD^+-bound (9), and 2-deoxy-D-glucitol 6-phosphate-bound (9). Figure 3.21 depicts the motion of this loop.

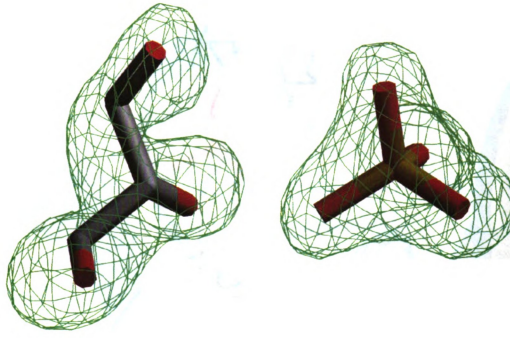


Figure 3.20 7σ simulated annealing omit electron map of the MIP synthase/NADH active site with a phosphate and glycerol modeled in.

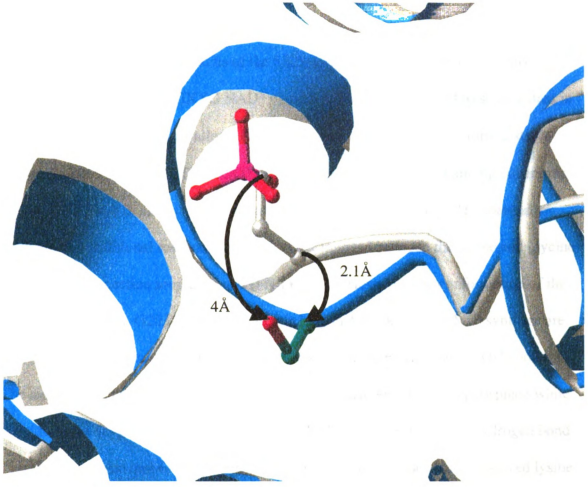


Figure 3.21 Conformational change of the phosphate-binding loop in the NADH-bound structure in cyan compared with the NAD^+-bound structure in silver. Note: The conformation of this loop in the structure of the *apo* MIP synthase, the low occupancy NAD^+-bound MIP synthase(9), and the MIP synthase/NAD^+/2-deoxy-D-glucitol 6-phosphate complex (9) are all identical to that of the NAD^+-bound structure.

Numerous interactions were observed between the enzyme active site and phosphate and glycerol (Figure 3.22). The phosphate oxygens make hydrogen bonds with the main chain nitrogen atoms of the S323-G324-Q325-T326 sequence. This sequence mimics the GXGGXXG NAD^+ pyrophosphate-binding motif to some extent in that the glycine residue G324 allows for a tight turn from β13 to α12 to form a stable phosphate-binding pocket. This SGQT motif is absolutely conserved among eukaryotic species from yeast *S. cerevisiae* to human (Figure 1.10). However, in *M. tuberculosis*, there are two additional resides, Q201 and V202, inserted between the serine and glycine making this connection loop extended (Figure 3.23). However, the conformation of the side chains of S200, G203-A204-T205 in this loop in *M. tuberculosis* MIP synthase are similar to those seen in the structure of the *S. cerevisiae* MIP synthase/NADH complex. The side chain nitrogen atom of Q325 makes a hydrogen bond with the phosphate while the side chain oxygen atom of Q325 buttresses K412 in its position via a hydrogen bond. The phosphate also makes salt bridge interactions with three absolutely conserved lysine residues, K373, K412, K489. The putative glycerol molecule makes interactions with N350, L352, D356, L360, K369, K373, I402, K412, and D438. It is positioned close to the nicotinamide C-4 where the oxidation and reduction occur. It is important to point out that the putative divalent ion chelates the water molecule that holds two conserved aspartate residues, D356 and D438, together for hydrogen bonding with the glycerol hydroxyl oxygens.

Clearly, binding of the phosphate and glycerol mimics the substrate binding to the enzyme active site, and the ordered active site is due to the binding of these two molecules. This is consistent with the idea that the substrate folds the enzyme active site

by encapsulating itself within the enzyme active site. On the other hand, the position of the phosphate is completely different from that of 2-deoxy-D-glucitol 6-phosphate in the previously reported structure (9). Figure 3.24 shows the discrepancy between the two structures. The unambiguous position of the phosphate in the MIP synthase/NADH complex structure calls into question the mechanism proposed previously based on the structure of the MIP synthase/NAD^+/2-deoxy-D-glucitol 6-phosphate complex (9). In addition, D356, which was flipped out of the enzyme active site in the structure of the MIP synthase/NAD^+/2-deoxy-D-glucitol 6-phosphate complex (9), is now swung into the position of the phosphate of the inhibitor 2-deoxy-D-glucitol 6-phosphate. The conformation of D356 in the NADH-bound structure is the same as that seen in the structures of the MIP synthase/NAD^+ complexes where this residue is ordered.

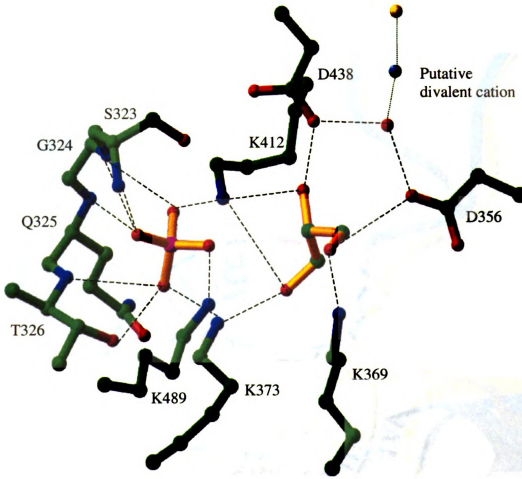


Figure 3.22 Interactions in the enzyme active site observed from the MIP synthase/NADH complex structure. The phosphate and glycerol are in gold; the putative divalent cation is in cyan. The fourth coordination ligand of the putative divalent cation, which was modeled as a water molecule in the MIP synthase/NADH complex structure, is in yellow.

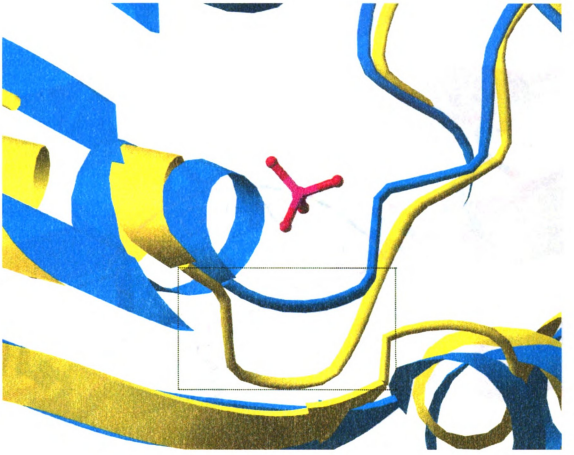


Figure 3.23 Overlay of the *S. cerevisiae* MIP synthase/NADH complex structure in cyan and the *M. tuberculosis* MIP synthase/NAD^+ complex structure (17) in yellow. In the dotted box are the phosphate-binding loops.

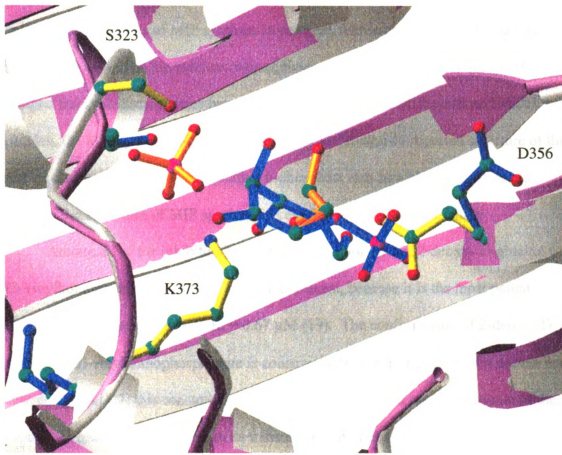


Figure 3.24 Overlay of the MIP synthase/NADH complex structure in silver and the MIP synthase/NAD^+/2-deoxy-D-glucitol 6-phosphate complex structure in lavender. Gold, the phosphate and glycerol in the NADH-bound structure; blue, 2-deoxy-D-glucitol 6-phosphate in the previous inhibitor-bound structure (9); yellow, side chains of the NADH-bound structure; cyan, sides chains of the previous inhibitor-bound structure(9).

3.7 Structure of the MIP synthase/NAD^+/2-deoxy-D-glucitol 6-(*E*)-vinylhomophosphonate complex

In order to elucidate completely the mechanism for the reaction catalyzed by MIP synthase, the structures of MIP synthase in complex with various structural analogues of the substrate as well as the reaction intermediates are essential. Several questions emerged from the discrepancies between the structures of the MIP synthase/NADH complex, the MIP synthase/NAD^+ complex, and the previously reported structure of the MIP synthase/NAD^+/2-deoxy-D-glucitol 6-phosphate also brought about a need to elucidate the structures of MIP synthase in complex with various inhibitors.

Among all of the inhibitors used in crystallization so far, 2-deoxy- D-glucitol 6-(*E*)-vinylhomophosphonate was of particular interest, because it is the most potent inhibitor synthesized so far with K_i = 0.67 μM (19). The conformation of 2-deoxy- D-glucitol 6-(*E*)-vinylhomophosphonate is constrained by the introduction of a double bond between C6 and C7; this constraint allows the molecule to be a structural mimic of the substrate D-glucose 6-phosphate in its transoid conformation.

Obtaining crystals of MIP synthase with fully occupied inhibitor 2-deoxy- D-glucitol 6-(*E*)-vinylhomophosphonate was an unexpected challenge. Crystals were obtainable from co-crystallization setups with the inhibitor concentrations lower than 1.5 mM. However, electron density maps showed no density for bound inhibitor molecules. Soaking experiments with inhibitor concentrations as high as 10 mM also did not produce electron density maps with a fully occupied inhibitor, although scattered pieces of electron density were observed in the enzyme active site. This was surprising given the high binding affinity of the inhibitor 2-deoxy- D-glucitol 6-(*E*)-vinylhomophosphonate

(K_i = 0.67 μM). When the inhibitor concentration was increased to 13.5 mM and the pH of the soaking stabilizer was 4.5, there was unambiguous electron density for phosphate, and a clear electron density feature was present where the glycerol molecule was modeled in the NADH-bound structure but with a bit different shape, indicating that the inhibitor molecule was bound within the enzyme active site but not fully occupied. The breaking point was another soaking experiment with the same concentration (13.5 mM) of inhibitor but at pH 5.5, which produced an electron density map indicating a fully occupied inhibitor in one of two molecules in the asymmetric unit. It is important to realize that MIP synthase loses activity at pH lower than 7.2. At low pH, many of the acidic side chains become protonated affecting inhibitor/substrate binding. At higher pH (5.5), some of the acidic side chains are deprotonated, facilitating binding of the inhibitor/substrate.

3.7.1 Overall structure of the MIP synthase/NAD^+/2-deoxy-D-glucitol 6-(*E*)-vinylhomophosphonate complex

The overall structure of the MIP synthase/NAD^+/2-deoxy-D-glucitol 6-(*E*)-vinylhomophosphonate complex is similar to that of the MIP synthase/NADH complex described in section 3.5. The RMSD between the two structures was only 0.49 Å, which is the lowest RMSD between two structures of MIP synthase seen so far. All of the active site residues are ordered in both molecules in the asymmetric unit and the side chains of the active site residues overlap with those of the NADH-bound structure very well. C436 is the only exception in that its Cα moved 2.4 Å and its side chain sulfur moved 3.5 Å away from their positions in the NADH-bound structure (Figure 3.25). In

fact, the position of C436 is identical to that of the *apo*, NAD^+-bound, and 2-deoxy-D-glucitol 6-phosphate-bound structures (9). The reason for this conformational change in the NADH-bound structure is unclear.

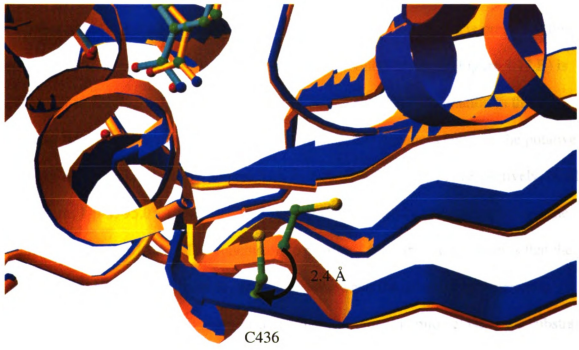


Figure 3.25 Overlay of the MIP synthase/NAD^+/2-deoxy-D-glucitol 6-(*E*)-vinylhomophosphonate complex structure in blue with the MIP synthase/NADH complex structure in gold.

3.7.2 Conformation of NAD^+

The conformation of NAD^+ is similar to that of NADH in the NADH-bound structure; there was no hydrogen bond pinching the nicotinamide and phosphodiester together. However, the distance between the amide and phosphodiester is a bit closer than that of NADH (3.2 Å versus 3.8 Å). The putative divalent cation is present in this structure as well, but the fourth ligand in the NADH-bound structure (Figure 3.17 B) is not present. The fourth ligand is now S439 O, identical to the coordination in the *M. tuberculosis* MIP synthase structure (17). Interatomic distances between the putative divalent cation and its four ligands are 2.26 Å, 2.58 Å, 2.52 Å, 2.83 Å respectively, bond angles are: 1-0-3: 95.0°; 1-0-4: 104.7°; 2-0-3: 92.01°; 2-0-4: 112.8° (numberings are the same as in Figure 3.18 A). The conclusion to be drawn from this observation is that the nicotinamide moves away from the phosphodiester when the substrate or a substrate analogue is bound in the active site, bringing the C4 of nicotinamide close to the substrate C5 to carry out the oxidation step. Therefore, this conformation of the cofactor may represent the enzyme in its active state. Also, changes seen between the NAD^+-bound structure and the NADH-bound structure are not purely due to the oxidation state of the cofactor.

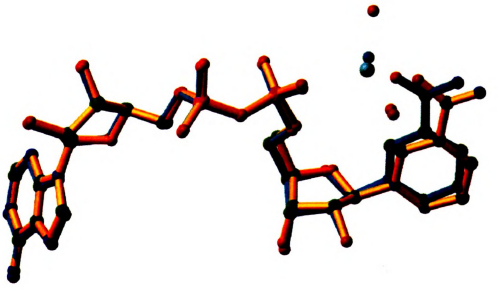


Figure 3.26 Overlay of cofactors in the MIP synthase/NAD^+/2-deoxy-D-glucitol 6-(*E*)-vinylhomophosphonate complex structure in cyan and the MIP synthase/NADH complex structure in gold. The putative divalent cation is in aqua in the NADH-bound structure and in cyan in the 2-deoxy-D-glucitol 6-(*E*)-vinylhomophosphonate-bound structure. Shown in orange is the fourth ligand of the putative divalent cation in the NADH-bound structure. The third ligand water molecule is located at an identical position in both structures.

3.7.3 Structure of the inhibitor 2-deoxy-D-glucitol 6-(*E*)-vinylhomophosphonate

The inhibitor 2-deoxy-D-glucitol 6-(*E*)-vinylhomophosphonate is fully occupied in the enzyme active site as shown in the 1.6 σ simulated annealing omit map (Figure 3.27A). The inhibitor molecule is well nestled within the enzyme active site in an extended conformation, making hydrogen bond interactions as well as van der Waals interactions with the active site residues. Table 3.11 lists all the interactions observed. The phosphate group is in a transoid conformation relative to the inhibitor carbon backbone. It is making hydrogen bonds with the main chain nitrogen atoms of S323-G324-Q325-T326 as well as conserved lysine residues, K412 and K373. All of the hydroxyl groups of the inhibitor except O1 make hydrogen bonds with side chains of conserved active site residues. Figure 3.28 A depicts all of the interactions between the inhibitor molecule and the active site residues. It is important to note that the putative divalent cation chelates the water molecule as part of a hydrogen bond network that holds O2 and O3 of the inhibitor molecule in their positions. As opposed to the previously reported observation (9), the oxidation at C5 was not observed from the electron density map, with the distance from the inhibitor C5 to the nicotinamide C4 being 3.8 Å, which is a bit long for a direct hydride transfer.

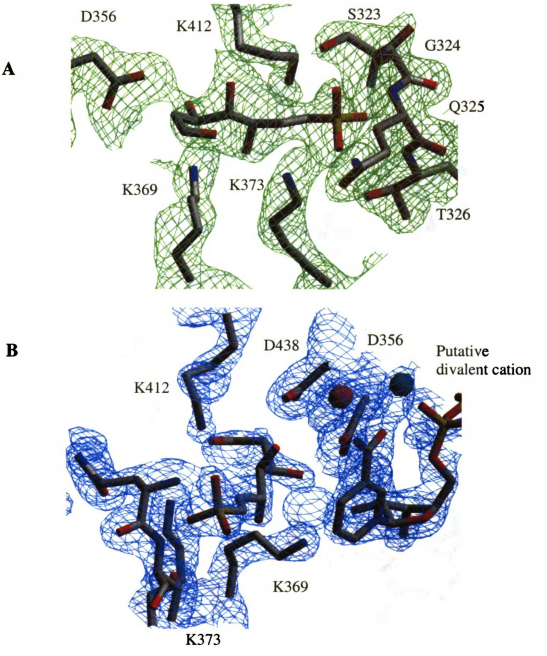


Figure 3.27 (A) Simulated annealing omit map of the MIP synthase/NAD^+/2-deoxy-D-glucitol 6-*(E)*-vinylhomophosphonate structure contoured at 1.6 σ. (B) 2Fo-Fc electron density map of the MIP synthase/NAD^+/2-deoxy-D-glucitol 6-*(E)*-vinylhomophosphonate structure contoured at 1.2 σ.

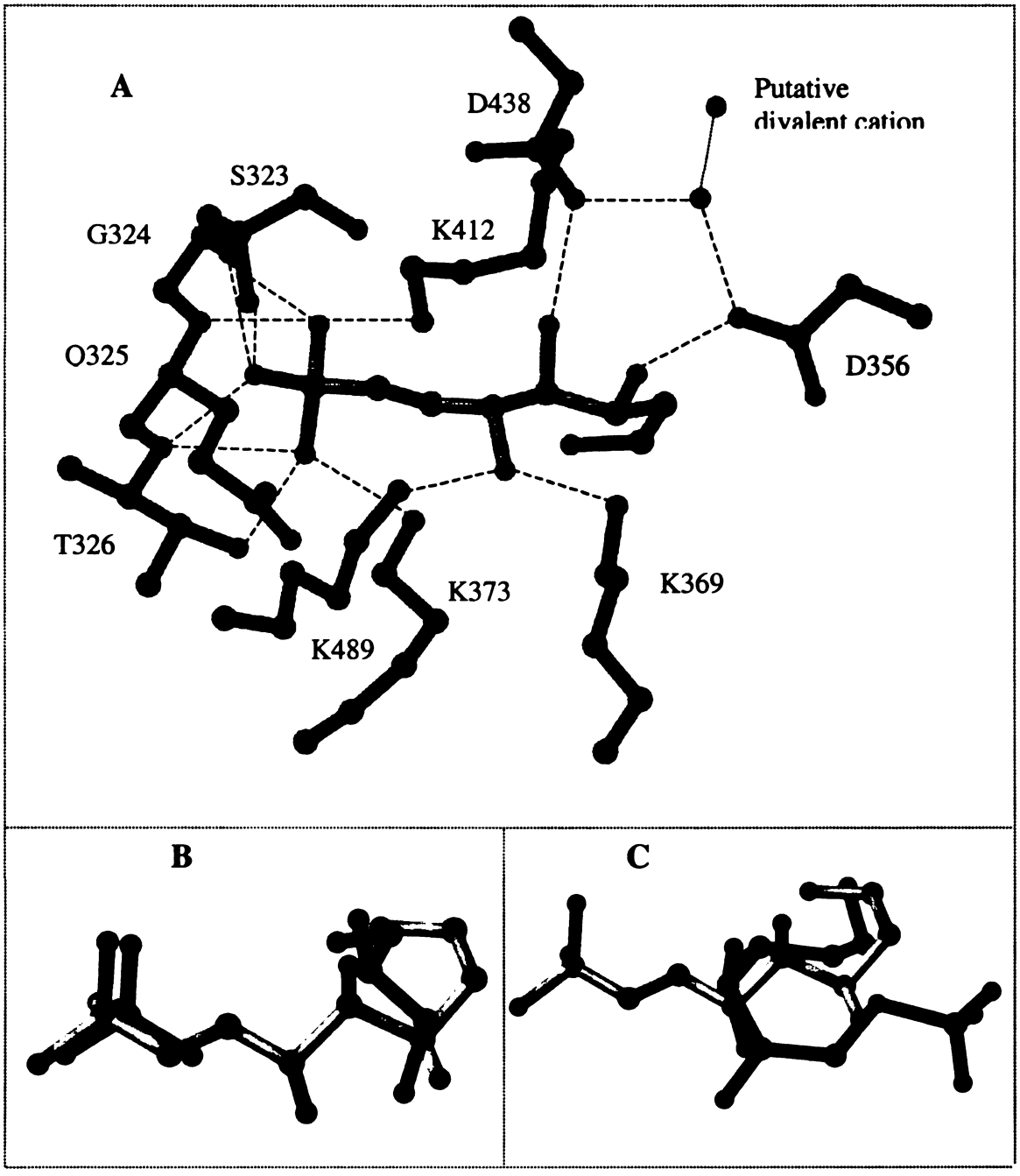


Figure 3.28 (A) Interactions between MIP synthase active site residues and the inhibitor 2-deoxy-D-glucitol 6-(*E*)-vinylhomophosphonate; the inhibitor is in lavender. (B) Overlay of 2-deoxy-D-glucitol 6-(*E*)-vinylhomophosphonate in gold and phosphate and glycerol in the NADH-bound structure in cyan. (C) Overlay of 2-deoxy-D-glucitol 6-(*E*)-vinylhomophosphonate in gold and 2-deoxy-D-glucitol 6-phosphate in the previously published inhibitor structure (9) in blue.

Table 3.11 Interactions between MIP synthase and the inhibitor 2-deoxy-D-glucitol 6-*(E)*-vinylhomophosphonate within a 3.5 Å cutoff.

2-deoxy-D-glucitol 6-*(E)*-vinylhomophosphonate atom	MIP synthase #	Residue	atom	Distance (Å)
C2	352	LEU	CD1	3.41
C2	360	LEU	CD1	3.26
C3	369	LYS	CD	3.49
O3	356	ASP	OD1	2.45
O3	369	LYS	NZ	3.28
O4	438	ASP	OD2	2.64
O5	369	LYS	NZ	2.99
O5	489	LYS	NZ	2.77
O1P	323	SER	N	3.08
O1P	324	GLY	N	2.53
O1P	325	GLN	N	3.46
O1P	326	THR	N	3.07
O2P	324	GLY	N	3.06
O2P	325	GLN	N	2.57
O2P	412	LYS	NZ	2.95
O3P	326	THR	N	2.88
O3P	326	THR	OG1	2.59
O3P	373	LYS	NZ	2.97

The constellation of new data from the NAD^+-bound, NADH-bound, and 2-deoxy-D-glucitol 6-(*E*)-vinylhomophosphonate-bound structures leads to the conclusion that the previous modeling of the inhibitor 2-deoxy-D-glucitol 6-phosphate is incorrect and the mechanism proposed should also be revised (9). First, D356 is in the same conformation in all of the above-mentioned structures, as opposed to the conformation in the previously reported MIP synthase/NAD^+/2-deoxy-D-glucitol 6-phosphate structure (9), where it is flipped out of the active site (Figure 3.24). Secondly, the phosphate moiety of 2-deoxy-D-glucitol 6-*(E)*-vinylhomophosphonate is in an identical position to that of the phosphate in the NADH-bound structure (Figure 3.28B), quite different from that of the inhibitor from the previously published MIP synthase/NAD^+/2-deoxy-D-glucitol 6-phosphate structure (Figure 3.28C) (9). Third, as shown in Figure 3.29, an analysis of surface charge distribution of the substrate-binding site indicates the new phosphate-binding pocket (where the phosphate moiety is located in both the MIP synthase/NADH complex and MIP synthase/NAD^+/2-deoxy-D-glucitol 6-*(E)*-vinylhomophosphonate complex structures) to be highly positive in charge, but in the MIP synthase/NAD^+/2-deoxy-D-glucitol 6-phosphate structure (9), the phosphate-binding site is negatively charged. Fourth, the structure of the MIP synthase/NAD^+/2-deoxy-D-glucitol 6-*(E)*-vinylhomophosphonate complex was determined at pH 5.5, closer to the active pH (7.2) than pH 4.5, at which the previous MIP synthase/NAD^+/2-deoxy-D-glucitol 6-phosphate structure was determined. Finally, in the structure of the MIP synthase/NAD^+/2-deoxy-D-glucitol 6-*(E)*-vinylhomophosphonate, the conformations of the substrate-interacting residues that are different from the previous MIP synthase/NAD^+/2-deoxy-D-glucitol 6-phosphate structure (9) agree well with that of

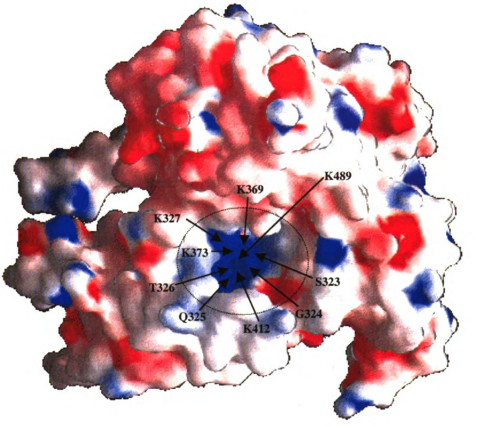


Figure 3.29 GRASP drawing (16) of the electrostatic potential surface of the MIP synthase. Circled dotted is the substrate-binding site centered at the phosphate-binding pocket. Blue, EPS> 6kcal/mol; red, EPS<-6 kcal/mol; white, EPS~0. For clarity, residues 190-200, 351-366 were removed in this figure.

Note: The substrate-binding surface is positively charged for encapsulation of the substrate. The pK_a's of D-glucose 6-phosphate are pK_{a1} =2.1, pK_{a2} = 6.8. At pH 7.0, the substrate phosphate is negatively charged; there must be strong polar interactions between the substrate and the substrate-binding site of MIP synthase.

the recently reported structure of MIP synthase from *M. tuberculosis* (17), especially the region surrounding the phosphate-binding pocket.

3.7.4 Modeling of the substrate and reaction intermediates

Based on the location and conformation of the inhibitor 2-deoxy-D-glucitol 6-*(E)*-vinylhomophosphonate, the substrate D-glucose 6-phosphate was modeled in its conformation necessary for cyclization (Figure 3.30). The phosphate portion was overlaid onto that of the inhibitor molecule with O6 at the position of the inhibitor C7; C6, C5, C4 were overlaid onto the inhibitor C6, C5, C4 respectively. The rest of the substrate molecule was modeled such that none of the backbone atoms and the hydroxyl oxygen atoms would collide with the side chains of active site residues.

The result of the modeling provided much information with respect to the interactions in the enzyme active site. All but O3 of the substrate hydroxyl groups make hydrogen bonds with the active site residue side chains. Figure 3.31 depicts most of the interactions seen.

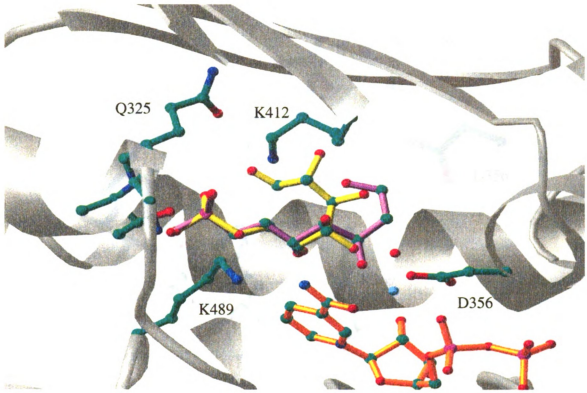


Figure 3.30 Modeling of the substrate D-glucose 6-phosphate (yellow) based on the structure of the inhibitor 2-deoxy-D-glucitol 6-*(E)*-vinylhomophosphonate (lavender). The putative divalent cation is in cyan.

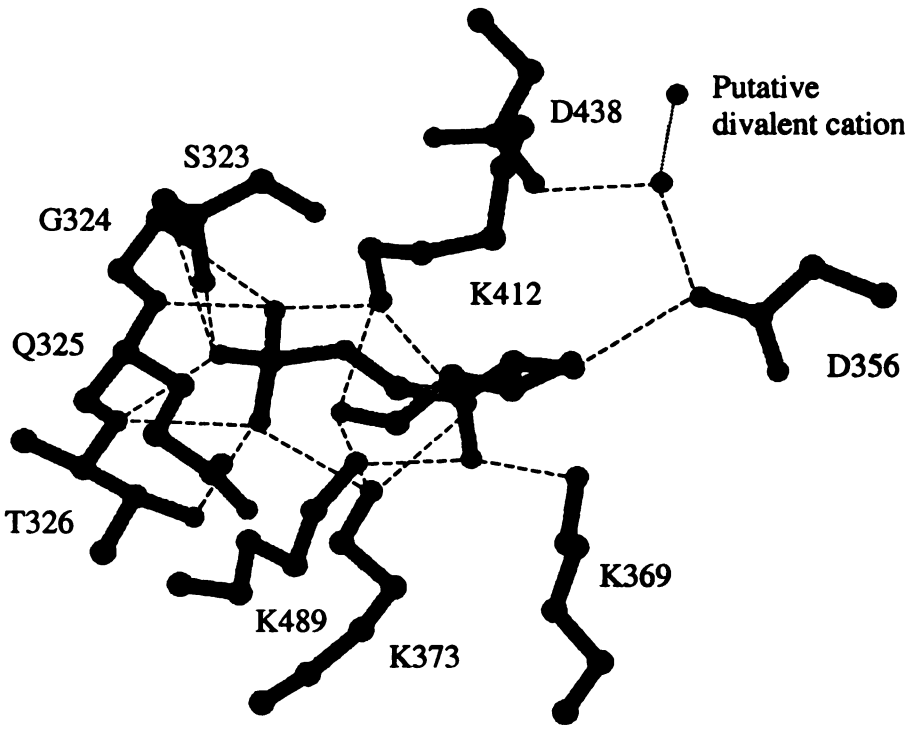


Figure 3.31 Interactions between MIP synthase and the modeled substrate D-glucose 6-phosphate.

Using the modeled substrate as a guide each of the reaction intermediates was also modeled. It is important to point out that once the hydride is transferred from the substrate C5 to the nicotinamide C4, there may be a slight re-position of the substrate in order to avoid the collision of the C5 carbonyl group with the nicotinamide ring. As a matter of fact, a modeling of the reaction intermediate 5-keto-D-glucose 6-phosphate without any change in the rest of the molecule resulted in the C5 carbonyl oxygen only 2.5 Å away from the nicotinamide C4. This would never be the case during the catalysis. A slight rotation of the phosphate portion about the C5-C6 axis resulted in a conformation where this unfavorable steric collision can be avoided. When the phosphate was rotated slightly, C4, C5, O5, and C6 were all kept in the same plane, since the enolization follows oxidation immediately. Figure 3.32 shows the modeling of 5-keto-D-glucose 6-phosphate (and the enolate intermediate). The slight rotation of the phosphate moiety did not disrupt the interaction between the substrate and the enzyme, instead there was an additional hydrogen bond between O3 and the D438 side chain as denoted by the red dashed line in Figure 3.33.

The final reaction intermediate, *myo*-2-inosose 1-phosphate was also modeled (Figure 3.34). This cyclic intermediate is in a conformation similar to the substrate but with an additional hydrogen bond between O3 and D438, which contributes to the stabilization of the cyclic conformation.

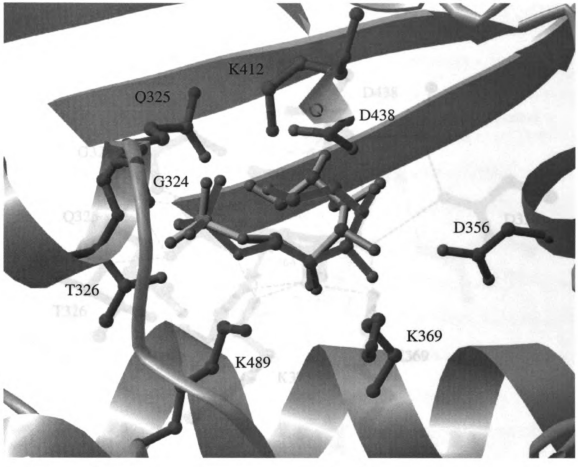


Figure 3.32 Modeling of the reaction intermediate 5-keto-D-glucose 6-phosphate. The substrate D-glucose 6-phosphate is in yellow, and the 5-keto-D-glucose 6-phosphate (and the enolate) is in blue.

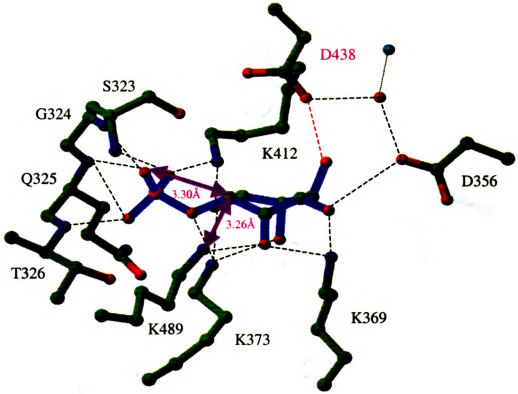


Figure 3.33 Interactions between the modeled 5-keto-D-glucose 6-phosphate (enolate) and the active site residues.

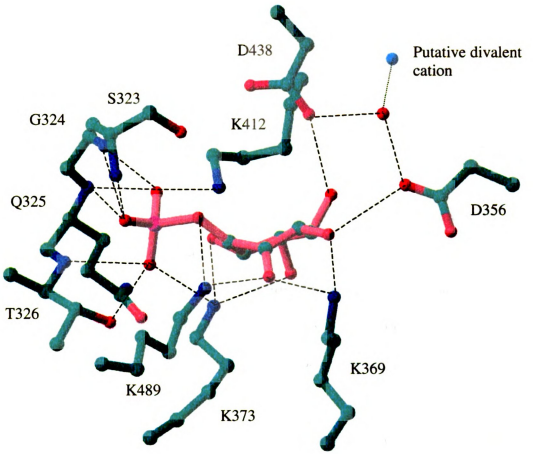


Figure 3.34 Interactions between the modeled *myo*-2-inosose 1-phosphate and the active site residues.

3.8 Proposed mechanism of MIP synthase

The apparent discrepancy between the structure of the MIP synthase/NAD^+/2-deoxy-D-glucitol 6-(*E*)-vinylhomophosphonate complex described above and the structure of the MIP synthase/NAD^+/2-deoxy-D-glucitol 6-phosphate complex published previously (9) (Figure 3.28 B) calls into question the mechanism proposed previously. Based on the new structure of MIP synthase/NAD^+/2-deoxy-D-glucitol 6-(*E*)-vinylhomophosphonate and modeling of the substrate and reaction intermediates, a new mechanism was proposed.

In the first step, the substrate is oxidized at C5 by NAD^+ (Figure 3.35). This involves a direct hydride transfer from the C5 of D-glucose 6-phosphate to the C4 of the nicotinamide moiety of NAD^+; this is confirmed by the crystal structure where the nicotinamide is located in a suitable orientation for hydride transfer to occur. In concert, a proton is lost from the C5 hydroxyl group of D-glucose 6-phosphate; this proton can be transferred to the K369 terminal nitrogen atom, which is 2.8 Å away from the O5 of the substrate. D320, adjacent to K369, could then accept this proton in a proton-shuffling system. It is also possible that the terminal nitrogen atom of K489, which is 2.92 Å away from the O5, pulls off the proton from O5.

The second step is the enolization. During the enolization, the *pro-R* hydrogen of C6 is eliminated. From the crystal structure, either the dibasic phosphate monoester (Figure 3.35) or K489 (Figure 3.36) may act as the base at the enolization step. The phosphate monoester acting as the base at the enolization step is similar to that proposed for 3-dehydroquinate synthase (20,21). This mechanism takes advantage of phosphate in a transoid conformation relative to the carbon backbone of the substrate. From the

modeling of the 5-keto-D-glucose 6-phosphate based on the crystal structure, K489 is also in a suitable position to remove the *pro-R* hydrogen of C6. The negative charge of the enol can be stabilized by K369.

In the aldol condensation step, the phosphate could transfer the proton abstracted from C6 to O1. If the *pro-R* hydrogen of C6 was removed by K489 at the previous step, then K412 may deliver a proton to O1 in a proton-shuffling system with N350, and the developing negative charge on O1 can be stabilized by K373. It is also possible that MIP synthase utilizes the type I aldolase mechanism where K369 could form a Schiff base with C5, followed by the aldol condensation step, and the developing negative charge on O1 could then be stabilized by K412.

The last step is the reduction by NADH. The hydride that was transferred in the first step to the nicotinamide C4 returns to the C5 of the intermediate *myo*-2-inosose 1-phosphate. Using the same proton-shuffling system, a proton could then be transferred to the C5 ketone oxygen from D320, via K369.

At this point, there is still not enough structural evidence regarding whether the substrate binds in its cyclic form followed by ring opening catalyzed by the enzyme, or binds in its acyclic form, which constitutes less than 0.4 % of D-glucose 6-phosphate in solution.

Verification of the mechanism proposed above requires mutational investigation and further structural investigation of MIP synthase in complex with various structural analogues of the reaction intermediates.

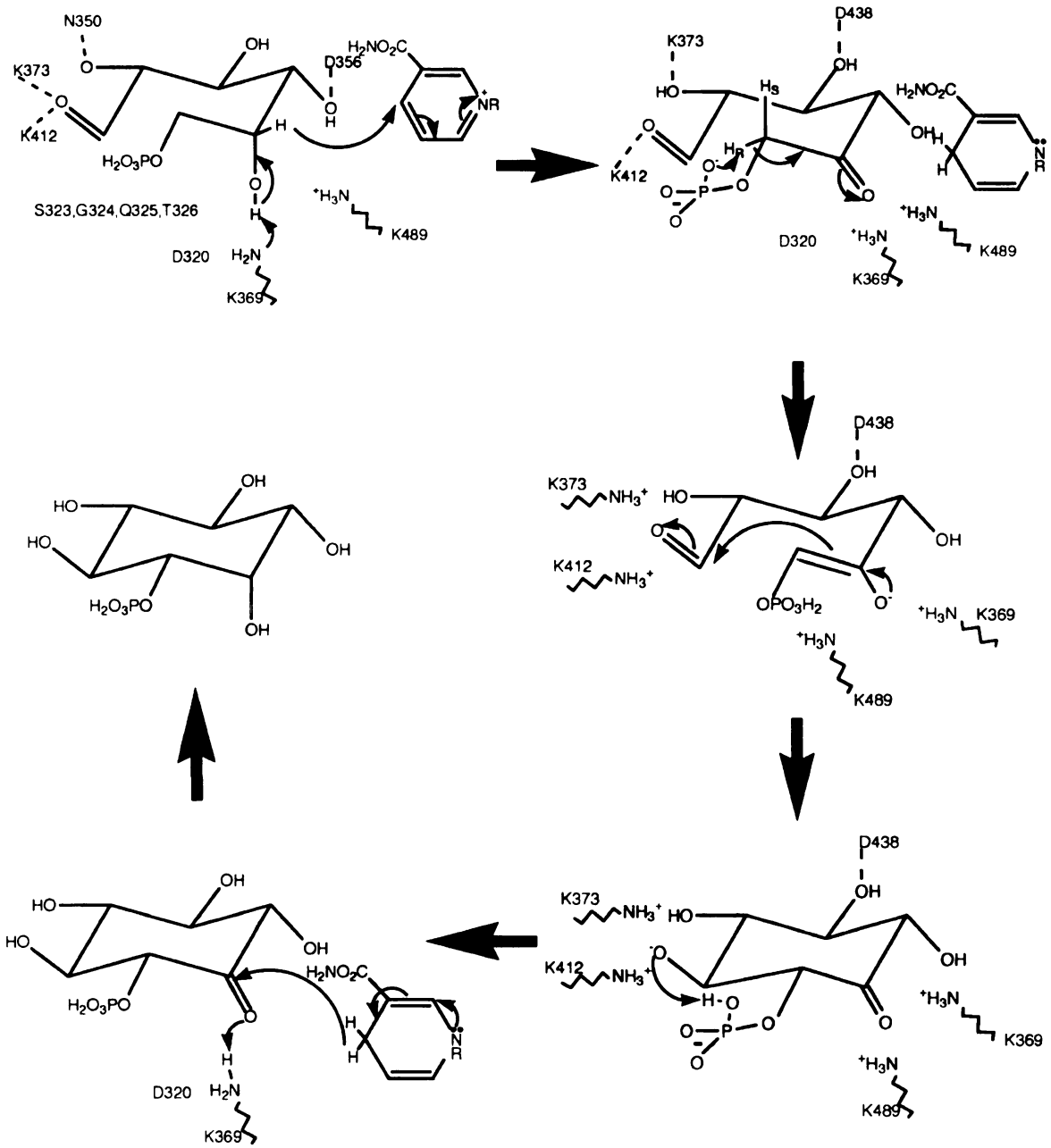


Figure 3.35 Proposed mechanism of MIP synthase where the phosphate monoester acts as the base in the enolization step.

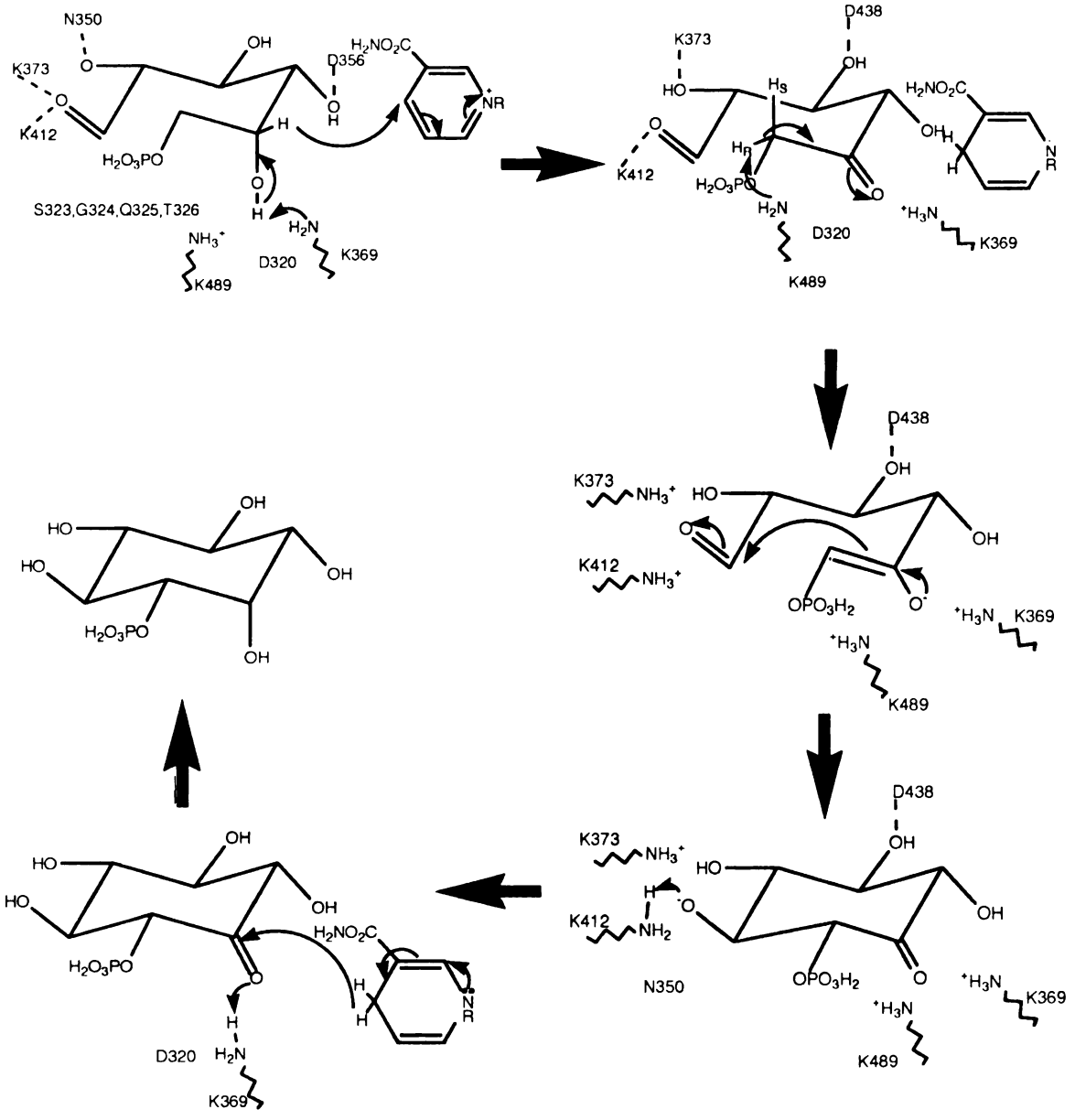


Figure 3.36 An alternative mechanism of MIP synthase where K489 acts as the base in the enolization step.

3.9 Conclusions

The structures of *S. cerevisiae* MIP synthase in its *apo* form, NAD^+-bound form, NADH-bound form, and in complex with an inhibitor, 2-deoxy-D-glucitol 6-*(E)*-vinylhomophosphonate were newly determined. It was observed previously from the structures of *S.cerevisiae* MIP synthase in the presence of partially occupied NAD^+, and in complex with fully occupied NAD^+ and an inhibitor 2-deoxy-D-glucitol 6-phosphate (9) that active site residues 351-409 were disordered in the presence of partially occupied NAD^+ but became ordered when NAD^+ and the inhibitor are bound with full occupancy. Therefore, it was concluded that the substrate binding folds the active site. Compared to the previously determined structures, the newly determined structures presented similarities as well as some significant differences, leading to the proposal of a new catalytic mechanism of MIP synthase.

In the structure of *apo* MIP synthase residues 351-375 were missing; there were no structural differences within the Rossmann fold NAD^+-binding domain between the *apo* structure and the previous low occupancy NAD^+-bound structure (9). This indicates that the NAD^+-binding domain of MIP synthase is very rigid, while the enzyme active site is very flexible. Compared to the previous low occupancy NAD^+-bound structure (9), residues 376-409 became ordered in the structure of *apo* MIP synthase; this increased order appears to be puzzling.

The structure of MIP synthase in the presence of fully occupied NAD^+ represents the state of MIP synthase before the oxidation of the substrate or after the reduction, which leads to the formation and release of the product. Residues 362-379 are disordered in 5 out of 6 molecules in the asymmetric units in both the $P2_1$ and C2 forms of the MIP

synthase/NAD^+ complex crystal structures. Thus, it appears that the previous observation that residues 351-409 are disordered in the presence of partially occupied NAD^+ (9) could be due to not fully occupied NAD^+. The fully folded active site of molecule C in the structure of the $P2_1$ crystal form and the presence of a significant electron density feature in its active site affirms the previous notion (9) that small molecule binding causes the folding of the active site.

Consistent with the assumption that the enzyme is at its catalytically active state after NAD^+ is reduced to NADH, the structure of the MIP synthase/NADH complex presented a phenomenal picture that is completely different from all structures seen thus far. In this structure all the active site residues are completely ordered, and the conformation of the nicotinamide changed significantly with a putative divalent cation eliminating the tight hydrogen bond between the nicotinamide and the phosphodiester that was present in the NAD^+-bound structure. Addition of EDTA in the soaking stabilizer eliminated the electron density feature between the nicotinamide and the phosphodiester, indicating that it is very likely to be a divalent cation. Two small molecules, phosphate and glycerol, are bound in the enzyme active site mimicking the substrate binding. The unambiguous position of the phosphate-like anion is completely different from that of the inhibitor molecule in the previous structure of the MIP synthase/NAD^+/2-deoxy-D-glucitol 6-phosphate complex, calling into question some of the details in the previously published model (9).

The structure of the MIP synthase/NAD^+/2-deoxy-D-glucitol 6-*(E)*-vinylhomophosphonate complex provided insights into the catalytic mechanism of MIP synthase. The inhibitor molecule is bound in the enzyme active site in an extended

conformation. The position of the phosphate moiety of the inhibitor is identical to that of the modeled phosphate anion in the NADH-bound structure, as opposed to the previously published 2-deoxy-D-glucitol 6-phosphate-bound structure (9), where the phosphate moiety of the inhibitor 2-deoxy-D-glucitol 6-phosphate sits in the opposite side of the enzyme active site where D356 is located in all newly determined structures. According to the new structural data, modeling of the substrate in its pseudocyclic conformation necessary for the aldol cyclization was performed, and a new mechanism proposed. The fact that 2-deoxy-D-glucitol 6-*(E)*-vinylhomophosphonate, a structural analogue of the substrate in its transoid conformation, is the most potent inhibitor synthesized so far strongly supports the mechanism where the phosphate monoester can act as the base in the enolization step to pull the *pro-R* hydrogen off the C6. From the crystallographic structural data, the possibility that MIP synthase could act as a type I aldolase could not be ruled out.

In concert, it can be concluded that MIP synthase experiences conformational changes at various stages during the reaction. At the start of each turnover, the enzyme is tightly bound to NAD^+ since NAD^+ is a prosthetic group, and the active site is disordered as seen in the NAD^+-bound structure. Once the substrate is bound in the enzyme active site, the entire active site folds to completely encapsulate the substrate within its active site. The nicotinamide changes to the conformation seen in the structures of the MIP synthase/NADH complex and the MIP synthase/NAD^+/2-deoxy-D-glucitol 6-*(E)*-vinylhomophosphonate complex. This motion of the nicotinamide brings the substrate C5 close to the nicotinamide C4 to facilitate the oxidation. All non-redox reactions occur within the same active site, NADH then reduces the cyclized reaction intermediate, *myo-*

2-inososose 1-phosphate to produce MIP, and the mobile region of the active site will open up to release the product as seen in the NAD^+-bound structure ready for the next cycle of turnover.

Further experiments are needed to verify the mechanism proposed above. These may include: (1) Elucidation of the structures of MIP synthase in complex with different structural analogues of the substrate as well as reaction intermediates, among which the structure in complex with one of the reaction intermediates, *myo*-2-inososose 1-phosphate, would be of particular interest; (2) Elucidation of the structures of MIP synthase in complex with the previously utilized inhibitors at a higher pH; (3) Elucidation of the structures of MIP synthase mutants, namely S323A, D356A, K369A, K373A, K412A, D438A, and K489A, in complex with the substrate or a substrate analogue; (4) Evaluation of MIP synthase activity in the presence of various divalent cations, including Zn^{2+}, Mn^{2+}, Mg^{2+}, Ni^{2+}, etc.

3.10 Literature cited

1. Jack, A., and Levitt, M. (1978) *Acta Crystallogr A* **34**, 931

2. Press, W. H., Flannery, B. P., Teukolosky, S. A., and Vetterling, W. T. (1986) *Numerical Recipes*, Cambridge University Press, Cambridge

3. Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P. J. (1983) *Science* **220**, 671

4. Metropolis, N., Rosenbluth, M., Rosenbluth, A., Teller, A., and Teller, E. (1953) *J. Chem. Phys.* **21**, 1087

5. Verlet, L. (1967) *Phys. Rev.* **159**, 98

6. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr. A* **47**, 110-119

7. Stein, A. J., and Geiger, J. H. (2000) *Acta Crystallogr D Biol Crystallogr* **56 (Pt 3)**, 348-350

8. Otwinowski, Z., and Minor, W. (1996) *Methods Enzymol* **276**, 307-326

9. Stein, A. J., and Geiger, J. H. (2002) *J Biol Chem* **277**, 9484-9491

10. Jones, T. A. (1985) *Methods Enzymol* **115**, 157-171

11. Brunger, A. T., Adams, P. D., Clore, G. M., Delano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1997) *Crystallography & NMR System, version 1.0*, Yale University

12. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr D Biol Crystallogr* **54 (Pt 5)**, 905-921

13. Navaza, J. (1994) *Acta Crystallogra. A* **50**, 157-163

14. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Cryst.* **26**, 283-291

15. Nicholls, A., Sharp, K. A., and Honig, B. (1991) *Proteins* **11**, 281-296

16. Nicholls, A. *Graphical Representation and Analysis of Structural Properties*

17. Norman, R. A., McAlister, M. S., Murray-Rust, J., Movahedzadeh, F., Stoker, N. G., and McDonald, N. Q. (2002) *Structure (Camb)* **10**, 393-402

18. Maeda, T., and Eisenberg, F., Jr. (1980) *J Biol Chem* **255**, 8458-8464

19. Tian, F., Migaud, M. E., and Frost, J. W. (1999) *J. Am. Chem. Soc.* **121**, 5795-5796

20. Carpenter, E. P., Hawkins, A. R., Frost, J. W., and Brown, K. A. (1998) *Nature* **394**, 299-302

21. Bender, S. L., Widlanski, T., and Knowles, J. R. (1989) *Biochemistry* **28**, 7560-7572