

TRANSCRIPTION FACTOR BINDING SITE DETECTION AND COMPARATIVE
MODELING AS DISCOVERY APPROACHES IN *PYROCOCCUS*

by

DARIN COWART

(Under the Direction of Robert A. Scott)

ABSTRACT

The molecular machinery of basal archaeal transcription is similar but simpler than that of eukaryotes, while many elements of archaeal transcriptional regulation resemble those of bacteria. Together, these factors make archaeal transcription as a useful model for studying transcription. This work describes two different computational methods to study different aspects of archaeal transcription, in particular within *Pyrococcus*. First, the UORdb database was developed that enables transcription factor binding site searches in promoter regions on a genome-wide scale through the enrichment of the search space with prospective binding sites. The UORdb database has been expanded as a web portal for quick and flexible TFBS searching (http://scottgroup.rcc.uga.edu/scottgrp/uor_search/py/home), using both a global internal search method and external motif finders such as BioProspector, CUBIC, and MEME. This database application was crucial in identifying the *Pyrococcus furiosus* transcriptional regulator protein, dubbed SurR for Sulfur-response regulator, and in validating some of its regulatory binding sites throughout the *P. furiosus* genome. Second, a comparative model was constructed for the *P. furiosus* RNAP complex composed of seven protein subunits and based on the 2.8Å resolution X-ray crystallographic structure of the eukaryal open-clamp *Saccharomyces cerevisiae* RNAPII complex. Despite limitations in the comparative model, comparison with archaeal RNAP structures, including the 3.4Å resolution *Sulfolobus solfataricus* RNAP structure confirm its accuracy in modeling an archaeal RNAP, particularly with the RNAP active site region and the dock domain, an interaction site for TFIIB crucial in the process of transcription initiation.

INDEX WORDS: Archaea, *Pyrococcus furiosus*, RNAP, transcription regulation, SurR, data mining, regulatory binding site, operon prediction, comparative modeling, homology modeling, motif, transcription factor, *Saccharomyces cerevisiae*, *Sulfolobus solfataricus*

TRANSCRIPTION FACTOR BINDING SITE DETECTION AND COMPARATIVE
MODELING AS DISCOVERY APPROACHES IN *PYROCOCCUS*

by

DARIN COWART

B.S., Florida State University, 1992

B.S., The University of Georgia, 2000

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial
Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2009

© 2009

Darin Cowart

All Rights Reserved

TRANSCRIPTION FACTOR BINDING SITE DETECTION AND COMPARATIVE
MODELING AS DISCOVERY APPROACHES IN *PYROCOCCUS*

by

DARIN COWART

Major Professor: Robert A. Scott

Committee: Liming Cai
Robert J. Woods
Ying Xu

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
May 2009

DEDICATION

To my parents Dan and Louise Cowart and my brothers Marlon and Monte for all the love and support they have shown me through my life and for teaching me the value of hard work.

ACKNOWLEDGEMENTS

First, I thank my research advisor Dr. Scott for mentoring me as a scientist and supporting my research, both in my undergraduate and graduate years at the University of Georgia. Dr. Scott constantly challenged my own expectations and led to me broaden my interests and skills far beyond what I thought I was capable, for which I will always be grateful.

I thank the faculty members of Department of Chemistry who have provided advice and direction in my development as a scientist, in particular Dr. Majetich, for professional development and support through the years. I also thank the members of my advisory committee: to Dr. Cai and Dr. Xu, for expanding my horizons in the area of bioinformatics; and to Dr. Woods, for introducing me to computational approaches in studying macromolecules and for providing the knowledge and means to develop tools to aid in that study. I thank Dr. Xu and Dr. Olman for the use of the bioinformatics tools and resources developed in their labs, most particularly for CUBIC and the use of operon predictions.

Many thanks belong to the members of the Scott research group, for advice and support through the years, especially for feedback during the development of the UORdb database and web application. In particular, I thank my friend and colleague Gina Lipscomb, a constant source of motivation and support throughout my career as a graduate researcher; her suggestions and feedback during the validation of SurR as a transcriptional regulator were an indispensable aid. I also thank my good friend Denny Elking, for inspiration in developing and using computational tools and expanding my knowledge of programming.

Last, I thank my family for all the moral support they have given me during the years.
Above all, I thank God for my life and that innate sense of curiosity that led me to science.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	v
CHAPTER	
1 ARCHAEOAL TRANSCRIPTION AND ITS REGULATION.....	1
1.1 Introduction to Archaea.....	1
1.2 Elements of basal archaeal transcription	2
1.3 Description of basal archaeal transcription initiation.....	5
1.4 Transcriptional regulators.....	6
1.5 Motif finding as a pathway to discovery	10
1.6 Molecular modeling of RNAP	12
2 THE UORDB DATABASE: A RESOURCE FOR PUTATIVE TFBS DETECTION.....	24
2.1 UORdb as a research tool.....	24
2.2 Goals and design philosophy.....	24
2.3 Selection criteria for searching UORs.....	26
2.4 UORdb search methods.....	27
2.5 Construction of UORdb.....	30
3 UORDB IN THE DISCOVERY OF REGULATORY TRANSCRIPTION FACTOR	
SURR.....	45
3.1 SurR, a sulfur-responsive regulatory transcription factor for <i>Pyrococcus</i>	
<i>furiosus</i>	45

3.2 Initial SELEX investigation of SurR binding sites	48
3.3 UORdb in SurR binding site identification throughout the <i>P. furiosus</i> genome	49
3.4 Further validation of UORdb search results and evidence of SurR binding sites	52
4 MOLECULAR MODELING OF <i>PYROCOCCUS FURIOSUS</i> RNA POLYMERASE	
(RNAP).....	76
4.1 Modeling an archaeal RNAP.....	76
4.2 Selection of modeling template	77
4.3 Selection alignments of <i>S. cerevisiae</i> RNAPII templates and <i>P. furiosus</i> RNAP	
subunits.....	78
4.4 Comparative modeling and assembly of model subunits into the RNAP	
complex	82
4.5 Evaluation of model quality	84
4.6 Comparison to archaeal structures	87
4.7 The conserved RNAP active site and dock domains.....	89
4.8 Conclusion of modeling exercise	92
5 CONCLUSION AND FUTURE APPROACHES.....	141
5.1 Two computational to aid transcriptional research	141
5.2 UORdb, a simple but versatile bioinformatics tool.....	141
5.3 Comparative modeling for archaeal RNAP.....	143
REFERENCES	145
APPENDICES	153
A Abbreviations.....	153
B UORdb input file examples	155

C	Sequence alignments.....	161
---	--------------------------	-----

CHAPTER 1

ARCHAEAL TRANSCRIPTION AND ITS REGULATION

1.1 Introduction to Archaea

Archaea are distinguished among the immense diversity of organisms on Earth by their distinctive and unusual features, leading to their emergence as a separate domain of life [1, 2]. Archaea are among the most prevalent forms of life on earth and found in a wide diversity of locales [3, 4]. They are probably best known for those members known as extremophiles. These resilient microorganisms have been discovered in harsh environments known for extremes in pH, salinity, and temperature: acid and alkaline waters, deep-sea volcanic vents, solfataric fields, saturated salt seas, boiling springs, Antarctic lakes, etc. [5] Their hardiness under such inhospitable conditions has sparked much interest within agricultural, chemical, and pharmaceutical industries, both in the organisms and their biochemistry, particularly with “extremozymes” [6, 7]. Interest in Archaea extends far beyond industrial applications of extremophiles. Archaea are among the contemporary organisms most closely related to an original prokaryotic ancestor. Originally categorized as “archaeobacteria”, Archaea are similar to Bacteria with respect to metabolic processes and basic morphology. However, many elements of their genomic information-processing machinery—including translation, transcription, and replication pathways—have significant parallels with those of eukaryotes [8, 9]. Archaea still share some characteristics of transcription and transcriptional regulation with Bacteria; as such, they provide a bridge between the different domains of life [10]. Together, these two aspects

make Archaea a simpler but useful model for studying transcription and may lead to better comprehension of more elaborate eukaryotic transcriptional mechanisms.

The archaeon we and others have studied is *Pyrococcus furiosus*, a hyperthermophile and a member of the genus *Thermococcus*. Hyperthermophiles are a divergent group of organisms—spanning 34 genera and 10 orders from Archaea and Bacteria—and are defined by their optimal growth temperatures of 80 °C or above, with growth inhibited at temperatures of 50 °C or below [11]. *P. furiosus* has been noted for its resistance to thermal breaks in its genomic DNA and its ability to retain chromosomal integrity at temperatures over 100 °C [12]. Discovered by Fiala and Stetter in hydrothermal marine sediments off the Italian coast, *P. furiosus* has an optimal growth temperature of 100 °C and is capable of using carbohydrates and peptides as carbon sources [13]. *P. furiosus*, and to a lesser degree its relatives *Pyrococcus abyssi*, *Pyrococcus horikoshii*, and *Thermococcus kodakaraensis*, are the primary focus of this work [14-16].

This dissertation describes the application of two different computational methods to investigate different aspects of transcription in *P. furiosus*. First, the mining of genomic sequence databases in search of potential transcription factor binding sites (TFBSs) can aid the discovery of regulatory networks. Second, the construction of a comparative molecular model of the *P. furiosus* RNA polymerase complex – using the eukaryal *Saccharomyces cerevisiae* homolog as a template – may provide insight into the structure and function of a key component of the transcriptional machinery.

1.2 Elements of basal archaeal transcription

Transcription is the synthesis of RNA using DNA as an information template. This process involves “reading” the sequence of DNA nucleotides and transcribing it into a

complementary sequence of RNA as the first step in protein synthesis. The primary engine in this mechanism is the DNA-dependent RNA polymerase (RNAP) holoenzyme that directs gene transcription in a site-specific manner. RNAP, together with the general transcription factors (TFs) and promoter elements of the genomic DNA, form the basal transcriptional machinery: the minimum components necessary for transcription to proceed. In addition to the basal transcriptional apparatus, other regulatory TF proteins may control the expression of some genes by either activating or repressing transcription, thereby influencing the protein composition of the cell. Transcriptional regulation provides a means by which the organism can maintain its internal make-up and plays a crucial role in cellular response to external stress. By examining these different elements of the transcriptional machinery, we gain greater insight to the biochemical workings of both a particular organism and life in general.

RNAP enzymes are massive multisubunit protein complexes (up to 0.6 MDa) known to all three domains of life but with some notable differences among them (Figure 1.1). Eukaryotes have three different RNAPs, each known for catalyzing a separate class of transcription: RNAP I (also known as Pol I) for the synthesis of ribosomal RNA (rRNA), the chief component of the ribosome; RNAP II (or Pol II) for the synthesis of messenger RNA (mRNA) which serves as the template for the protein product of translation; and RNAP III (or Pol III) for the synthesis of transfer RNA (tRNA) which is used in template codon recognition and is essential for protein translation. In contrast, Bacteria and Archaea have a single RNAP to catalyze all transcription products needed for the cell. The archaeal RNAP actually resembles the eukaryal RNAP II complex more than the bacterial RNAP, with most of its subunits having sequence similarity to eukaryal RNAP II subunits, inferring both homology and structural similarity. Archaeal RNAP

is composed of 11-12 protein subunits. For *Pyrococcus furiosus*, the subunits are A', A'', B, D, E', E'', H, K, L, M, N, and P with A', A'', B, D, and K constituting the RNAP core [17].

The less complex bacterial RNAP has a core complex of 5 protein subunits: α I, α II, β' , β'' , and ω . [18] These core subunits have counterparts among the eukaryal and archaeal RNAP complexes but have a lower degree of sequence and structural similarity with them [17].

Although these bacterial subunits form a minimal core capable of synthesizing RNA products, it still cannot initiate transcription from promoters. This requires an additional σ factor for specific promoter recognition, specifically by interaction within the promoter at the -10 and -35 nucleotide positions, relative to the gene's transcription start site (TSS) at +1 [18]. Most bacterial transcription is initiated by one σ factor. However, alternate σ factors may be activated in response to external stress, growth cycles, and changes in morphology, thus playing a role in transcriptional regulation. In addition, the σ factor is necessary for recruitment of the bacterial RNAP to the promoter [19].

In addition to RNAP, other proteins – such as the general transcription factors – are required for basal transcription initiation. In eukaryotes, there are at least five general transcription factors (TFIIB, -D, -E, -F, and -H) in addition to the enormous Mediator complex (>1 MDa) composed of over 20 protein subunits [20]. In contrast to this complex assembly, Archaea only require two general transcription factors for basal transcription to proceed: the TATA-binding protein (TBP) and transcription factor B (TFB). Both have homology to eukaryotic transcription factors: TFB to eukaryal TFIIB, and archaeal TBP to eukaryal TBP, the latter combining with TBP-associating factors (TAFs) to form the much larger complex TFIID [21]. Although not required for *in vitro* transcription, an additional transcription factor TFE – homologous to the N-terminal region of the α -subunit of eukaryal TFIIE – can enhance

recognition of promoter elements by the basal transcriptional machinery under certain conditions [22]. Figure 1.1 illustrates the similarities and differences in RNAP subunits and general transcription factors among the three domains of life.

The final component of the basal transcription machinery is the promoter, the region of DNA that expedites the transcription of a given gene (Figure 1.2). Both eukaryal and archaeal promoters share two common features: the AT-rich TATA box centered at -25 bases relative to the transcription start site (TSS), and the 7-bp purine-rich BRE (TBB Recognition Element) located ~2-3 bp upstream of the TATA box [23]. The TATA box and BRE serve as binding sites for the general transcription factors TBP and TFB. Both are key components in the formation of the preinitiation complex and subsequent recruitment of RNAP. Although not well conserved, the eukaryal TATA box has been defined by the consensus sequence TATA(A/T)AA(G/A) [24]. The archaeal TATA box sequence is even more varied; however, consensus sequences have been described for groups within Archaea including *Pyrococcus* [25]. The BRE fulfills several roles in archaeal transcription. The binding to BRE and the TATA box by TFB and TBP is the chief determinant of archaeal promoter strength [26]. In addition, the BRE-TFB interaction is the principal factor in specifying the orientation of archaeal transcription [27]. In Bacteria, promoter composition differs from the Eukarya-Archaea model, involving two 5-6 bp regions centered at the -10 and -35 positions (relative to the TSS); these interact with different domains of the σ factor as the first step of bacterial transcription initiation [18].

1.3 Description of basal archaeal transcription initiation

The mechanism of transcription initiation in Archaea resembles that of Eukarya. TBP first recognizes the TATA box, binding to the minor groove and imparting a ~65° bend in the

DNA; this allows for cooperative binding of TFB [28]. TFB interacts with both TBP and the BRE, both stabilizing the protein-DNA complex and determining the orientation of transcription [26, 27]. Together, TBP and TFB recruit RNAP to the TSS through interactions between the N-terminal Zn ribbon domain of TFB and subunits of the RNAP complex (specifically, subunit B in *Pyrococcus*) [29]. TFB and RNAP define the position of the TSS, 25-30 bp downstream of the TATA box [24]. In addition, TFE interacts with the E and F subunits of RNAP, possibly involving the region between the TATA box and the TSS; this further stabilizes the initiation complex [30]. Figure 1.3 illustrates some of the key events in the process of transcription initiation.

1.4 Transcriptional regulators

Although the components and the mechanism of archaeal basal transcription more closely resemble that of Eukarya than Bacteria, the known examples of transcription regulation in Archaea are bacteria-like: archaeal regulators often have homologous counterparts in Bacteria and share similar methods of activation or repression [31]. The most common DNA-binding protein motif found in prokaryotes is the helix-turn-helix (HTH) domain, including important variations such as the winged HTH [32]. Comparative analyses of gene products from archaeal and bacterial genomes predict the presence of many DNA-binding HTH/wHTH archaeal proteins; this infers a stronger link between archaeal and bacterial regulation than to eukaryal regulation [33].

Only a few archaeal transcriptional regulators have been described in depth, the bulk of which are known to be repressors of gene transcription [34]. The two observed mechanisms of action for archaeal repressors appear to be similar to that of bacterial repressors. The repressor

may bind to DNA sites that, in turn, prevent binding of the general TFs to the TATA box and the BRE. Alternately, the bound repressor protein may block recruitment of the RNAP complex to the TSS [35].

Among the earliest archaeal regulators described was the transcriptional repressor MDR1 of *Archaeoglobus fulgidus* [36]. This homolog of bacterial repressors DtxR and SirR is a metal-dependent autoregulator, repressing the transcription of its own gene. MDR1 binds in a metal-dependent fashion to its recognition site overlapping the TSS and blocking recruitment of RNAP in the presence of metal ions.

A number of archaeal regulators from the Lrp/AsnC family of bacterial transcriptional regulators have been discovered; this regulator family is primarily linked to the regulation of amino acid metabolism [37]. PF1601, a homolog of LrpA in *Pyrococcus furiosus*, is an autoregulator, repressing transcription of its gene by binding to DNA overlapping the TSS, thus abrogating recruitment of RNAP [38, 39]. Other Lrp-like regulators have been reported in Archaea (e.g., Lrs14 from *Sulfolobus solfataricus*) and they act as negative autoregulators. Unlike LrpA, the Lrs14 mechanism of regulation relies on interfering with TBP and TFB binding. The Lrs14 recognition site overlaps the BRE and the TATA box, thus preventing binding of TFB and TBP and the transcription of the Lrs14 gene [40]. Several other Lrp putative regulators have been reported, although these have not been strictly classified as activators or repressors. These include Sa-Lrp from *Sulfolobus acidocaldarius* [41], Ss-Lrp from *S. solfataricus* [42], and Ss-LrpB also from *S. solfataricus* [43].

Bacterial regulators of the Lrp family are known to be sensitive to small molecule effectors including a variety of amino acids [37]. The only archaeal Lrp-like regulator known to be responsive to a specific effector is LysM from *S. solfataricus*, a regulator linked to *lysWXJK*,

a gene cluster related to lysine biosynthesis. LysM binds just upstream of the BRE and TATA box region in the *lysW* promoter in the absence of lysine. LysM is most likely a transcriptional activator. In the presence of lysine, the *lysWXJK* cluster has lower expression levels *in vivo*; likewise, LysM has a lower binding affinity to the *lysW* promoter *in vitro* [44].

Some of the discovered archaeal regulators are unique to Archaea such as the transcriptional repressor TrmB. TrmB was first reported in *Thermococcus litoralis*, involving the repression of the *mal* operon – a gene cluster encoding the trehalose/maltose ABC transporter – with the sugars maltose and trehalose acting as regulation inducers [45]. A similar regulator system was found in *P. furiosus*, but acting on a gene cluster encoding a separate maltodextrin ABC transporter, with sugars such as malotriose acting as inducer but not trehalose or maltose. Thus, TrmB is a bifunctional transcriptional regulator capable of acting on two different promoters and under different conditions of induction [46]. The recently solved crystal structure of an N-terminal truncated derivative of TrmB reveals a unique sugar-binding fold [47].

Two other TrmB-like archaeal regulators have been reported: TrmBL1 from *P. furiosus* [48] and Tgr from *Thermococcus kodakaraensis* [49]. TrmBL1 is noteworthy for presumably serving as a global transcriptional regulator – acting on *P. furiosus* genes of sugar-transport systems and sugar-metabolizing enzymes [48]. These particular genes had previously been described as a putative regulon due to a common sequence motif in their promoter regions, now known to be part of TrmBL1 recognition site [25].

The paralogous regulator Tgr binds to the same sequence motif as TrmBL1. Experimental work with Tgr augmented the aforementioned *in vitro* research from *P. furiosus* by providing *in vivo* microarray expression data from an available genetic system, *T. kodakaraensis*. These experiments compared a Tgr-deletion strain of *T. kodakaraensis* with a wild-type under

glycolytic and gluconeogenic conditions. Under gluconeogenic conditions, the Tgr-deletion strain grew at significantly lower rates compared with the wild-type strain, while under glycolytic conditions, comparable growth rates were observed. Also noteworthy, the deletion mutant displayed defects in the activation of gluconeogenic genes under gluconeogenic conditions, thus inferring that Tgr can function as a repressor as well as an activator. Interestingly, a previously reported sequence motif, the Thermococcales glycolytic motif (TGM) is found in different positions relative to the TATA/BRE in gluconeogenic and glycolytic promoters. This system of position- and effector-dependent regulation can be extended to TrmBL1 and is the first case of positive and negative control in archaeal transcriptional regulation. Although this is common in bacterial regulation, the prevalence of this system among archaeal regulators is not known [50].

Phr, a heat shock response regulator in *P. furiosus*, exhibits an interesting mixture of bacterial and eukaryal features. Phr represses transcription of its own gene and of several heat shock response-related genes. It acts as transcriptional repressor by binding to its recognition site overlapping the TSS and blocking recruitment of RNAP [51]. The solved crystal structural revealed the presence of two distinctive domains with homology to Bacteria and Archaea. The N-terminal winged helix DNA-binding domain shows fold similarity with domains of the bacterial SmtB/ArsR regulator family, while the C-terminal antiparallel coiled coil helical domain is remotely homologous to the heat shock response-related eukaryal BAG domain [52]. Few archaeal activators have been described in depth in comparison to archaeal transcriptional repressors, but Ptr2 provides a notable exception. Ptr2, a member of the Lrp/AsnC regulator family, is known to activate *in vitro* transcription of the *rb2* and *fdxA* genes in the thermophilic methanogen *Methanocaldococcus jannaschii* by stimulating recruitment of TBP to the TATA

box [53-55]. Ptr2 homologs in *Methanothermococcus thermolithotrophicus* and *Methanococcus maripaludis* were also examined for indications of transcriptional activation. The *M. thermolithotrophicus* homolog showed activation compared to *M. jannaschii* Ptr2; the *M. maripaludis* protein required mutation of some of residues to match the *M. jannaschii* Ptr2 before transcription was activated [56]. Detailed analysis of the Ptr2 promoter architecture revealed that alignment of the upstream activating site was crucial for *in vitro* positive regulation and recruitment of TBP is the primary method of activation for Ptr2 [57].

1.5 Motif finding as a pathway to discovery

Numerous sequence motif detection methods have emerged in recent years that can aid in the identification of regulatory transcription factor binding sites (TFBSs). These methods typically search for high information content (IC) regions in the promoter regions of genes – more generally, in the UORs, or upstream of ORF (open reading frame) regions. This is based on the observation that regulatory TFBSs are evolutionarily conserved components of the underlying regulatory transcriptional machinery. [58]

A common strategy is to search a small set of promoter sequences from genes with similar expression profiles. Prospective binding sites are highlighted against a low IC sequence background, resulting in a consensus sequence motif for each predicted TFBS. This is not generally applicable to genome-wide motif detection because of the inherent noisy background of the input UORs. However, consensus motifs from prior searches on small search spaces of UORs can serve as input for other genome-wide search methods, including regular expressions to account for sequence variation. Selection criteria can also be used to enrich or highlight those UORs more likely to contain TFBSs. Those UORs containing prospective TFBSs—and their

corresponding genes (ORFs)—can be subjected to experimental validation, including high-throughput validation by genome-wide location analyses such ChIP-chip [59] or DIP-chip [60].

Experimental methods directly involving a putative or previously characterized transcription factor can also supply sequence motifs for genome-wide detection. Footprinting methods rely on enzymatic or chemical reactions that cleave or modify DNA. The change in the pattern of reactivity along the DNA sequence in the absence and presence of a DNA-binding protein is called the "footprint" of the protein on the DNA. For example, DNase I or hydroxyl radical produced by chemical reactions cleave DNA more or less non-sequence specifically. The presence of a TF across its BS will inhibit the cleavage in the BS region, leaving a "footprint" of missing cleavages in this region. SELEX (systematic evolution of ligands by exponential enrichment) can be used as a technique that identifies a “consensus” binding site for a given TF [61, 62]. The transcription factor of interest is used to select (usually by electrophoretic mobility shift assays, EMSA) high-affinity binding sites from a randomized library of DNA sequences. These selected sequences are then amplified, and several rounds of selection and amplification result in enrichment of only those DNA sequences that have recognized TFBSs. This sequence may not be identical to any of those found in true genomic binding sites but should closely resemble them.

The desire to combine genome-wide regular expression searches and UOR selection criteria with efficient motif-detection algorithms led to the construction of UORdb. This database is intended to provide quick, user-friendly motif searches within *Pyrococcus* genomes and can be extended to other archaeal and bacterial genomes. UORdb is discussed in detail in Chapter 2 of this dissertation.

1.6 Molecular modeling of RNAP

In studying transcription, it is important to remember the oft-repeated dictum: in order to understand function, we must understand structure [20]. As the largest component of the archaeal transcription machinery and the primary engine of transcription, the RNAP multisubunit complex provides an obvious target for structural investigation. Until recently, no structural information was available for any archaeal RNAP, except for the 1.75Å-resolution crystal structure of the *Methanococcus jannaschii* RNAP subcomplex E/F, an archaeal homolog of the eukaryal RNAP subunits RPB7 and RPB4 [63]. However, the sequence similarity between homologous archaeal RNAP and eukaryal RNAP II subunits ranges from ~25-75% [64], suggesting that a comparative model could be created using a known eukaryal RNAP II structure as model template. Structures for the *Saccharomyces cerevisiae* RNAP II (crystallized as a complex of ten subunits) have been solved in two forms with respect to the “clamp” module, a mobile portion of the complex formed by the N- and C-terminal regions of RPB1 and the C-terminal region of RPB2. The first form was determined at 2.8Å resolution and has the clamp in an open state (PDB code 1I50); the second form was determined at 3.1Å resolution with the clamp closed (PDB code 1I3Q). Figure 1.5 illustrates the 2.8Å RNAP II structure.

The construction of a comparative model of the *P. furiosus* RNAP involved the sequence alignment of its protein subunits with their homologous counterparts from *S. cerevisiae*, followed by molecular modeling using the lowest resolution *S. cerevisiae* RNAP II structure (1I50) as a structural template. The individual subunits were then refined and assembled into a complex suitable for comparison with other RNAP structures and for investigation of its internal structure.

Since the construction of this model, experimental structural information has become available for archaeal RNAPs. A cryoelectron microscopic structure of the *P. furiosus* RNAP has

been determined at 13Å resolution [65]. In addition, the first X-ray crystal structure for an archaeal RNAP – from *Sulfolobus solfataricus* – was solved at 3.4Å resolution [66]. Both are useful for comparing to the *P. furiosus* model and assessing its structural validity. Our *P. furiosus* RNAP modeling exercise is described in detail in Chapter 4 of this dissertation.

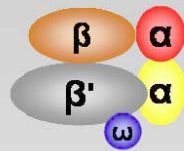
Figure 1.1 Comparison of basal transcriptional machinery across three domains of life.

Individual RNAP subunits (top) and general transcription factor (TF) proteins (bottom) are colored to indicate homologous relationships, with corresponding subunit/TF names as labeled.

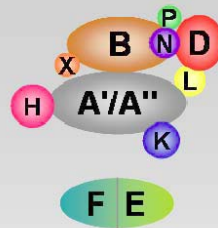
This comparison is limited to components required for *in vitro* basal transcription. Adapted from [17].

RNA
Polymerase

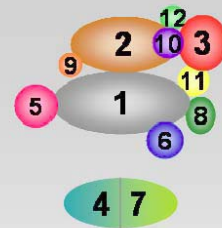
Bacteria



Archaea



Eukarya (Class II)



Transcription
factors



Figure 1.2 Core promoter elements of the three domains. The position of each promoter element relative to the transcription start site is indicated by the scale at the top of the figure. There are no distance indicators between the transcription start site and the translation start site because this distance varies by context. Sequence elements are shown with boxes and are indicated by their abbreviations: the BRE (TFB Recognition Element); the TATA Box (TBP binding element); and the -10 and -35 boxes (bacterial promoter elements.) Archaeal promoter elements are outlined with dotted lines to indicate minor deviations in sequence consensus as compared to the homologous elements in Eukarya. Figure adapted from [9].

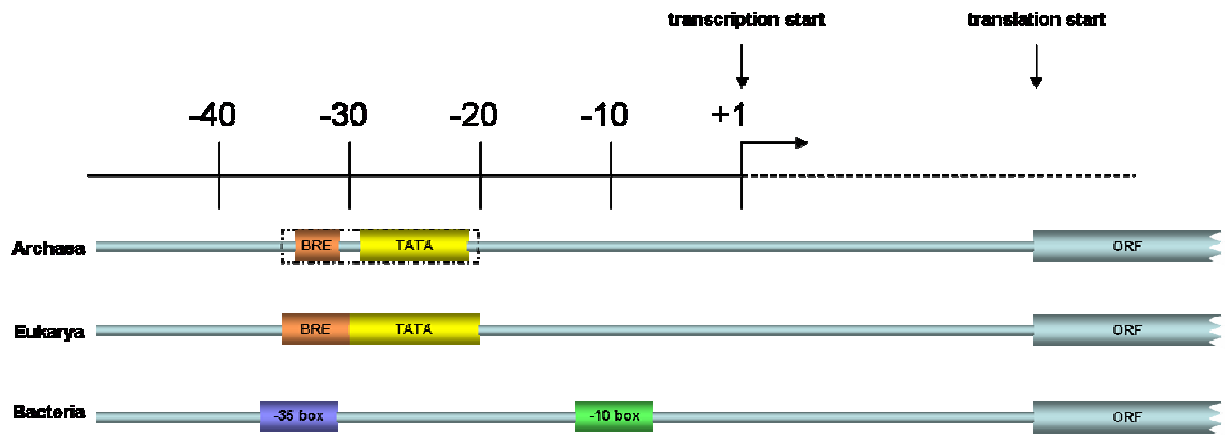


Figure 1.3 Process of archaeal transcription initiation. An archaeal core promoter (top) consists of a BRE (orange-brown box), TATA box (yellow box), and transcription initiation site (gray bent arrow). TBP (light blue) first binds to the TATA box, causing a sharp bend in the DNA, and TFB (yellow) stabilizes the ternary complex (middle). This ternary complex of TBP, TFB and promoter DNA recruits the 12-subunit RNAP to the transcription start site (bottom), forming the preinitiation complex. Adapted from [31].

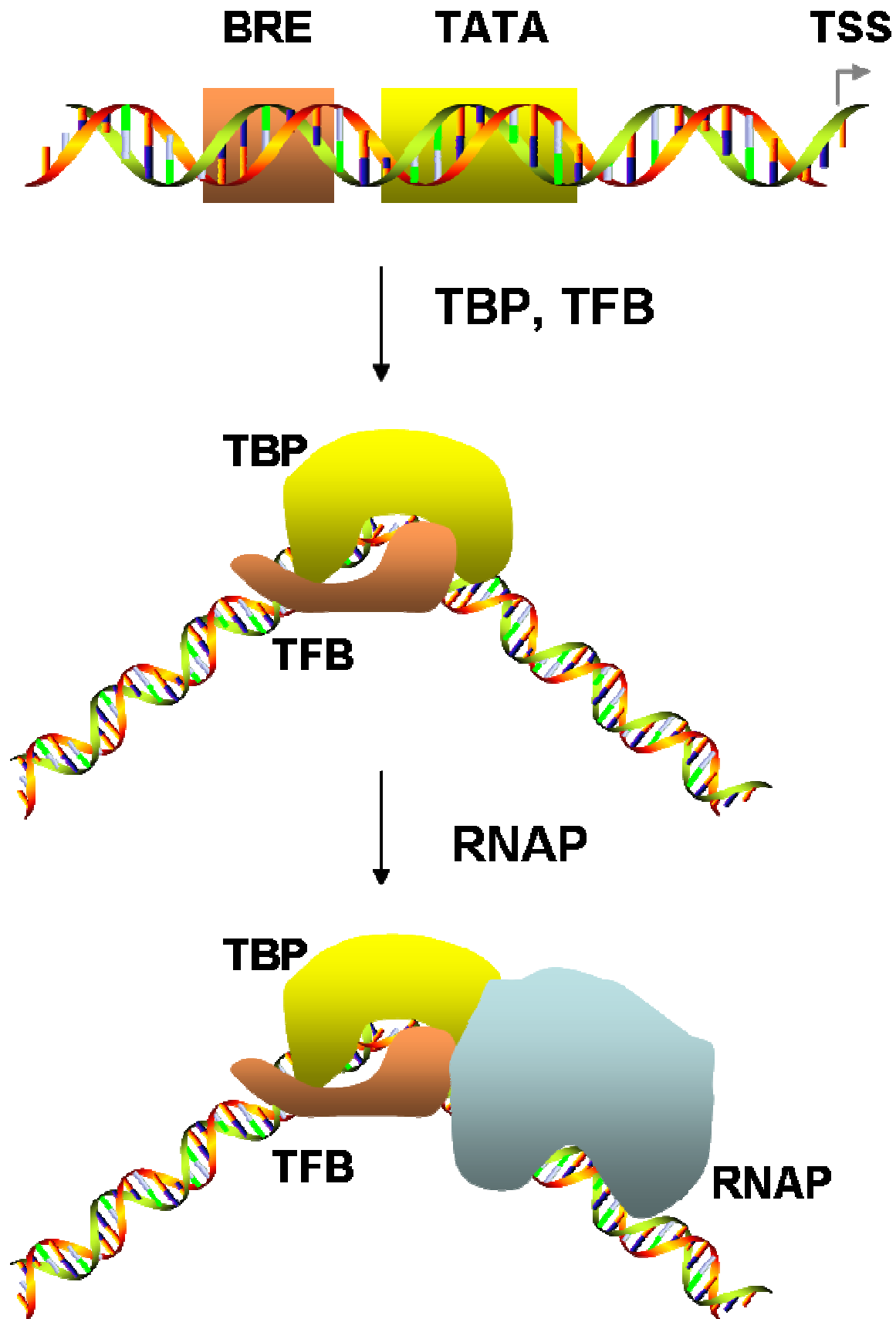
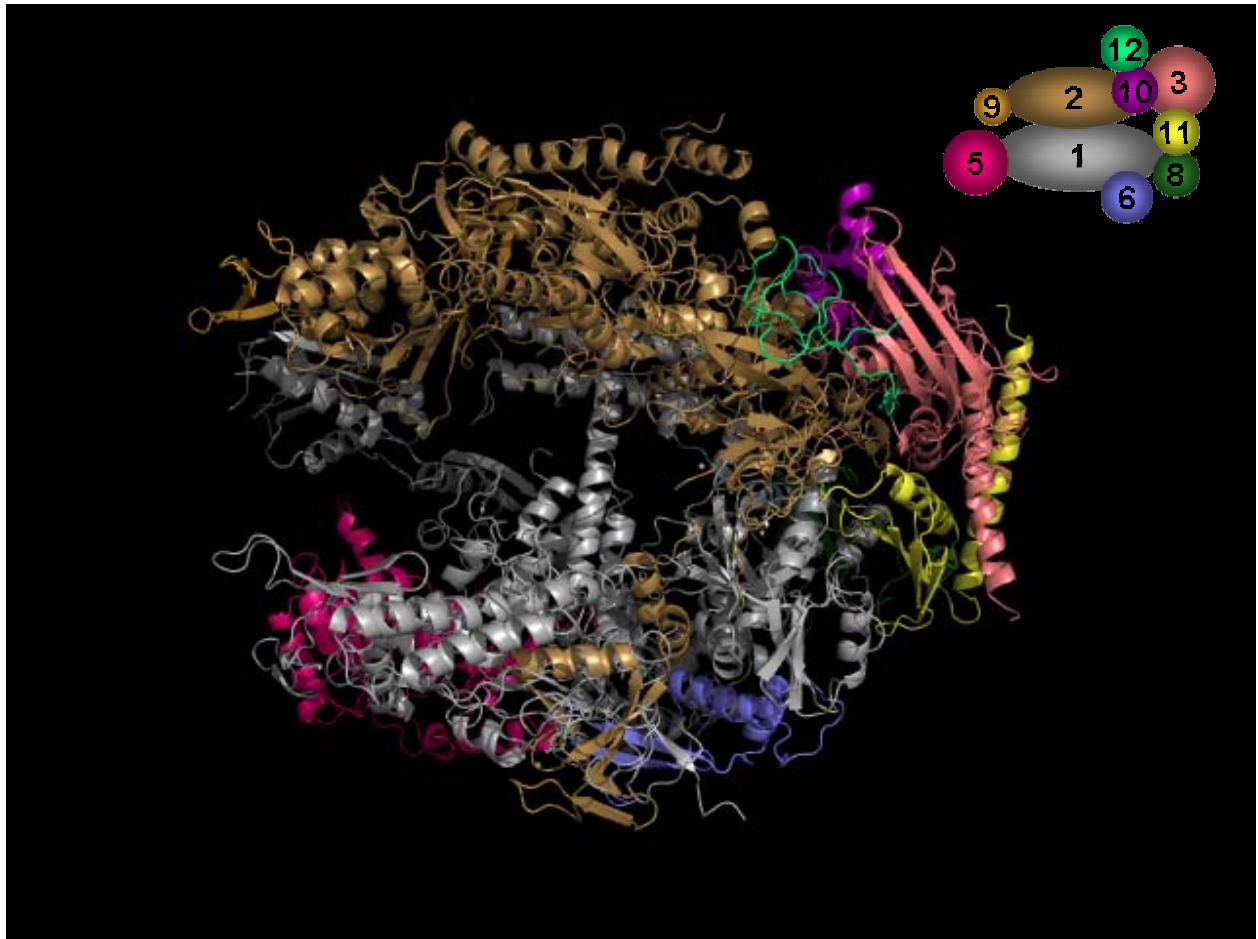


Figure 1.4 Archaeal mechanisms of transcriptional repression and activation. **A.** An archaeal core promoter consisting of a BRE (orange-brown bar), TATA box (yellow bar), and transcription initiation site (gray bent arrow). **B-D.** Diagrams depicting several possible archaeal transcriptional regulatory mechanisms with the regulatory transcription factor shaded in green, TBP shaded in yellow, TFB shaded in orange-brown, and their corresponding binding sites designated by bars of like colors beneath the DNA. The RNAP complex is shaded light blue. Two mechanisms for transcription repression (represented by a black arrow with red X) include: binding of the regulatory transcription factor to a binding site overlapping: **(B)** the BRE/TATA region thus preventing the binding of TBP and subsequently TFB; and **(C)** the transcription initiation site thus preventing the recruitment of RNAP by the TBP/TFB/DNA ternary complex. Transcription activation (represented by a black arrow alone) has been shown to occur through at least one mechanism: **(D)** binding of regulatory transcription factor to a position upstream of the TATA box and facilitating the recruitment of TBP to the core promoter. This accelerates the entire process of transcription initiation, thus increasing the rate of transcription from the promoter. Note that the $\sim 65^\circ$ bend caused by the binding of TBP to the TATA box is not illustrated in this figure. Adapted from [67].



Figure 1.5 Structure of *Saccharomyces cerevisiae* RNAPII at 2.8Å resolution. The complex is shown in cartoon representation, along with color-coded legend of the RNAP subunits at the top right. This structure does not include the RPB4/RPB7 subcomplex – homologous to archaeal RNAP subunits, E and F. Adapted from [20, 68] with protein structure created with PyMOL [69].



CHAPTER 2

THE UORDB DATABASE: A RESOURCE FOR PUTATIVE TFBS DETECTION

2.1 UORdb as a research tool

UORdb is a dynamically generated database created for researchers to search for prospective transcription factor binding sites (TFBSs) in DNA sequences upstream of ORFs ("upstream-of-ORF regions", UORs) on up to a genome-wide scale in a timely fashion. It relies on a core module written in Python [70] to organize genomic information to allow for motif finding with both internal and external search algorithms with a web interface implemented with `mod_python` [71]. The main focus for UORdb has been the clade of *Pyrococcus* genomes; however, UORdb is capable of searching other prokaryotic genomes and has recently been extended to include the bacterial *Shewanella oneidensis* MR-1. This chapter describes the design, creation, and operation of UORdb; example results are described in Chapter 3.

2.2 Goals and design philosophy

From its inception, UORdb was constructed to meet four central design goals. The first of these goals was to increase predictability in TFBS searches through the application of selection criteria. Specifically, our premise is that the UORs should be enriched to some degree with TFBS-containing UORs (i.e., we would filter out UORs less likely to contain TFBSs) by selection criteria so that the signal is not overshadowed by the background. These selection criteria are described in detail below.

The second design goal for UORdb was to provide an array of search methods for locating prospective TFBSs. Numerous motif-finding algorithms have emerged in recent years, most available as standalone applications. Most of the methods rely on descriptions of motif width but do not use a sequence of nucleotides as input. Moreover, the success of these methods is typically dependent on discerning high information content (IC) sequences against a low IC background. These algorithms can be applied to select UORs supplied by UORdb but are not generally conducive to searching genome-wide. Prospective TFBSs may be known to researchers from the literature or elucidated through SELEX or footprinting experiments. So an additional search method to search for specific nucleotide sequences was developed to augment the external search algorithms. This internal search method is especially suited for genome-wide searches because it does not require a low-IC background in the search space.

Another key design goal for UORdb was that it would require relatively little preparation of the input to build the database. Each genome to be included in the database will require three separate classes of input: a genomic sequence, a set of ORF definitions to delimit the endpoints – the start and stop codons - for each predicted gene, and a means to collate ORFs into predicted operons. These input files should be readily available and easily updated.

Finally, UORdb should be universally applicable to prokaryotic genomes. The only limitation is that each genome must have available the aforementioned classes of input. Genome sequences and ORF definitions are inherent to all annotated genomes, while operon predictions have become available for many prokaryotic genomes with advent of operon prediction databases [3-5].

2.3 Selection criteria for searching UORs

The intent of UOR selection criteria is to improve predictability for large-scale motif searching across genomes. Three general criteria were adopted in the construction of UORdb to increase the chances for successful TFBS hits. The first and simplest criterion essentially defines the UOR sequence at its most basic level: candidate sequences for TFBS searches should be sequences extracted immediately upstream of a given ORF up to -300bp with respect to the ORF translation initiation site (TIS). These regions are more likely to contain promoter elements of transcribed genes and possibly contain regulatory TFBSs. In general the locations of transcript start sites (TSSs) for genes are unknown, requiring the use of TIS for base numbering. Usually the defined UOR will not overlap with its corresponding ORF downstream; overlap with other ORFs upstream is allowed. Given that regulatory transcription factors bind to DNA, binding sites may be effective anywhere around the start of the coding region, so we allow the possibility that some TFBSs may be found just inside the ORF sequence. Therefore, BS searches are allowed up to +50bp with respect to the ORF TIS. The longest possible UOR would have the endpoints (-300, +50) with respect to the TIS. These endpoints are selectable parameters in the UORdb query page.

The second criterion for UOR selection is based on the set of ORF definitions for a given genome. UORdb accepts as input ORF definitions from among the genome databases GenBank [72], Refseq [73], and TIGR Comprehensive Microbial Resource (CMR) [74]. Each database employs different methods of TIS prediction and can result in a single ORF having three different predicted start codons (but always the same stop); thus, an ORF can potentially have three different lengths. Figure 2.1 provides one example, ORF locus SO_0134 from *Shewanella oneidensis* MR-1. Experimental validation of TISs in prokaryotic genomes is far from complete,

but computational estimates of TIS predictive reliability with different databases, particularly Refseq, reveal some variations in accuracy for some genomes. [75] This second criterion selects for UORs corresponding to the shortest possible ORF – the “minimal” ORF – from among the three possible ORF definitions. Provided that the true TIS for a given gene is not downstream of any of the three predicted TISs, the UOR will thus cover upstream sequence of any TIS.

The last criterion for UOR selection involves the use of operon predictions. We assume that TFBSs are more likely to be found in the upstream region of an operon, rather than upstream of ORFs within an operon (since there should be no promoters in these latter regions). This upstream region of the operon is the same as the UOR for the first ORF in that operon. If an ORF is not located in any predicted operon, then it is considered to represent an operon of one ORF. Unlike the first two criteria, the predicted operon selection criterion is not destructive. It highlights those UORs that are upstream regions for their corresponding operons but does not remove the UORs within operons from the search space. This option was chosen during the initial design of UORdb as a safeguard against suspect predictions for *Pyrococcus* operons, although the recent emergence of more reliable operon prediction methods has eliminated some of this concern [5, 10].

2.4 UORdb search methods

UORdb provides access to two general classes of TFBS search methods. The first is a method intended for users with some knowledge of a prospective TFBS sequence, possibly obtained from footprinting or SELEX experimental data. A motif search method based on regular expressions was adopted that accepts nucleotide sequences coded in accordance with IUPAC and IUB recommendations [76]. One useful advantage of regular expressions is that it

could be used for variable gaps between left- and right-side blocks of a suspected TFBS obtained through experimental data, e.g. ATTN{3,5}TAA (i.e., ATTnnnTAA, ATTnnnnTAA, or ATTnnnnnTAA). This in-house UORdb regular expression-based method is capable of searching both strands (via reverse complements) and recognizing overlapping motifs. Motifs are first parsed for IUPAC-IUB nucleotide symbols and translated into groups of the core symbols (A, T, C, G), then compiled by the regular expression engine before UOR searching. The use of IUPAC-IUB symbols allows for variability in the target motif: the results of these may be further refined or the related UORs submitted to external motif searchers.

Table 2.1 IUPAC-IUB Nucleotide Code Translation for UORdb motif input [76]

Nucleotide Code	Translation for Regular Expression Search
A	A
B	C or G or T
C	C
D	A or G or T
G	G
H	A or C or T
K	G or T
M	A or C
N	A or C or G or T
R	A or G
S	C or G
T	T
V	A or C or G
W	A or T
Y	C or T

The second class of UORdb search methods does not require any knowledge of sequence and includes motif finders external to the core Python module. Three widely used and reliable motif search algorithms – CUBIC, MEME, and BioProspector – are employed to search UORs.

These can be used individually or, perhaps more usefully, in parallel for comparison of motif prediction. The search space for an external motif finder is typically limited to those UORs suspected of having common TFBSs and may number no more than a few dozen. Most motif searching algorithms rely on the detection of high-IC regions against a low noise, low-IC background of nucleotide bases.

The first of these external search methods, CUBIC, treats all subsequences (defined by input parameters) as data points for clustering, using a distance function inversely proportional to sequence similarity among the different fragments [77, 78]. To identify clusters, a minimal spanning tree (MST) is constructed to represent these data, and this in turn is mapped into a linear sequence; data clusters correspond to subtrees of MST, essentially reducing cluster identification to a tree partitioning problem. Clusters corresponding to high-IC motifs will be detected as “valleys” in the final linearly sequential representation. Input parameters including motif length and length of IC-rich left motif subsequence allow the user to specify a prospective motif to be a single block or two IC-rich blocks with a defined internal gap sequence.

Another external search method, MEME, has long been utilized in the detection of high-IC sites in genome sequences [79]. MEME relies on the application of expectation maximization to a finite mixture model fitted to the input sequences. This model includes a motif component composed of similar fixed-width sequences and a background component of its complement in the sequences. The algorithm takes multiple rounds adjusting the parameters of the probabilistic model to find a maximum likelihood fitting the input data. MEME only requires a motif length parameter to search sequences; this can also be applied as a minimum and a maximum length for prospective motifs, making it useful for “blind” searches of UORs.

A third external search method, BioProspector, employs a Gibbs sampling strategy to detect motifs in genome sequences [80]. Using a background Markov model based on intergenic sequences from *S. cerevisiae*, a motif probability matrix is initialized based on a random alignment of the input sequences. Gibbs sampling improves the motif matrix in an iterative fashion until convergence is met, yielding the highest scoring motifs from the input sequences. Similar to CUBIC, BioProspector accepts input parameters defining either a single block-motif or a two-block motif with a variable internal gap sequence.

The two kinds of UORdb search methods can be used to complement one another and maximize discovery. One possible strategy is to first search UORs genome-wide with the in-house regular expression-based method, essentially casting a wide net for prospective binding sites. Those UORs containing motifs can then be input to the external search methods that focus on information content and may further elucidate the motif sequence. A motif sequence must be input to the in-house method, requiring some prior knowledge of the suspected binding site beforehand, possibly based on a known transcription regulator or from experimental data such as SELEX or footprinting. Another technique is to first search a small subset of UORs with the external motif search methods to obtain a prospective TFBS sequence. The found motif may then be translated into a regular expression to search genome-wide. Either strategy can be implemented to refine or expand on initial searches.

2.5 Construction of UORdb

As mentioned above, the core of UORdb is a module written in Python, an object-oriented interpreted language. The database is implemented by instantiating a UORdb object for a set of genomes, loosely-modeled on the python file object with an explicit close method for

properly closing all internal file-like resources. Initialization of UORdb requires two input sources for each genome: the genome sequence and a set of ORF definitions. The sequence file can be prepared from FASTA-formatted genome sequence files with suffix .fna from the NCBI ftp site (<ftp://ftp.ncbi.nih.gov/genomes/Bacteria/>); this file is reformatted for UORdb by removing the first title line of the file and removing all other whitespace from the remainder. The sequence file for each genome is opened as a read-only, shared memory-map object (via POSIX-compliant mmap) and allows for safe access by separate processes and reduces memory usage. The use of Python memory-mapped objects provides an additional benefit: they can be directly searched with regular expressions without the creation of an intermediate string object. A UOR is essentially a substring of the memory-mapped object. The endpoints for each UOR are taken from user input relative to its ORF TIS, e.g. (-200, +1) or (-250,+50). This allows the user some flexibility in quickly changing the UOR endpoints.

ORF definitions are provided by a tab-delimited input file containing fields for each ORF: organism (or genome), the definition origin (or database), the ORF locus, the ORF start position in the genome, the ORF stop position, and any annotation present. With the exception of the organism and database fields, this information can be downloaded from Refseq, GenBank, and TIGR/CMR in tab-delimited format. These can be collated into one file with a spreadsheet program; the organism and database fields can easily be updated, and the resulting spreadsheet can be output to deliver the tab-delimited ORF definition file. The Python standard module csv can quickly process this file and allow for immediate application of the second UOR selection criterion for minimal-length ORF definitions.

The genome sequence and the set of ORF definitions are minimal requirements for UORdb initialization. After initialization, operon predictions can then be applied as a lookup

table for each genome entered in the UORdb object. Although operon predictions provide one of the most important UOR selection criteria, the earlier lack of reliable predictions for all prokaryotic genomes led to their removal as a strict minimal requirement to initializing UORdb. A variety of operon predictions have been selected for use during the development of UORdb. Among the first of these was a prediction set based on the likelihood that a gene pair belong to the same operon using features such as intergenic distance and conservation of gene environment [81]; this prediction algorithm serves as a foundation for the database OperonDB [82]. A more accurate prediction set used by MicrobesOnline was adopted for use with UORdb, again based on gene pairs but incorporating phylogenetic information from COG [83, 84]. The latest and most reliable operon predictions used by UORdb have been provided by the DOOR database [85]. Its underlying algorithm includes additional features to those mentioned above [86]: similarity of Gene Ontology (GO) functional information between gene pairs [87] and frequencies of specific DNA motifs in intergenic regions. The DOOR operon predictions are given as gene pairs and group genes into operons by an identifier. A table indexed by the corresponding ORF locus of a given UOR allows for quick application to search results and evaluation by the user.

The top-level UORdb object is organized in a hierarchy, mirroring genome structure. At the pinnacle of the hierarchy, this database object acts as a container for Genome objects corresponding to each input genome. The Genome object contains the genome sequence as the memory-mapped file described above. This mid-level object also serves as a container for Orf objects which enclose all defined ORFs for its corresponding genome. These base-level objects contain the ORF locus, start and stop positions, annotation, and database source (i.e. Refseq, GenBank, or TIGR/CMR). After the initialization of UORdb, searches can be conducted with

the top-level UORdb interface by delegating each genome-specific search within its Genome object. Operon predictions can be applied to highlight any motif hits meeting the third UOR selection criterion.

UORdb was first implemented as a Python module for command line-oriented scripting. Once the top-level UORdb object is initialized with genome input, multiple searches could be conducted without the need for a SQL database backend. Although useful for UNIX-style command interpreters, a more user-friendly and web-accessible interface was introduced. To this end, the UORdb search page (http://scottgroup.rcc.uga.edu/scottgrp/uor_search/py/home) was constructed using mod_python and Apache HTTP Server 2.2.3 for dynamic webpage generation from a virtual server running Red Hat Enterprise Linux 5.3. The web interface allows the searches with the UORdb in-house method (Search by Sequence) or with the external motif finders (Search by Conserved Motifs). Search by Sequence defaults to genome-wide searching and includes parameters for the input motif sequence, direction of search (same direction as UOR corresponding ORF), choice of genome, and endpoints for UOR relative to the ORF translation start site (left-side range: -300 to -50, right-side range: +1 to +50) (see Figure 2.2). Search for Conserved Motifs requires the user to input UORs designated by corresponding ORF locus as text or options in a selection list extracted from all genomes available to UORdb; the form also allows the same selection parameters for UOR endpoints as Search by Sequence (see Figure 2.3). Separate fields for external motif finders – CUBIC, MEME, and BioProspector – include parameters specific to the underlying detection algorithm. All three have some combination of parameters for block and gap length or motif length (see Figure 2.4).

Motif hits for both pages are output in the form of HTML tables capable of sorting by column via ECMAScript (toggled by clicking the table header cells). Columns for these tables

include data such as UOR name (ORF locus), motif sequence, motif position relative to the genome and to the ORF for the UOR, annotation, its predicted operon, and the UOR sequence itself. Motifs are highlighted in the UOR sequence by coloring those found in the forward direction (5' to 3') red and those found in the reverse direction (3' to 5') green. Sorting by column allows the user to group together motif hits by one feature; for example, motif hits may be partitioned into subgroups by motif position relative to the ORF. Search for Conserved Motifs includes a separate table for each consensus motif returned by each selected external motif finder, along with a table for the motif hit score and other motif finder-specific data. Users may also download a CSV (comma-separated) version of the output for each table, which may easily be input to a spreadsheet program.

Figure 2.1 Discrepancies in TIS prediction in ORF definitions. SO_0134, an ORF locus from *Shewanella oneidensis* MR-1, is defined by the public databases Refseq, GenBank, and TIGR/CMR to have the same stop site – 142309 with respect to nucleotide position in the genome - but three different start sites: 141383 (Refseq), 141416 (GenBank), and 141212 (TIGR/CMR). The second UOR selection criterion selects for the minimum-length ORF definition from the three genome databases; in this case, GenBank provides the shortest of the ORF definitions. This choice of UOR (bordered in red and dashed – the UOR length is unspecified) increases the chances that it will contain elements of the true promoter region and putative TFBSs.

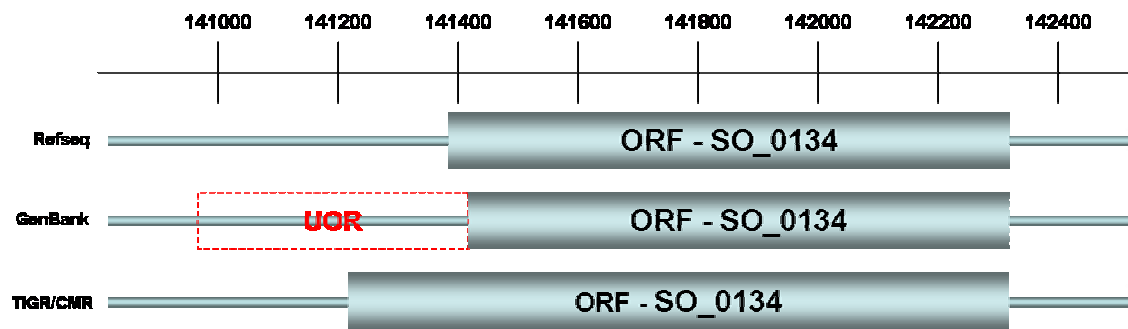


Figure 2.2 Web interface for searching UORdb by input sequence. The UORdb Search by Sequence web portal includes fields for the input motif sequence, direction to search with respect to the UOR and corresponding ORF, choice of genome (defaults to all genomes available to UORdb), and endpoints for the UOR sequence relative to ORF translation start site. Output will be returned as HTML tables below the form illustrated (see Figure 2.5 for example).

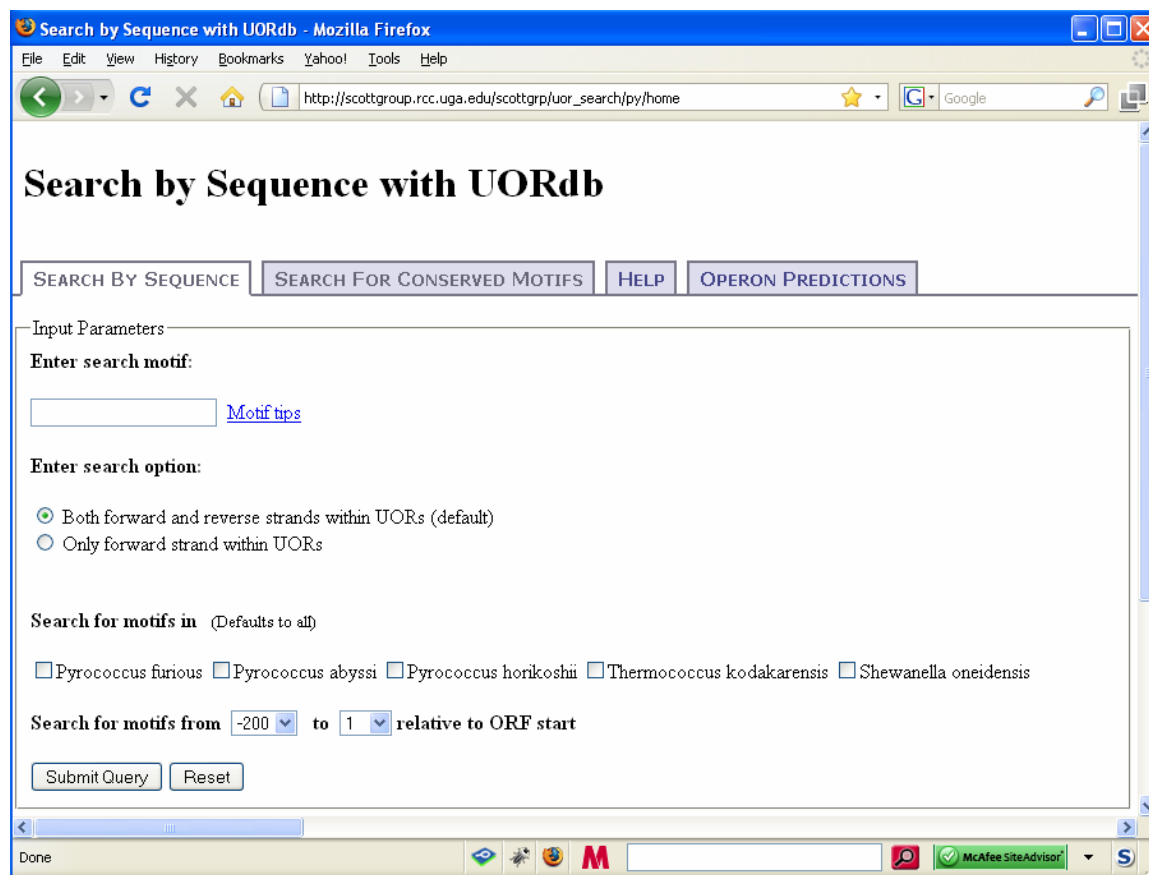


Figure 2.3 Web interface (top section) for searching UORdb with external motif finders.

The UORdb Search for Conserved Motifs web portal includes fields for those UORs to be searched (denoted by ORF locus) as either text or from a selection list along with selection lists for endpoints of the UOR sequence relative to ORF translation start site.

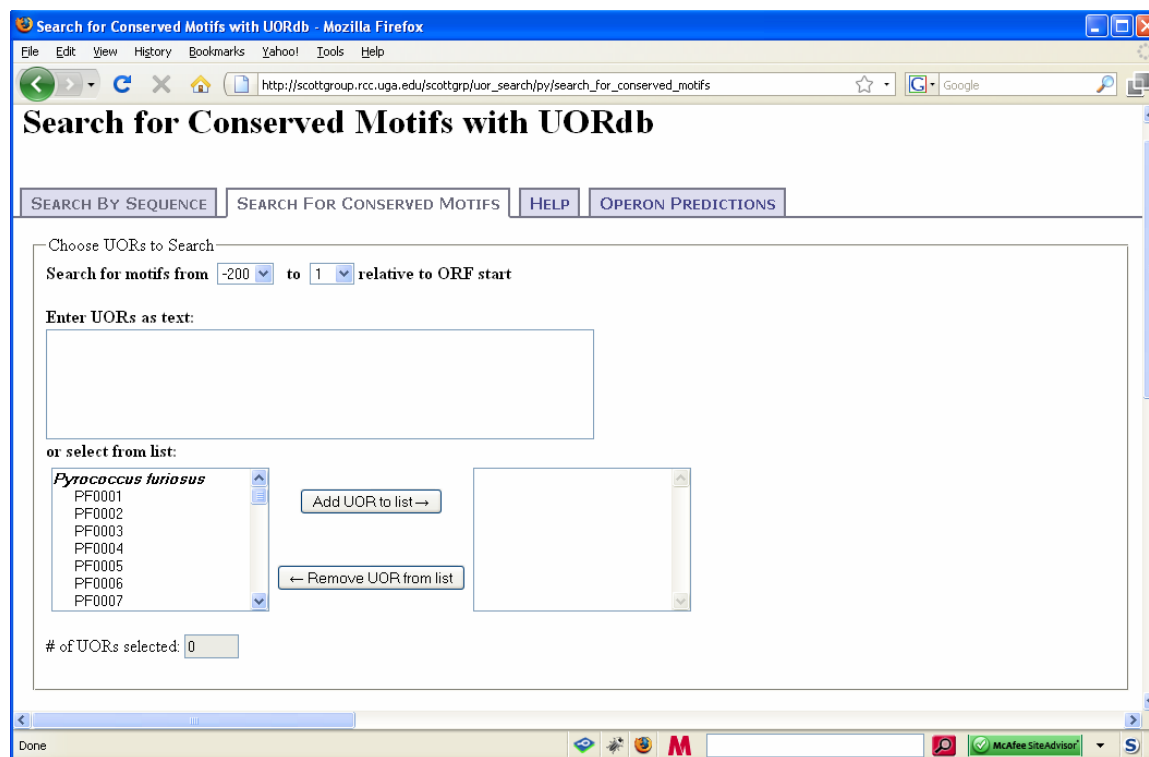


Figure 2.4 Web interface (lower section) for searching UORdb with external motif finders.

Motif finder-specific fields for CUBIC, MEME, and BioProspector include parameters describing motif block and gap lengths or total motif length. All fields are required for a given motif finder. Output will be returned as HTML tables below the form illustrated (see Figure 2.5 for example).

Search for Conserved Motifs with UORdb - Mozilla Firefox

File Edit View History Bookmarks Yahoo! Tools Help

http://scottgroup.rcc.uga.edu/scottgrp/uor_search/py/search_for_conserved_motifs

☒ Use CUBIC [Help](#)

Block1: 10
Gap: 0
Block2: 0

Motif Length: 10

Find 3 consensus motifs

of seq with motif 3 to 3

☒ Use MEME [Help](#)

Block1: Not used
Gap: Not used
Block2: Not used

Motif Length: 8 to 50

Find 3 consensus motifs

Palindromes only ☐ Yes ☒ No

☒ Use BioProspector [Help](#)

Block1: 10
Gap: 0 to 0
Block2: 0

Motif Length: 10 to 10

Find 3 consensus motifs

Figure 2.5 HTML output table for UORdb and Search by Sequence page. The output table includes columns for UOR name (designated by ORF locus), the specific motif found by the detection method, motif position relative to the ORF translation start site and relative to the entire genome, annotation, predicted operons, and the UOR sequence (only partially shown due to column width). Motifs are highlighted in the UOR sequence by color: red for those found in the same direction as the UOR, green for those found in the reverse direction as reverse complements.

UOR Name	Motif	Motif Coordinates (Relative to ORF) From	Motif Coordinates (Relative to ORF) To	Motif Coordinates (Absolute) From	Motif Coordinates (Absolute) To	Annotation	Predicted Operon	Sequence
PF0034	GTTGAGAAC	-169	-161	41144	41152	hypothetical protein	PF0034 PF0033	ATTATCAACTGCCACAACITTTACCCITTTGGGGTTGAGAACG
PF0254	GTTGAGAAC	-71	-63	263369	263377	hypothetical protein	PF0258 PF0257 PF0256 PF0255 PF0254 PF0253 PF0252	GGAGGAGGGTAAGAACTGACTGAAAAAGCCTGCAGACACTG
PF1795	GTTGAGAAC	-80	-72	1667079	1667071	sarcosine oxidase subunit alpha	PF1795 PF1796 PF1797 PF1798	AGTCTCTGAAGACGAAGATTGGAAAGTTATAAAGTATTTGC

CHAPTER 3

UORDB IN THE DISCOVERY OF REGULATORY TRANSCRIPTION FACTOR SURR

3.1 SurR, a sulfur-responsive regulatory transcription factor for *Pyrococcus furiosus*

One of the initial applications of UORdb was in the discovery of a regulatory transcription factor in *P. furiosus*. The translated protein of *P. furiosus* ORF locus PF0095, dubbed SurR for Sulfur-response Regulator, was first identified and characterized as a regulatory transcription factor by R. A. Scott laboratory colleague Gina Lipscomb as part of an experimental protocol to target putative transcription factors [88]. The first step of this approach was the identification of genes that are highly up- or down-regulated under specific growth conditions, in particular in the presence or absence of elemental sulfur (S^0), through the use of genome-wide microarray expression profiling experiments. With the premise that such gene regulation occurs at the level of transcription and that the promoter regions of highly regulated genes may contain regulatory transcription factor binding sites (TFBSs), the UORs of the corresponding genes served as probes for DNA affinity protein capture experiments, and thus used to “capture” regulatory transcription factors.

Examination of the microarray array expression data for *P. furiosus* genes up- and down-regulated when grown in the presence of S^0 revealed numerous targets for further investigation [89, 90], including PF1423, the leading ORF in the membrane-bound hydrogenase- (*mbh*-) encoding operon which was significantly down-regulated. The specific activities of the translated membrane-bound hydrogenase and related cytoplasmic hydrogenases are severely

reduced in the presence of S^0 , implying that these hydrogenases are not responsible for appreciable S^0 reduction under these conditions and leaving their exact role in S^0 metabolism unclear [91]. To better elucidate its role, the UOR of PF1423 was targeted for DNA affinity protein capture, a discovery technique described and developed by Scott laboratory co-workers Yu Chen [92] and Meiyao Wang [93]. This method involves immobilizing a biotinylated DNA target sequence containing the UOR (and promoter region) of the ORF of interest with magnetic beads coated with streptavidin. The immobilized DNA target acts as “bait” for capturing DNA-binding proteins when exposed to soluble cell extract. Any proteins not bound to the DNA fragment can be removed after the beads are isolated from the solution by a magnet. Heparin can be used as competitor for non-specific DNA-binding proteins; this filters the captured proteins to enhance the presence of sequence-specific DNA-binding proteins and should increase the chances for the discovery of transcription factors.

DNA affinity protein capture experiments using the *mbh1* probe (including the UOR of PF1423) and subsequent peptide mass fingerprinting [94, 95] of gel-separated captured proteins identified PF0095, annotated by GenBank and Refseq as a “conserved hypothetical protein.” Conserved Domain searches of the NCBI Conserved Domain Database were conducted to elaborate further on possible functions for PF0095 [96-98], with results shown in Figure 3.1. The best scoring conserved domain hit was for COG1777 from the Cluster of Orthologous Genes database [99]. This COG, annotated as “predicted transcriptional regulators”, is composed of 10 protein sequences from 8 members of Euryarchaeota. Additionally, a 56-residue region near the N-terminus of PF0095 provided another conserved domain hit with HTH_ARSR, a subfamily of prokaryotic helix-turn-helix regulatory transcription proteins and containing the regulatory transcription factor Arsenical Resistance Operon Repressor, ArsR [100]. A PSI-BLAST

(Position-Specific Iterated BLAST [101]) search of the NCBI non-redundant protein sequences databases (nr) resulted in 29 sequence hits with E-values below the threshold on the previous iteration and ranging from 1×10^{-99} to 0.004. Most of these sequences are from archaeal sources and are annotated as either transcriptional regulators or conserved hypothetical proteins. A distal tree of the PSI-BLAST search results may be found in Figure 3.2.

Lipscomb further validated the sequence-specific DNA-binding of PF0095 protein for the aforementioned *mbh1* DNA sequence [88]. The protein product of PF0095, SurR, was expressed and purified as both his-tagged and cleavable his-tagged recombinants from *E. coli* for use with *in vitro* EMSAs (electrophoretic mobility shift assays). For comparison of binding specificity, other DNA probes were included including promoter-containing sequences from *sipB* (from PF2026 which was strongly up-regulated in the presence of S^0) and *lrpA* (containing the promoter region of LrpA regulator [39]) along with ORF-only intragenic DNA sequences from *mbh1*. *In vitro* EMSAs revealed that SurR was shifted more completely with *mbh1* than either *sipB* or *lrpA*; moreover, SurR was shown to have far higher binding affinity with the *mbh1* promoter-containing sequence than with the corresponding intragenic probe. To localize the SurR binding sites in the *P. furiosus* genome, Lipscomb employed DNAase I footprinting [102, 103] to localize the SurR binding in its promoter region. It is noteworthy that PF0094 (*pdo*), the gene upstream of *surR* but oriented in the opposite direction, and encoding a protein disulfide oxidoreductase (PDO), was highly up-regulated in the presence of S^0 from the aforementioned DNA microarray results [89, 90]. Footprinting results from the DNA probe containing the *pdo-surR* intergenic region yielded a single 83bp footprint region immediately upstream from the *pdo* gene, from -4 to -86 on the *pdo* coding strand and from -8 to -90 on the opposite strand. Interestingly, when footprinting was applied to a probe derived from the UOR of the first ORF in

the *mbh1* operon, two separate footprints emerged. One site is proximal to the *mbh1* translation start, a 39bp region from -5 to -43 on the coding strand and a 30bp region from -10 to -39 on the opposite strand; further downstream lies a second footprint, a 53bp region from -107 to -159 on the coding strand and a -112 to -164 region on the opposite strand. Titrated SurR footprinting revealed that the longer distal site was occupied first, then the shorter proximal site, thereby demonstrating that the distal site has a greater binding affinity for SurR than the proximal site.

3.2 Initial SELEX investigation of SurR binding sites

The expression and purification of SurR allowed for the use of SELEX (briefly described in Chapter 1) to elucidate a consensus DNA binding sequence for SurR [61, 62]. SELEX began with a pool of 30bp synthetic DNA sequences, centrally randomized except for primers at the 5' and 3' ends. Upon exposure to SurR, those DNA sequences that were specifically bound to SurR were selected by EMSA, then amplified by PCR and purified for the next round of selection. Subsequent rounds of selection were applied to enrich the selection pool with DNA sequences of higher binding affinity for SurR. After the concluding round of selection, the remaining DNA sequences were cloned and sequenced, and motif detection software was employed to determine a consensus binding motif. The initial application of SELEX to SurR used six rounds of selection to yield 19 DNA sequences. From this final pool, Lipscomb employed MEME [79] to elucidate the consensus motif GTTN₃AAC. The motif was most prominent in 11 of the final 19 sequences (see Figure 3.3), although only six contained the perfect palindromic consensus sequence. Further selection rounds may have obtained a stronger motif with less variability in each motif position; however, it was less desirable to obtain a consensus sequence of the highest

binding affinity than a more variable motif that might reflect the actual binding sites found in the *P. furiosus* genome.

However, multiple occurrences of the perfect palindrome GTTN₃AAC were found within the earlier obtained DNAase I footprints of the *surR-pdo* and *mbh1* DNA probes: two within the 83bp *surR-pdo* footprint and one each for both *mbh1* footprints. Closer examination of the footprints revealed that one motif in the *surR-pdo* footprint and both motifs in the *mbh1* footprint have nearby inverted repeats that may be involved in SurR binding, as described by the sequence GTTN₃AACN₅GTT. With both the *surR-pdo* and *mbh1* DNA probes, multiple SurR-DNA binding complexes of differing electrophoretic mobility had been observed. This led to speculation that SurR binding sites may have additional degenerate motifs with multiple occurrences of binding with SurR, perhaps in a cooperative fashion to affect transcription of the affected genes. Although the consensus sequence GTTN₃AAC appeared to be a necessary portion of the SurR binding site, additional DNA regions may be a sufficient requirement for specific SurR-DNA binding.

3.3 UORdb in SurR binding site identification throughout the *P. furiosus* genome

UORdb provided the opportunity to search for putative SurR binding sites on a genome-wide scale. Using the consensus sequence obtained from the SELEX results as the input motif to the UORdb Search by Sequence portal with UOR limits set to [-200, +1], the search application found 148 UORs with the GTTN₃AAC motif with 123 non-redundant motifs (see Table 3.1 below). Although this is a large number of occurrences for the 9bp palindrome, it is far smaller than the 944 motifs obtained from searching the *P. furiosus* genome without any restriction to UORs.

It is doubtful that all 123 motif occurrences are actually true SurR binding sites; however, sequence information gleaned from the footprinting experiments may enable searches for a more extended motif containing other sequence features involved in SurR-DNA binding. The earlier-described motif GTTN₃AACN₅GTT found within the *mbh1* and *pdo-surR* footprints contains an additional right-side block positioned 5bp from the right-side block of the original SELEX-determined motif. This extended motif is essentially a fusion of the palindromes GTTN₃AAC and AACN₅GTT, raising the possibility that adjoining SurR motifs separated at regular gap lengths bind to multiple SurR proteins, as suggested by the different-sized footprints observed in the *mbh1* and *surR-pdo* footprints. A new global search with this extended motif sequence as input yielded 23 UORs with 20 non-redundant motif occurrences (see Table 3.2 below.) The resulting sequences were further submitted to MEME [79] by the UORdb Search by Conserved Motif webpage; the resulting sequence logo (created by WebLogo [104]) contains the GTTN₃AACN₅GTT motif most prominently in the MEME-determined consensus sequence GTTN₃AACN₃ANGTT. 20 of the 23 UORs are predicted to be UORs of the first ORF of its predicted operon, meeting the third UORdb selection criterion and thus more likely to contain transcription factor binding sites. Of additional interest is that 16 of the 23 UORs form 8 groups of divergently oriented gene pairs which share (at least) partially overlapping UORs. Lipscomb also found a strong correlation between the ORFs corresponding to these 23 UORs and those ORFs affected within 10 min after S⁰ addition in kinetic experiments from the aforementioned DNA microarray analysis [90] which correspond to a primary response of *P. furiosus* to the presence of S⁰. In addition, none of these 23 ORFs matched those ORFs that were implicated in a secondary response to S⁰ (those ORFs affected after 30 min after S⁰ addition). Together, these

provide more evidence for the role of SurR in gene regulation by inferring that SurR is directly involved in the primary response of *P. furiosus* to the presence of S⁰.

The UORdb searches with the GTTN₃AACN₅GTT motif reveal a pattern of 3bp-length IC-rich blocks separated between alternating discrete intervals of low IC gaps (3₋₃-3₋₅-3, where subscripted numbers represent low-IC sequence). Although the GTTN₃AAC motif was the detected consensus sequence because of its relatively high levels of information content, it is possible that the AACN₅GTT motif is as relevant to SurR-DNA binding even though it exhibits more sequence variability than the GTTN₃AAC motif. In addition, the multiple SurR-DNA complexes observed in the *mbh1* and *surR-pdo* footprinting EMSAs together, suggest that multiple degenerate sites may be necessary for SurR to affect gene transcription. It was originally proposed that the GTTN₃AAC sequence may be repeated at 5bp intervals in the SurR binding sites; however, a more appropriate description may be that the AACN₅GTT motif is repeated at 3bp intervals. This may be more sensible physically for protein-DNA binding: the length of this motif plus the 3bp interval is 14bp, which is greater than the length required for a full turn (10bp) in B-DNA; thus, it is possible that multiple SurR protein bind to the DNA binding sites by wrapping around the central DNA axis adjacent to one another. A genome-wide search of *P. furiosus* UORs using the UORdb Search by Sequence webpage results in 132 UOR hits with 110 non-redundant occurrences of the AACN₅GTT motif (see Table 3.3), which is comparable to the GTTN₃AAC search results (148 UOR hits with 123 non-redundant motif occurrences.)

The intersection of both UORdb search results for GTTN₃AAC and AACN₅GTT with respect to the same UOR name resulted in 27 UORs containing some combination of the two motifs (see Table 2.4.) These encompass the earlier UORdb search results for

GTTN₃AACN₅GTT but also include UORs for PF0443, PF1684, PF1911, and PF1912, the last two corresponding to a pair of divergently oriented ORFs. 22 of these 27 UORs meet the third selection criterion for UORdb searches and are directly upstream of the first ORF in their predicted operon. In light of the multiple SurR-DNA binding complexes observed in the *mbh1* and *surR-pdo* EMSAs, these reveal a more complete picture of the structure of true SurR binding sites: they may include multiple occurrences of either GTTN₃AAC or AACN₅GTT alone (such as PF1453), a “fusion” of the two motifs as GTTN₃ACCN₅GTT (such as PF1423), or a non-overlapping combination of the two motifs separated at discrete intervals (such as PF1911.) UORdb was able to provide immediate search results as well as a target list of ORFs for experimental validation for SurR binding sites. The UORdb selection criteria, particularly the use of operon predictions, infuse a reasonable level of confidence in initiating experimental techniques such as DNAase I footprinting.

3.4 Further validation of UORdb search results and evidence of SurR binding sites

Footprinting and EMSAs earlier confirmed that PF0094 (using the *surR-pdo* DNA probe) and PF1423 (using the *mbh1* DNA probe) contained true SurR binding sites. Other targets for binding sites focus on those that were known to be regulated in the presence of S⁰ by microarray analysis [89, 90]. The target set included the significantly down-regulated ORFs PF0559, PF0891, and PF0531 and the highly up-regulated ORFs PF1186, PF1453, and PF2051. PF0559 is annotated as “hydrogenase expression/formation regulatory protein”. Although not currently predicted to be at the start of its operon, it had been previously; it was also one of the most down-regulated ORFs in the *P. furiosus* microarray experiments [89]. PF0891 is the leading ORF in its predicted operon, which is responsible for encoding a bifunctional sulfhydrogenase

[105, 106]. PF0531 is predicted to be the leading ORF in its operon and is annotated as “cobalamin biosynthesis protein m”; two ORFs in its operon (PF0529 AND PF0530) are known to be significantly down-regulated in the presence of S^0 . PF1186 has been recently studied and is known to encode the protein NAD(P)H elemental sulfur oxidoreductase, dubbed NSR [90]. PF1453 is the first ORF in a 14-ORF operon which encodes for a 14-subunit transmembrane complex known to be associated with a Ni–Fe hydrogenase [107]. PF2051 is annotated as an “ArsR transcriptional regulator” which is implicated in the secondary S^0 -response of *P. furiosus* [90].

Lipscomb ran EMSAs for UOR DNA probes of the down-regulated ORFs PF0559, PF0891, and PF0531 revealing the presence of four-to-five, three, and two SurR-DNA complexes, respectively, while those for the up-regulated ORFs PF1186, PF1453, and PF2051 displayed two, six, and three protein-DNA complexes. All these were indicative of specific binding by SurR. The UOR of PF1453, which is known to contain two occurrences of the high-IC GTTN₃AAC motif separated by 35bp, gave the most complexes, comparable those obtained for the *mbh1* and *surR-pdo* DNA probes. DNAase I footprinting for PF0559, PF0891, and PF0531 yielded footprints of lengths 73bp, 42bp, and 32bp, respectively, while the footprints for PF1186, PF1453, and PF2051 had lengths of 53bp, 95bp, and 54bp, respectively. These results confirm a correlation between the number of protein-DNA complexes and footprint length.

Another target of interest, PF1911, was not listed as a S^0 -regulated ORF in the S^0 regulation microarray experiments with *P. furiosus* [89, 90]. PF1911 contains two putative SurR binding motifs (GTTCACAAC positioned at -46bp from the PF1911 start site and AACCTTTGGTT positioned at -80bp from the PF1911 start site) and is predicted to be the first ORF in its operon. PF1911 may actually be up- or down-regulated in the presence of S^0 , but it

simply may not meet the threshold for detection of regulation. PF1911 is annotated as “ferredoxin-NADP(+) reductase subunit alpha” by Refseq and as a “putative hydrogenase” by GenBank. The prior-established link between SurR and the *mbh1* operon, which encodes the membrane-bound hydrogenase, implies that, while it may not be significantly S^0 -regulated, nonetheless it may contain true SurR binding sites and should be subject to experimental validation. Lipscomb conducted EMSA with the UOR sequence (containing the promoter region) of PF1911 and showed a strong electromobility shift along with four to five separate SurR-DNA binding complexes. In addition, DNAase I footprinting revealed a 65bp footprint, from -33bp to -97bp relative to the PF1911 start site. Moreover, the footprint contained four 3-5-3 motif blocks corresponding to the AACN₅GTT motif (but not all exactly matching that sequence and thus not detected by UORdb or other motif finders.) These may correspond to the multiple protein-DNA complexes observed in the EMSA. It should be noted that if PF1911 is not subject to S^0 -regulation and is not related to the *P. furiosus* primary S^0 -response, this may represent another role for the SurR regulator. In time, more experimental validation of the UORdb motif hits will complete the story.

These validation experiments give some confidence in UORdb as a bioinformatics tool. Input motifs can be quickly modified between searches to accommodate putative extended (or even shortened) motifs. Searches with different motifs such as GTTN₃AAC and AACN₅GTT can be outputted in CSV formats and easily combined or intersected in spreadsheet programs. UOR selection criteria, particularly those involving operon predictions, were an aid in highlighting UORs likely to contain true binding sites; however, these depend on the strength of the original prediction and should be used with caution. It should be noted that supplemental

information from experiments such as the S^0 regulation microarray expression data are particularly helpful in targeting genes for investigation of putative binding sites.

Figure 3.1 NCBI Conserved Domain search for PF0095. The best scoring conserved domain hit was for the 79-bp COG1777 from the COG database [99], annotated as “predicted transcriptional regulators”. The other conserved domain hit included a 56-bp region near the N-terminus of PF0095 with HTH_ARSR, a subfamily of prokaryotic helix-turn-helix regulatory transcription proteins and containing the regulatory transcription factor Arsenical Resistance Operon Repressor, ArsR [100].

Figure 3.2 Distal tree for NCBI PSI-BLAST search results for PF0095. A PSI-BLAST (Position-Specific Iterated BLAST [101]) search of the NCBI non-redundant protein sequences databases (nr) resulted in 29 sequence hits with E-values below the threshold on the previous iteration and ranging from 1×10^{-99} to 0.004. The distal tree for the search results reveals many related sequences annotated as transcriptional regulators, particularly from the ArsR family [100].

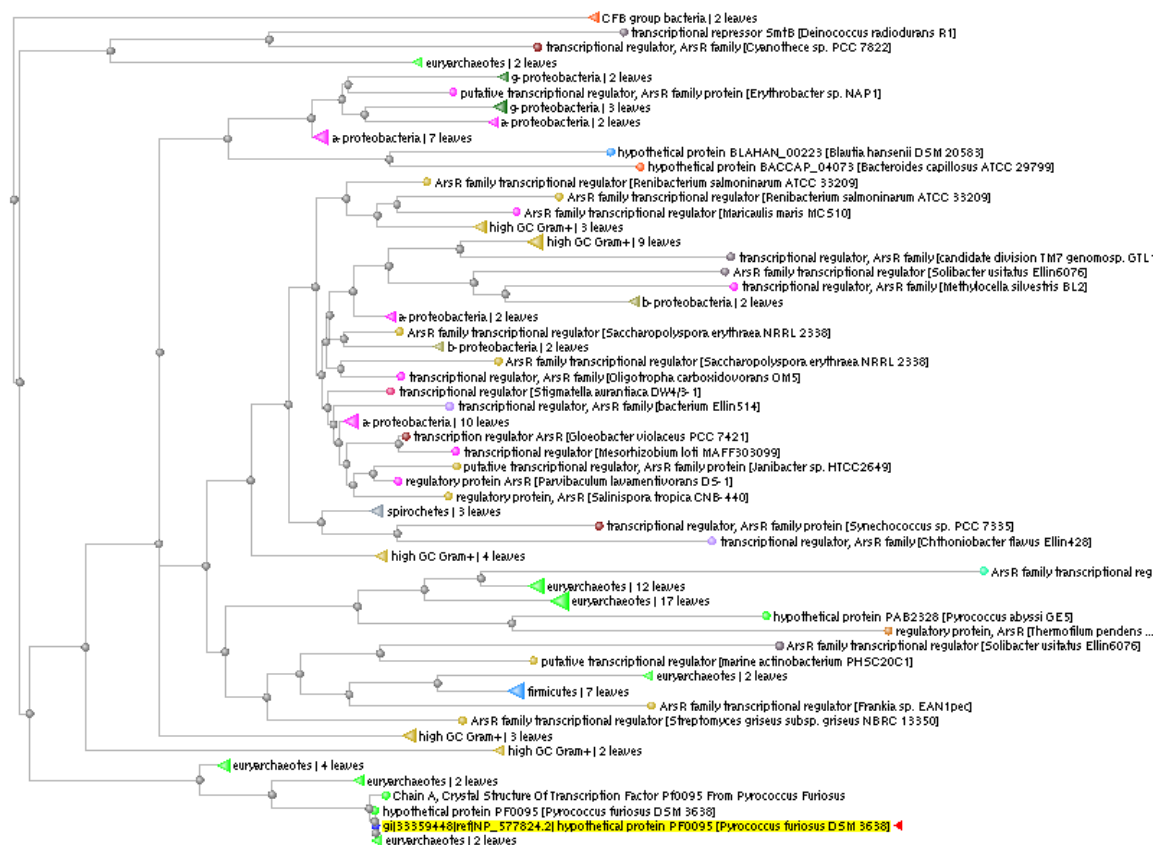
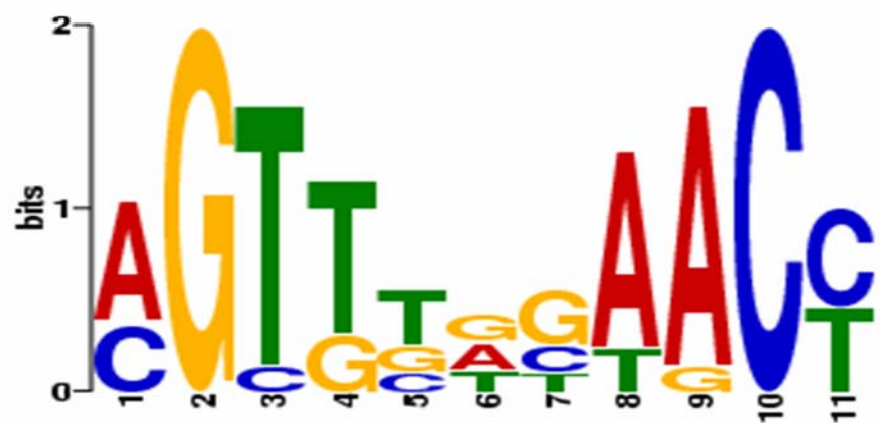


Figure 3.3 SurR consensus binding motif as characterized by SELEX. A. Consensus sequence logo (created by WebLogo [104]) for SurR binding site motif as elucidated by MEME [79] from the resulting 19 SELEX DNA sequences. **B.** Multiple alignment from the MEME results (generated by Jalview 2.4 [108]) The most prominent occurrences of the palindromic GTTN₃AAC motif were found in this subset of the SELEX DNA sequences.

A



B

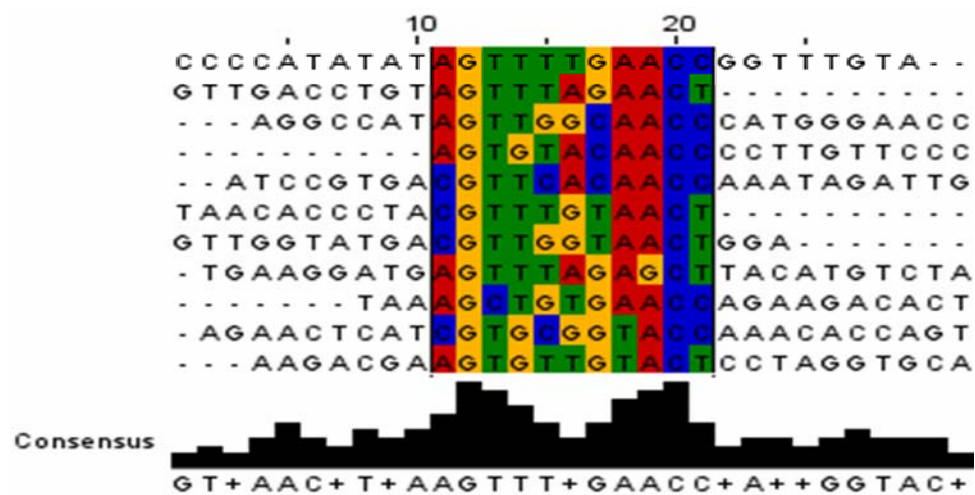


Figure 3.4 Extended binding motif for SurR. **A.** Examination of the DNAase I footprinting results with the *surr-pdo* DNA probe revealed an extended putative binding motif for SurR, represented here as GTTN₃AACN₅GTT. Motif occurrences found by UORdb (for 200-bp length UORs) were further submitted to MEME [79] with the resulting sequence logo (created by WebLogo [104]). **B.** Multiple alignment from the MEME results (generated by Jalview 2.4 [108].)

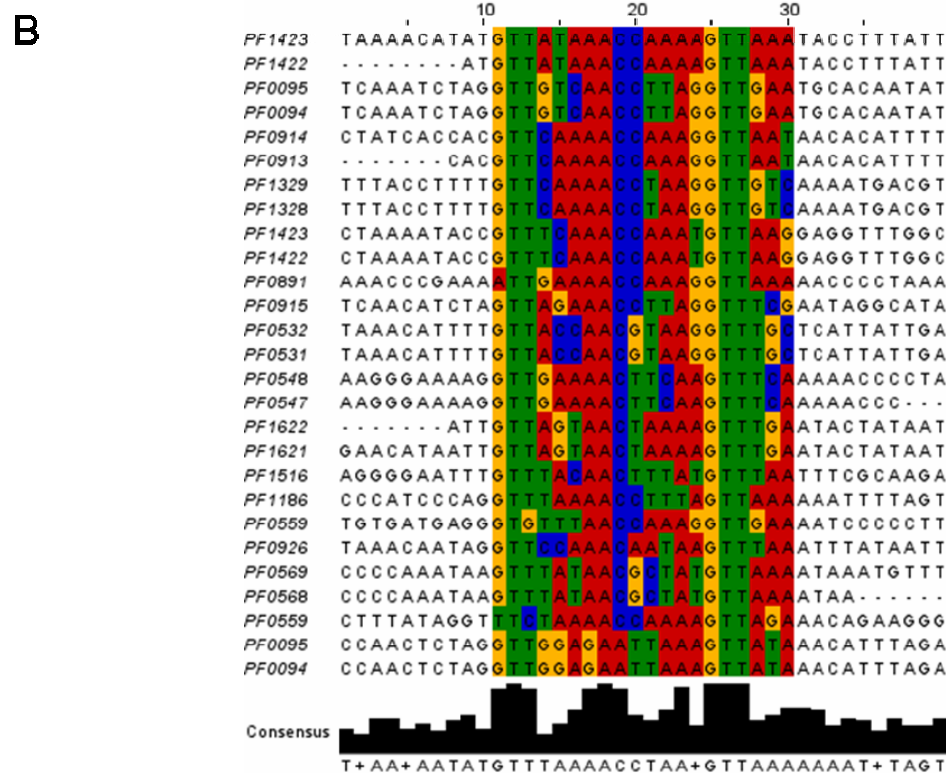
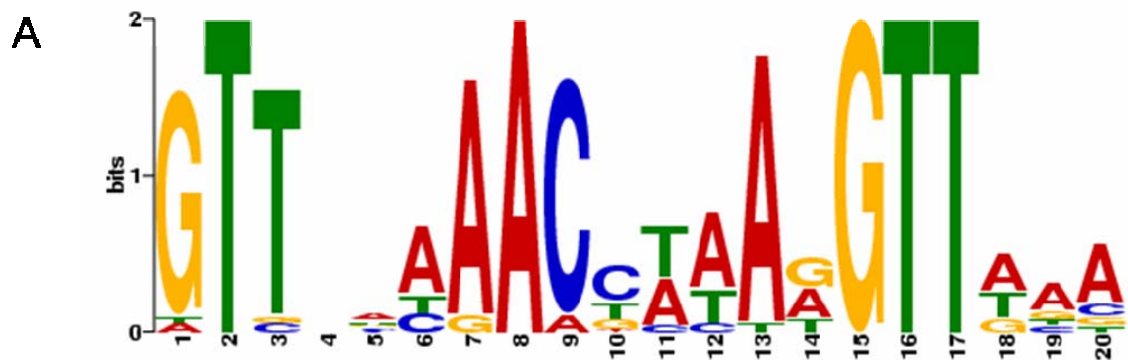


Table 3.1 UORdb Results for GTTN₃AAC Motif

UOR Name ^a	Motif	UOR Start ^b	UOR Stop ^b	Genome Start ^c	Genome Stop ^c	Annotation ^d
PF0003	GTTAGTAAC	-112	-104	3444	3452	moeB-like protein
PF0034	GTTGAGAAC	-169	-161	41144	41152	hypothetical protein
PF0038	GTTTAGAAC	-106	-98	45607	45615	glyoxalase II family member
PF0047	GTTTGTAAC	-190	-182	55974	55982	hypothetical protein
PF0055	GTTTTTAAC	-58	-50	62333	62341	putative HTH transcription regulator
PF0094	GTTGACAAC	-74	-66	103265	103273	glutaredoxin-like protein
PF0094	GTTATAAAC	-30	-22	103221	103229	glutaredoxin-like protein
PF0095	GTTTATAAC	-112	-104	103229	103221	hypothetical protein
PF0095	GTTGTCAAC	-68	-60	103273	103265	hypothetical protein
PF0147	G TTCATAAC	-199	-191	156648	156656	potassium channel related protein
PF0157	GTTTTTAAC	-41	-33	168839	168847	D-fructose-6-phosphate amidotransferase
PF0177	GTTTGAAAC	-51	-43	186065	186057	V-type ATP synthase subunit I
PF0207	GTTGTGAAC	-181	-173	222673	222665	argininosuccinate synthase
PF0254	GTTGAGAAC	-71	-63	263369	263377	hypothetical protein
PF0283	GTTGTAAAC	-119	-111	293740	293732	putative acylphosphatase
PF0315	GTTTTCAAC	-148	-140	330297	330305	hypothetical protein
PF0335	GTTAAAAAC	-84	-76	350529	350537	flagella-related protein d, putative
PF0339	GTTAATAAC	-111	-103	353523	353531	methyltransferase
PF0342	GTTGGGAAC	-123	-115	355692	355700	hypothetical protein
PF0350	GTTTGCAAC	-171	-163	364077	364085	putative cation antiporter
PF0360	GTTTATAAC	-82	-74	373902	373894	oligopeptide ABC transporter
PF0369	GTTGAAAAC	-115	-107	384503	384511	putative endoglucanase
PF0388	GTTGTAAAC	-54	-46	397599	397607	hypothetical protein
PF0389	GTTTACAAC	-102	-94	397607	397599	transposase
PF0393	GTTTTCAAC	-90	-82	401354	401346	hypothetical protein
PF0439.3n	GTTGATAAC	-155	-147	451153	451145	hypothetical protein
PF0443	GTTGGGAAC	-69	-61	460874	460882	putative membrane transport protein
PF0470	GTTAAAAAC	-31	-23	485912	485920	hypothetical protein
PF0529	GTTTGGAAC	-48	-40	546842	546850	hypothetical protein
PF0531	GTTGGTAAC	-84	-76	547743	547751	cobalamin biosynthesis protein m
PF0532	GTTACCAAC	-101	-93	547751	547743	hypothetical protein
PF0538	G TTCATAAC	-121	-113	558421	558429	hypothetical protein
PF0547	GTTTTCAAC	-182	-174	566575	566583	hypothetical protein
PF0548	GTTGAAAAC	-68	-60	566583	566575	hydrogenase expression/formation protein
PF0557	GTTCCGAAC	-79	-71	574517	574525	3-ketoacyl-(acyl-carrier-protein) reductase
PF0558	GTTCCGAAC	-117	-109	574525	574517	cation efflux system protein
PF0559	GTTTCTAAC	-49	-41	575553	575545	hydrogenase expression/formation regulatory protein
PF0568	GTTTATAAC	-24	-16	590847	590855	hypothetical protein
PF0569	GTTATAAAC	-27	-19	590855	590847	hypothetical protein
PF0576	GTTTTCAAC	-173	-165	596353	596361	hypothetical protein
PF0576	GTTTCTAAC	-134	-126	596314	596322	hypothetical protein

UOR Name ^a	Motif	UOR Start ^b	UOR Stop ^b	Genome Start ^c	Genome Stop ^c	Annotation ^d
PF0583	GTTGAAAAC	-106	-98	603286	603294	daunorubicin resistance ATP-binding protein
PF0588	GTTTTTAAC	-40	-32	608204	608212	phospho-sugar mutase
PF0597	GTTCTTAAC	-177	-169	617545	617537	hypothetical iaa-amino acid hydrolase 1 precursor
PF0606	GTTCATAAC	-60	-52	628097	628089	hypothetical protein
PF0656	GTTTTCAAC	-102	-94	669051	669059	hypothetical protein
PF0660	GTTTTCAAC	-109	-101	672080	672088	iron-sulfur cluster binding protein
PF0673	GTTGAAAAC	-39	-31	683729	683721	putative acetyl-CoA carboxylase biotin carboxyl carrier protein subunit
PF0675	GTTAGTAAC	-25	-17	685443	685435	hypothetical protein
PF0694	GTTTCAAAC	-139	-131	702563	702555	flavoprotein
PF0768	GTTAAAAAC	-99	-91	761500	761508	acetyl / acyl transferase related protein
PF0781	GTTGATAAC	-54	-46	768401	768393	hypothetical protein
PF0799	GTTCATAAC	-195	-187	785146	785154	unknown
PF0799	GTTGACAAC	-65	-57	785016	785024	unknown
PF0800	GTTGTCAAC	-125	-117	785024	785016	transposase
PF0801	GTTGTCAAC	-134	-126	785024	785016	hypothetical protein
PF0804	GTTATAAAC	-145	-137	788735	788743	methanol dehydrogenase regulator
PF0840	GTTGCAAAC	-189	-181	813851	813843	hypothetical nucleotidyltransferase
PF0865	GTTAAAAAC	-44	-36	840293	840301	aromatic acid decarboxylase
PF0866	GTTTTTAAC	-75	-67	840301	840293	hypothetical protein
PF0891	GTTTTTAAC	-143	-135	863620	863612	sulfhydrogenase beta subunit
PF0913	GTTCAAAAC	-197	-189	884574	884582	hypothetical protein
PF0913	GTTAATAAC	-183	-175	884560	884568	hypothetical protein
PF0914	GTTCAAAAC	-81	-73	884574	884582	hypothetical protein
PF0914	GTTAATAAC	-67	-59	884560	884568	hypothetical protein
PF0915	GTTAGAAAC	-58	-50	885386	885394	hypothetical protein
PF0920	GTTTCCAAC	-198	-190	888142	888150	nucleotidyltransferase
PF0926	GTTTGGAAC	-81	-73	892098	892106	hypothetical protein
PF0947	GTTAGCAAC	-185	-177	911549	911557	hypothetical protein
PF1027	GTTACTAAC	-138	-130	983677	983685	hypothetical protein
PF1063	GTTATAAAC	-83	-75	1016376	1016368	hypothetical protein
PF1063	GTTTTTAAC	-31	-23	1016428	1016420	hypothetical protein
PF1065	GTTTATAAC	-176	-168	1018769	1018777	hypothetical protein
PF1087	GTTGTTAAC	-31	-23	1036410	1036418	hypothetical protein
PF1098	GTTTCAAAC	-159	-151	1045905	1045913	hypothetical protein
PF1100	GTTTGCAAC	-109	-101	1047031	1047023	hypothetical protein
PF1126	GTTGTAAC	-152	-144	1075478	1075486	hypothetical protein
PF1135	GTTGTGAAC	-120	-112	1084791	1084783	heat shock protein HtpX
PF1147	GTTGACAAC	-72	-64	1101031	1101039	putative monovalent cation/H ⁺ antiporter subunit D
PF1164	GTTTGGAAC	-99	-91	1110574	1110566	putative Na ⁺ /H ⁺ antiporter
PF1186	GTTTTAAAC	-26	-18	1132897	1132905	NADH oxidase
PF1204	GTTACAAAC	-91	-83	1145394	1145386	seryl-tRNA synthetase

UOR Name ^a	Motif	UOR Start ^b	UOR Stop ^b	Genome Start ^c	Genome Stop ^c	Annotation ^d
PF1227	GTTAGAAAC	-136	-128	1161999	1161991	transposase
PF1238	GTTGAAAAC	-153	-145	1171247	1171239	putative ABC transporter
PF1253	GTTTATAAC	-42	-34	1188453	1188461	aspartate aminotransferase
PF1255	GTTGGAAC	-171	-163	1189998	1189990	hypothetical protein
PF1269	GTTTGAAAC	-105	-97	1201641	1201633	methionine synthase
PF1316	GTTCATAAC	-104	-96	1239717	1239709	hypothetical protein
PF1319	GTTTTGAAC	-135	-127	1242031	1242023	hypothetical protein
PF1324	GTTAGTAAC	-71	-63	1245432	1245424	hypothetical protein
PF1328	GTTTTGAAC	-79	-71	1249933	1249941	ferredoxin--NADP(+) reductase subunit alpha
PF1329	GTTCAAAAC	-89	-81	1249941	1249933	H-II hydrogenase subunit beta
PF1366	GTTGAAAAC	-34	-26	1285122	1285130	transposase
PF1367	GTTTTCAAC	-48	-40	1285130	1285122	50S ribosomal protein L7Ae
PF1399	GTTTGGAAC	-87	-79	1312685	1312677	putative ATPase, vanadate-sensitive
PF1412	GTTCCAAAC	-200	-192	1327897	1327889	dipeptide transport ATP-binding protein
PF1412	GTTCCAAAC	-149	-141	1327948	1327940	dipeptide transport ATP-binding protein
PF1414	GTTGGAAC	-113	-105	1329475	1329467	flap endonuclease-1
PF1422	GTTTGAAAC	-126	-118	1337382	1337390	thioredoxin reductase
PF1422	GTTTATAAC	-11	-3	1337267	1337275	thioredoxin reductase
PF1423	GTTATAAAC	-147	-139	1337275	1337267	putative monovalent cation/H ⁺ antiporter subunit E
PF1423	GTTTCAAAC	-32	-24	1337390	1337382	putative monovalent cation/H ⁺ antiporter subunit E
PF1453	GTTTAAAAC	-103	-95	1360135	1360143	putative monovalent cation/H ⁺ antiporter subunit E
PF1453	GTTGTTAAC	-61	-53	1360093	1360101	putative monovalent cation/H ⁺ antiporter subunit E
PF1454	GTTAACAAC	-154	-146	1360101	1360093	hypothetical protein
PF1454	GTTTTAAAAC	-112	-104	1360143	1360135	hypothetical protein
PF1467	GTTGGCAAC	-107	-99	1373568	1373560	putative ABC transporter
PF1479	GTTAAAAAC	-96	-88	1382832	1382824	putative oxidoreductase, Fe-S subunit
PF1490	GTTTTAAAAC	-34	-26	1391897	1391905	hypothetical protein
PF1491	GTTTAAAAC	-55	-47	1391905	1391897	30S ribosomal protein S17e
PF1516	GTTTACAAC	-58	-50	1415022	1415030	GMP synthase subunit B
PF1537	GTTTTTAAAC	-57	-49	1436963	1436971	hypothetical protein
PF1561	GTTGACAAC	-129	-121	1455816	1455824	50S ribosomal protein L30e
PF1569	GTTTGGAAC	-153	-145	1466829	1466837	exosome complex RNA-binding protein Rrp4
PF1617	GTTTGGAAC	-177	-169	1509509	1509501	hypothetical protein
PF1621	GTTACTAAC	-70	-62	1513968	1513976	hypothetical protein
PF1622	GTTAGTAAC	-197	-189	1513976	1513968	n-type ATP pyrophosphatase superfamily
PF1652	GTTATAAAC	-124	-116	1537640	1537632	hypothetical protein
PF1665	GTTATGAAC	-30	-22	1548857	1548849	histidinol-phosphate aminotransferase
PF1679	GTTAGCAAC	-10	-2	1561854	1561846	3-isopropylmalate dehydratase large subunit

UOR Name ^a	Motif	UOR Start ^b	UOR Stop ^b	Genome Start ^c	Genome Stop ^c	Annotation ^d
PF1684	GTTCAAAAC	-86	-78	1565631	1565623	acetylglutamate kinase
PF1698	GTTAAAAAC	-117	-109	1578254	1578246	putative ribose ABC transporter
PF1715	GTTATAAAC	-43	-35	1595766	1595774	pyrroline-5-carboxylate reductase
PF1749	GTTTATAAC	-81	-73	1626536	1626528	putative sulfate transport integral membrane protein
PF1752	GTTTAGAAC	-160	-152	1629141	1629133	transposase
PF1795	GTTGAGAAC	-80	-72	1667079	1667071	sarcosine oxidase subunit alpha
PF1815	GTTTGAAAC	-90	-82	1679926	1679934	30S ribosomal protein S17P
PF1864	GTTAGAAAC	-137	-129	1721104	1721112	transposase
PF1878	GTTAAAAAC	-197	-189	1730945	1730937	methylated-DNA--protein-cysteine methyltransferase
PF1896	GTTAGAAAC	-33	-25	1746846	1746838	l-isoaspartyl protein carboxyl methyltransferase
PF1898	GTTTTAAAC	-179	-171	1748875	1748883	hypothetical protein
PF1906	GTTAATAAC	-28	-20	1757229	1757221	adenosylmethionine-8-amino-7-oxononanoate aminotransferase
PF1911	GTTCAACAAC	-54	-46	1762897	1762905	ferredoxin--NADP(+) reductase subunit alpha
PF1912	GTTGTGAAC	-95	-87	1762905	1762897	hypothetical protein
PF1920	GTTAGAAAC	-52	-44	1771957	1771965	triosephosphate isomerase
PF1927	GTTCTTAAC	-41	-33	1778257	1778265	hypothetical protein
PF1928	GTTAAGAAC	-29	-21	1778265	1778257	ATP-binding protein phnp
PF1940.1n	GTTTATAAC	-42	-34	1793571	1793579	hypothetical protein
PF1975	GTTTAAAAC	-61	-53	1825360	1825352	hypothetical protein
PF1981	GTTAAAAAC	-40	-32	1832121	1832129	hypothetical protein
PF1982	GTTTTTAAC	-41	-33	1832129	1832121	hypothetical protein
PF2011	GTTATAAAC	-143	-135	1858487	1858479	hypothetical protein
PF2047.1n	GTTAAAAAC	-122	-114	1889574	1889582	hypothetical protein
PF2048	GTTTTTAAC	-51	-43	1889582	1889574	peptidase
PF2050	GTTTGGAAC	-101	-93	1892777	1892785	hypothetical protein
PF2051	GTTCCAAAC	-27	-19	1892785	1892777	arsr family transcriptional regulator
PF2060.1n	GTTTAAAAC	-12	-4	1901228	1901236	hypothetical arylsulfatase regulatory protein
PF2064	GTTATTAAC	-67	-59	1905977	1905985	arylsulfatase regulatory protein, putative

^aUORs are denoted by the loci of their corresponding downstream ORFs.

^bUOR starts and stops are relative to the corresponding ORF, where -1 is the first nucleotide immediately upstream of the ORF start.

^cGenome starts and stops are relative to first nucleotide of the sequenced genome.

^dAnnotation is derived from the ORF definitions input to UORdb from the source databases Refseq [73], GenBank [72], and TIGR/CMR [74].

Table 3.2 UORdb Results for GTTN₃AACN₅GTT Motif

UOR Name ^a	Motif	UOR Start ^b	UOR Stop ^b	Genome Start ^c	Genome Stop ^c	Annotation ^d	Predicted Operon ^e
PF0094	GTTGTCAACCTTAGGTT	-82	-66	103265	103281	glutaredoxin-like protein	PF0094
PF0095	GTTGTCAACCTTAGGTT	-68	-52	103281	103265	hypothetical protein	PF0095 - PF0097
PF0531	GTTGGTAACAAAATGTT	-84	-68	547735	547751	cobalamin biosynthesis	PF0531 - PF0528
PF0532	GTTACCAACGTAAGGTT	-92	-76	547743	547759	protein m	PF0532 - PF0534
	GTTACCAACGTAAGGTT	-101	-85	547759	547743	hypothetical protein	
	GTTGGTAACAAAATGTT	-109	-93	547751	547735		
PF0547	GTTGAAAACCTCAAGTT	-190	-174	566575	566591	hypothetical protein	PF0547
PF0548	GTTGAAAACCTCAAGTT	-68	-52	566591	566575	hydrogenase expression/formation protein	PF0548 - PF0549
PF0559	GTTTCTAACTTTTGTT	-49	-33	575561	575545	hydrogenase expression/formation regulatory protein	PF0558 - PF0559
PF0568	GTTTATAACGCTATGTT	-24	-8	590839	590855	hypothetical protein	PF0568 - PF0564
PF0569	GTTTATAACGCTATGTT	-35	-19	590855	590839	hypothetical protein	PF0569
PF0891	GTTTTTAACCTTTGTT	-143	-127	863628	863612	sulphydrogenase beta subunit	PF0891 - PF0894
PF0913	GTTCAAAACCAAAGGTT	-197	-181	884566	884582	hypothetical protein	PF0914 - PF0913
	GTTATTAACCTTTGTT	-191	-175	884560	884576		
PF0914	GTTCAAAACCAAAGGTT	-81	-65	884566	884582	hypothetical protein	PF0914 - PF0913
	GTTATTAACCTTTGTT	-75	-59	884560	884576		
PF0915	GTTAGAAACCTTAGGTT	-58	-42	885378	885394	hypothetical protein	PF0917 - PF0915
	GTTTCTAACTAGATGTT	-66	-50	885386	885402		
PF0926	GTTTGGAACCTATTGTT	-81	-65	892090	892106	hypothetical protein	PF0926
	GTTCCAAACAATAAGTT	-89	-73	892098	892114		
PF1100	GTTTGCAACTCGTAGTT	-109	-93	1047039	1047023	hypothetical protein	PF1100
PF1186	GTTTAAAACCTTTAGTT	-34	-18	1132897	1132913	NADH oxidase	PF1186
PF1328	GTTCAAAACCTAAGGTT	-87	-71	1249933	1249949	ferredoxin--NADP(+) reductase subunit alpha	PF1328 - PF1327
PF1329	GTTCAAAACCTAAGGTT	-89	-73	1249949	1249933	H-II hydrogenase subunit beta	PF1329 - PF1332
PF1422	GTTATAAACCAAAAAGTT	-19	-3	1337267	1337283	thioredoxin reductase	PF1422
	GTTTCAAACCAAAATGTT	-134	-118	1337382	1337398		
PF1423	GTTATAAACCAAAAAGTT	-147	-131	1337283	1337267	putative monovalent cation/H ⁺ antiporter subunit E	PF1423 - PF1436
	GTTTCAAACCAAAATGTT	-32	-16	1337398	1337382		

UOR Name ^a	Motif	UOR Start ^b	UOR Stop ^b	Genome Start ^c	Genome Stop ^c	Annotation ^d	Predicted Operon ^e
PF1516	GTTTACAACTTTATGTT	-58	-42	1415014	1415030	GMP synthase subunit B	PF1516 - PF1515
PF1621	GTTAGTAACTAAAAGTT	-78	-62	1513968	1513984	hypothetical protein	PF1621
PF1622	GTTAGTAACTAAAAGTT	-197	-181	1513984	1513968	n-type ATP pyrophosphatase superfamily	PF1622

^aUORs are denoted by the loci of their corresponding downstream ORFs.

^bUOR starts and stops are relative to the corresponding ORF, where -1 is the first nucleotide immediately upstream of the ORF start.

^cGenome starts and stops are relative to first nucleotide of the sequenced genome.

^dAnnotation is derived from the ORF definitions input to UORdb from the source databases Refseq [73], GenBank [72], and TIGR/CMR [74].

^eOperon predictions are derived from the DOOR database (<http://csb11.bmb.uga.edu/OperonDB/>) [85]. Highlighted UORs signify that the current UOR is upstream of the first ORF in its predicted operon.

Table 3.3 UORdb Results for AACN₅GTT Motif

UOR Name ^a	Motif	UOR Start ^b	UOR Stop ^b	Genome Start ^c	Genome Stop ^c	Annotation ^d
PF0023	AAC TTC AAG TT	-99	-89	26727	26737	hypothetical protein
PF0041	AAC TTC AGG TT	-63	-53	48431	48441	putative oxidative cyclase
PF0074	AAC ATG AAG TT	-151	-141	79422	79412	alcohol dehydrogenase, short chain
PF0079	AAC TTT TGG TT	-94	-84	85409	85419	cyclic 2,3-diphosphoglycerate-synthetase
PF0080	AAC CAAA AGT T	-41	-31	85419	85409	hypothetical protein
PF0094	AAC CTA AGG TT	-82	-72	103271	103281	glutaredoxin-like protein
PF0094	AAC TCT AGG TT	-52	-42	103241	103251	glutaredoxin-like protein
PF0095	AAC CTAG AGT T	-92	-82	103251	103241	hypothetical protein
PF0095	AAC CTT AGG TT	-62	-52	103281	103271	hypothetical protein
PF0119	AAC GGAT TGT T	-53	-43	125178	125188	putative periplasmic sugar binding protein
PF0134	AAC CTAT AGT T	-62	-52	144544	144554	hypothetical protein
PF0135	AAC TAT AGG TT	-64	-54	144554	144544	lysyl-tRNA synthetase
PF0144	AAC AAT GGG TT	-177	-167	153555	153545	hypothetical protein
PF0155	AAC CGGG GGT T	-165	-155	163948	163958	asparaginyl-tRNA synthetase
PF0238	AAC CAG CGG TT	-176	-166	249992	250002	hypothetical protein
PF0275	AAC AAAG CGT T	-195	-185	285599	285589	Na ⁺ /H ⁺ antiporter
PF0279	AAC GTC AAG TT	-124	-114	292413	292423	putative ABC transporter
PF0296	AAC TTT AAG TT	-154	-144	312041	312051	cobalamin biosynthesis protein
PF0306	AAC GTAA AGT T	-98	-88	323019	323029	sua5 superfamily-related protein
PF0312	AAC GAAA AGT T	-47	-37	329099	329109	ADP-specific phosphofructokinase
PF0322	AAC TTA ACG TT	-49	-39	337801	337811	nickel responsive regulator
PF0432.3n	AAC TCA AAG TT	-74	-64	443332	443342	hypothetical protein
PF0443	AAC CTA AAG TT	-200	-190	461003	461013	putative membrane transport protein
PF0485	AAC TTA ACG TT	-86	-76	503277	503287	cell division inhibitor minD - like protein
PF0486	AAC GTTA AGT T	-39	-29	503287	503277	hypothetical protein
PF0487	AAC TTT GTG TT	-76	-66	505223	505233	related to bacterial autolysins
PF0488	AAC ACA AAG TT	-125	-115	505233	505223	30S ribosomal protein S6e
PF0531	AAC CTT ACG TT	-92	-82	547749	547759	cobalamin biosynthesis protein m
PF0531	AAC AAA ATG TT	-78	-68	547735	547745	cobalamin biosynthesis protein m
PF0532	AAC ATTT TGT T	-109	-99	547745	547735	hypothetical protein
PF0532	AAC GTA AGG TT	-95	-85	547759	547749	hypothetical protein
PF0545	AAC AGG CAG TT	-195	-185	563768	563778	hypothetical protein
PF0546	AAC CAT TAG TT	-125	-115	564950	564960	hypothetical protein
PF0547	AAC TTGA AGT T	-190	-180	566581	566591	hypothetical protein
PF0548	AAC TTC AAG TT	-62	-52	566591	566581	hydrogenase expression/formation protein
PF0552	AAC CAT TGG TT	-59	-49	570056	570066	arsenical-resistance protein acr3
PF0553	AAC CAAT GGT T	-93	-83	570066	570056	arsenate reductase
PF0559	AAC CAA AGG TT	-75	-65	575529	575519	hydrogenase expression/formation regulatory protein
PF0559	AAC TTT TGG TT	-43	-33	575561	575551	hydrogenase expression/formation regulatory protein
PF0563	AAC AAAG CGT T	-177	-167	580071	580061	hypothetical protein
PF0568	AAC GCT ATG TT	-18	-8	590839	590849	hypothetical protein

UOR Name ^a	Motif	UOR Start ^b	UOR Stop ^b	Genome Start ^c	Genome Stop ^c	Annotation ^d
PF0569	AACATAGCGTT	-35	-25	590849	590839	hypothetical protein
PF0578	AACGAAGAGTT	-108	-98	597050	597040	hypothetical protein
PF0594	AACTTCACGTT	-189	-179	616280	616290	ornithine carbamoyltransferase
PF0603	AACTTTGAGTT	-186	-176	625539	625529	hypothetical protein
PF0632	AACGGTTGGTT	-128	-118	645533	645543	hypothetical protein
PF0632	AACTTAGCGTT	-25	-15	645430	645440	hypothetical protein
PF0636	AACGAAAGGTT	-175	-165	650300	650310	DEXX-box atpase
PF0646	AACAACAAGTT	-83	-73	659203	659213	hypothetical protein
PF0687.1n	AACTCAAAGTT	-122	-112	695583	695573	hypothetical protein
PF0688	AACTAAGTGTT	-125	-115	698965	698955	subtilisin-like protease
PF0704	AACAGGAAGTT	-178	-168	709163	709173	hypothetical protein
PF0715	AACAAGAAGTT	-139	-129	715746	715756	NAD(P)H oxidase
PF0736	AACAAGTTGTT	-25	-15	732381	732391	hypothetical protein
PF0740	AACATTTGGTT	-79	-69	735957	735947	heavy-metal transporting cpx-type atpase
PF0758	AACAAAATGTT	-49	-39	753429	753439	hypothetical protein
PF0783	AACTCTCGGTT	-115	-105	769447	769437	hypothetical protein
PF0806	AACCTAATGTT	-35	-25	789610	789620	hypothetical protein
PF0807	AACCAAAAGTT	-82	-72	791342	791352	hypothetical protein
PF0859	AACTCCAAGTT	-35	-25	834813	834823	hypothetical protein
PF0860	AACTTGGAGTT	-31	-21	834823	834813	btpA family protein
PF0891	AACCTTTGGTT	-137	-127	863628	863618	sulfhydrogenase beta subunit
PF0913	AACCAAAGGTT	-191	-181	884566	884576	hypothetical protein
PF0914	AACCAAAGGTT	-75	-65	884566	884576	hypothetical protein
PF0915	AACATCTAGTT	-66	-56	885392	885402	hypothetical protein
PF0915	AACCTTAGGTT	-52	-42	885378	885388	hypothetical protein
PF0926	AACTTATTGTT	-89	-79	892104	892114	hypothetical protein
PF0926	AACCTATTGTT	-75	-65	892090	892100	hypothetical protein
PF0933	AACTTTTCGTT	-125	-115	897405	897415	DNA repair helicase rad3, putative
PF0952	AACCGTACGTT	-76	-66	913874	913864	hypothetical protein
PF0958	AACTTCATGTT	-127	-117	919394	919404	hypothetical protein
PF0971	AACCAAGGGTT	-119	-109	931105	931115	pyruvate ferredoxin oxidoreductase subunit gamma
PF0972	AACCCTTGGTT	-131	-121	931115	931105	hypothetical protein
PF0981	AACTTTTAGTT	-88	-78	937534	937524	hypothetical efflux transport protein
PF0984	AACAAGAAGTT	-142	-132	940468	940478	hypothetical protein
PF1031	AACCAGGGGTT	-97	-87	988107	988117	divalent cation tolerance protein
PF1047	AACGATGAGTT	-149	-139	1000260	1000250	hypothetical protein
PF1057	AACAACATGTT	-88	-78	1011386	1011376	phosphoglycerate kinase
PF1088	AACAGCATGTT	-134	-124	1036847	1036857	hypothetical protein
PF1100	AACTCGTAGTT	-103	-93	1047039	1047029	hypothetical protein
PF1117	AACTTTGAGTT	-129	-119	1065907	1065897	hypothetical protein
PF1129	AACATATAGTT	-87	-77	1080421	1080431	hypothetical protein

UOR Name ^a	Motif	UOR Start ^b	UOR Stop ^b	Genome Start ^c	Genome Stop ^c	Annotation ^d
PF1161	AACAGAGGGTT	-40	-30	1110387	1110397	hypothetical protein
PF1163	AACCCTCTGTT	-79	-69	1110397	1110387	hypothetical protein
PF1186	AACTAAAGGTT	-34	-24	1132903	1132913	NADH oxidase
PF1188	AACTCTTAGTT	-93	-83	1133979	1133969	pyruvate kinase
PF1202	AACATAGAGTT	-109	-99	1143352	1143342	hypothetical protein
PF1203	AACATAGAGTT	-179	-169	1143352	1143342	formaldehyde:ferredoxin oxidoreductase
PF1247.ln	AACTTTTGTGTT	-154	-144	1183168	1183178	hypothetical protein
PF1248	AACTTTTGTGTT	-157	-147	1183168	1183178	transposase
PF1266	AACTCCATGTT	-150	-140	1198529	1198519	cystathionine gamma-synthase
PF1268	AACGAGGTGTT	-49	-39	1200768	1200758	5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase
PF1321	AACGATAGGTT	-26	-16	1242871	1242861	DEXX-box atpase
PF1328	AACCTTAGGTT	-87	-77	1249939	1249949	ferredoxin--NADP(+) reductase subunit alpha
PF1329	AACCTAAGGTT	-83	-73	1249949	1249939	H-II hydrogenase subunit beta
PF1364	AACATAAGGTT	-43	-33	1283874	1283884	hypothetical protein
PF1365	AACTTCGTGTT	-105	-95	1283735	1283725	hypothetical protein
PF1377	AACTAGCAGTT	-52	-42	1292942	1292952	transcription initiation factor IIB
PF1378	AACTGCTAGTT	-107	-97	1292952	1292942	ribonuclease P protein component 2
PF1408	AACTTTTGTGTT	-154	-144	1322564	1322554	putative dipeptide-binding protein
PF1420	AACATTAAGTT	-155	-145	1334098	1334088	hypothetical protein
PF1422	AACATTTGGTT	-134	-124	1337388	1337398	thioredoxin reductase
PF1422	AACTTTTGGTT	-19	-9	1337273	1337283	thioredoxin reductase
PF1423	AACCAAAAGTT	-141	-131	1337283	1337273	putative monovalent cation/H ⁺ antiporter subunit E
PF1423	AACCAAAATGTT	-26	-16	1337398	1337388	putative monovalent cation/H ⁺ antiporter subunit E
PF1477	AACTTTAGGTT	-61	-51	1382311	1382321	methylmalonyl-CoA mutase
PF1478	AACCTAAAGTT	-97	-87	1382321	1382311	hypothetical protein
PF1481	AACCTTTGGTT	-61	-51	1385346	1385336	hypothetical protein
PF1516	AACTTTATGTT	-52	-42	1415014	1415024	GMP synthase subunit B
PF1520	AACGATGCGTT	-94	-84	1419855	1419865	ABC transporter integral membrane protein
PF1568	AACGGGGAGTT	-66	-56	1465971	1465981	exosome complex exonuclease Rrp41
PF1621	AACTTTTAGTT	-78	-68	1513974	1513984	hypothetical protein
PF1622	AACTAAAAGTT	-191	-181	1513984	1513974	n-type ATP pyrophosphatase superfamily
PF1684	AACATAATGTT	-74	-64	1565645	1565635	acetylglutamate kinase
PF1716	AACAATGGGTT	-156	-146	1596360	1596370	hypothetical protein
PF1747	AACTAACTGTT	-79	-69	1624581	1624571	hypothetical protein
PF1792	AACCAAAAGTT	-147	-137	1664951	1664941	hypothetical 5'-cyclic-nucleotide phosphodiesterase cpda
PF1794	AACTGGAGGTT	-20	-10	1666511	1666501	inosine-5'-monophosphate dehydrogenase related protein I
PF1831	AACAAAAGGTT	-26	-16	1689350	1689340	archaeal histone a1
PF1837	AACCGATAGTT	-164	-154	1693366	1693376	hypothetical protein
PF1838	AACCTTTCGTT	-102	-92	1694729	1694739	hypothetical protein

UOR Name ^a	Motif	UOR Start ^b	UOR Stop ^b	Genome Start ^c	Genome Stop ^c	Annotation ^d
PF1839	AACGAAAGGTT	-39	-29	1694739	1694729	CTP synthetase
PF1851	AACTCAGTGTT	-47	-37	1709264	1709274	hypothetical protein
PF1881	AACATTTTCGTT	-38	-28	1733860	1733870	chromatin protein
PF1899	AACTAATCGTT	-179	-169	1748565	1748555	putative 2-deoxy-d-gluconate 3-dehydrogenase
PF1911	AACCTTTGGTT	-90	-80	1762931	1762941	ferredoxin--NADP(+) reductase subunit alpha
PF1912	AACCAAAGGTT	-61	-51	1762941	1762931	hypothetical protein
PF1959	AACAAAGCGTT	-112	-102	1808800	1808790	cofactor-independent phosphoglycerate mutase
PF1960	AACTTAGCGTT	-64	-54	1810191	1810181	aldose reductase
PF1978	AACAAGAAGTT	-181	-171	1829066	1829056	nicotinate-nucleotide pyrophosphorylase
PF1988	AACGGGGGGTT	-18	-8	1839097	1839087	cell division protein FtsZ
PF2034	AACAAAGCGTT	-135	-125	1877906	1877916	hypothetical protein

^aUORs are denoted by the loci of their corresponding downstream ORFs.

^bUOR starts and stops are relative to the corresponding ORF, where -1 is the first nucleotide immediately upstream of the ORF start.

^cGenome starts and stops are relative to first nucleotide of the sequenced genome.

^dAnnotation is derived from the ORF definitions input to UORdb from the source databases Refseq [73], GenBank [72], and TIGR/CMR [74].

Table 3.4 Combined UORdb Results for GTTN₃AAC and AACN₅GTT Motif

UOR Name ^a	Motif	UOR Start ^b	UOR Stop ^b	Genome Start ^c	Genome Stop ^c	Annotation ^d	Predicted Operon ^e
PF0094	GTTGACAAC	-74	-66	103265	103273	glutaredoxin-like protein	PF0094
	GTTATAAAC	-30	-22	103221	103229		
	AACCTAAGGTT	-82	-72	103271	103281		
	AACTCTAGGTT	-52	-42	103241	103251		
PF0095	GTTTATAAC	-112	-104	103229	103221	hypothetical protein	PF0095 - PF0097
	GTTGTCAAC	-68	-60	103273	103265		
	AACCTAGAGTT	-92	-82	103251	103241		
	AACCTTAGGTT	-62	-52	103281	103271		
PF0443	GTTGGGAAC	-69	-61	460874	460882	putative membrane transport protein	PF0446 - PF0443
	AACCTCAAGTT	-200	-190	461003	461013		
PF0531	GTTGGTAAC	-84	-76	547743	547751	cobalamin biosynthesis protein m	PF0531 PF0528
	AACCTTACGTT	-92	-82	547749	547759		
	AACAAAATGTT	-78	-68	547735	547745		
PF0532	GTTACCAAC	-101	-93	547751	547743	hypothetical protein	PF0532 - PF0534
	AACATTTTGTT	-109	-99	547745	547735		
	AACGTAAGGTT	-95	-85	547759	547749		
PF0547	GTTTTCAAC	-182	-174	566575	566583	hypothetical protein	PF0547
	AACCTGAAGTT	-190	-180	566581	566591		
PF0548	GTTGAAAAC	-68	-60	566583	566575	hydrogenase expression/formation protein	PF0548 - PF0549
	AACCTCAAGTT	-62	-52	566591	566581		
PF0559	GTTTCTAAC	-49	-41	575553	575545	hydrogenase expression/formation regulatory protein	PF0558 - PF0559
	AACCAAAGGTT	-75	-65	575529	575519		
	AACCTTTGGTT	-43	-33	575561	575551		
PF0568	GTTTATAAC	-24	-16	590847	590855	hypothetical protein	PF0568 - PF0564
	AACGCTATGTT	-18	-8	590839	590849		
PF0569	GTTATAAAC	-27	-19	590855	590847	hypothetical protein	PF0569
	AACATAGCGTT	-35	-25	590849	590839		
PF0891	GTTTTTAAC	-143	-135	863620	863612	sulfhydrogenase beta subunit	PF0891 - PF0894
	AACCTTTGGTT	-137	-127	863628	863618		
PF0913	GTTCAAAAC	-197	-189	884574	884582	hypothetical protein	PF0914 - PF0913
	GTTAATAAC	-183	-175	884560	884568		
	AACCAAAGGTT	-191	-181	884566	884576		
	GTTCAAAAAC	-81	-73	884574	884582		
PF0914	GTTAATAAC	-67	-59	884560	884568	hypothetical protein	PF0914 - PF0913
	AACCAAAGGTT	-75	-65	884566	884576		
	GTTAGAAAC	-58	-50	885386	885394		
PF0915	AACATCTAGTT	-66	-56	885392	885402	hypothetical protein	PF0917 - PF0915
	AACCTTAGGTT	-52	-42	885378	885388		
PF0926	GTTTGGAAC	-81	-73	892098	892106	hypothetical protein	PF0926
	AACCTATTGTT	-89	-79	892104	892114		
	AACCTATTGTT	-75	-65	892090	892100		
PF1100	GTTTGCAAC	-109	-101	1047031	1047023	hypothetical protein	PF1100
	AACCTCGTAGTT	-103	-93	1047039	1047029		
PF1186	GTTTTAAAC	-26	-18	1132897	1132905	NADH oxidase	PF1186
	AACTAAAGGTT	-34	-24	1132903	1132913		

UOR Name ^a	Motif	UOR Start ^b	UOR Stop ^b	Genome Start ^c	Genome Stop ^c	Annotation ^d	Predicted Operon ^e
PF1328	GTTTTGAAC	-79	-71	1249933	1249941	ferredoxin--NADP(+) reductase subunit alpha	PF1328 - PF1327
	AACCTTAGGTT	-87	-77	1249939	1249949		
PF1329	GTTCAAAAC	-89	-81	1249941	1249933	H-II hydrogenase subunit beta	PF1329 - PF1332
	AACCTAAGGTT	-83	-73	1249949	1249939		
PF1422	GTTTGAAAC	-126	-118	1337382	1337390	thioredoxin reductase	PF1422
	GTTTATAAC	-11	-3	1337267	1337275		
	AACATTTGGTT	-134	-124	1337388	1337398		
	AACTTTTGGTT	-19	-9	1337273	1337283		
PF1423	GTTATAAAC	-147	-139	1337275	1337267	putative monovalent cation/H+ antiporter subunit E	PF1423 - PF1436
	GTTTCAAAC	-32	-24	1337390	1337382		
	AACCAAAAAGTT	-141	-131	1337283	1337273		
	AACCAAATGTT	-26	-16	1337398	1337388		
PF1516	GTTTACAAC	-58	-50	1415022	1415030	GMP synthase subunit B	PF1516 - PF1515
	AACTTTATGTT	-52	-42	1415014	1415024		
PF1621	GTTACTAAC	-70	-62	1513968	1513976	hypothetical protein	PF1621
	AACTTTATGTT	-78	-68	1513974	1513984		
PF1622	GTTAGTAAC	-197	-189	1513976	1513968	n-type ATP pyrophosphatase superfamily	PF1622
	AACATAAAGTT	-191	-181	1513984	1513974		
PF1684	GTTCAAAAC	-86	-78	1565631	1565623	acetylglutamate kinase	PF1681 - PF1686
	AACATAATGTT	-74	-64	1565645	1565635		
PF1911	GTTTACAAC	-54	-46	1762897	1762905	ferredoxin--NADP(+) reductase subunit alpha	PF1911 - PF1910
	AACCTTTGGTT	-90	-80	1762931	1762941		
PF1912	GTTGTGAAC	-95	-87	1762905	1762897	hypothetical protein	PF1912 - PF1915
	AACCAAAGGTT	-61	-51	1762941	1762931		

^aUORs are denoted by the loci of their corresponding downstream ORFs.

^bUOR starts and stops are relative to the corresponding ORF, where -1 is the first nucleotide immediately upstream of the ORF start.

^cGenome starts and stops are relative to first nucleotide of the sequenced genome.

^dAnnotation is derived from the ORF definitions input to UORdb from the source databases Refseq [73], GenBank [72], and TIGR/CMR [74].

^eOperon predictions are derived from the DOOR database (<http://csb11.bmb.uga.edu/OperonDB/>) [85].

CHAPTER 4

MOLECULAR MODELING OF *PYROCOCCUS FURIOSUS* RNA POLYMERASE (RNAP)

4.1 Modeling an archaeal RNAP

As described in Chapter 1, the archaeal RNAP and other transcription initiation components are more closely related to the eukaryal class II transcription components, than to the bacterial system. Kornberg's Nobel Prize-winning work on structural characterization of the yeast RNAP system revealed a new level of detail of eukaryal transcription initiation. Our group had previously shown that the assembly of the archaeal RNAP and general transcription factors on promoter DNA showed both similarities and differences compared to the yeast system [109]. We reasoned that examination of the structure of the archaeal RNAP could help elucidate the structural basis of these differences. Without a crystallographic characterization of the archaeal system (at the time), our approach consisted of building a homology model of the *P. furiosus* RNAP. This chapter describes this effort, which was terminated once Murakami et al. solved the crystal structure of the *Sulfolobus solfataricus* RNAP [66].

Constructing a comparative model of the multisubunit *P. furiosus* RNAP is a multistep process. First, appropriate templates with reasonable sequence similarity to the target subunit genes must be selected to ensure structural similarity in the final model. Second, we require sequence alignments of the *P. furiosus* protein subunits with their homologous counterparts from the template. Once the template and target subunits are aligned, a model for each subunit can be created using comparative modeling software. The structures of the individual subunits of the

model can then be refined and assembled into a RNAP complex, which can then be compared to other RNAP structures for assessment of its reliability, followed by interpretation of any structural differences.

4.2 Selection of modeling template

As discussed in Chapter 1, the homology between archaeal and eukaryal transcriptional machinery can provide useful structural templates for the comparative modeling of a *P. furiosus* RNAP complex. Until most recently, the most complete sources of structural information for any archaeal or eukaryal RNAP were the RNAPII structures deposited in 2001 by Kornberg, Cramer, *et al* [68]. Two 10-subunit X-ray crystal structures of *Saccharomyces cerevisiae* RNAP II were derived, revealing two different forms with respect to access to the active site for transcription initiation. From comparison of these two forms, the overall structure of the RNAPII complexes can be subdivided into different internal modules. Approximately half of the entire complex is comprised of a central "core" module containing the active site regions of RPB1 and RPB2, which are homologous to *P. furiosus* subunits A, A', and B, along with RNAPII subunits RPB3, RPB10, RPB11, and RPB12, which are homologous to *P. furiosus* RNAP subunits D, N, L and P, respectively. Three other modules – the “jaw-lobe”, the “shelf”, and the “clamp” – were found to be mobile with respect to the core module. The jaw-lobe module is composed of domains and domain-like regions from RPB1, RPB2, and RPB9, the first two of which are homologous to *P. furiosus* RNAP subunits A, A', B, and P. The shelf module contains separate regions from RPB1, RPB5, and RPB6 (the homologous counterparts to *P. furiosus* RNAP subunits A, A', H, and K.) The clamp module is formed from the N- and C-terminal regions of RPB1 and the C-terminal region of RPB2. The two deposited RNAPII

structures correspond to different states of the clamp module. A closed form, where the clamp swings over the RNAPII active center, was determined at 3.1Å resolution (PDB code: 1I3Q). An open form, where the clamp swings back and allows entry of promoter DNA for transcription initiation, was determined at 2.8Å resolution (PDB code: 1I50). The clamp is highly mobile, with some residues of RPB1 having a C_α displacement over 14Å in the transition between these two states.

A description of the interactions between these modules and DNA in the transcribing RNAP complex has been suggested from these structures and earlier models [110]. DNA may enter the complex through a “cleft” region in the shelf module, passing between two “jaw” regions – an “upper jaw” from the jaw-lobe module and a “lower jaw” from the shelf. Beyond the active center, where the DNA-RNA hybrid would form, is a “wall” of protein; this wall forces outgoing RNA to exit the active center region perpendicular to the path of incoming DNA. The mobile clamp in its open state allows promoter DNA to enter the complex in an essentially straight orientation before forming a transcription bubble (DNA-RNA hybrid) necessary for transcription initiation. Interest in the open-clamp structure of RNAPII led to its selection as the protein template for the *P. furiosus* RNAP comparative model.

4.3 Sequence alignments of *S. cerevisiae* RNAPII templates and *P. furiosus* RNAP subunits

Of the ten subunits present in the open-clamp structure of *S. cerevisiae* RNAPII, seven exhibited some degree of sequence similarity allowable for comparative modeling with eight of the thirteen annotated *P. furiosus* RNAP subunits. Not surprisingly, the similarity between these eukaryal and archaeal sequences is still low, with several subunits and their homologous counterparts near or just below suggested lower bounds of percent identities for model building

[111] , such as that of RNAP subunit D. A possible remedy for this is the use of multiple template structures; however, at the time of model construction, there were no other homologous RNAP structures available as templates. Table 4.1 summarizes some of the alignment statistics between the target and template proteins.

Table 4.1 Sequence Alignment Statistics between *P. furiosus* RNAP and *S. cerevisiae* RNAPII using BLAST [112], WU BLAST2 [113], and MPsrch [114]

<i>P. furiosus</i> subunit	<i>S. cerevisiae</i> subunit	NCBI BLAST		WU BLAST2		EBI MPsrch
		% Identity	% Positives	% Identity	% Positives	% Identity
A	RPB1	34	54	42	61	44.5
A'	RPB1	40	55	44	60	43.6
B	RPB2	39	57	50	67	41.3
D	RPB3	28	43	34	55	42.9
H	RPB5	36	53	46	70	46.5
K	RPB6	47	66	47	66	51.0
L	RPB11	32	48	34	49	34.6
N	RPB10	52	73	52	73	52.4

Percentages only reported for largest segments of the alignment between each subunit.

Only these eight were included in the final *P. furiosus* RNAP model; subunits M and P have some sequence similarity with RPB9 and RPB12 but were deemed to be too low for effective modeling and are omitted from this model. Sequence alignment reveals an important difference between archaeal and eukaryal RNAP complexes: the largest subunit of *S. cerevisiae* RNAPII – RPB1 – has sequence similarity with two of the *P. furiosus* RNAP subunits – A and A'. The eukaryal subunit sequence is essentially “split” between the archaeal subunits, with the archaeal RNAP subunit A' aligning with a N-terminal region from the eukaryal RPB1 while the archaeal RNAP subunit A is sequence-similar to a C-terminal region of RPB1. Pairwise and

multiple alignments with putative prokaryotic homologs were conducted with various alignment methods including NCBI BLASTP [112], WU BLAST2 [113], the MPsrch [114], and Multalin [115]. Predicted secondary structure assignments are included in the alignments from PHD [116], PSIPRED [117], and MOE (Molecular Operating Environment 2004.3) computational software [118]. These alignments can be found in Appendix C, Figures C.1-C.95. MOE was used to create the primary alignment considered for comparative modeling alignment. As a modification of the classic Needleman-Wunsch alignment algorithm [119], this alignment method optimizes a function based on residue similarity scores (computed from an amino acid substitution table such as BLOSUM62) and gap penalties (for initiating and extending alignment gaps). In addition, the alignment can adjust both the similarity scores and gap penalties by using actual secondary structural assignments from the template and predicted secondary structure for the target. For its pairwise alignment, MOE employed the BLOSUM62 substitution matrix, a gap start penalty of 7, and a gap extend penalty of 1 with secondary structural matching enabled.

A residue-by-residue conservation scoring function was applied to the aligned sequences in Figures A8-A102 to judge the degree of similarity between regions of the template and target proteins [120]. This conservation scoring scheme was calculated by a Perl script that incorporated three areas of concentration: amino acid diversity at given position, normalized for redundant sequences using a modified Shannon entropy function; stereochemical diversity, using a “distance” function in conjunction with the BLOSUM62 substitution matrix to gauge similarity between different residues; and alignment gaps, using a simple gap penalty function. Conservation scores were plotted against each residue number for each model target sequence, as illustrated in Figures 4.1-4.3.

The bulk of the functional domains in the RNAP complex can be found in the subunits A, A', and B, which contain some of the longest regions of high conservation. The highest levels of conservation between *P. furiosus* subunit A and RPB1 can be found in *P. furiosus* residues 64-105, which align with *S. cerevisiae* residues 1062-1102 in the cleft region of the shelf module in RPB1, where DNA is proposed to enter the RNAP complex during transcription. A region of low conservation near residue 206 in subunit A corresponds to several deletions in the jaw domain of RPB1. RNAP subunit A' has three principal regions of high conservation with eukaryal RPB1: residues 199-245 which align with *S. cerevisiae* residues 223-268 in the clamp-head and clamp-core domains, residues 304-360 and 406-462 which align with *S. cerevisiae* residues 328-384, and residues 429-485 that encompass most of the active site domain and the dock domain, which is the primary binding target of TFIIB [121]. The dock domain is expected to be highly conserved for the key role of mediating the recruitment of RNAP to the promoter DNA by the TFIIB/TBP/DNA complex during transcription initiation [30]. Subunit A' also has several regions of low conservation: residues 150-169 have scores 0-0.1 and overlap with a 12-residue deletion in the clamp-head domain of RPB1, a subsequence near residue 587 overlaps with insertions in the pore1 domain of RPB1, and the C-terminal subsequence near residue 882 that corresponds with the foot domain in RPB1 which is mostly missing in the archaeal subunit. *P. furiosus* subunit B has several long regions of high conservation throughout its alignment with RPB2. One noteworthy region of low sequence similarity near *P. furiosus* residue 390 corresponds to an insertion in the protrusion domain of RPB2.

The remaining *P. furiosus* subunits are much smaller than the three core subunits and have smaller regions of high conservation. The largest of these, subunit D, has a highly conserved region from residues 29-69, corresponding to residues 30-70 in eukaryal RPB3,

covering portions of the dimerization and Domain2 regions. Regions of low conservation include the C-terminal region from residues 162-261, containing several deletions from the loop domain, and a short sequence near residue 79, corresponding to the Zn2 zinc loop in RPB3. RNAP subunit H is aligned with the C-terminal assembly domain of eukaryal RPB5 with mid-range conservation scores throughout; the N-terminal jaw domain of RPB5 is missing in the *P. furiosus* subunit. Subunit K has medium-to-high ranges of conservation though most of its alignment with the assembly domain of RPB6; an exception is at residue 31 adjacent to an alignment gap. Subunit L is aligned with the dimerization region of RPB11, with levels of conservation somewhat lower than subunit K; a pair of low scoring residues 52-53 are astride an alignment gap. Subunit N is the most conserved of all the smaller subunits, aligning with the Zn5 zinc ribbon of RPB10 and involving four highly conserved cysteine residues implicated in zinc binding.

4.4 Comparative modeling and assembly of model subunits into the RNAP complex

Homology modeling of the *P. furiosus* RNAP subunits was again conducted with the aid of MOE 2004.3 [118]. The MOE-Homology module used three major steps to construct a final comparative model. First, coordinates were copied from regions in the template to specify initial baseline geometry for model building. All atomic coordinates were copied from residues where sequence identity was conserved; for all other residues, only atomic coordinates from the backbone atoms were used for the initial geometry. The second step involved the building of intermediate comparative models using a Boltzmann-weighted randomized modeling procedure adapted from Levitt [122] in conjunction with a specialized strategy of handling insertions and deletions between the template and target proteins [123]. A large number (500-2000) of

intermediate models were used for each *P. furiosus* RNAP model subunit to increase the chances of a good quality structural model. The intermediate models were evaluated by a residue packing quality function sensitive to how much non-polar side-chain groups were buried and how often hydrogen bonding opportunities were satisfied in each model. Each intermediate model was subjected to minimization using the AMBER94 all-atom force field [124] sufficient to relieve any steric clashes. The final step of model construction involved the selection of the best model as determined by the packing quality function. The final model for each RNAP subunit was then further refined by “coarse” minimization, described by the MOE documentation as a series of conjugate gradient minimization steps using the Truncated Newton optimization algorithm, with RMS gradient test of 10. All final model structures for each RNAP subunit were then saved in PDB format. Further refinement of the individual model subunits was not pursued prior to assembly of the RNAP model complex. Applying additional minimization or molecular dynamics to the models could have resulted in proteins of better stereochemical quality but might also lessen structural similarity to the original template, which was the only point of reference for evaluating the model structure at the time of its construction. This also could have induced significant structural changes resulting in bad contacts and atomic clashes with other subunits when the RNAP complex was assembled.

Assembly of the separate *P. furiosus* RNAP model subunits into the model complex entailed superposing each subunit onto its homologous subunit in the *S. cerevisiae* RNAPII complex. The atomic coordinates for all subunits of the RNAPII template were first “frozen” to provide a consistent frame of reference for the superposition of each model subunit. Superpositioning was achieved using the MOE-ProSuperpose module, based on a method developed by Shapiro [125], by matching the four protein main-chain atoms (N, O, C, and C α)

between the template and target according to the original sequence alignment used in the model construction and minimizing the root-mean-square deviation (RMSD) of the distance between the aligned atoms. After all RNAP model subunits were superposed onto their RNAPII template subunits, the *S. cerevisiae* RNAPII complex was then removed from the molecular workspace, leaving only the assembled *P. furiosus* RNAP complex, which was subsequently saved in PDB format. The structure of the model RNAP complex is represented in Figure 4.4, with individual subunit structural representations in Figures 4.5-4.12.

4.5 Evaluation of model quality

Because the RNAP model subunits were only subjected to two separate rounds of minimization and were based on homologous eukaryal templates of low-to-medium sequence similarity, it was expected the models would have corresponding low-to-medium levels of stereochemical quality. One element of quality – correctness – can be measured by Ramachandran analysis [126] of the phi and psi dihedral angles for model residues (where the phi angle is the dihedral between the C_α and amino N atoms and the psi angle is the dihedral between the C_α and carboxylic C atoms of an amino acid.) The correctness of a model structure has been defined as the degree to which a given residue does not have an impossible structure e.g. bad phi and psi angles and atom clashes [127]. This is particularly important for model analysis when no other structure is available for comparison. Accuracy, another element of judging model quality, has been defined as how well the model fits its template [127]. One gauge of accuracy is the calculated RMSD between select atoms of aligned model and template structures, often using the C_α atom, or less frequently the four backbone atoms (C_α , N, C, and O), of each amino acid residue as measurement endpoints in the calculation. Further confidence in

the accuracy of a model may be gained by matching regions of secondary structure in the model and template proteins.

PDB structures of each individual RNAP model subunit were submitted to MolProbity [128] for the Ramachandran analysis. Ramachandran plots for each subunit are provided in Figures 4.13-4.20. Only model subunit H met the required percentage (>99.8%) of residues having phi and psi angles in allowed regions at 100% with no outliers while still falling short at 94.9% for the expected percentage (98%) of phi and psi angles in favored regions of the Ramachandran plot. The remainder of the subunits exhibit substandard percentages with respect to phi and psi values in favored and allowed regions. This is probably most acute in the largest subunits A, A', and B, which together form much of the core of RNAP. Although subunit A has the lowest percentage of residues with phi and psi values in allowed regions, subunits A' and B have 6 and 7 proline outliers respectively, a clear indicator of poor stereochemical quality and model incorrectness because of the limited number of conformations available to proline residues. As discussed above, further refinement involving minimization could remedy these defects. Because this was not pursued, caution should be exercised in applying predictions from this RNAP model.

RMSD determinations for each subunit were calculated using PyMOL [69] by first aligning the model and its template sequence, then applying superpositioning to minimize the RMSD between the corresponding protein structures. Calculations were based on both C $_{\alpha}$ and main chain atomic distances between aligned model and template amino acid residues. RMSDs for all subunits were determined to be less than 1 Å; because of the relative low-to-middle level of sequence similarity between the models and templates, these satisfactory results may be more appropriately credited to the MOE-Homology module in assigning atomic coordinates and

judging a best model from the pool of intermediate models generated during the comparative modeling procedure. Table 4.2 summarizes the MolProbity Ramachandran analysis for the individual RNAP model subunits along with the RMSD calculations between the superpositioned model and its *S. cerevisiae* template.

Table 4.2 Summary of Ramachandran analysis and RMSD calculations for *P. furiosus*

RNAP model subunits

<i>P. furiosus</i> subunit	Ramachandran analysis ^a				RMSD (Å) ^b		
	Total no. of residues	Total outliers	% favored ^c	% allowed ^d	Template	C _α ^e	Main chain ^f
A	387	24	79.1	93.8	RPB1	0.627	0.649
A'	905	32	84.9	96.5	RPB1	0.547	0.589
B	1115	41	83.6	96.3	RPB2	0.529	0.557
D	259	16	81.5	93.8	RPB3	0.599	0.642
H	78	0	94.9	100	RPB5	0.445	0.461
K	55	1	85.5	98.2	RPB6	0.406	0.475
L	93	1	89.2	98.9	RPB11	0.499	0.546
N	62	1	87.1	98.4	RPB10	0.395	0.452

^aRamachandran plots and analysis were generated by MolProbity [128].

^bRMSD calculations for superpositioned model and template made by PyMOL [69].

^cThe percent of residue expected to be in favored regions of the plot (98%).

^dThe percent of residue expected to be in allowed regions of the plot (>99.8%).

^eSuperpositioning and RMSD calculation based on C_α atom distances between aligned model and template.

^fSuperpositioning and RMSD calculation based on the protein main chain (C_α, N, C, and O) atomic distances between aligned model and template.

The results of the RMSD determinations impart some confidence in the accuracy of the model subunits in resembling the template structures; however, they do not overcome the poor results of the Ramachandran analyses of the models. As discussed above, any conclusions drawn from this model RNAP complex should be limited in scope.

4.6 Comparison to archaeal RNAP structures

When this *P. furiosus* RNAP model was built, there were no other RNAP structures of close homology available as templates with the exceptions of the open- and closed-clamp structures of *S. cerevisiae* RNAPII (PDB codes 1I50 and 1I3Q, respectively.) Recently, a cryo-electron microscopic structure of the *P. furiosus* RNAP has been determined at 13Å resolution [65]. At this resolution a quantitative structural comparison with the cryo-EM map cannot be attempted, but some quantitative comparisons can be made. As noted in the discussion above of template-model sequence alignments, there are important structural differences between the archaeal RNAP and eukaryal RNAPII, most likely with functional implications in transcription process. Both the model and the EM map structures reveal some common features. Subunit A has several deletions in the jaw domain of RPB1 and completely lacks the C-terminal repeat domain (CTD), a feature exclusive to eukaryal RNAPII and known to interact with the Mediator complex and numerous RNA processing enzymes. In subunit A', the region around the active site and dock domains and much of the clamp domain is highly conserved; both, in comparison to RPB1, reveal insertions in the pore1 domain, a mostly absent foot domain, and a deletion in the clamp head. The EM map and the homology model both have insertions in the protrusion domain in comparison to eukaryal RPB2. For subunit D, the model and EM structures show deletions from the zinc loop domain of RPB3. In subunit H of both structures, the jaw domain of the N-terminal portion of RPB5 is missing. The N-terminal region of RPB6 is absent in subunit K of both model and EM map structures. Both subunits L and N appear to be very structurally similar in both model and EM map structures.

The first of any archaeal RNAP crystal structures was deposited recently for the *Sulfolobus solfataricus* RNAP complex, solved as a dimer at 3.4Å resolution (PDB code 2PMZ)

[66]. Not surprisingly, sequence alignments of homologous RNAP subunits between *P. furiosus* and *S. solfataricus* using BLAST [112] highlight the greater sequence similarity of the two archaeal RNAP complexes, in contrast to the eukaryal RNAPII. Superpositioning of the homologous subunits was applied with PyMOL [69] and RMSDs were determined between the pairs of superpositioned structures. Tables 4.3 and 4.4 display a summary of these results.

Table 4.3 Sequence similarity and RMSD calculations for *P. furiosus* model subunits superpositioned with *S. solfataricus* RNAP subunits

<i>P. furiosus</i> subunit	<i>S. solfataricus</i> subunit ^a	NCBI BLAST ^b		RMSD (Å) ^c	
		% Identity	% Positives	C _α ^d	Main chain ^e
A	A(A'')	51	71	2.632	2.813
A'	A'(A)	56	71	1.711	1.767
B	B	55	73	2.286	2.355
D	D	34	50	1.199	1.439
H	H	46	63	1.171	1.182
K	K	49	67	1.073	1.212
L	L	43	62	0.898	0.970
N	N	62	87	1.144	1.355

^aAlternate names for archaeal RNAP subunits homologous to eukaryal RPB1 are provided in parenthesis; this is due to inconsistencies in naming conventions for these subunits.

^bAlignments conducted using NCBI BLAST [112]. Percentages were only reported for largest segments of the alignment between each subunit.

^cRMSD calculations for superpositioned model and template made by PyMOL [69].

^dSuperpositioning and RMSD calculation based on C_α atom distances between aligned model and template.

^eSuperpositioning and RMSD calculation based on the protein main chain (C_α, N, C, and O) atomic distances between aligned model and template.

The sequence and structural similarity of the *S. solfataricus* RNAP to the model RNAP suggest that it would have been a preferable template in the modeling of *P. furiosus* RNAP. This alleviated some of the serious stereochemical problems present in the model structure before resorting to further energy minimization. The *S. solfataricus* RNAP is in the closed-clamp form,

but its highly conserved active site and domain dock regions in subunit A should be structurally similar to the open-clamp forms of the *P. furiosus* model and its template *S. cerevisiae* RNAPII structure.

Table 4.4 Sequence similarity and RMSD calculations for *S. cerevisiae* RNAPII subunits superpositioned with *S. solfataricus* RNAP subunits

<i>S. cerevisiae</i> subunit	<i>S. solfataricus</i> subunit ^a	NCBI BLAST ^b		RMSD (Å) ^c	
		% Identity	% Positives	C _α ^d	Main chain ^e
RPB1	A(A'')	42	60	1.586	1.696
RPB1	A'(A)	30	50	2.577	2.683
RPB2	B	40	58	1.760	1.889
RPB3	D	24	43	1.093	1.177
RPB5	H	43	63	1.022	1.028
RPB6	K	42	64	0.762	0.794
RPB11	L	40	51	3.473	3.253
RPB10	N	46	73	1.044	1.036

^aAlternate names for archaeal RNAP subunits homologous to eukaryal RPB1 are provided in parenthesis; this is due to inconsistencies in naming conventions for these subunits.

^bAlignments conducted using NCBI BLAST [112]. Percentages were only reported for largest segments of the alignment between each subunit.

^cRMSD calculations for superpositioned structures made by PyMOL [69].

^dSuperpositioning and RMSD calculation based on C_α atom distances between aligned structures.

^eSuperpositioning and RMSD calculation based on the protein main chain (C_α, N, C, and O) atomic distances between aligned structures.

4.7 The conserved RNAP active site and dock domains

One of the most conserved regions of the RNAP complex encompasses the active site domain and the dock domain of RPB1, containing both the catalytic heart of the transcriptional engine and primary binding target of TFIIB [121]. The dock domain has high sequence conservation in RNAP complexes and plays an essential role in mediating the recruitment of RNAP to the promoter DNA by the TFIIB/TBP/DNA complex during transcription initiation [30]. The eukaryal active site domain comprises residues 346-374 and 436-507 of RPB1 and corresponds to residues 322-350 and 413-484 in the *P. furiosus* RNAP subunit A', and residues

319-347 and 411-482 in the *S. solfataricus* RNAP subunit A'. The eukaryal dock domain comprises residues 375-435 of RPB1 and corresponds to residues 351-412 in the *P. furiosus* RNAP subunit A', and residues 348-410 in the *S. solfataricus* RNAP subunit A'. Structure among all three is expected to be very similar, as demonstrated by the low RMSDs with the *P. furiosus* model: for *S. cerevisiae* (C_{α} , 0.439Å; main chain, 0.482Å) and for *S. solfataricus* (C_{α} , 0.815Å; main chain, 0.828Å). Figure 4.21 illustrates several views of the two domains within all three superposed structures.

The Mg^{2+} binding site of eukaryal RPB1 of the RNAP active center has been identified as the RNAP active center [68, 110]. The metal A binding site is centrally comprised of three highly conserved Asp residues from RPB1 that chelate the metal ion: D481, D483, and D485 which are homologous to *S. solfataricus* RNAP A' residues D456, D458 and D460 and *P. furiosus* RNAP A' residues D458, D460 and D462, with the trio of Asp residues separated by one amino acid. Figure 4.22 gives a close-up of the metal A binding site. Superpositioning of the model with the other crystallographic structures reveals a very similar placement with the Asp side-chains among all three homologs, with the side-chains oriented toward the metal position and chelating oxygens within 2-3Å of the metal. These structures are derived from both the open- and closed-clamp structures of RNAP/RNAPII complexes and show no major conformational changes between the two forms.

The binding site region of the dock domain, from eukaryal RNAPII RPB1 residues 409-419 has been identified as a binding site for the N-terminal zinc ribbon motif of TFIIB [121], most likely interacting with the hydrophilic residues in this highly conserved region near the RNAP active site. The relatively high level of sequence and structural conservation in this region suggest a similar binding interaction involving archaeal TFB. RPB1 residues 409-419

correspond to *S. solfataricus* RNAP A' residues 382-392 and *P. furiosus* RNAP A' residues 385-395. This multiple sequence alignment with CLUSTALW [129] reveals insertions near one end of this region as revealed by the structural deviations near RBP1 residue 419 adjacent to the partial helix assigned, as illustrated in Figure 4.23.

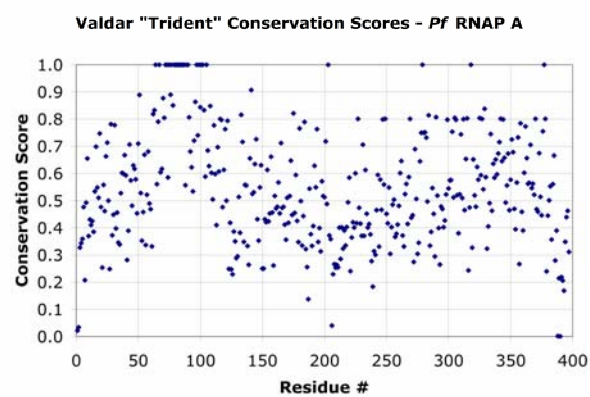
Eukaryal dock domain residue R412 has been previously targeted for mutagenesis by former R.A. Scott laboratory member H.T. Chen [130, 131] in conjunction with photocrosslinking and hydroxyl radical cleavage experiments that confirmed TFIIB interactions with the current dock domain sequence of interest. When eukaryal RPB1 residues 416-420 were replaced by the protein sequence GSG, a lethal mutation ensued. Interestingly, mutation of RPB1 residue R412, when mutated to Ala or His, led to a cold sensitive growth phenotype similar to those observed from mutations of TFIIB Zn ribbon [132, 133]. This resulted in no detectable interactions between TFIIB and RNAPII, and further reduced recruitment of TFIIB and RNAPII to the pre-initiation complex and lowered levels of *in vitro* transcriptional activity. RPB1 residue R412 is conserved with *P. furiosus* subunit A' residue R388 and *S. solfataricus* subunit A' residue R385, with all three having similar orientations. One striking feature of this locale surrounding RPB1 residue R412 is the abundance of Arg and Asp residues from both eukaryal and archaeal RNAPs, with the additional hydrophilic residues most likely involved in binding to the Zn ribbon region of TFIIB (or archaeal TFB). A close-up of the region immediately in the vicinity of RPB1 residue R412 is provided in Figure 4.24. As with the metal A binding site of the active site domain, both open- and closed-forms of the eukaryal and archaeal RNAPs show little difference in the structures of the active site and dock domain.

4.8 Conclusion of modeling exercise

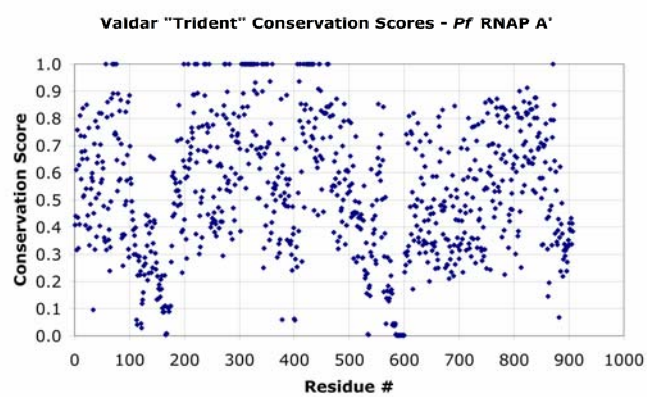
Despite the limitations of this comparative model, it did prove capable in modeling two of the most important and most conserved domains present in any RNAP, using the *S. solfataricus* RNAP structure as a comparative reference. However, the limitations extend beyond the aforementioned stereochemical problems and warrant caution in applying predictions from it. The assembly of the model RNAP subunits into the final complex is only an approximation, one that does not account for attractive interactions such as salt bridges between the different subunits. A domain-by-domain comparison would be best, particularly without a more similar modeling template. In addition, because the model and the template are open-clamped forms of the RNAP complex, the model is missing the E/F subunits (homologous to RPB7/RPB4); fortunately, these are present in the *S. solfataricus* RNAP structure. Indeed, the *S. solfataricus* RNAP structure represents the current best template for modeling archaeal RNAPs in the closed-clamp state, one that could be supplemented by the *S. cerevisiae* RNAPII as a secondary template. For now, this model allows a glimpse into the true structure of an open-clamp archaeal RNAP complex.

Figure 4.1 Conservation score in multiple sequence alignment with *P. furiosus* RNAP subunits A, A', and B. The tripartite conservation scoring scheme formulated by Valdar [120] is used to assess the sequence alignment for **A.** *P. furiosus* RNAP subunit A as shown in the Appendix, Figures C.1-C.17; for **B.** *P. furiosus* RNAP subunit A' as shown in the Appendix, Figures C.18-C.42; and for **C.** *P. furiosus* RNAP subunit B shown in the Appendix, Figures C.43-C.76.

A.



B.



C.

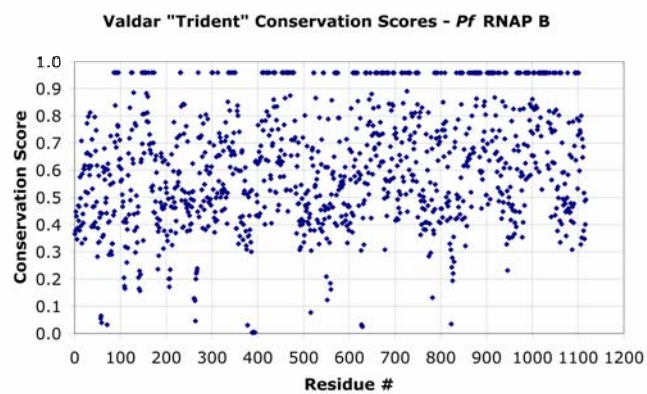
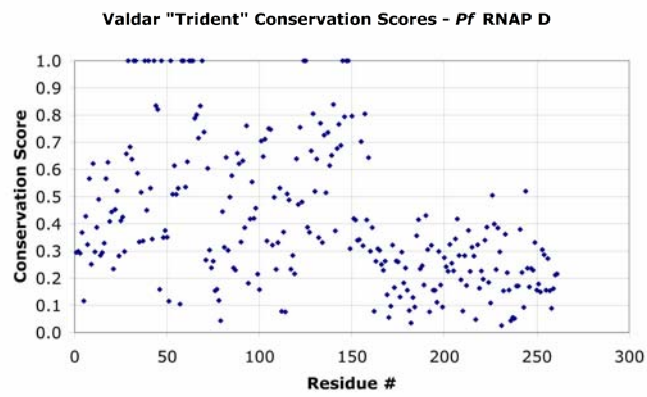
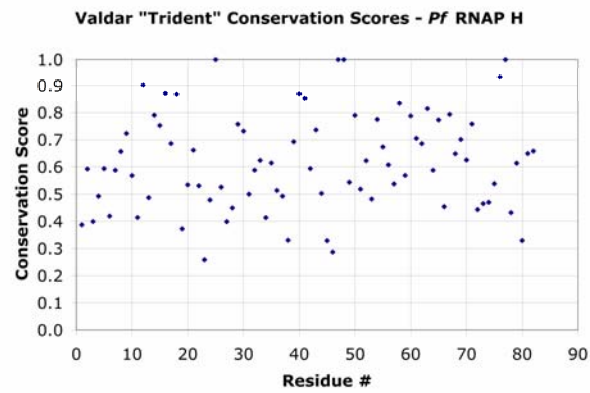


Figure 4.2 Conservation score in multiple sequence alignment with *P. furiosus* RNAP subunits D, H, and K. The tripartite conservation scoring scheme formulated by Valdar [120] is used to assess the sequence alignment for **A.** *P. furiosus* RNAP subunit D as shown in the Appendix, Figures C.77-C.85; for **B.** *P. furiosus* RNAP subunit H as shown in the Appendix, Figures C.86-C.87; and for **C.** *P. furiosus* RNAP subunit K as shown in the Appendix, Figures C.88-C.89.

A.



B.



C.

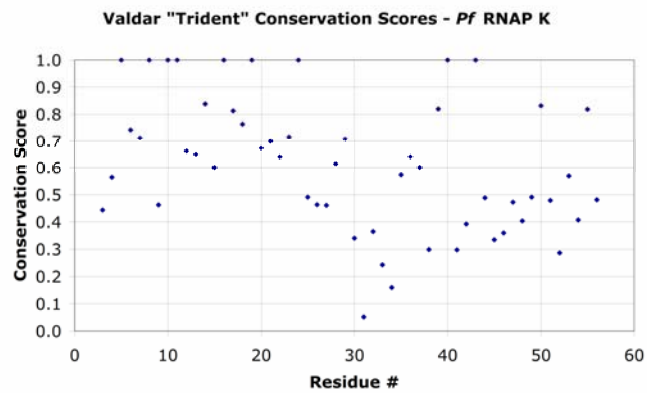
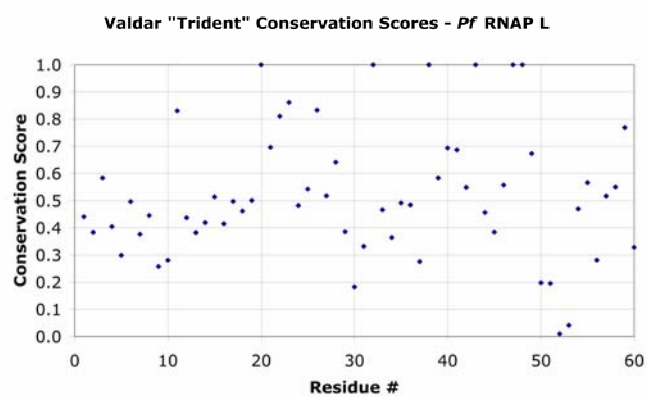


Figure 4.3 Conservation score in multiple sequence alignment with *P. furiosus* RNAP subunits L and N. The tripartite conservation scoring scheme formulated by Valdar [120] is used to assess the sequence alignment for **A.** *P. furiosus* RNAP subunit L as shown in the Appendix, Figures C.90-C.92; and for **B.** *P. furiosus* RNAP subunit N as shown in the Appendix, Figures C.93-C.95.

A.



B.

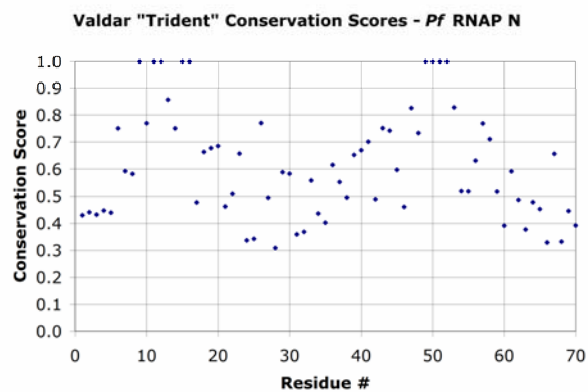


Figure 4.4 The assembled *P. furiosus* RNAP model complex. The RNAP complex was assembled by superposing the comparative models of nine RNAP subunits on the original 2.8Å-resolution RNAPII structure from *S. cerevisiae* (PDB code: 1I50) using MOE 2004.3 software [118]. Each individual model unit was structurally superposed on its template subunit while leaving all remaining proteins frozen. Once all model subunits were superposed, the template RNAPII subunits were removed, leaving the assembled *P. furiosus* RNAP complex. A color-coded legend of the RNAP subunits is provided at the top right. This image was created using PyMOL [69].

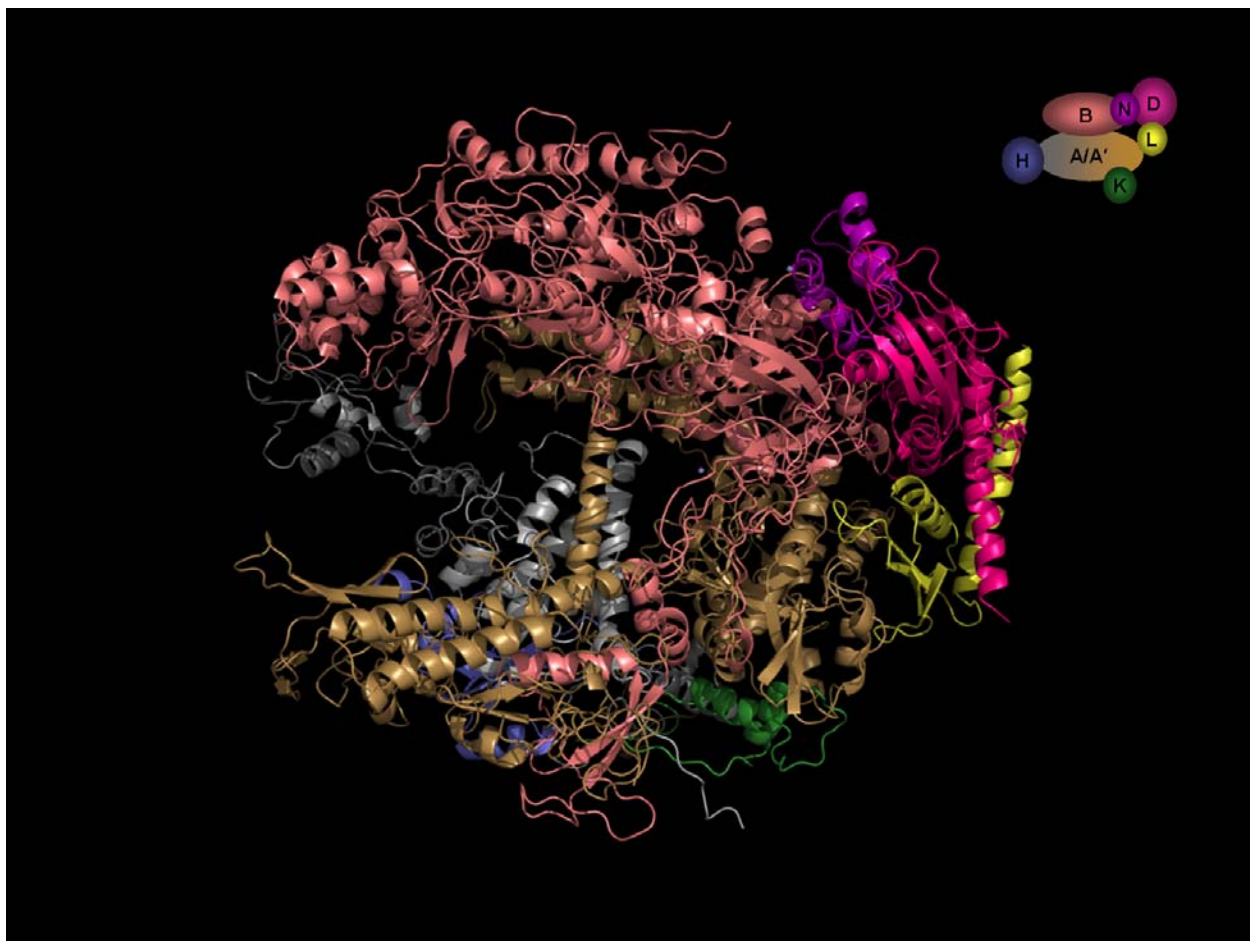


Figure 4.5 *Pyrococcus furiosus* RNAP model subunit A. The structure is colored with respect to secondary structural assignments as calculated by PyMOL [69]: loop regions as green, α -helices as blue, and β -sheets as yellow. This image was created using PyMOL [69].

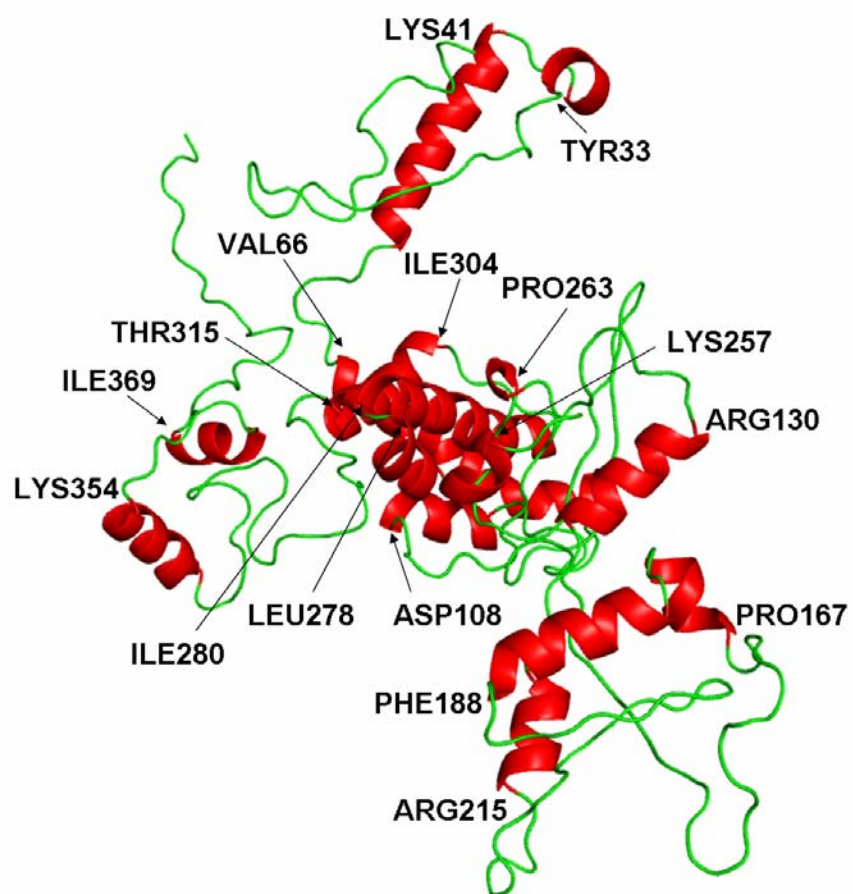


Figure 4.6 *Pyrococcus furiosus* RNAP model subunit A'. The structure is colored with respect to secondary structural assignments as calculated by PyMOL [69]: loop regions as green, α -helices as blue, and β -sheets as yellow. The figure includes front and reverse views of the protein model. This image was created using PyMOL [69].

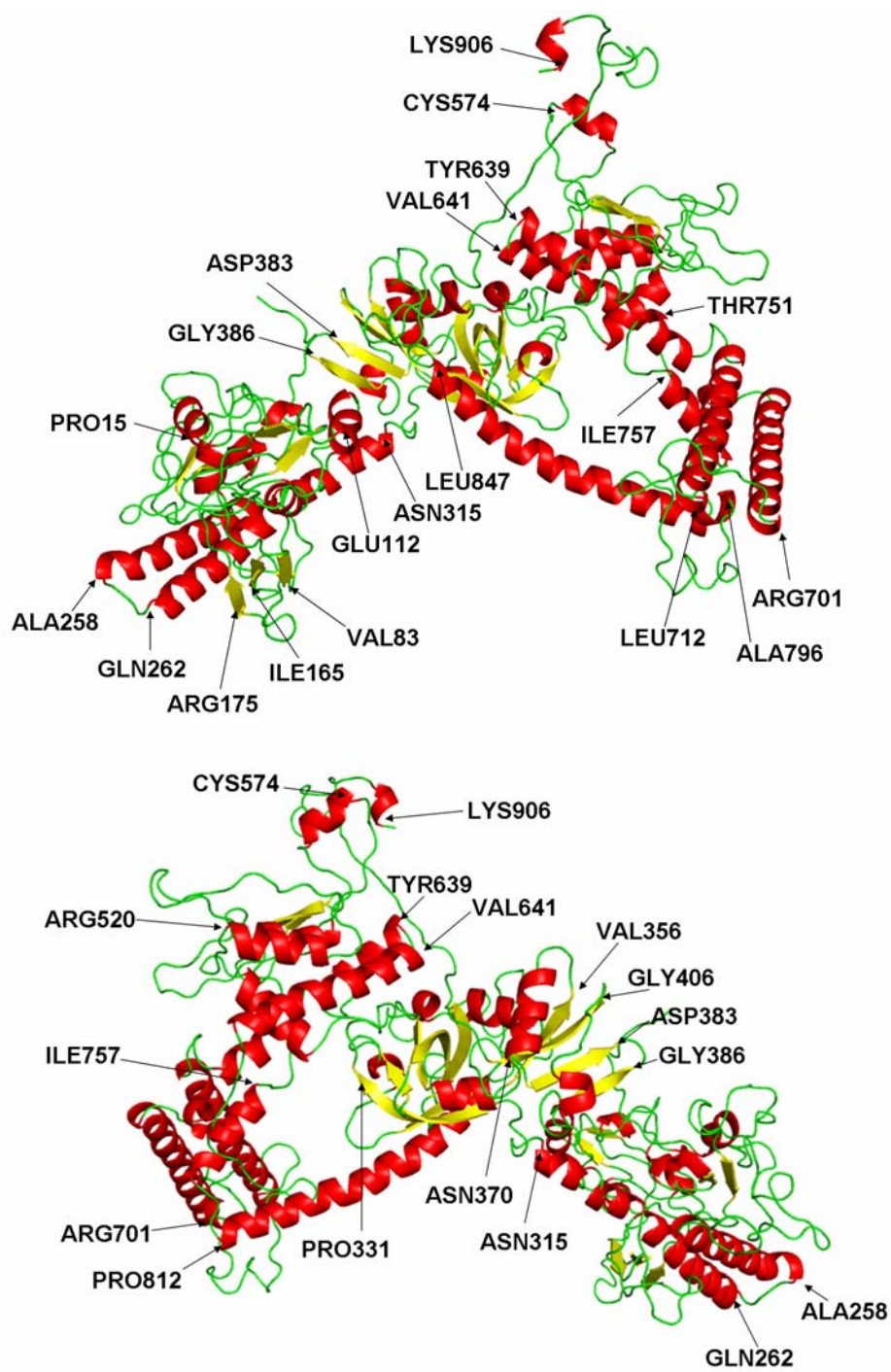


Figure 4.7 *Pyrococcus furiosus* RNAP model subunit B. The structure is colored with respect to secondary structural assignments as calculated by PyMOL [69]: loop regions as green, α -helices as blue, and β -sheets as yellow. The figure includes front and reverse views of the protein model. This image was created using PyMOL [69].

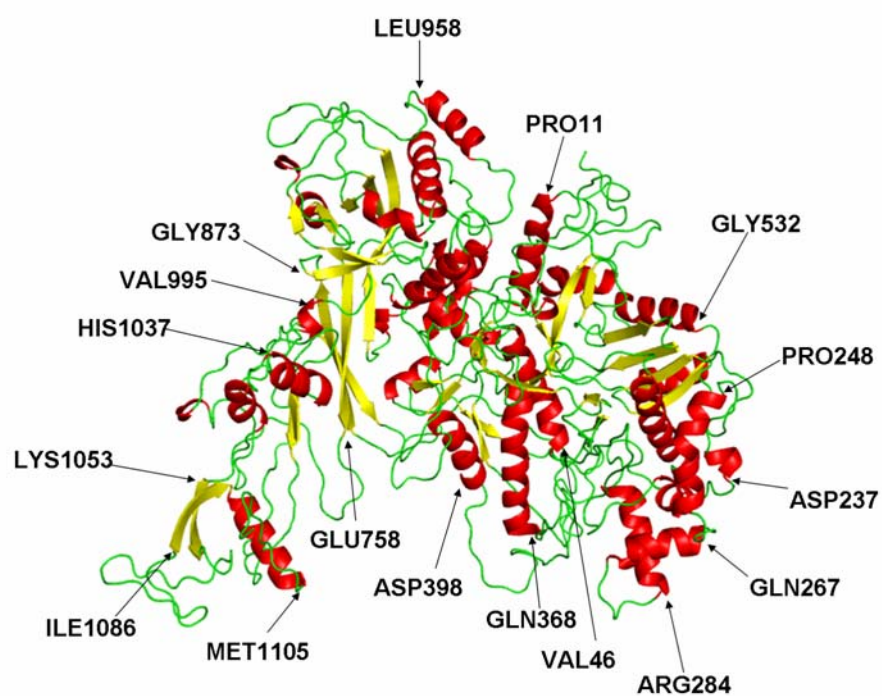
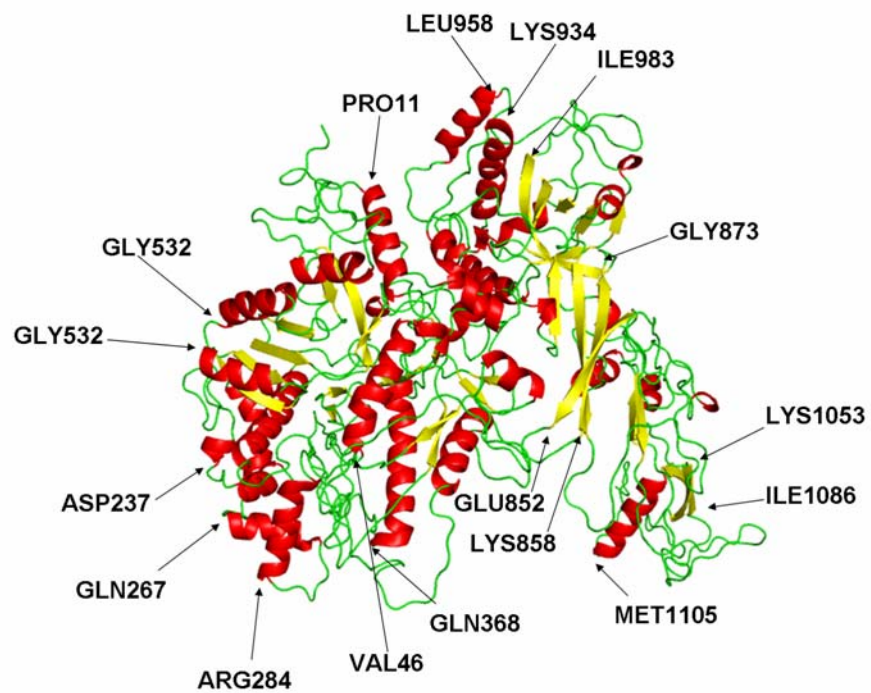


Figure 4.8 *Pyrococcus furiosus* RNAP model subunit D. The structure is colored with respect to secondary structural assignments as calculated by PyMOL [69]: loop regions as green, α -helices as blue, and β -sheets as yellow. This image was created using PyMOL [69].

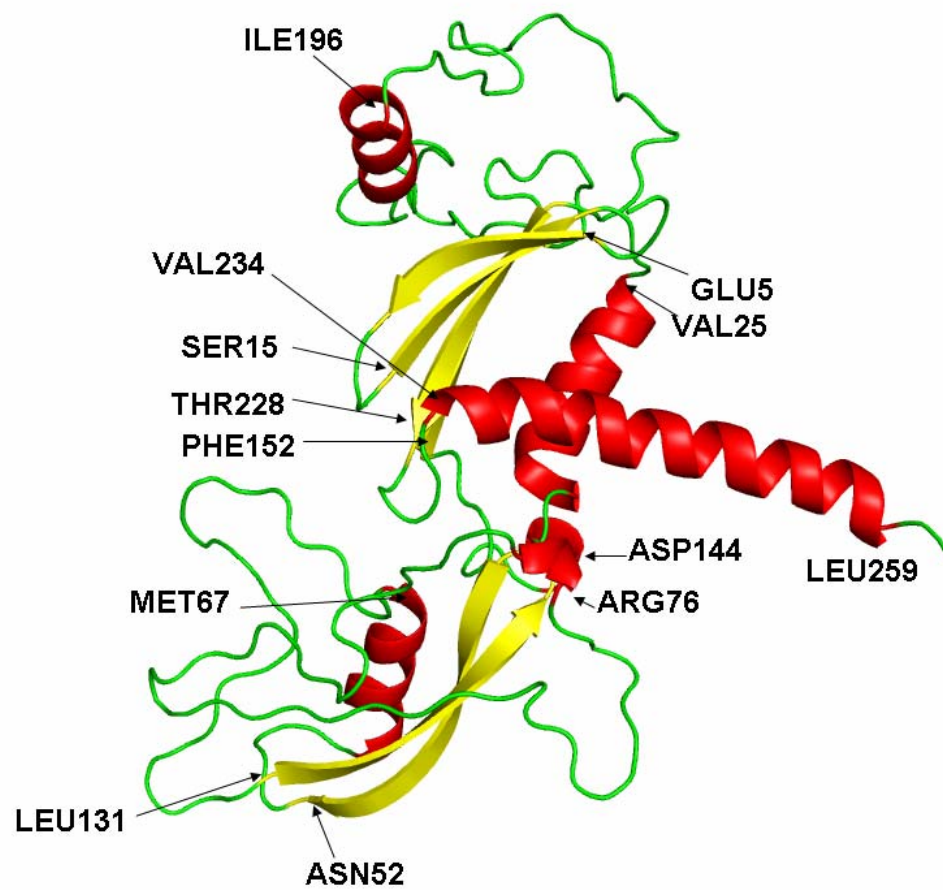


Figure 4.9 *Pyrococcus furiosus* RNAP model subunit H. The structure is colored with respect to secondary structural assignments as calculated by PyMOL [69]: loop regions as green, α -helices as blue, and β -sheets as yellow. This image was created using PyMOL [69].

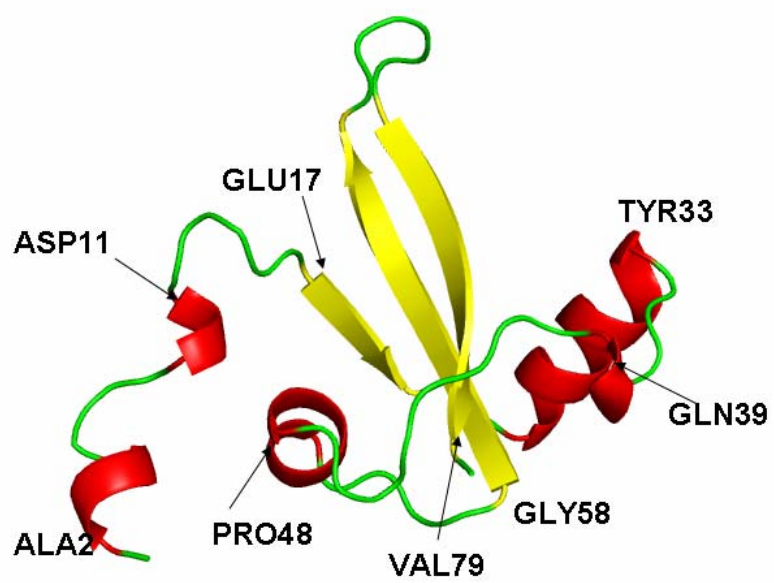


Figure 4.10 *Pyrococcus furiosus* RNAP model subunit K. The structure is colored with respect to secondary structural assignments as calculated by PyMOL [69]: loop regions as green, α -helices as blue, and β -sheets as yellow. This image was created using PyMOL [69].

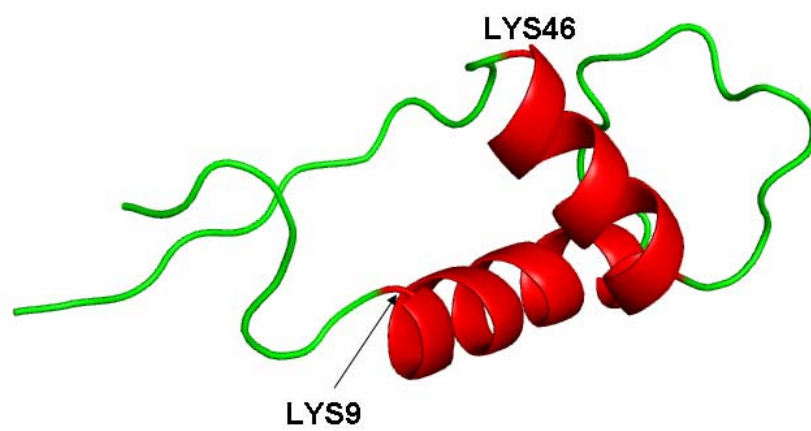


Figure 4.11 *Pyrococcus furiosus* RNAP model subunit L. The structure is colored with respect to secondary structural assignments as calculated by PyMOL [69]: loop regions as green, α -helices as blue, and β -sheets as yellow. This image was created using PyMOL [69].

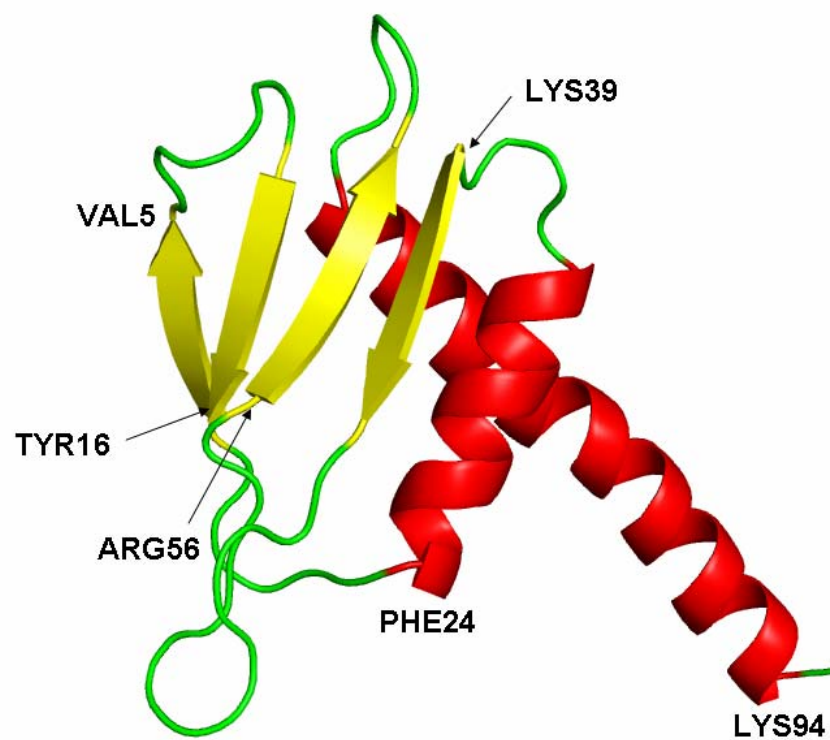


Figure 4.12 *Pyrococcus furiosus* RNAP model subunit N. The structure is colored with respect to secondary structural assignments as calculated by PyMOL [69]: loop regions as green, α -helices as blue, and β -sheets as yellow. This image was created using PyMOL [69].

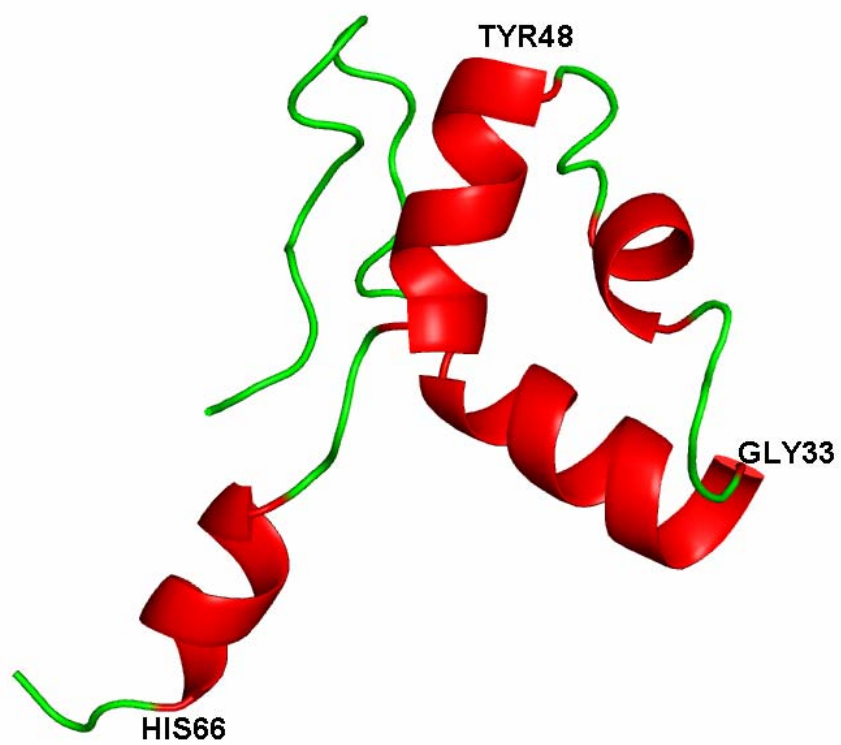
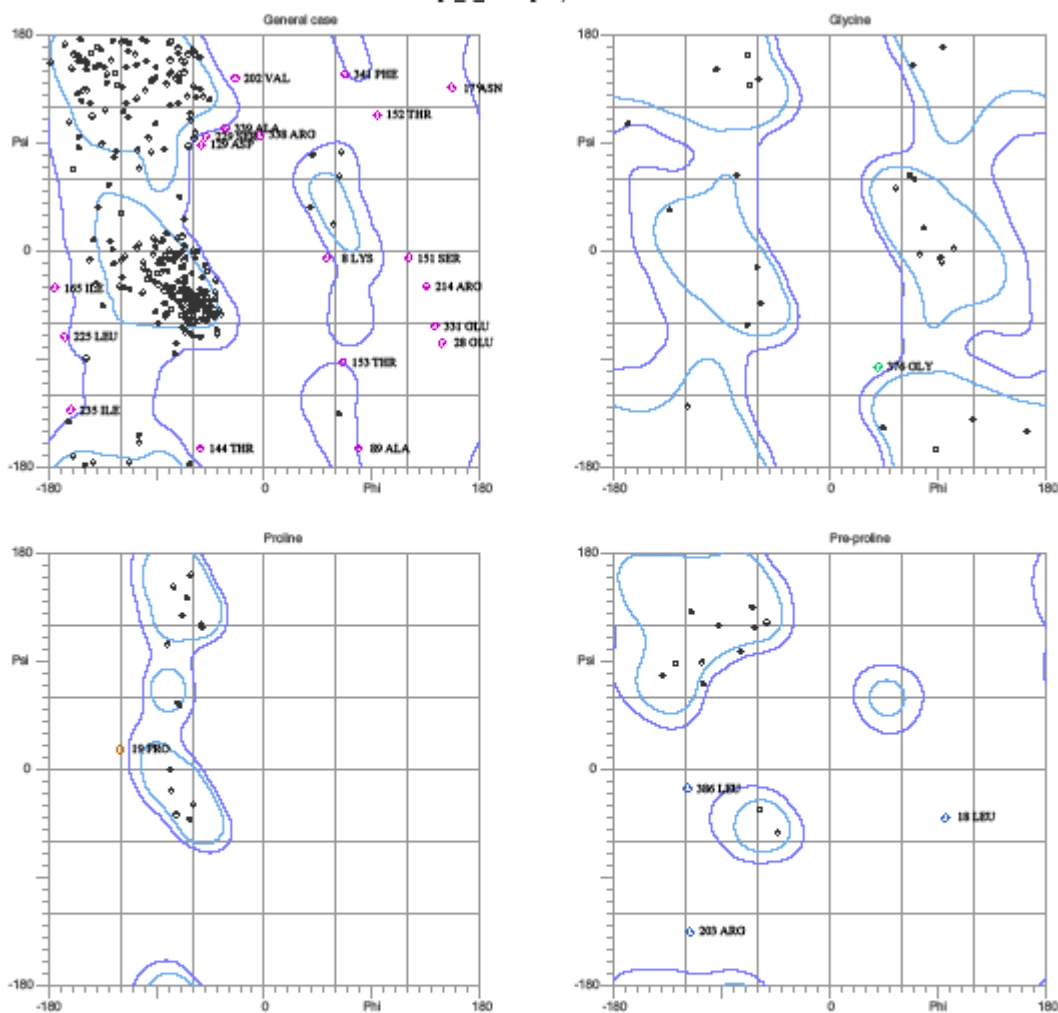


Figure 4.13 Ramachandran plot for *P. furiosus* RNAP model subunit A. Plots of the phi and psi dihedral angles for general case residues, glycines, prolines, and pre-prolines in model subunit A was generated by MolProbity [128]. The percentage of all residues in favored and allowed regions, along with a list of outlier residues with phi and psi angles included.

MolProbity Ramachandran analysis

pf_a_clean.pdb, model 1



79.1% (306/387) of all residues were in favored (98%) regions.
93.8% (363/387) of all residues were in allowed (>99.8%) regions.

There were 24 outliers (phi, psi):

8 LYS (52.8, -5.0)
17 ASN (156.6, 137.2)
18 LEU (96.3, -40.7)
19 PRO (-121.2, 17.2)
28 GLU (148.2, -76.6)
89 ALA (78.6, -164.7)
129 ASP (-53.8, 89.4)
144 THR (-54.8, -165.0)
151 SER (120.4, -5.6)
152 THR (94.9, 114.2)

153 THR (65.7, -92.2)
165 ILE (-176.0, -30.6)
202 VAL (-25.5, 145.4)
203 ARG (-117.5, -136.0)
214 ARG (135.6, -29.0)
225 LEU (-167.4, -71.8)
229 SER (-49.4, 96.5)
235 ILE (-162.6, -132.4)
331 GLU (142.4, -62.9)
338 ARG (-4.3, 97.4)
339 ALA (-33.1, 103.5)
341 PHE (67.7, 148.1)
376 GLY (40.3, -96.6)
386 LEU (-119.4, -15.6)

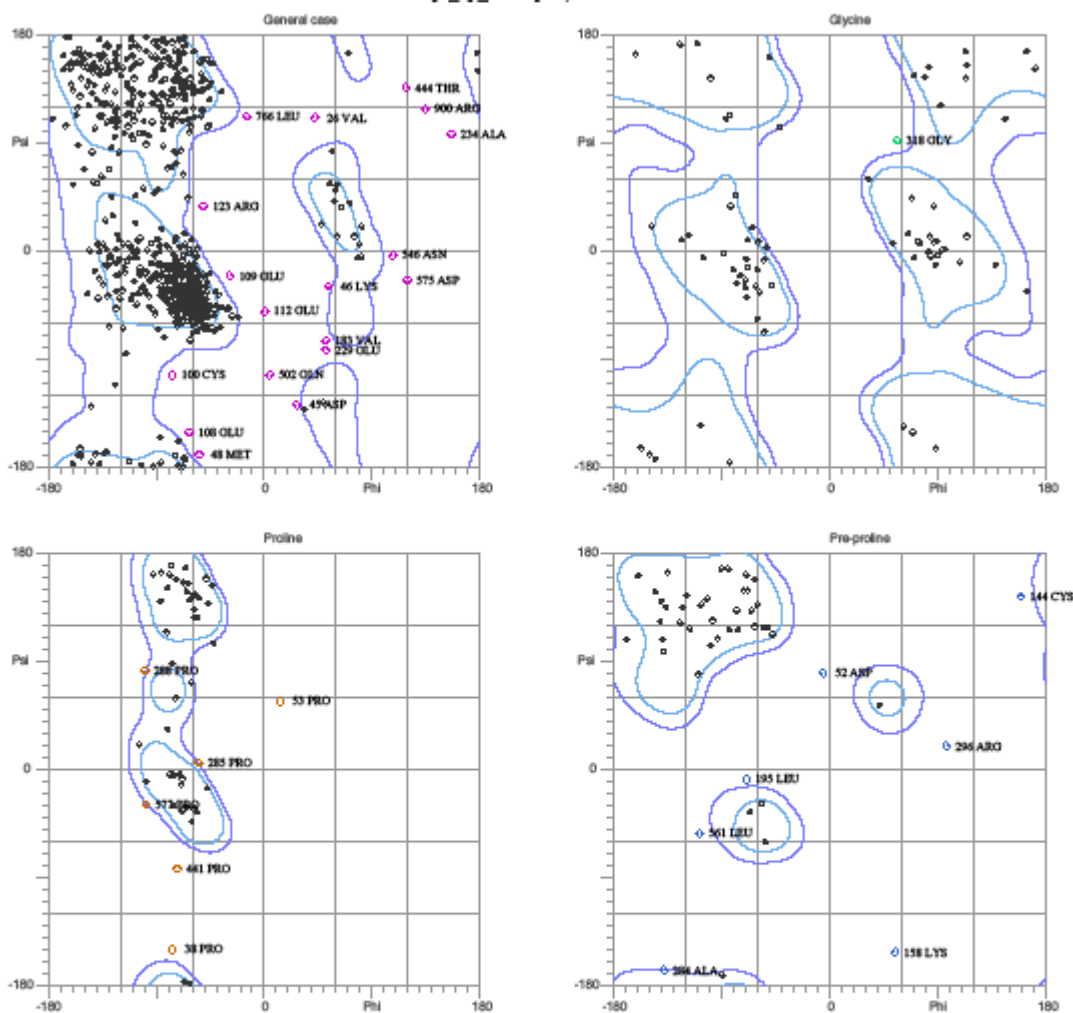
<http://kinemage.biochem.duke.edu>

Lovell, Davis, et al. Proteins 50:437 (2003)

Figure 4.14 Ramachandran plot for *P. furiosus* RNAP model subunit A'. Plots of the phi and psi dihedral angles for general case residues, glycines, prolines, and pre-prolines in model subunit A' was generated by MolProbity [128]. The percentage of all residues in favored and allowed regions, along with a list of outlier residues with phi and psi angles included.

MolProbity Ramachandran analysis

pf_ap_clean.pdb, model 1



84.9% (768/905) of all residues were in favored (98%) regions.
96.5% (873/905) of all residues were in allowed (>99.8%) regions.

There were 32 outliers (phi, psi):

26 VAL (42.6, 112.9)
38 PRO (-77.5, -150.6)
45 ASP (27.5, -129.6)
46 LYS (53.9, -29.2)
48 MET (-55.1, -169.8)
52 ASP (-6.6, 81.0)
53 PRO (13.7, 57.8)
109 CYS (-77.7, -105.5)
108 GLU (-63.5, -151.6)
109 GLU (-29.7, -20.2)
112 GLU (8.4, -50.8)
123 ARG (-52.0, 38.5)
144 CYS (159.6, 145.7)
158 LYS (54.7, -152.4)

183 VAL (51.8, -74.0)
195 LEU (-71.0, -8.9)
229 GLU (51.7, -82.5)
234 ALA (156.1, 98.9)
284 ALA (-129.1, -167.6)
285 PRO (-55.1, 6.1)
288 PRO (-100.5, 83.9)
296 ARG (97.5, 26.2)
318 GLY (56.2, 99.6)
441 PRO (-73.1, -82.9)
444 THR (118.1, 137.4)
502 GLN (4.1, -105.4)
546 ASN (107.8, -3.1)
561 LEU (-109.1, -53.8)
575 ASP (119.6, -24.6)
577 PRO (-99.6, -39.2)
768 LEU (-15.8, 113.7)
909 ARG (34.4, 119.9)

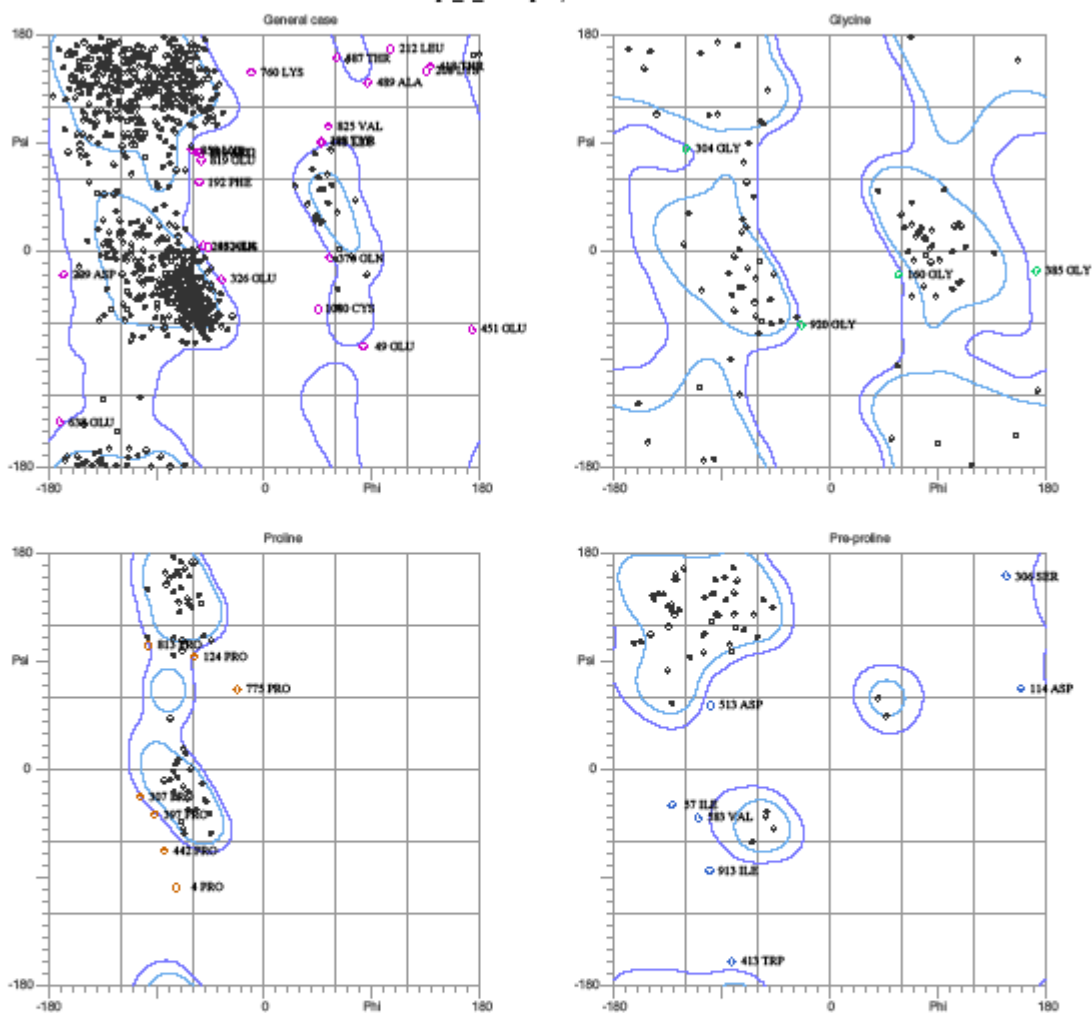
<http://kinemage.biochem.duke.edu>

Lovell, Davis, et al. Proteins 50:437 (2003)

Figure 4.15 Ramachandran plot for *P. furiosus* RNAP model subunit B. Plots of the phi and psi dihedral angles for general case residues, glycines, prolines, and pre-prolines in model subunit B was generated by MolProbity [128]. The percentage of all residues in favored and allowed regions, along with a list of outlier residues with phi and psi angles included.

MolProbity Ramachandran analysis

pf_b_clean.pdb, model 1



83.6% (32/37) of all residues were in General (GPA) regions
94.7% (35/37) of all residues were in allowed (G+D-EN) regions.

There were 61 authors (94%, *p* < .01).

4 FRO (-74.8, -76.0)
59 OLS (-82.8, -80.8)
57 LSE (-102.3, -20.2)
81 ABO (-56.0, 83.7)
114 AEP (-59.5, -68.2)
116 APO (-59.5, 95.4)
142 OLS (-37.5, -18.0)
190 FRS (-55.8, 58.1)
199 TTR (-67.7, 82.1)
199 ABO (-54.7, 82.8)
208 LSE (-35.5, 101.5)
209 AEP (-54.1, -19.1)
212 LEP (-65.1, 540.0)
245 AEA (-83.8, 5.2)
254 OLS (-130.3, 86.4)
258 TTR (-67.3, 140.5)
259 FRO (-106.4, -22.5)
262 ABO (-26.3, 23.8)

[illegible]

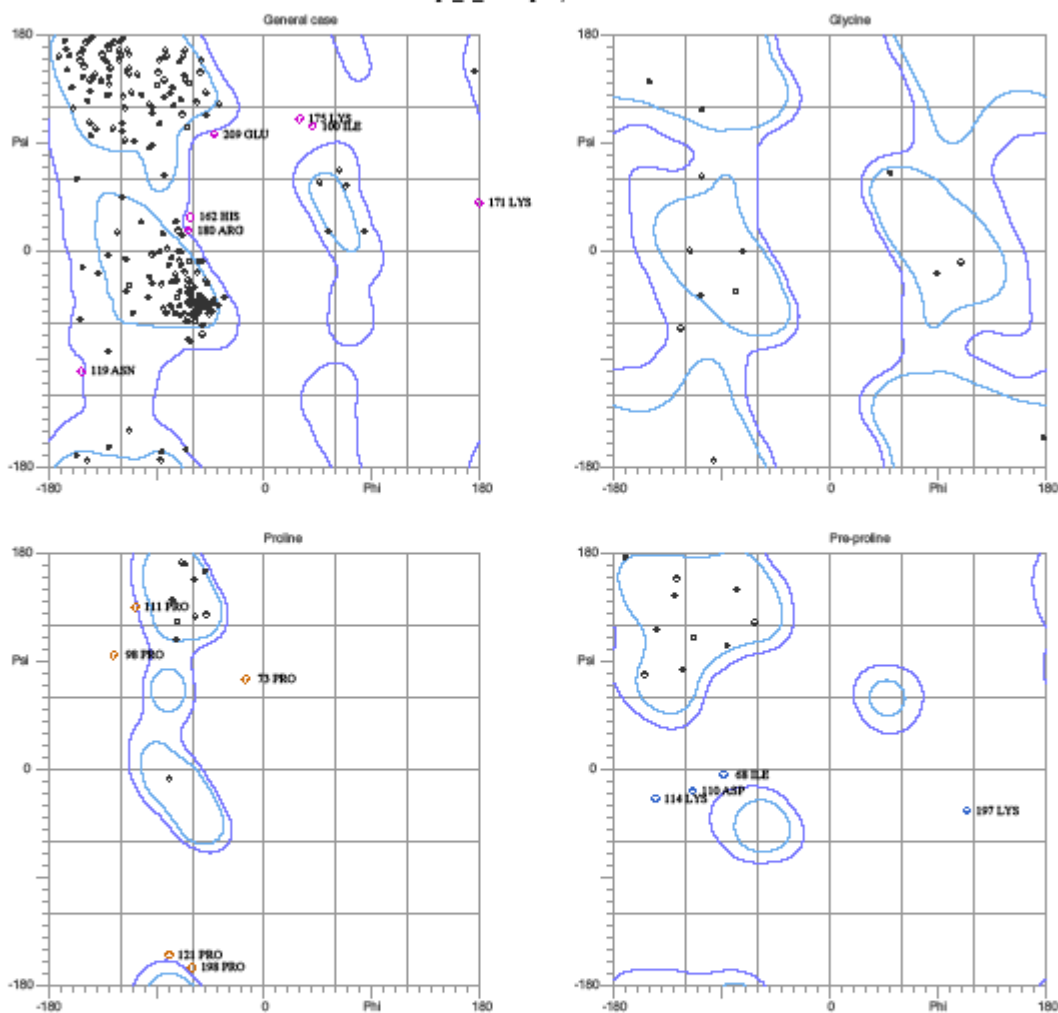
<http://kinemage.biochem.duke.edu>

Lovell, Davis, et al. Proteins 50:437 (2003)

Figure 4.16 Ramachandran plot for *P. furiosus* RNAP model subunit D. Plots of the phi and psi dihedral angles for general case residues, glycines, prolines, and pre-prolines in model subunit D was generated by MolProbity [128]. The percentage of all residues in favored and allowed regions, along with a list of outlier residues with phi and psi angles included.

MolProbity Ramachandran analysis

pf_d_clean.pdb, model 1



81.5% (211/259) of all residues were in favored (98%) regions.
93.8% (243/259) of all residues were in allowed (>99.8%) regions.

There were 16 outliers (phi, psi):

68 ILE (-89.7, -4.7)
73 PRO (-16.1, 76.9)
98 PRO (-126.4, 96.2)
100 ILE (40.3, 105.5)
110 ASP (-115.5, -17.4)
111 PRO (-108.1, 136.3)

114 LYS (-146.7, -24.3)
119 ASN (-153.7, -100.9)
121 PRO (-80.7, -154.7)
162 HIS (-62.8, 29.8)
171 LYS (179.4, 41.6)
175 LYS (29.1, 111.6)
180 ARG (-65.0, 18.5)
197 LYS (114.8, -34.0)
198 PRO (-61.6, -165.6)
209 GLU (-42.2, 98.3)

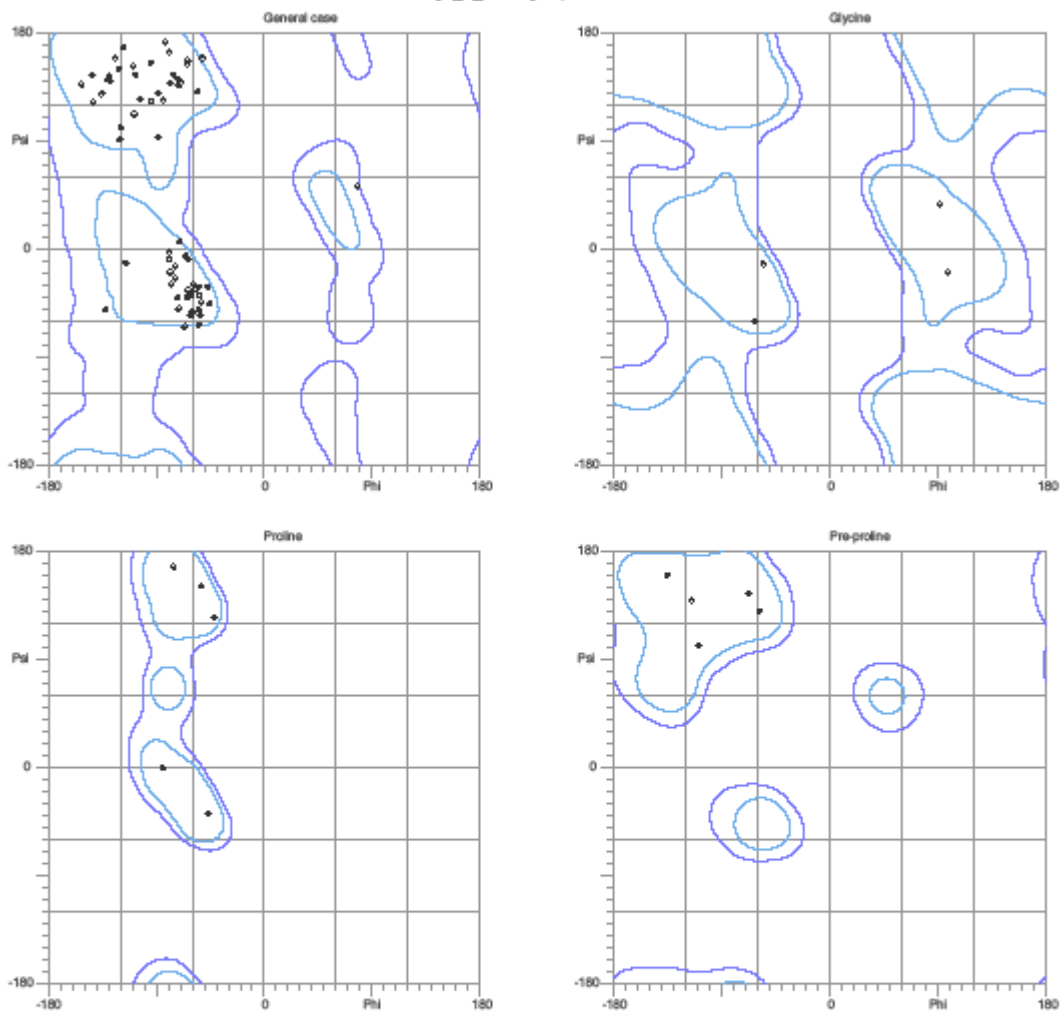
<http://kinemage.biochem.duke.edu>

Lovell, Davis, et al. Proteins 50:437 (2003)

Figure 4.17 Ramachandran plot for *P. furiosus* RNAP model subunit H. Plots of the phi and psi dihedral angles for general case residues, glycines, prolines, and pre-prolines in model subunit H was generated by MolProbity [128]. The percentage of all residues in favored and allowed regions, along with a list of outlier residues with phi and psi angles included.

MolProbity Ramachandran analysis

pf_h_clean.pdb, model 1



94.9% (74/78) of all residues were in favored (98%) regions.
100.0% (78/78) of all residues were in allowed (>99.8%) regions.

There were no outliers.

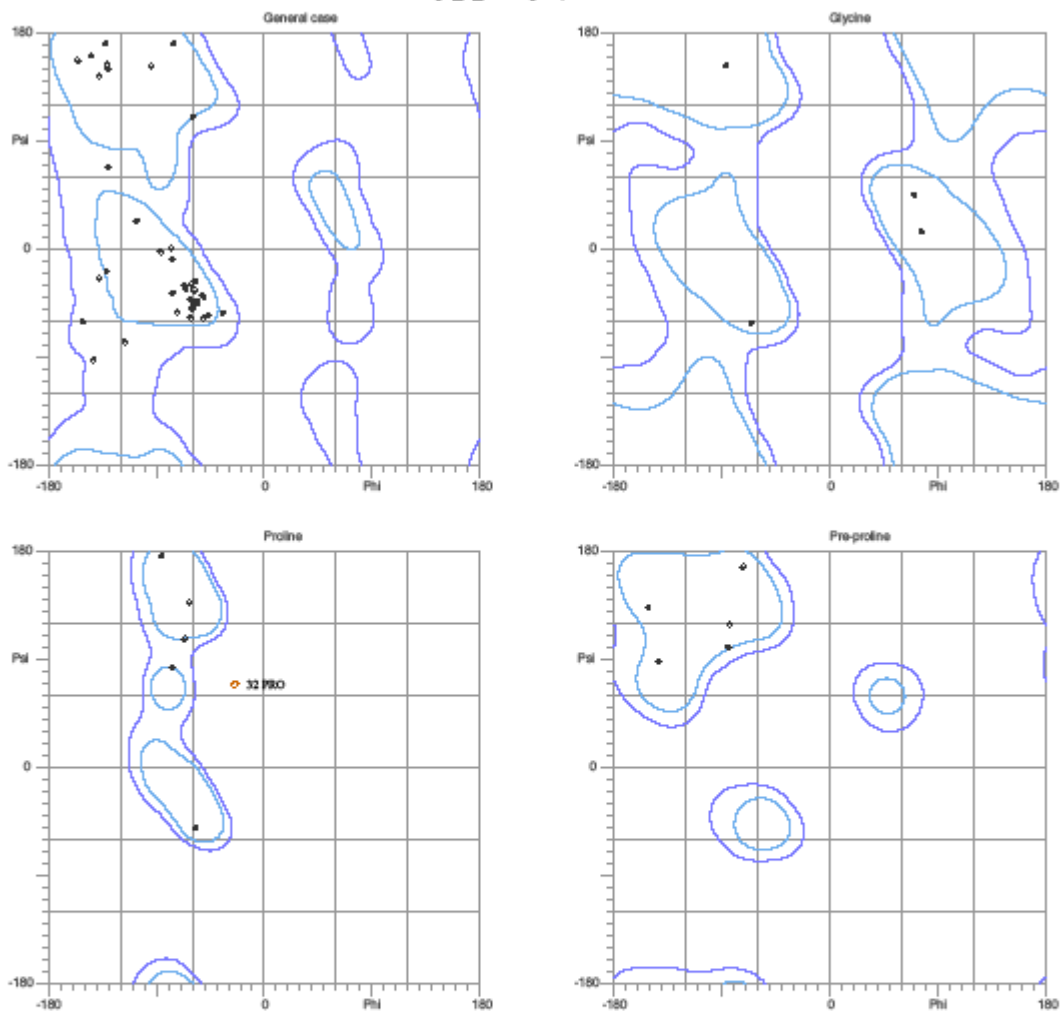
<http://kinemage.biochem.duke.edu>

Lovell, Davis, et al. Proteins 50:437 (2003)

Figure 4.18 Ramachandran plot for *P. furiosus* RNAP model subunit K. Plots of the phi and psi dihedral angles for general case residues, glycines, prolines, and pre-prolines in model subunit K was generated by MolProbity [128]. The percentage of all residues in favored and allowed regions, along with a list of outlier residues with phi and psi angles included.

MolProbity Ramachandran analysis

pf_k_clean.pdb, model 1



85.5% (47/55) of all residues were in favored (98%) regions.
98.2% (54/55) of all residues were in allowed (>99.8%) regions.

There were 1 outliers (phi, psi):
32 PRO (-25.8, 70.2)

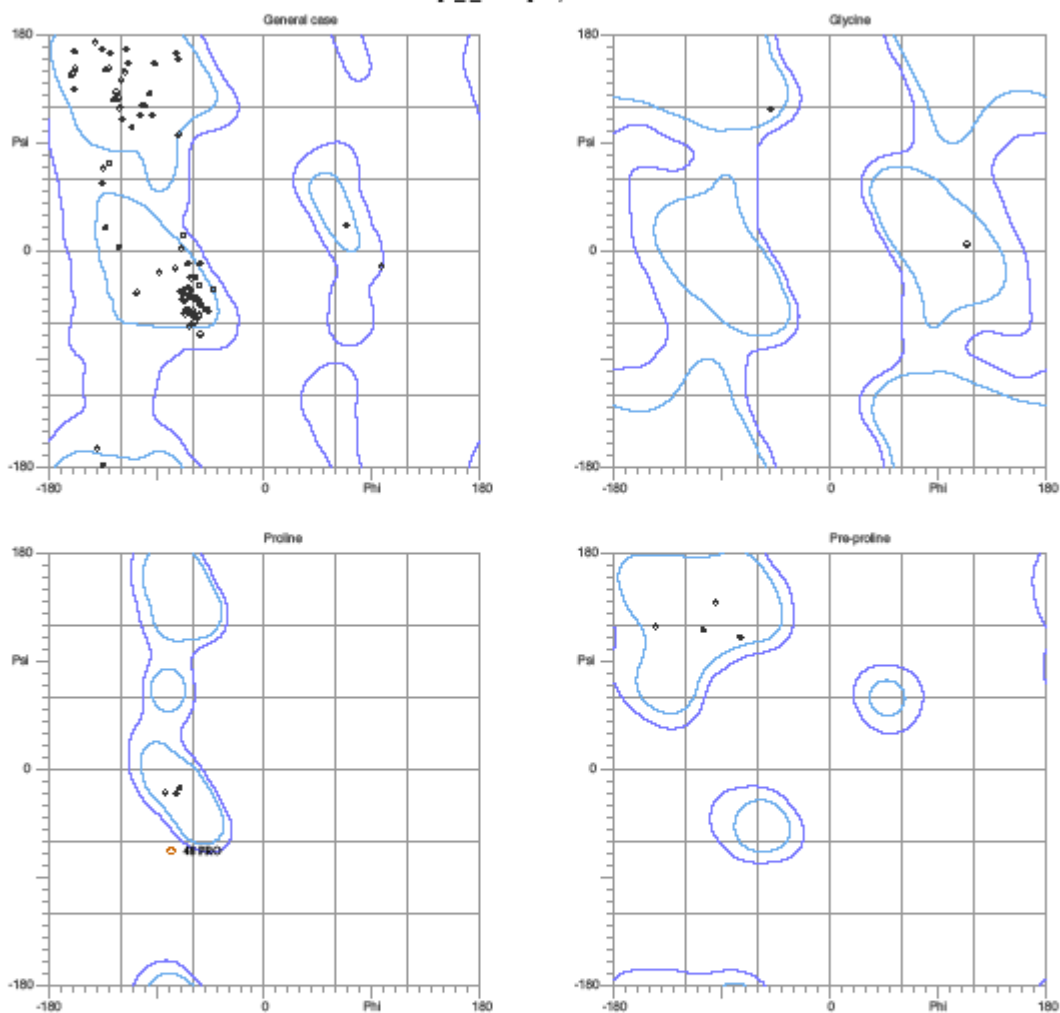
<http://kinemage.biochem.duke.edu>

Lovell, Davis, et al. Proteins 50:437 (2003)

Figure 4.19 Ramachandran plot for *P. furiosus* RNAP model subunit L. Plots of the phi and psi dihedral angles for general case residues, glycines, prolines, and pre-prolines in model subunit L was generated by MolProbity [128]. The percentage of all residues in favored and allowed regions, along with a list of outlier residues with phi and psi angles included.

MolProbity Ramachandran analysis

pf_1_clean.pdb, model 1



89.2% (83/93) of all residues were in favored (98%) regions.
98.9% (92/93) of all residues were in allowed (>99.8%) regions.

There were 1 outliers (phi, psi):
48 PRO (-78.4, -67.8)

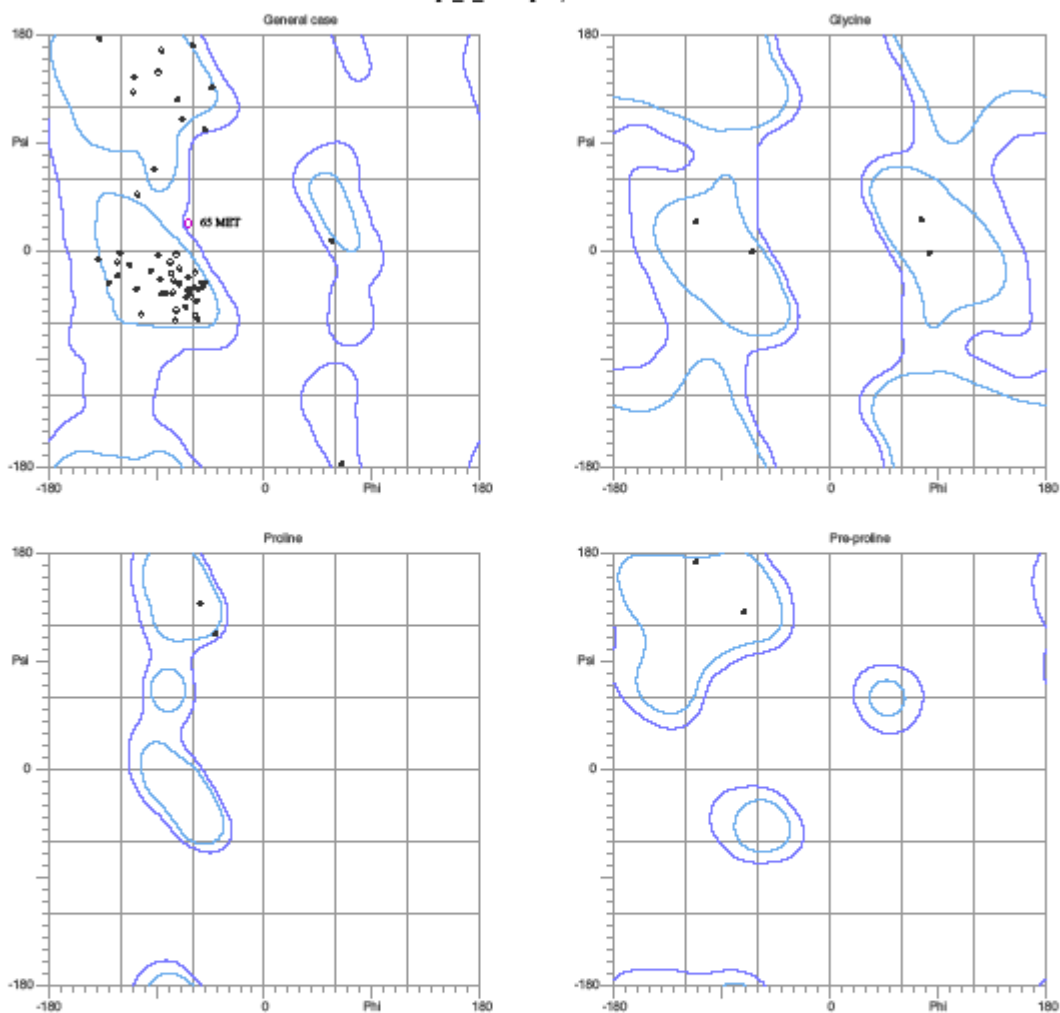
<http://kinemage.biochem.duke.edu>

Lovell, Davis, et al. Proteins 50:437 (2003)

Figure 4.20 Ramachandran plot for *P. furiosus* RNAP model subunit N. Plots of the phi and psi dihedral angles for general case residues, glycines, prolines, and pre-prolines in model subunit N was generated by MolProbity [128]. The percentage of all residues in favored and allowed regions, along with a list of outlier residues with phi and psi angles included.

MolProbity Ramachandran analysis

pf_n_clean.pdb, model 1



87.1% (54/62) of all residues were in favored (98%) regions.
98.4% (61/62) of all residues were in allowed (>99.8%) regions.

There were 1 outliers (ϕ , ψ):
65 MET (-64.9, 24.8)

<http://kinemage.biochem.duke.edu>

Lovell, Davis, et al. Proteins 50:437 (2003)

Figure 4.21 Superposed active site and dock domains from *S. cerevisiae* RNAPII, *P.*

***furiosus* RNAP, and *S. solfataricus* RNAP.** Structures of the open-clamp *S. cerevisiae* RNAPII RPB1 (PDB code 1I50) [68], the closed-clamp dimeric structure of *S. solfataricus* RNAP (PDB code 2PMZ) [66], and the *P. furiosus* comparative model were superposed in MOE [118] and PyMOL [69]. The eukaryal active site domain comprises residues 346-374 and 436-507 of RPB1 and corresponds to residues 322-350 and 413-484 in the *P. furiosus* RNAP subunit A', and residues 319-347 and 411-482 in the *S. solfataricus* RNAP subunit A'. In the active site domain, the metal A ion is colored as darkviolet, *S. cerevisiae* active site protein are purple-blue, *S. solfataricus* as light blue, and *P. furiosus* as a medium "TV blue". The eukaryal active site domain comprises residues 375-435 of RPB1 and corresponds to residues 351-412 in the *P. furiosus* RNAP subunit A', and residues 348-410 in the *S. solfataricus* RNAP subunit A'. In the dock domain, *S. cerevisiae* protein is colored as chartreuse, *S. solfataricus* as pale green, and *P. furiosus* as dark forest green. Side-chains implicated in metal A binding or interacting with TFIIB are shown. **A.** View of the two domains with the active site to the front. **B.** Reverse view with the dock domain to the front left. **C.** Another view of the dock domain with the dock domain to the front right, further exposing residues involved in TFIIB/TFB binding interactions. This picture was created with PyMOL.

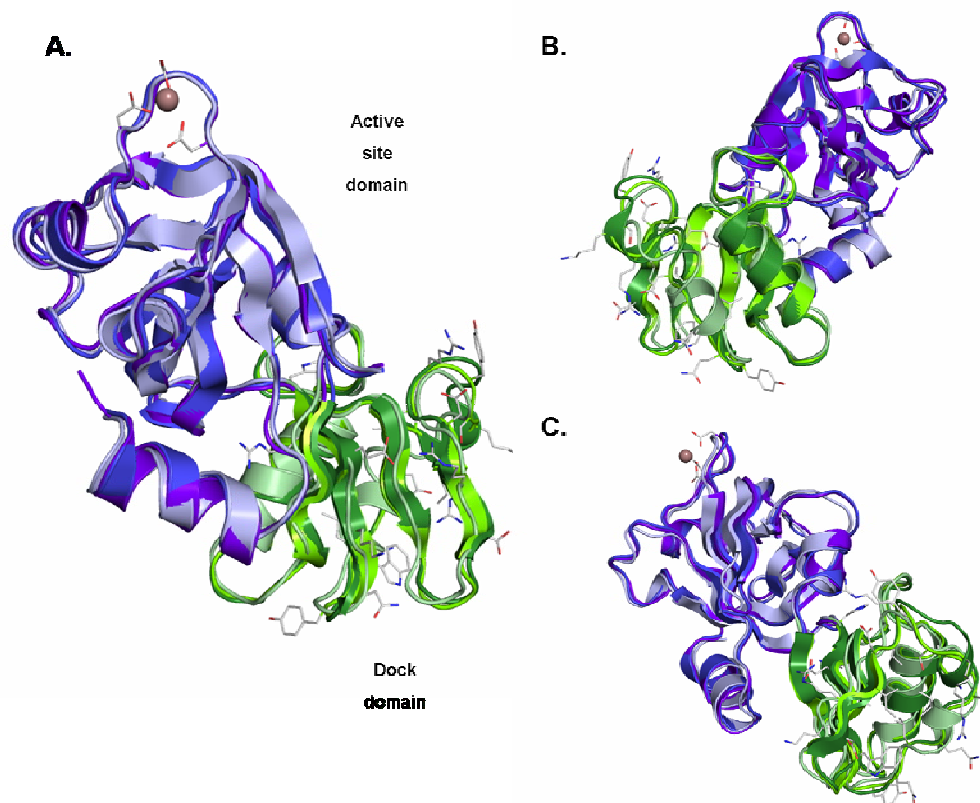
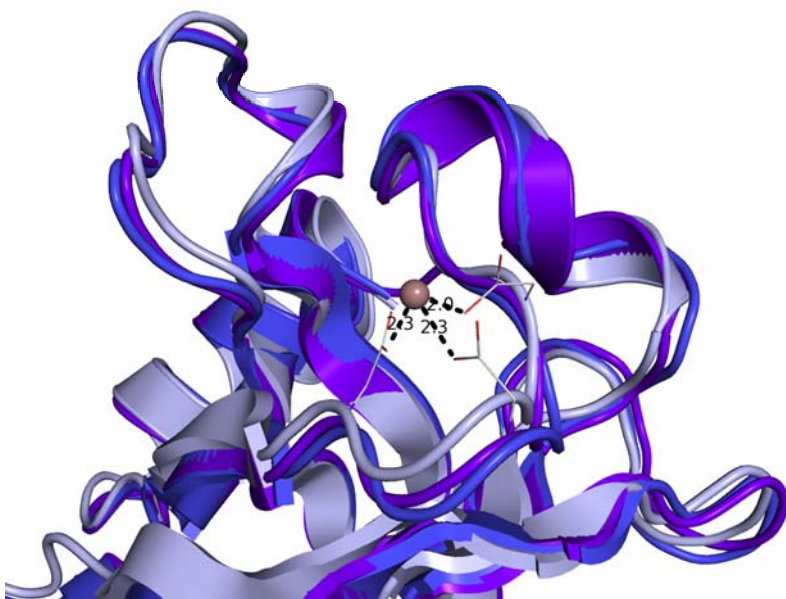


Figure 4.22 Superposed metal A-binding sites of active center domains of *S. cerevisiae* RNAPII, *P. furiosus* RNAP, and *S. solfataricus* RNAP. The Mg^{2+} binding site of eukaryal RPB1 of the RNAP active center has been identified and localized as the RNAP active center [68, 110]. The metal A binding site contains three highly conserved residues from RPB1 that chelate the metal ion: D481, D483, and D485 (homologous to *S. solfataricus* RNAP A' residues D456, D458 and D460 and *P. furiosus* RNAP A' residues D458, D460 and D462.) Structures of the open-clamp *S. cerevisiae* RNAPII RPB1 (PDB code 1I50) [68], the closed-clamp dimeric structure of *S. solfataricus* RNAP (PDB code 2PMZ) [66], and the *P. furiosus* comparative model were superposed in MOE [118] and PyMOL [69]. The metal A ion is colored as darkviolet, *S. cerevisiae* protein as purple-blue, *S. solfataricus* as light blue, and *P. furiosus* as a medium “TV blue”. **A.** Distances from the Asp side chain oxygens are approximate. **B.** A closer view of the metal A binding site with the side chains of the chelating residues for *S. solfataricus* and *P. furiosus* shown. This picture was created with PyMOL.

A.



B.

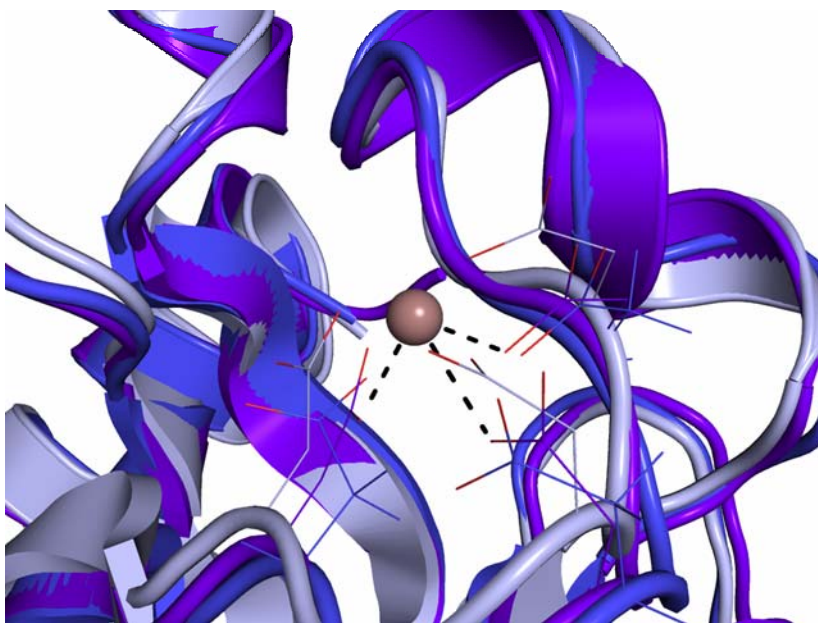
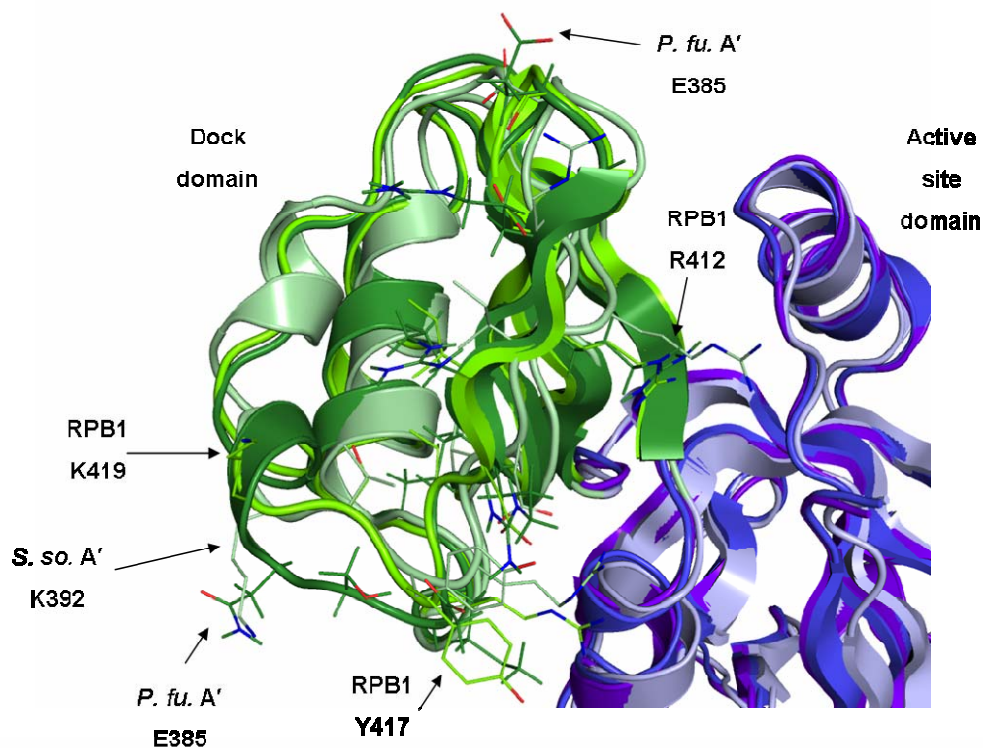


Figure 4.23 Superposed TFIIB interaction regions in the dock domain from *S. cerevisiae* RNAPII, *P. furiosus* RNAP, and *S. solfataricus* RNAP. The TFIIB interaction region of the dock domain, from eukaryal RNAPII RPB1 residues 409-419, has been identified as a binding site for the N-terminal zinc ribbon motif of TFIIB [121], most likely interacting with the hydrophilic residues in this highly conserved region near the RNAP active site. The relatively high level of sequence and structural conservation in this region suggest a similar binding interaction involving archaeal TFB. **A.** Structures of the open-clamp *S. cerevisiae* RNAPII RPB1 (PDB code 1I50) [68], the closed-clamp dimeric structure of *S. solfataricus* RNAP (PDB code 2PMZ) [66], and the *P. furiosus* comparative model were superposed in MOE [118] and PyMOL [69]. RPB1 residues 409-419 correspond to *S. solfataricus* RNAP A' residues 382-392 and *P. furiosus* RNAP A' residues 385-395. RPB1 residue R412 is conserved with *P. furiosus* subunit A' residue R388 and *S. solfataricus* subunit A' residue R385. *S. cerevisiae* protein is colored as chartreuse, *S. solfataricus* as pale green, and *P. furiosus* as dark forest green. This picture was created with PyMOL. **B.** Multiple sequence alignment with CLUSTALW [129] reveals insertions near one end of this region as revealed by the structural deviation near RBP1 residue 419. Sequence alignment image was created with the JalView alignment editor [108].

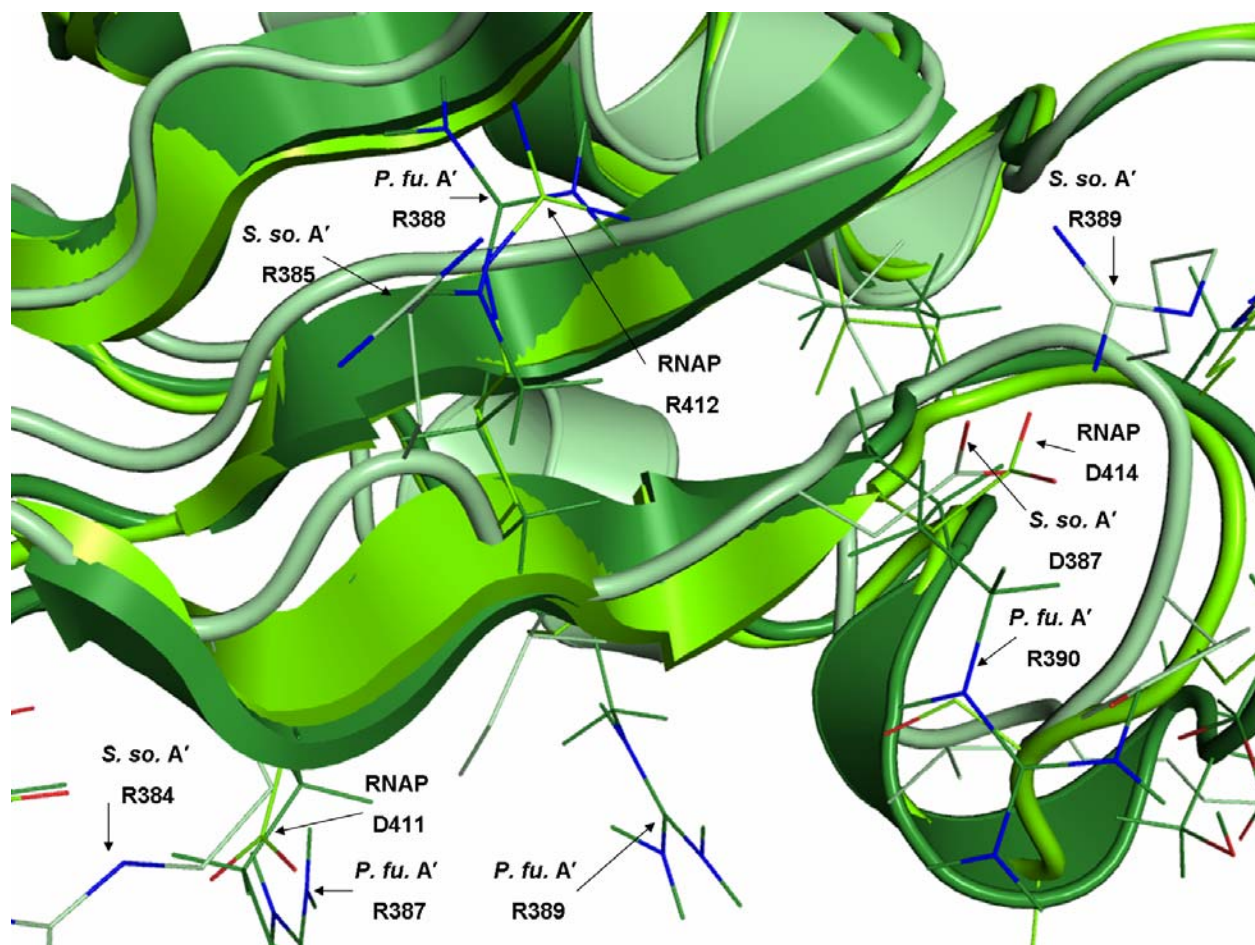
A.



B.

<i>S. cerevisiae</i> /346-507	346	DFSARTVISGDPNLELDQVGVPKSI AKTLTYPEVVTPYNI DRLTQLVRNGPNEHP	400
<i>P. fustus</i> /322-484	322	NFSARTVISPDPMISINEVGVPVEVAMELTVPEKVTEFN IERLRKMVLNGPDKYP	378
<i>S. solfataricus</i> /319-482	319	DFSRTVISPDPNISIDEVGVPEI IAKTLTVPERITPWNIEKL RQFVINGPDKWP	373
<i>S. cerevisiae</i> /346-507	401	GAKYVIRDSGDRIDL RYS. KRAGDIQLQYGWKVERHIMDNDPVL FNRQPSLHKM	453
<i>P. fustus</i> /322-484	377	GANYVIDPEGRRRRIMDS. NKETLANQLDIGWTVERHLMDDIVLFNRQPSLHRM	430
<i>S. solfataricus</i> /319-482	374	GANYVIRPDGRRIDL RYVKDRKELASTLAPGYI IERHLIDGDIVLFNRQPSLHR I	428
<i>S. cerevisiae</i> /346-507	454	SMMAHVRKVIPYSTFRNL SVTSPYNADFDGDEMNLHV PQSEETRAELSOLCAV	507
<i>P. fustus</i> /322-484	431	SIMAHVRVRMPYRTFRNL AVCPPYNADFDGDEMNLHV PQTEEAQAEARILMEV	484
<i>S. solfataricus</i> /319-482	429	SMMAHVRVLKGLTFRNL LVCPPYNADFDGDEMNLHV PQSEEAIAEAKEIMLV	482

Figure 4.24 The molecular neighborhood of RPB1 residue R412 in the TFIIB binding site of the dock domain. A closer view of one side of the putative TFIIB-dock domain interaction site domain is provided, partly spanning eukaryal RNAPII RPB1 residues 411-416, and centers around residue R412. This residue has been demonstrated to be involved in TFIIB binding by mutagenesis experiments [130, 131]. In particular, RPB1 residue R412 is conserved with *P. furiosus* subunit A' residue R388 and *S. solfataricus* subunit A' residue R385, with all three have very similar side-chain orientations. This locale is rich in Arg and Asp residues from both eukaryal and archaeal RNAPs. Structures of the open-clamp *S. cerevisiae* RNAPII RPB1 (PDB code 1I50) [68], the closed-clamp dimeric structure of *S. solfataricus* RNAP (PDB code 2PMZ) [66], and the *P. furiosus* comparative model were superposed in MOE [118] and PyMOL [69]. RPB1 residues 409-419 correspond to *S. solfataricus* RNAP A' residues 384-389 and *P. furiosus* RNAP A' residues 387-392. *S. cerevisiae* protein is colored as chartreuse, *S. solfataricus* as pale green, and *P. furiosus* as dark forest green. This picture was created with PyMOL.



CHAPTER 5

CONCLUSION AND FUTURE APPROACHES

5.1 Two computational methods to aid transcriptional research

Although the computational approaches described in this thesis, genome-wide motif detection and comparative modeling of an RNAP complex, differed in degrees of success and were limited in scope to *Pyrococcus*, they do stand as examples of the relative ease in supplementing experimental researchers by using *in silico* methods as a pathway to discovery. Both methods achieve some success in elucidating different aspects of the archaeal transcriptional machinery, but they have some limitations that may be remedied by future researchers.

5.2 UORdb, a simple but versatile bioinformatics tool

The implementation of UORdb as a data-mining application and its use by the R.A. Scott lab demonstrate its utility as a facile bioinformatics tool. The regular expression-based search method is quite versatile for researchers engaged in footprinting and SELEX experiments, enabling them to “browse” a genome on-the-fly with different motif input guesses based on real-time experimental data. It also complements external motif finders such as MEME [79], CUBIC [77, 78], BioProspector[80]. Although these external motif finders are not practical for genome-level “fishing” expeditions due to excessive background noise, the UORdb in-house search method easily provides a first filter for such global searches. When knowledge of the motif is

limited to sequence lengths, external motif finders can supply an initial motif guess which can then be expanded genome-wide by the UORdb. UORdb is easily applied to any sequenced prokaryotic genome; the bacterial genome *Shewanella oneidensis* has recently been included in the UORdb web server. The database itself is dynamically generated from simply prepared input files and does not require the use of an additional relational database server, making maintenance a minimal consideration.

One of the great strengths of UORdb has been its use of operon predictions in the UOR selection criteria in highlighting UORs likely to contain prospective binding sites. Unfortunately, the sources and reliability of these predictions has varied a great deal since the initial construction of UORdb and reliance on less-than-confident predictions has been an Achilles heel for the database. Recent examples from the MicrobesOnline [83] and the DOOR database [85] have largely eliminated these concerns with reliable predictions and standardized formats. Indeed, the DOOR database has superseded much of UORdb by easily extracting UORs from a given genome for searching with external motif finders such as CUBIC and MEME.

UORdb does demonstrate that a simple searching application can be a great boon to researchers in the midst of investigating a given transcription factor and its associated binding sites. UORdb search results are useful in targeting genes for experiments such as EMSAs and footprinting which can be used to conclusively establish binding by transcription factors. Researchers can try a large array of motif choices deduced from experimental data or literature sources. Because it can operate at a genome-wide scale, UORdb may be useful for high-throughput strategies such as ChIP-chip [59] and DIP-chip [60]. UORdb, along with more complex applications such as the DOOR database and MicrobesOnline, can quickly and easily

provide aid in one of the foremost goals of our research: the delineation of regulatory transcriptional networks in an organism.

5.3 Comparative modeling for archaeal RNAP

Although the construction of a comparative model of the *P. furiosus* RNA polymerase complex requires the alignment and modeling of several large protein subunits, it is a relatively simple task when compared to arduous experimental structural elucidations such as X-ray crystallography or protein nuclear magnetic resonance spectroscopy. A recent example of its application was the comparative model built to supplement the 13Å resolution cryoelectron microscopic structure of *P. furiosus* RNAP [65] discussed in Chapter 4; the template for this model was also based on the eukaryal *S. cerevisiae* RNAPII. With the publication of archaeal RNAP structures such as the 3.44Å resolution dimeric structure of *S. solfataricus* RNAP (PDB code 2PMZ) [66], archaeal RNAP models can be created and assembled that may be more structurally accurate and correct with respect to the target. This, in turn, may aid in inferring function related to transcription including binding interactions with transcription factors. Although this structure was not available when the *P. furiosus* RNAP model in this dissertation was built, it would have been one of the likely structural templates for homology modeling, possibly as primary template, possibly with a high-resolution eukaryal RNAPII as a secondary template. The greater sequence similarity and the relatively low RMSDs between the *P. furiosus* and *S. solfataricus* RNAP subunits demonstrate their potential as primary templates for comparative modeling of archaeal RNAPs.

Although the model was lacking in good stereochemical quality, it was accurate enough to allow study of the highly conserved active site and dock domains of the RNAP machinery in

conjunction with the *S. solfataricus* RNAP structure. These two domains are among the most important present in the RNAP complex for both transcription initiation and catalyzing the transcription process. The *S. solfataricus* RNAP structure provides a means of comparison in validating the comparative model structure, as a test of its quality and the success of the modeling process as a whole. These comparisons reveal that the high sequence similarity of these regions does impart similar structure and most likely similar function. Homology modeling of different regions of the RNAP complex should provide some aid to researchers planning experiments studying transcription factor binding during transcription initiation.

Comparative models of RNAP have potential applications beyond qualitative comparisons of structures, including simulations using molecular dynamics (MD) software such as AMBER [134]. One of the original goals in the comparative modeling project was to investigate the open-clamp form of the *P. furiosus* RNAP structure using AMBER. Unfortunately, prohibitive internal constraints within AMBER7 limiting the simulation to less than 500000 atoms led to insurmountable difficulties in conducting MD simulations of the model complex solvated with TIP3P water models [135]; hence, this approach was abandoned. With fewer memory restrictions in place, this could be attempted again with a *P. furiosus* RNAP model to simulate the relative mobility of the clamp module and other structural changes between the open- and closed-clamp forms of the RNAP complex. In general, this might also aid in the structural refinement of new archaeal RNAP comparative models.

REFERENCES

1. Woese, C.R. and G.E. Fox, *Phylogenetic structure of prokaryotic domain - primary kingdoms*. Proceedings of the National Academy of Sciences of the United States of America, 1977. **74**(11): p. 5088-5090.
2. Kandler, O. and H. König, *Chemical composition of peptidoglycan-free cell-walls of methanogenic bacteria*. Archives of Microbiology, 1978. **118**(2): p. 141-152.
3. Ochsenreiter, T., et al., *Diversity and abundance of Crenarchaeota in terrestrial habitats studied by 16S RNA surveys and real time PCR*. Environmental Microbiology, 2003. **5**(9): p. 787-797.
4. Robertson, C.E., et al., *Phylogenetic diversity and ecology of environmental Archaea*. Current Opinion in Microbiology, 2005. **8**(6): p. 638-642.
5. Chaban, B., S.Y.M. Ng, and K.F. Jarrell, *Archaeal habitats - from the extreme to the ordinary*. Canadian Journal of Microbiology, 2006. **52**(2): p. 73-116.
6. Rothschild, L.J. and R.L. Mancinelli, *Life in extreme environments*. Nature, 2001. **409**(6823): p. 1092-1101.
7. Vieille, C. and G.J. Zeikus, *Hyperthermophilic enzymes: Sources, uses, and molecular mechanisms for thermostability*. Microbiology and Molecular Biology Reviews, 2001. **65**(1): p. 1-+.
8. Olsen, G.J. and C.R. Woese, *Archaeal genomics: An overview*. Cell, 1997. **89**(7): p. 991-994.
9. Soppa, J., *Transcription initiation in Archaea: facts, factors and future aspects*. Molecular Microbiology, 1999. **31**(5): p. 1295-1305.
10. Bell, S.D. and S.P. Jackson, *Transcription and translation in Archaea: A mosaic of eukaryal and bacterial features*. Trends in Microbiology, 1998. **6**(6): p. 222-228.
11. Stetter, K.O., *Hyperthermophiles in the history of life*. Philosophical Transactions of the Royal Society B-Biological Sciences, 2006. **361**(1474): p. 1837-1842.
12. Peak, M.J., F.T. Robb, and J.G. Peak, *Extreme resistance to thermally-induced dna backbone breaks in the hyperthermophilic archaeon Pyrococcus-furiosus*. Journal of Bacteriology, 1995. **177**(21): p. 6316-6318.
13. Fiala, G. and K.O. Stetter, *Pyrococcus-furiosus sp-nov represents a novel genus of marine heterotrophic archaebacteria growing optimally at 100-degrees c*. Archives of Microbiology, 1986. **145**(1): p. 56-61.
14. Fukui, T., et al., *Complete genome sequence of the hyperthermophilic archaeon Thermococcus kodakaraensis KOD1 and comparison with Pyrococcus genomes*. Genome Research, 2005. **15**(3): p. 352-363.
15. Gonzalez, J.M., et al., *Pyrococcus horikoshii sp. nov., a hyperthermophilic archaeon isolated from a hydrothermal vent at the Okinawa Trough*. Extremophiles, 1998. **2**(2): p. 123-130.
16. Cohen, G.N., et al., *An integrated analysis of the genome of the hyperthermophilic archaeon Pyrococcus abyssi*. Molecular Microbiology, 2003. **47**(6): p. 1495-1512.

17. Cramer, P., *Multisubunit RNA polymerases*. Current Opinion in Structural Biology, 2002. **12**(1): p. 89-97.
18. Borukhov, S. and E. Nudler, *RNA polymerase holoenzyme: structure, function and biological implications*. Current Opinion in Microbiology, 2003. **6**(2): p. 93-100.
19. Gruber, T.M. and C.A. Gross, *Multiple sigma subunits and the partitioning of bacterial transcription space*. Annual Review of Microbiology, 2003. **57**: p. 441-466.
20. Kornberg, R.D., *The molecular basis of eukaryotic transcription*. Proceedings of the National Academy of Sciences of the United States of America, 2007. **104**(32): p. 12955-12961.
21. Hausner, W., et al., *Two transcription factors related with the eucaryal transcription factors TATA-binding protein and transcription factor IIB direct promoter recognition by an archaeal RNA polymerase*. Journal of Biological Chemistry, 1996. **271**(47): p. 30144-30148.
22. Bell, S.D., et al., *The archaeal TFIIE alpha homologue facilitates transcription initiation by enhancing TATA-box recognition*. Embo Reports, 2001. **2**(2): p. 133-138.
23. Bartlett, M.S., *Determinants of transcription initiation by archaeal RNA polymerase*. Current Opinion in Microbiology, 2005. **8**(6): p. 677-684.
24. Smale, S.T. and J.T. Kadonaga, *The RNA polymerase II core promoter*. Annual Review of Biochemistry, 2003. **72**: p. 449-479.
25. van de Werken, H.J.G., et al., *Identification of a glycolytic regulon in the archaea Pyrococcus and Thermococcus*. Fems Microbiology Letters, 2006. **260**(1): p. 69-76.
26. Qureshi, S.A. and S.P. Jackson, *Sequence-specific DNA binding by the S-shibatae TFIIB homolog, TFB, and its effect on promoter strength*. Molecular Cell, 1998. **1**(3): p. 389-400.
27. Bell, S.D., et al., *Orientation of the transcription preinitiation complex in Archaea*. Proceedings of the National Academy of Sciences of the United States of America, 1999. **96**(24): p. 13662-13667.
28. Kosa, P.F., et al., *The 2.1-angstrom crystal structure of an archaeal preinitiation complex: TATA-box-binding protein/transcription factor (II)B core/TATA-box*. Proceedings of the National Academy of Sciences of the United States of America, 1997. **94**(12): p. 6042-6047.
29. Goede, B., et al., *Protein-protein interactions in the archaeal transcriptional machinery - Binding studies of isolated RNA polymerase subunits and transcription factors*. Journal of Biological Chemistry, 2006. **281**(41): p. 30581-30592.
30. Werner, F. and R.O.J. Weinzierl, *Direct modulation of RNA polymerase core functions by basal transcription factors*. Molecular and Cellular Biology, 2005. **25**(18): p. 8344-8355.
31. Bell, S.D., *Archaeal transcriptional regulation - variation on a bacterial theme?* Trends in Microbiology, 2005. **13**(6): p. 262-265.
32. Aravind, L., et al., *The many faces of the helix-turn-helix domain: Transcription regulation and beyond*. Fems Microbiology Reviews, 2005. **29**(2): p. 231-262.
33. Aravind, L. and E.V. Koonin, *DNA-binding proteins and evolution of transcription regulation in the archaea*. Nucleic Acids Research, 1999. **27**(23): p. 4658-4670.
34. Geiduschek, E.P. and M. Ouhammouch, *Archaeal transcription and its regulators*. Molecular Microbiology, 2005. **56**(6): p. 1397-1407.
35. Ouhammouch, M., *Transcriptional regulation in Archaea*. Current Opinion in Genetics & Development, 2004. **14**(2): p. 133-138.

36. Bell, S.D., et al., *Transcriptional regulation of an archaeal operon in vivo and in vitro*. Molecular Cell, 1999. **4**(6): p. 971-982.
37. Brinkman, A.B., et al., *The Lrp family of transcriptional regulators*. Molecular Microbiology, 2003. **48**(2): p. 287-294.
38. Dahlke, I. and M. Thomm, *A Pyrococcus homolog of the leucine-responsive regulatory protein, LrpA, inhibits transcription by abrogating RNA polymerase recruitment*. Nucleic Acids Research, 2002. **30**(3): p. 701-10.
39. Brinkman, A.B., et al., *An Lrp-like transcriptional regulator from the archaeon Pyrococcus furiosus is negatively autoregulated*. Journal of Biological Chemistry, 2000. **275**(49): p. 38160-38169.
40. Bell, S.D. and S.P. Jackson, *Mechanism of autoregulation by an archaeal transcriptional repressor*. Journal of Biological Chemistry, 2000. **275**(41): p. 31624-31629.
41. Enoru-Eta, J., et al., *Purification and characterization of Sa-lrp, a DNA-binding protein from the extreme thermoacidophilic archaeon Sulfolobus acidocaldarius homologous to the bacterial global transcriptional regulator Lrp*. Journal of Bacteriology, 2000. **182**(13): p. 3661-3672.
42. Enoru-Eta, J., et al., *High resolution contact probing of the Lrp-like DNA-binding protein Ss-Lrp from the hyperthermoacidophilic crenarchaeote Sulfolobus solfataricus P2*. Molecular Microbiology, 2002. **45**(6): p. 1541-1555.
43. Peeters, E., et al., *Ss-LrpB, a novel Lrp-like regulator of Sulfolobus solfataricus P2, binds cooperatively to three conserved targets in its own control region*. Molecular Microbiology, 2004. **54**(2): p. 321-336.
44. Brinkman, A.B., et al., *The Sulfolobus solfataricus Lrp-like protein LysM regulates lysine biosynthesis in response to lysine availability*. Journal of Biological Chemistry, 2002. **277**(33): p. 29537-29549.
45. Lee, S.J., et al., *TrmB, a sugar-specific transcriptional regulator of the trehalose/maltose ABC transporter from the hyperthermophilic archaeon Thermococcus litoralis*. Journal of Biological Chemistry, 2003. **278**(2): p. 983-990.
46. Lee, S.J., et al., *TrmB, a sugar sensing regulator of ABC transporter genes in Pyrococcus furiosus exhibits dual promoter specificity and is controlled by different inducers*. Molecular Microbiology, 2005. **57**(6): p. 1797-1807.
47. Krug, M., et al., *Crystal structure of the sugar binding domain of the archaeal transcriptional regulator TrmB*. Journal of Biological Chemistry, 2006. **281**(16): p. 10976-10982.
48. Lee, S.J., et al., *Characterization of the TrmB-like protein, PF0124, a TGM-recognizing global transcriptional regulator of the hyperthermophilic archaeon Pyrococcus furiosus*. Molecular Microbiology, 2007. **65**(2): p. 305-318.
49. Kanai, T., et al., *A global transcriptional regulator in Thermococcus kodakaraensis controls the expression levels of both glycolytic and gluconeogenic enzyme-encoding genes*. Journal of Biological Chemistry, 2007. **282**(46).
50. Roy, S., S. Garges, and S. Adhya, *Activation and repression of transcription by differential contact: Two sides of a coin*. Journal of Biological Chemistry, 1998. **273**(23): p. 14059-14062.
51. Vierke, G., et al., *A novel archaeal transcriptional regulator of heat shock response*. Journal of Biological Chemistry, 2003. **278**(1): p. 18-26.

52. Liu, W., et al., *Crystal structure of the archaeal heat shock regulator from Pyrococcus furiosus: A molecular chimera representing eukaryal and bacterial features*. Journal of Molecular Biology, 2007. **369**(2): p. 474-488.
53. Ouhammouch, M., et al., *Activation of archaeal transcription by recruitment of the TATA-binding protein*. Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**(9): p. 5097-5102.
54. Ouhammouch, M., et al., *A fully recombinant system for activator-dependent archaeal transcription*. Journal of Biological Chemistry, 2004. **279**(50): p. 51719-51721.
55. Ouhammouch, M. and E.P. Geiduschek, *A thermostable platform for transcriptional regulation: the DNA-binding properties of two Lrp homologs from the hyperthermophilic archaeon Methanococcus jannaschii*. Embo Journal, 2001. **20**(1-2): p. 146-156.
56. Ouhammouch, M. and E.P. Geiduschek, *An expanding family of archaeal transcriptional activators*. Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(43): p. 15423-15428.
57. Ouhammouch, M., et al., *Promoter architecture and response to a positive regulator of archaeal transcription*. Molecular Microbiology, 2005. **56**(3): p. 625-637.
58. Newberg, L.A., et al., *A phylogenetic Gibbs sampler that yields centroid solutions for cis-regulatory site prediction*. Bioinformatics, 2007. **23**(14): p. 1718-1727.
59. Ren, B., et al., *Genome-wide location and function of DNA binding proteins*. Science, 2000. **290**(5500): p. 2306-+.
60. Liu, X., et al., *DIP-chip: Rapid and accurate determination of DNA-binding specificity*. Genome Research, 2005. **15**(3): p. 421-427.
61. Tuerk, C. and L. Gold, *Systematic Evolution of Ligands by Exponential Enrichment - RNA Ligands to Bacteriophage-T4 DNA-Polymerase*. Science, 1990. **249**(4968): p. 505-510.
62. Ellington, A.D. and J.W. Szostak, *Invitro selection of RNA molecules that bind specific ligands*. Nature, 1990. **346**(6287): p. 818-822.
63. Todone, F., et al., *Structure of an archaeal homolog of the eukaryotic RNA polymerase II RPB4/RPB7 complex*. Molecular Cell, 2001. **8**(5): p. 1137-1143.
64. Langer, D., et al., *Transcription in Archaea - similarity to that in Eucarya*. Proceedings of the National Academy of Sciences of the United States of America, 1995. **92**(13): p. 5768-5772.
65. Kusser, A.G., et al., *Structure of an archaeal RNA polymerase*. Journal of Molecular Biology, 2008. **376**(2): p. 303-307.
66. Hirata, A., B.J. Klein, and K.S. Murakami, *The X-ray crystal structure of RNA polymerase from Archaea*. Nature, 2008. **451**(7180): p. 851-U12.
67. Lipscomb, G.L.P., *A novel redox-sensitive transcriptional regulator involved in Pyrococcus Furiosus sulfur response*. 2007, University of Georgia. p. 20-21.
68. Cramer, P., D.A. Bushnell, and R.D. Kornberg, *Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution*. Science, 2001. **292**(5523): p. 1863-1876.
69. DeLano, W.L., *The PyMOL Molecular Graphics System*. 2008, DeLano Scientific: Palo Alto, California, USA.
70. van Rossum, G., *Python Reference Manual*, F.L. Drake, Jr., Editor. 2008, Python Software Foundation: Hampton, NH, USA.
71. Trubetskoy, G., *Mod_python Manual*. 2007.
72. Benson, D.A., et al., *GenBank*. Nucleic Acids Research, 2009. **37**: p. D26-D31.

73. Pruitt, K.D., T. Tatusova, and D.R. Maglott, *NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins*. Nucleic Acids Research, 2007. **35**: p. D61-D65.
74. Peterson, J.D., et al., *The Comprehensive Microbial Resource*. Nucleic Acids Research, 2001. **29**(1): p. 123-125.
75. Hu, G.Q., et al., *Computational evaluation of TIS annotation for prokaryotic genomes*. BMC Bioinformatics, 2008. **9**.
76. Cornish-Bowden, A., *Nomenclature for incompletely specified bases in nucleic-acid sequences: recommendations 1984 - Nomenclature Committee of the International Union of Biochemistry (NC-IUB)*. Journal of Biological Chemistry, 1986. **261**(1): p. 13-17.
77. Olman, V., D. Xu, and Y. Xu, *CUBIC: identification of regulatory binding sites through data clustering*. J Bioinform Comput Biol, 2003. **1**(1): p. 21-40.
78. Olman, V., et al., *CUBIC: Search for Binding Sites*, in *Proceedings of the 2004 IEEE Computational Systems Bioinformatics Conference*. 2004, IEEE Computer Society.
79. Bailey, T.L. and C. Elkan, *Fitting a mixture model by expectation maximization to discover motifs in biopolymers*. Proc Int Conf Intell Syst Mol Biol, 1994. **2**: p. 28-36.
80. Liu, X., D.L. Brutlag, and J.S. Liu, *BioProspector: discovering conserved DNA motifs in upstream regulatory regions of co-expressed genes*. Pac Symp Biocomput, 2001: p. 127-38.
81. Ermolaeva, M.D., O. White, and S.L. Salzberg, *Prediction of operons in microbial genomes*. Nucleic Acids Research, 2001. **29**(5): p. 1216-1221.
82. Perte, M., et al., *OperonDB: a comprehensive database of predicted operons in microbial genomes*. Nucleic Acids Research, 2009. **37**: p. D479-D482.
83. Price, M.N., et al., *A novel method for accurate operon predictions in all sequenced prokaryotes*. Nucleic Acids Research, 2005. **33**(3): p. 880-892.
84. Tatusov, R.L., et al., *The COG database: new developments in phylogenetic classification of proteins from complete genomes*. Nucleic Acids Research, 2001. **29**(1): p. 22-28.
85. Mao, F.L., et al., *DOOR: a database for prokaryotic operons*. Nucleic Acids Research, 2009. **37**: p. D459-D463.
86. Dam, P., et al., *Operon prediction using both genome-specific and general genomic information*. Nucleic Acids Research, 2007. **35**(1): p. 288-298.
87. Pellegrini, M., et al., *Assigning protein functions by comparative genome analysis: Protein phylogenetic profiles*. Proceedings of the National Academy of Sciences of the United States of America, 1999. **96**(8): p. 4285-4288.
88. Lipscomb, G.L.P., *A novel redox-sensitive transcriptional regulator involved in Pyrococcus Furiosus sulfur response*. 2007, University of Georgia.
89. Schut, G.J., J.Z. Zhou, and M.W.W. Adams, *DNA microarray analysis of the hyperthermophilic archaeon Pyrococcus furiosus: Evidence for a new type of sulfur-reducing enzyme complex*. Journal of Bacteriology, 2001. **183**(24): p. 7027-7036.
90. Schut, G.J., S.L. Bridger, and M.W.W. Adams, *Insights into the metabolism of elemental sulfur by the hyperthermophilic archaeon Pyrococcus furiosus: Characterization of a coenzyme A-dependent NAD(P)H sulfur oxidoreductase*. Journal of Bacteriology, 2007. **189**(12): p. 4431-4441.

91. Adams, M.W.W., et al., *Key role for sulfur in peptide metabolism and in regulation of three hydrogenases in the hyperthermophilic archaeon Pyrococcus furiosus*. Journal of Bacteriology, 2001. **183**(2): p. 716-724.
92. Chen, Y., *The Fishing Expedition: Searching for Unidentified Archaeal Transcription Factors*. 2003, University of Georgia, Athens.
93. Wang, M., *Discovery of new transcription factors in Pyrococcus furiosus with DNA-affinity protein capture and mass spectrometry*. 2006, University of Georgia, Athens.
94. Perkins, D.N., et al., *Probability-based protein identification by searching sequence databases using mass spectrometry data*. Electrophoresis, 1999. **20**(18): p. 3551-3567.
95. Clauser, K.R., P. Baker, and A.L. Burlingame, *Role of accurate mass measurement (+/- 10 ppm) in protein identification strategies employing MS or MS MS and database searching*. Analytical Chemistry, 1999. **71**(14): p. 2871-2882.
96. Marchler-Bauer, A. and S.H. Bryant, *CD-Search: protein domain annotations on the fly*. Nucleic Acids Research, 2004. **32**: p. W327-W331.
97. Marchler-Bauer, A., et al., *CDD: a conserved domain database for protein classification*. Nucleic Acids Research, 2005. **33**: p. D192-D196.
98. Marchler-Bauer, A., et al., *CDD: a conserved domain database for interactive domain family analysis*. Nucleic Acids Research, 2007. **35**: p. D237-D240.
99. Tatusov, R.L., et al., *The COG database: an updated version includes eukaryotes*. BMC Bioinformatics, 2003. **4**.
100. Busenlehner, L.S., M.A. Pennella, and D.P. Giedroc, *The SmtB/ArsR family of metalloregulatory transcriptional repressors: structural insights into prokaryotic metal resistance*. FEMS Microbiology Reviews, 2003. **27**(2-3): p. 131-143.
101. Altschul, S.F., et al., *Gapped BLAST and PSI-BLAST: a new generation of protein database search programs*. Nucleic Acids Research, 1997. **25**(17): p. 3389-3402.
102. Galas, D.J. and A. Schmitz, *DNAase footprinting a simple method for the detection of protein-DNA binding specificity*. Nucleic Acids Research, 1978. **5**(9): p. 3157-3170.
103. Wilson, D.O., P. Johnson, and B.R. McCord, *Nonradiochemical DNase I footprinting by capillary electrophoresis*. Electrophoresis, 2001. **22**(10): p. 1979-1986.
104. Crooks, G.E., et al., *WebLogo: A sequence logo generator*. Genome Research, 2004. **14**(6): p. 1188-1190.
105. Ma, K.S. and M.W.W. Adams, *Hydrogenases I and II from Pyrococcus furiosus*, in *Hyperthermophilic Enzymes, Pt B*. 2001. p. 208-216.
106. Ma, K., et al., *Hydrogenase of the hyperthermophile Pyrococcus furiosus is an elemental sulfur reductase or sulfhydrogenase: evidence for a sulfur-reducing hydrogenase ancestor*. Proceedings of the National Academy of Sciences of the United States of America, 1993. **90**(11): p. 5341-5344.
107. Silva, P.J., et al., *Enzymes of hydrogen metabolism in Pyrococcus furiosus*. European Journal of Biochemistry, 2000. **267**(22): p. 6541-6551.
108. Clamp, M., et al., *The Jalview Java alignment editor*. Bioinformatics, 2004. **20**(3): p. 426-427.
109. Renfrow, M.B., et al., *Transcription factor B contacts promoter DNA near the transcription start site of the archaeal transcription initiation complex*. Journal of Biological Chemistry, 2004. **279**(4): p. 2825-2831.
110. Cramer, P., et al., *Architecture of RNA polymerase II and implications for the transcription mechanism*. Science, 2000. **288**(5466): p. 640-649.

111. Baker, D. and A. Sali, *Protein structure prediction and structural genomics*. Science, 2001. **294**(5540): p. 93-96.
112. Altschul, S.F., et al., *Basic local alignment search tool*. Journal of Molecular Biology, 1990. **215**(3): p. 403-410.
113. Lopez, R., et al., *WU-Blast2 server at the European Bioinformatics Institute*. Nucleic Acids Research, 2003. **31**(13): p. 3795-3798.
114. Labarga, A., et al., *Web services at the European Bioinformatics Institute*. Nucleic Acids Research, 2007. **35**: p. W6-W11.
115. Corpet, F., *Multiple sequence alignment with hierarchical-clustering*. Nucleic Acids Research, 1988. **16**(22): p. 10881-10890.
116. Rost, B., C. Sander, and R. Schneider, *PHD - an automatic mail server for protein secondary structure prediction*. Computer Applications in the Biosciences, 1994. **10**(1): p. 53-60.
117. Bryson, K., et al., *Protein structure prediction servers at University College London*. Nucleic Acids Research, 2005. **33**: p. W36-W38.
118. *Molecular Operating Environment (MOE)*. 2004, Chemical Computing Group Inc.: Montreal, Canada.
119. Needleman, S.B. and C.D. Wunsch, *A general method applicable to the search for similarities in the amino acid sequence of two proteins*. Journal of Molecular Biology, 1970. **48**(3): p. 443-&.
120. Valdar, W.S.J., *Scoring residue conservation*. Proteins-Structure Function and Bioinformatics, 2002. **48**(2): p. 227-241.
121. Bushnell, D.A., et al., *Structural basis of transcription: An RNA polymerase II-TFIIB cocrystal at 4.5 angstroms*. Science, 2004. **303**(5660): p. 983-988.
122. Levitt, M., *Accurate modeling of protein conformation by automatic segment matching*. Journal of Molecular Biology, 1992. **226**(2): p. 507-533.
123. Fichteler, T., U. Dengler, and D. Schomburg, *Prediction of protein 3-dimensional structures in insertion and deletion regions - a procedure for searching data-bases of representative protein-fragments using geometric scoring criteria*. Journal of Molecular Biology, 1995. **253**(1): p. 114-131.
124. Cornell, W.D., et al., *A second generation force field for the simulation of proteins, nucleic acids, and organic molecules (vol 117, pg 5179, 1995)*. Journal of the American Chemical Society, 1996. **118**(9): p. 2309-2309.
125. Shapiro, A., et al., *A method for multiple superposition of structures*. Acta Crystallographica Section A, 1992. **48**: p. 11-14.
126. Ramachandran, G.N., C. Ramakrishnan, and V. Sasisekharan, *Stereochemistry of polypeptide chain configurations*. Journal of Molecular Biology, 1963. **7**(1): p. 95-&.
127. Rhodes, G., *Crystallography Made Crystal Clear: A Guide for Users of Macromolecular Models*. 3rd ed. 2006, San Diego: Academic Press.
128. Davis, I.W., et al., *MolProbity: all-atom contacts and structure validation for proteins and nucleic acids*. Nucleic Acids Research, 2007. **35**: p. W375-W383.
129. Thompson, J.D., D.G. Higgins, and T.J. Gibson, *CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice*. Nucleic Acids Research, 1994. **22**(22): p. 4673-4680.

130. Chen, H.T. and S. Hahn, *Binding of TFIIB to RNA polymerase II: Mapping the binding site for the TFIIB zinc ribbon domain within the preinitiation complex*. Molecular Cell, 2003. **12**(2): p. 437-447.
131. Chen, H.T. and S. Hahn, *Mapping the location of TFIIB within the RNA polymerase II transcription preinitiation complex: A model for the structure of PIC*. Cell, 2004. **119**(2): p. 169-180.
132. Bangur, C.S., T.S. Pardee, and A.S. Ponticelli, *Mutational analysis of the D1/E1 core helices and the conserved N-terminal region of yeast transcription factor IIB (TFIIB): Identification of an N-terminal mutant that stabilizes TATA-binding protein-TFIIB-DNA complexes*. Molecular and Cellular Biology, 1997. **17**(12): p. 6784-6793.
133. Pardee, T.S., C.S. Bangur, and A.S. Ponticelli, *The N-terminal region of yeast TFIIB contains two adjacent functional domains involved in stable RNA polymerase II binding and transcription start site selection*. Journal of Biological Chemistry, 1998. **273**(28): p. 17859-17864.
134. Pearlman, D.A., et al., *AMBER, a package of computer-programs for applying molecular mechanics, normal-mode analysis, molecular-dynamics and free-energy calculations to simulate the structural and energetic properties of molecules*. Computer Physics Communications, 1995. **91**(1-3): p. 1-41.
135. Jorgensen, W.L., et al., *Comparison of simple potential functions for simulating liquid water*. Journal of Chemical Physics, 1983. **79**(2): p. 926-935.

APPENDICES

Appendix A Abbreviations

Table A.1 Table of abbreviations

Abbreviation Definition	
<i>Transcriptional machinery</i>	
BRE	TFB recognition element
ORF	Open reading frame
RNAP	RNA polymerase
TATA box	A/T-rich promoter element, TBP recognition site
TBP	TATA binding protein
TF	Transcription factor
TFB	Transcription factor B
TFBS	Transcription factor binding site
<i>Protein names</i>	
PDO	Protein disulfide oxidoreductase of <i>P. furiosus</i>
Phr	Heat shock regulator of <i>P. furiosus</i>
SurR	Sulfur response regulator of <i>P. furiosus</i>
<i>Gene names</i>	
<i>mbh1</i>	First ORF in <i>P. furiosus</i> membrane-bound hydrogenase operon (PF1423)
<i>pdo</i>	Gene encoding protein disulfide oxidoreductase (PF0094)
<i>surr</i>	Gene encoding sulfur response regulator of <i>P. furiosus</i> (PF0095)
<i>Experimental methods</i>	
EMSA	Electromobility shift assay
SELEX	Systematic evolution of ligands by exponential enrichment
ChIP-chip	Chromatin immunoprecipitation with microarray detection
DIP-chip	DNA immunoprecipitation with microarray detection

Abbreviation Definition

Miscellaneous terms

BS	Binding site
EM	Electron microscopic
HTH	Helix-turn-helix DNA-binding domain
IC	Information content
MD	Molecular dynamics
MST	Minimal spanning tree
PDB	Protein Data Bank
RMSD	Root-mean-square deviation
S ⁰	Elemental sulfur
TIS	Translation initiation site
TSS	Transcription start site
UOR	Upstream of ORF region

Units

Å	Angstrom
bp	base pair
Da	Dalton
mDa	megaDalton

Appendix B. UORdb input file formats

locus	start	stop	annotation
PF0001	927	94	hypothetical protein
PF0002	2243	939	hypothetical protein
PF0003	3340	2240	moaB-like protein
PF0004	4126	3353	hypothetical protein
PF0005	4926	4123	ABC transporter, ATP-binding protein
PF0006	5308	6300	hypothetical ABC transporter
PF0007	6464	7951	hypothetical protein
PF0008	8054	8548	hit family protein
PF0009	8538	9230	hypothetical protein
PF0010	9916	10905	D-cysteine desulhydrase
PF0011	11762	10956	putative ABC transporter, transmembrane component
PF0012	11810	12613	hypothetical protein
PF0013	13322	12573	transposase
PF0014	13736	13350	hypothetical protein
PF0015	14479	13751	hypothetical protein
PF0016	14581	15354	5'-methylthioadenosine phosphorylase II
PF0017	16236	17498	cell division control protein 6
PF0018	17498	19339	DNA polymerase II small subunit
PF0019	19330	23130	DNA polymerase II large subunit
PF0020	23182	23814	hypothetical protein
PF0021	23795	24502	DNA repair and recombination protein RadB
PF0022	25908	24499	putative RNase E
PF0023	26638	25862	hypothetical protein
PF0024	32164	30812	putative glycylate kinase
PF0025	32207	32998	hypothetical protein
PF0026	34340	32979	tRNA CCA-pyrophosphorylase
PF0027	35099	34545	putative integral membrane transport protein
PF0028	35838	35386	putative acetyl transferase
PF0029	37252	35819	hypothetical protein
PF0030	38330	37233	hypothetical protein
PF0031	39247	38333	threonine synthase
PF0032	39735	40019	hypothetical permease
PF0033	40429	40169	hypothetical protein
PF0034	40983	40405	hypothetical protein
PF0035	41109	42077	damorubicin resistance ATP-binding protein
PF0036	42064	42855	damorubicin resistance membrane protein
PF0037	44922	43228	hypothetical protein
PF0038	45509	44919	glyoxalase II family member
PF0039	46013	45552	xanthine-guanine phosphoribosyltransferase
PF0040	46613	46038	type II phosphatidic acid phosphatase
PF0041	48378	46636	putative oxidative cyclase
PF0042	49118	48375	metal-dependent hydrolase
PF0043	51636	49183	phosphoenolpyruvate synthase
PF0044	53145	51760	hypothetical protein
PF0045	54169	53135	hypothetical protein
PF0046	55124	54156	hypothetical protein
PF0047	55792	55124	hypothetical protein
PF0048	55848	56570	succinoglycan biosynthesis regulator
PF0049	57459	56554	cytosine-specific DNA-methyltransferase
PF0050	57747	57460	DNA-directed RNA polymerase subunit L

Figure B.1 Refseq ORF definitions as tab-delimited file (single-page example for *P. furiosus*)

locus	start	stop	annotation
PF0001	927	94	hypothetical protein
PF0002	2243	939	hypothetical protein
PF0003	3340	2240	moaB-like protein
PF0004	4126	3353	hypothetical protein
PF0005	4926	4123	ABC transporter, ATP-binding protein
PF0006	5308	6300	ABC transporter (ATP-binding)
PF0007	6464	7951	hypothetical protein
PF0008	8054	8548	hit family protein (hit)
PF0009	8538	9230	thiamine biosynthesis protein
PF0010	9841	10905	1-aminocyclopropane-1-carboxylate deaminase
PF0011	11750	10956	putative ABC transporter, transmembrane component
PF0012	11810	12613	hypothetical protein
PF0013	13322	12573	transposase
PF0014	13736	13350	hypothetical protein
PF0015	14479	13751	hypothetical protein
PF0016	14581	15354	5'-methylthioadenosine phosphorylase I
PF0017	16236	17498	origin recognition protein subunit 1
PF0018	17498	19339	DNA polymerase II subunit 1
PF0019	19339	23130	DNA polymerase II subunit 2
PF0020	23182	23814	hypothetical protein
PF0021	23795	24502	recombinase
PF0022	25908	24499	putative RNase E
PF0023	26638	25862	hypothetical protein
PF0024	32164	30812	putative glyceralate kinase
PF0025	32207	32998	hypothetical protein
PF0026	34340	32979	tRNA nucleotidyltransferase (cca)
PF0027	35099	34545	putative integral membrane transport protein
PF0028	35838	35386	putative acetyl transferase
PF0029	37252	35819	hypothetical protein
PF0030	38330	37233	hypothetical protein
PF0031	39247	38333	threonine synthase
PF0032	39735	40019	hypothetical protein
PF0033	40429	40169	hypothetical protein
PF0034	40983	40405	hypothetical protein
PF0035	41109	42077	daunorubicin resistance ATP-binding protein
PF0036	42064	42855	daunorubicin resistance membrane protein
PF0037	44922	43228	hypothetical protein
PF0038	45509	44919	glyoxalase II family member
PF0039	46013	45552	xanthine-guanine phosphoribosyltransferase
PF0040	46613	46038	type II phosphatidic acid phosphatase
PF0041	48378	46636	putative oxidative cyclase
PF0042	49118	48375	metal-dependent hydrolase
PF0043	51636	49183	phosphoenolpyruvate synthase (pyruvate, water dikinase)
PF0044	53145	51760	hypothetical protein
PF0045	54169	53135	hypothetical protein
PF0046	55124	54156	hypothetical protein
PF0047	55792	55124	hypothetical protein

Figure B.2 GenBank ORF definitions as tab-delimited file (single-page example for *P. furiosus*)

locus	start	stop	annotation
PF0001	927	94	hypothetical protein
PF0002	2243	939	hypothetical protein
PF0003	3340	2240	moeB-like protein
PF0004	4126	3353	hypothetical protein
PF0005	4926	4123	ABC transporter, ATP-binding protein
PF0006	5308	6300	ABC transporter (ATP-binding)
PF0007	6464	7951	hypothetical protein
PF0008	8054	8548	hit family protein (hit)
PF0009	9538	9230	thiamine biosynthesis protein
PF0010	9841	10805	L-aminocyclopropane-L-carboxylate deaminase
PF0011	11750	10956	putative ABC transporter, transmembrane component
PF0012	11810	12613	hypothetical protein
PF0013	13322	12573	transposase
PF0014	13736	13350	hypothetical protein
PF0015	14479	13751	hypothetical protein
PF0016	14581	15354	5-methylthioadenosine phosphorylase I
PF0017	16236	17498	origin recognition protein subunit 1
PF0018	17498	19339	DNA polymerase II subunit 1
PF0019	19339	23130	DNA polymerase II subunit 2
PF0020	23182	23814	hypothetical protein
PF0021	23795	24502	recombinase
PF0022	25908	24499	putative RNase E
PF0023	26638	25862	hypothetical protein
PF0024	32164	30812	putative glycerate kinase
PF0025	32207	32998	hypothetical protein
PF0026	34340	32979	tRNA nucleotidyltransferase (cca)
PF0027	35099	34545	putative integral membrane transport protein
PF0028	35838	35386	putative acetyl transferase
PF0029	37252	35819	hypothetical protein
PF0030	38330	37233	hypothetical protein
PF0031	39247	38333	threonine synthase
PF0032	39735	40019	hypothetical protein
PF0033	40429	40169	hypothetical protein
PF0034	40983	40405	hypothetical protein
PF0035	41109	42077	damorubicin resistance ATP-binding protein
PF0036	42064	42855	damorubicin resistance membrane protein
PF0037	44922	43228	hypothetical protein
PF0038	45509	44919	glyoxalase II family member
PF0039	46013	45552	xanthine-guanine phosphoribosyltransferase
PF0040	46613	46038	type II phosphatidic acid phosphatase
PF0041	48378	46636	putative oxidative cyclase
PF0042	49118	48375	metal-dependent hydrolase
PF0043	51636	49183	phosphoenolpyruvate synthase (pyruvate, water dikinase)
PF0044	53145	51760	hypothetical protein
PF0045	54169	53135	hypothetical protein
PF0046	55124	54156	hypothetical protein
PF0047	55792	55124	hypothetical protein
PF0048	55848	56570	succinoglycan biosynthesis regulator
PF0049	57459	56554	site-specific DNA-methyltransferase (cytosine-specific)

Figure B.3 TIGR/CMR ORF definitions as tab-delimited file (single-page example for *P. furiosus*)

organism	database	locus	start	stop	annotation
pf	refseq	PF0001	927	94	hypothetical protein
pf	refseq	PF0002	2243	939	hypothetical protein
pf	refseq	PF0003	3340	2240	mosB-like protein
pf	refseq	PF0004	4126	3353	hypothetical protein
pf	refseq	PF0005	4926	4123	ABC transporter, ATP-binding protein
pf	refseq	PF0006	5308	6300	hypothetical ABC transporter
pf	refseq	PF0007	6464	7951	hypothetical protein
pf	refseq	PF0008	8054	8548	hit family protein
pf	refseq	PF0009	8538	9230	hypothetical protein
pf	refseq	PF0010	9916	10905	D-cysteine desulfhydrase
pf	refseq	PF0011	11762	10956	putative ABC transporter, transmembrane component
pf	refseq	PF0012	11810	12613	hypothetical protein
pf	refseq	PF0013	13322	12573	transposase
pf	refseq	PF0014	13736	13350	hypothetical protein
pf	refseq	PF0015	14479	13751	hypothetical protein
pf	refseq	PF0016	14581	15354	5'-methylthioadenosine phosphorylase II
pf	refseq	PF0017	16236	17498	cell division control protein 6
pf	refseq	PF0018	17498	19339	DNA polymerase II small subunit
pf	refseq	PF0019	19330	23130	DNA polymerase II large subunit
pf	refseq	PF0020	23182	23814	hypothetical protein
pf	refseq	PF0021	23795	24502	DNA repair and recombination protein RadB
pf	refseq	PF0022	25908	24499	putative RNase E
pf	refseq	PF0023	26638	25862	hypothetical protein
pf	refseq	PF0024	32164	30812	putative glycerate kinase
pf	refseq	PF0025	32207	32998	hypothetical protein
pf	refseq	PF0026	34340	32979	tRNA CCA-pyrophosphorylase
pf	refseq	PF0027	35099	34545	putative integral membrane transport protein
pf	refseq	PF0028	35838	35386	putative acetyl transferase
pf	refseq	PF0029	37252	35819	hypothetical protein
pf	refseq	PF0030	38330	37233	hypothetical protein
pf	refseq	PF0031	39247	38333	threonine synthase
pf	refseq	PF0032	39735	40019	hypothetical permease
pf	refseq	PF0033	40429	40169	hypothetical protein
pf	refseq	PF0034	40983	40405	hypothetical protein
pf	refseq	PF0035	41109	42077	damorubicin resistance ATP-binding protein
pf	refseq	PF0036	42064	42855	damorubicin resistance membrane protein
pf	refseq	PF0037	44922	43228	hypothetical protein
pf	refseq	PF0038	45509	44919	glyoxalase II family member
pf	refseq	PF0039	46013	45552	xanthine-guanine phosphoribosyltransferase

Figure B.4 Combined database ORF definitions as tab-delimited file (single page example for *P. furiosus*)

```

>gi|18976372|ref|NC_003413.1| Pyrococcus furiosus DSM 3638, complete genome
AATTTCTTGT TTTGAGTCATAGTGTAAAAGGTATAAAATCAAAAATATGTTTGAGGTATTGTTCATGTCCAT
TTAAGCATCATTTTGTGATATCTTCACACTAATT TACCCATTCTCTCTAAATACAAAAC TAGTTCATTAG
TGTATCTACATGGATTAAATTTGCAATATAATCGTCGTA AAAAACCACCAATAGAACATAC TCC TAA GTT
AAGCATAGATCCGATGAGATAAAAAATTTTGCATAATATGGCCAGCTTCAAGGAGTAAATATCTATATCCT
CTGTCTCCATACTTCAACCTTATTTTAGAGTAATCCCCTGTAATTATTAATATTA AAC CAGCAGTTC TAA
TCCATTCTTGATTACCAAGAACCTTAACCAACGTATCTAAAAATTCATT TTTTGTGAAATTTGATACTC
TTCAATTTTAATTAATTCAATGTGTAGAGGATCATAATAATAAGATCCAAAACGTAAGTCATCTACATTG
AGCGATATCAAGTATACATGACAAGAAATTTAGCCCTCCTGCAGTTGGAAAGGATCTATGATAAATGGAGT
GATTTTCATCTTGATATTA AAC AACGCCACTTGACAGTTTAAGAAGTACAGAAAGTTCATTCAACTT
AATTTGGTTCATATGAAAATGTTCTGATAGATCTGCGAGTTACAATAGCCCGTACTAAATCATAAGACAAG
CTGGTTACATCAATTC TAGGTAGTTTATTTCTGTTCTATCTACATACTGTTTAGATGAAAATAAAAAAT
CATCAATGTTTACTGTCTGTAAAACTTTGTAT TCTCTCACTTGT TTC TTTGCCATTAGAGGGGTTTAT

```

Figure B.5 Extract of original genome sequence file prior to pre-processing for final memory-mapped UORdb input file, available from NCBI FTP (sample text from NC_003413.fna for *P. furiosus*)

OperonID	GI	Synonym	Start	End	Strand	Length	COG_number	
56599	18976373	PF0001	94	927	-	277	COG0778C	
	hypothetical protein							
56599	18976374	PF0002	939	2243	-	434	COG1944S	
	hypothetical protein							
56599	18976375	PF0003	2240	3340	-	366	COG0476H	moeB-
like protein								
56599	18976376	PF0004	3353	4126	-	257	-	
	hypothetical protein							
56599	18976377	PF0005	4123	4926	-	267	COG1121P	ABC
transporter, ATP-binding protein								
56600	18976380	PF0008	8054	8548	+	164	COG0537FGR	hit
family protein								
56600	18976381	PF0009	8538	9230	+	230	COG0476H	
	hypothetical protein							

Figure B.6 Extract of the tab-delimited operon predictions file, available from the DOOR operon predictions database (sample entries for *P. furiosus*)

Appendix C. Sequence Alignments

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	- - - - - , - - -	10 - - - - - , -	20 - - - - - , , , ,
Q8U0M5	P. furiosus	A	RMSD	M VSLST-I-KS	LIEKKGAN-L	PENVKS----
Q8U0M5	P. furiosus	A	PSIPRED	M VSLST-I-KS	LIEKKGAN-L	PENVKS----
Q8U0M5	P. furiosus	A	PHD	M VSLST-I-KS	LIEKKGAN-L	PENVKS----
P04050	S. cerevisiae	1	RMSD	FSDLTIKDIV LGVKDLQ-EN	LLVLRGKNEI	IQNAQRDAVT
P04050	S. cerevisiae	1	PSIPRED	FSDLTIKDIV LGVKDLQ-EN	LLVLRGKNEI	IQNAQRDAVT
P04050	S. cerevisiae	1	PHD	FSDLTIKDIV LGVKDLQ-EN	LLVLRGKNEI	IQNAQRDAVT
P04050	S. cerevisiae	1	MOE	TIKDIV LGVKD-LQEN	LLVLRGKNEI	IQNAQR----
P04050	S. cerevisiae	1	BLAST			
P04050	S. cerevisiae	1	BLAST2			
P04050	S. cerevisiae	1	MPsrch			
P04050	S. cerevisiae	1	Multa	----- -RGKN-E-II	QNAQRDAV-T	LFCCLL----
O35559	C. griseus		Multa	----- -NGDD-P-LS	RQAQENAT-L	LFNIHL----
P24928	H. sapiens	1	Multa	----- -NGDD-P-LS	RQAQENAT-L	LFNIHL----
P08775	M. musculus	1	Multa	----- -NGDD-P-LS	RQAQENAT-L	LFNIHL----
Q9AXN0	G. vacuolata	1	Multa	DGDQVAEVD	NSVSR-Q-LA	VEAQENAT-L
Q9XZS1	D. gaunche		Multa	----- -TGAD-R-IS	KQANENAT-L	LFQCCLI----
Q9XZS2	D. madeirensis		Multa	----- -TGAD-R-IS	KQANENAT-L	LFQCCLI----
Q9XZU9	D. subobscura		Multa	----- -TGAD-R-IS	KQANENAT-L	LFQCCLI----
P04052	D. melanogaster		Multa	----- -TGND-R-IS	KQANENAT-L	LFQCCLI----
Q9FRB0	O. sativa		Multa	----- -PGDD-D-IS	IEAQKNAT-L	FFNILL----
P18616	A. thaliana	1	Multa	----- -PGDD-A-LS	VEAQKNAT-L	FFNILL----
P36594	S. pombe	1	Multa	----- -RGSD-R-IT	RDVQNNAT-L	LFQIILL----
O96446	V. necatrix	1	Multa	----- TLKNL-I-EN	LP---IKN-L	LIEILI----
P14248	P. falciparum		Multa	FLEKLVIIKQ	INSND-T-LS	VEAQNNAT-I
P16356	C. elegans	1	Multa	----- -SGND-E-IS	KQAQYNAT-L	LMNILL----
Q25312	L. major		Multa	-FPSYHKDYN	GRFVN-V-LS	QQRVERAL-T
Q9NJR9	L. donovani		Multa	-FPSYHKDYS	GRFAN-V-LS	QQRVERAL-T
Q9BJ59	L. seymouri		Multa	-FPSYHKDYS	GRFLN-V-LS	QQRVKRAL-T
P17545	T. brucei brucei	1	Multa	-FPSYHRGGD	GYIRN-T-LS	RERIESAL-T
Q95W43	T. cruzi		Multa	-FPCYHKGVD	GRIHN-L-LS	RQRIENAL-T
O96452	N. locustae	1	Multa	----- VVDNA-L-QS	TSPNISED-S	FVHAFI----
Q9AVW8	G. theta	1	Multa	----INFKIS	FNQKK-S-DN	KQLINRDN-F
Q9HGZ5	A. oryzae	A	Multa	----- -RGDD-P-IS	QEAQENAT-L	LFKAQL----
Q27095	T. vaginalis		Multa	----- -KGND-N-IS	RTEQENAT-L	LLRIML----

*Method:
 BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
 PSIPRED and PHD = Structure Prediction
 Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.1 Sequence alignment for *P. furiosus* RNAP subunit A (sheet 1)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	-----30-----	-----40-----
Q8U0M5	P. furiosus	A	RMSD	----E-----	---LYEKL-- K--KYNEKYK LTKAEIEAII
Q8U0M5	P. furiosus	A	PSIPRED	----E-----	---LYEKL-- K--KYNEKYK LTKAEIEAII
Q8U0M5	P. furiosus	A	PHD	----E-----	---LYEKL-- K--KYNEKYK LTKAEIEAII
P04050	S. cerevisiae	1	RMSD	LFCCCL-----	---LRSRL-- ATRRVLQEYR LTKQAFDWVL
P04050	S. cerevisiae	1	PSIPRED	LFCCCL-----	---LRSRL-- ATRRVLQEYR LTKQAFDWVL
P04050	S. cerevisiae	1	PHD	LFCCCL-----	---LRSRL-- ATRRVLQEYR LTKQAFDWVL
P04050	S. cerevisiae	1	MOE	----DAVTLF	CCLLRSLAT R--RVLQEYR LTKQAFDWVL
P04050	S. cerevisiae	1	BLAST		VL
P04050	S. cerevisiae	1	BLAST2		QEYR LTKQAFDWVL
P04050	S. cerevisiae	1	MPsrch		
P04050	S. cerevisiae	1	Multa	----R-----	---SRLAT-- R--RVLQEYR LTKQAFDWVL
O35559	C. griseus		Multa	----R-----	---STLCS-- R--RMAEEFR LSGEAFDWLL
P24928	H. sapiens	1	Multa	----R-----	---STLCS-- R--RMAEEFR LSGEAFDWLL
P08775	M. musculus	1	Multa	----R-----	---STLCS-- R--RMAEEFR LSGEAFDWLL
Q9AXN0	G. vacuolata	1	Multa	----R-----	---AFLAS-- K--VILKEYR LDKKAFMHLL
Q9XZS1	D. gauchae		Multa	----R-----	---STLCT-- K--YVSEEFR LSTEAFEWLI
Q9XZS2	D. madeirensis		Multa	----R-----	---STLCT-- K--YVSEEFR LSTEAFEWLI
Q9XZU9	D. subobscura		Multa	----R-----	---STLCT-- K--YVSEEFR LSTEAFEWLI
P04052	D. melanogaster		Multa	----R-----	---STLCT-- K--YVSEEFR LSTEAFEWLV
Q9FRB0	O. sativa		Multa	----R-----	---STFAS-- K--RVLKEYR LTKAEFEWVI
P18616	A. thaliana	1	Multa	----R-----	---STLAS-- K--RVLEEYK LSREAFEWVI
P36594	S. pombe	1	Multa	----R-----	---SKFAV-- K--RVIMEYR LNKVAFEWIM
O96446	V. necatrix	1	Multa	----K-----	---YNLSI-- K--RILNEYK LSLEAYNWIL
P14248	P. falciparum		Multa	----R-----	---TYLNS-- K--LLTQTHK VSVKGLDWLL
P16356	C. elegans	1	Multa	----R-----	---STLCT-- K--NMCTKSK LNSEAFDWLL
Q25312	L. major		Multa	----R-----	---QILGS-- K--RVLKEYK LNDKAFEYLL
Q9NJR9	L. donovani		Multa	----R-----	---QILGS-- K--RVLKEYK LNDKAFEYLL
Q9BJ59	L. seymouri		Multa	----R-----	---QILGS-- K--RVLKEYK LNDKAFEYLL
P17545	T. brucei brucei	1	Multa	----R-----	---QLLAS-- K--RVLKEYK LNDRAFAYLL
Q95W43	T. cruzi		Multa	----R-----	---QLLAS-- K--RVLREYK LNDKAFEYLL
O96452	N. locustae	1	Multa	----R-----	---SCLRI-- K--KVIN-YR ISTLGVEWIA
Q9AVW8	G. theta	1	Multa	----R-----	---SSLSL-- K--IIYTKFN LNQHAIWNIV
Q9HGZ5	A. oryzae	A	Multa	----R-----	---SRLAF-- R--RLVTEYS MNKLAFAQHVI
Q27095	T. vaginalis		Multa	----Y-----	---SHLSA-- K--QCIFEYR LNEQAFKYIL

*Method:
 BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
 PSIPRED and PHD = Structure Prediction
 Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.2 Sequence alignment for *P. furiosus* RNAP subunit A (sheet 2)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	50-----	60-----	70-----	80-----
Q8U0M5	P. furiosus	A	RMSD	DDVVKEYERA	LVEPGEPVGT	VAAQSIGEPS	TQMTLNTFHY
Q8U0M5	P. furiosus	A	PSIPRED	DDVVKEYERA	LVEPGEPVGT	VAAQSIGEPS	TQMTLNTFHY
Q8U0M5	P. furiosus	A	PHD	DDVVKEYERA	LVEPGEPVGT	VAAQSIGEPS	TQMTLNTFHY
P04050	S. cerevisiae	1	RMSD	SNIEAQFLRS	VVHPGEMVGV	LAAQSIGEP	TQMTL-----
P04050	S. cerevisiae	1	PSIPRED	SNIEAQFLRS	VVHPGEMVGV	LAAQSIGEP	TQMTL-----
P04050	S. cerevisiae	1	PHD	SNIEAQFLRS	VVHPGEMVGV	LAAQSIGEP	TQMTL-----
P04050	S. cerevisiae	1	MOE	SNIEAQFLRS	VVHPGEMVGV	LAAQSIGEP	TQMTL-----
P04050	S. cerevisiae	1	BLAST	SNIEAQFLRS	VVHPGEMVGV	LAAQSIGEP	TQMTLNTFHF
P04050	S. cerevisiae	1	BLAST2	SNIEAQFLRS	VVHPGEMVGV	LAAQSIGEP	TQMTLNTFHF
P04050	S. cerevisiae	1	MPsrch	SNIEAQFLRS	VVHPGEMVGV	LAAQSIGEP	TQMTLNTFHF
P04050	S. cerevisiae	1	Multa	SNIEAQFLRS	VVHPGEMVGV	LAAQSIGEP	TQMTLNTFHF
O35559	C. griseus		Multa	GEIESKFNQA	IAHPGEMVGA	LAAQSIGEP	TQMTLNTFHY
P24928	H. sapiens	1	Multa	GEIESKFNQA	IAHPGEMVGA	LAAQSLGEP	TQMTLNTFHY
P08775	M. musculus	1	Multa	GEIESKFNQA	IAHPGEMVGA	LAAQSLGEP	TQMTLNTFHY
Q9AXN0	G. vacuolata	1	Multa	GEIENRFLHA	KCPPGEMIGA	VAPQSIGEP	TQMTLNTFHF
Q9XZS1	D. gauchae		Multa	GEIETRFQQA	QANPGEMVGA	LAAQSLGEP	TQMTLNTFHF
Q9XZS2	D. madeirensis		Multa	GEIETRFQQA	QANPGEMVGA	LAAQSLGEP	TQMTLNTFHF
Q9XZU9	D. subobscura		Multa	GEIETRFQQA	QANPGEMVGA	LAAQSLGEP	TQMTLNTFHF
P04052	D. melanogaster		Multa	GEIETRFQQA	QANPGEMVGA	LAAQSLGEP	TQMTLNTFHF
Q9FRB0	O. sativa		Multa	GEIESRFLQS	LVAPGEMIGC	VAAQSIGEP	TQMTLNTFHY
P18616	A. thaliana	1	Multa	GEIESRFLQS	LVAPGEMIGC	VAAQSIGEP	TQMTLNTFHY
P36594	S. pombe	1	Multa	GEVEARFQQA	VVSPGEMVGT	LAAQSIGEP	TQMTLNTFHY
O96446	V. necatrix	1	Multa	KEIKFKILKS	IISPENMVG	LAAQSVGEP	TQMTLNTFHL
P14248	P. falciparum		Multa	QEIEKIFYKS	LCHPGECVGA	LAAQSIGEP	TQMTLNTFHF
P16356	C. elegans	1	Multa	GEIESRFQQA	IAQPGEMVGA	LAAQSLGEP	TQMTLNTFHY
Q25312	L. major		Multa	KEIRTKYQQS	LITPGEIIGA	IAAQSCGEP	TQMTLNTFHN
Q9NJR9	L. donovani		Multa	KEIRTKYQES	LITPGEIIGA	IAAQSCGEP	TQMTLNTFHN
Q9BJ59	L. seymouri		Multa	KEIRTKYQQS	LITPGEIIGA	IAAQSCGEP	TQMTLNTFHN
P17545	T. brucei brucei	1	Multa	KEIRTKYHQS	LITPGENIGA	IAAQSCGEP	TQMTLNTFHN
Q95W43	T. cruzi		Multa	KEIRTKYLQS	LVSPGENIGA	IAAQSCGEP	TQMTLNTFHN
O96452	N. locustae	1	Multa	NEIKKKFVRG	IISTNEMVGT	LAAQSIGEP	TQMTLNTFHY
Q9AVW8	G. theta	1	Multa	NEICLSIRKS	LIQPGEMVGT	ISASIGQPA	TQMTLNTFHF
Q9HGZ5	A. oryzae	A	Multa	GAIESRFAKA	NAPAGEMVGV	LAAQSIGEP	TQMTLNTFHF
Q27095	T. vaginalis		Multa	GSIKDRFYRS	IVAPGEMVGT	IAGQSIGEPS	TQMTLNTFHF

*Method:

BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD

PSIPRED and PHD = Structure Prediction

Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.3 Sequence alignment for *P. furiosus* RNAP subunit A (sheet 3)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	90----	100-----	110-----
Q8U0M5	P. furiosus	A	RMSD	AGVAE-----	-----INVTL	GLPRIIEIVD ARKNPSTPMM
Q8U0M5	P. furiosus	A	PSIPRED	AGVAE-----	-----INVTL	GLPRIIEIVD ARKNPSTPMM
Q8U0M5	P. furiosus	A	PHD	AGVAE-----	-----INVTL	GLPRIIEIVD ARKNPSTPMM
P04050	S. cerevisiae	1	RMSD	-----NTFHF	AGVASKKVT	GVPRLKEILN VAKNMKTPSL
P04050	S. cerevisiae	1	PSIPRED	-----NTFHF	AGVASKKVT	GVPRLKEILN VAKNMKTPSL
P04050	S. cerevisiae	1	PHD	-----NTFHF	AGVASKKVT	GVPRLKEILN VAKNMKTPSL
P04050	S. cerevisiae	1	MOE	-----NTFHF	AGVASKKVT	GVPRLKEILN VAKNMKTPSL
P04050	S. cerevisiae	1	BLAST	AGVAS-----	-----KKVT	GVPRLKEILN VAKNMKTPSL
P04050	S. cerevisiae	1	BLAST2	AGVAS-----	-----KKVT	GVPRLKEILN VAKNMKTPSL
P04050	S. cerevisiae	1	MPsrch	AGVAS-----	-----KKVT	GVPRLKEILN VAKNMKTPSL
P04050	S. cerevisiae	1	Multa	AGVAS-----	-----KKVT	GVPRLKEILN VAKNMKTPSL
O35559	C. griseus		Multa	AGVSA-----	-----KNVT	GVPRLKEILN ISKKPKTPSL
P24928	H. sapiens	1	Multa	AGVSA-----	-----KNVT	GVPRLKEILN ISKKPKTPSL
P08775	M. musculus	1	Multa	AGVSA-----	-----KNVT	GVPRLKEILN ISKKPKTPSL
Q9AXN0	G. vacuolata	1	Multa	AGVSA-----	-----KNVT	GVPRLKEILN VAKHCKTPSL
Q9XZS1	D. gaunche		Multa	AGVSS-----	-----KNVT	GVPRLKEILN ISKKPKAPSL
Q9XZS2	D. madeirensis		Multa	AGVSS-----	-----KNVT	GVPRLKEILN ISKKPKAPSL
Q9XZU9	D. subobscura		Multa	AGVSS-----	-----KNVT	GVPRLKEILN ISKKPKAPSL
P04052	D. melanogaster		Multa	AGVSS-----	-----KNVT	GVPRLKEILN ISKKPKAPSL
Q9FRB0	O. sativa		Multa	AGVSA-----	-----KNVT	GVPRLKEILN VAKNIKTPSL
P18616	A. thaliana	1	Multa	AGVSA-----	-----KNVT	GVPRLKEILN VAKRIKTPSL
P36594	S. pombe	1	Multa	AGVSS-----	-----KNVT	GVPRLKEILN VAKNIKTPSL
O96446	V. necatrix	1	Multa	AGVSA-----	-----NITM	GVPRLKEILN VAKNIKTPCM
P14248	P. falciparum		Multa	AGVGS-----	-----KNVT	GVPRLKEILN IVKNVKTPTST
P16356	C. elegans	1	Multa	AGVSA-----	-----KNVT	GVPRLKEILN VSKTLKTPTSL
Q25312	L. major		Multa	AGISS-----	-----KNVT	GVPRLELLN VSKNQNASV
Q9NJR9	L. donovani		Multa	AGISS-----	-----KNVT	GVPRLELLN VSKNQNASV
Q9BJ59	L. seymouri		Multa	AGISS-----	-----KNVT	GVPRLELLN VSKNQNASV
P17545	T. brucei brucei	1	Multa	AGISS-----	-----KNVT	GVPRLELLN VSRNQKHASM
Q95W43	T. cruzi		Multa	AGISS-----	-----KNVT	GVPRLELLN VSRHQKHASM
O96452	N. locustae	1	Multa	AGVAQ-----	-----AITQ	GVPRLKEILN VAKNIKTPTSH
Q9AVW8	G. theta	1	Multa	AGVSA-----	-----KNVT	GVPRLKEIID IQKNIKSPSL
Q9HGZ5	A. oryzae	A	Multa	AGVSS-----	-----KNVT	GVPRLKEILN VATNIKTPTSM
Q27095	T. vaginalis		Multa	AGISA-----	-----HDVT	GVPRINELMN LAKHIRTPTSV

*Method:
 BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
 PSIPRED and PHD = Structure Prediction
 Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.4 Sequence alignment for *P. furiosus* RNAP subunit A (sheet 4)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	120-----	130-----	140-----150	-.....
Q8U0M5	P. furiosus	A	RMSD	TVYLD ~~EE~~ HRY	DRAKAE ~~EE~~ VAR	RIE ~~GT~~ TLENL	A-----
Q8U0M5	P. furiosus	A	PSIPRED	TVYLD ~~EE~~ HRY	DRAKAE ~~EE~~ VAR	RIE ~~GT~~ TLENL	A-----
Q8U0M5	P. furiosus	A	PHD	TVYLD ~~EE~~ HRY	DRAKAE ~~EE~~ VAR	RIE ~~GT~~ TLENL	A-----
P04050	S. cerevisiae	1	RMSD	TVYLEP ~~GH~~ AA	DQEQA ~~KL~~ IRS	AI ~~EH~~ TT ~~LK~~ SV	TIASEIYYDP
P04050	S. cerevisiae	1	PSIPRED	TVYLEP ~~GH~~ AA	DQEQA ~~KL~~ IRS	AI ~~EH~~ TT ~~LK~~ SV	TIASEIYYDP
P04050	S. cerevisiae	1	PHD	TVYLEP ~~GH~~ AA	DQEQA ~~KL~~ IRS	AI ~~EH~~ TT ~~LK~~ SV	TIASEIYYDP
P04050	S. cerevisiae	1	MOE	TVYLEP ~~GH~~ AA	DQEQA ~~KL~~ IRS	AI ~~EH~~ TT ~~LK~~ SV	TIASEIYYDP
P04050	S. cerevisiae	1	BLAST	TVYLEP ~~GH~~ AA	DQEQA ~~KL~~ IRS	AI ~~EH~~ TT ~~LK~~ SV	TIASEIYYDP
P04050	S. cerevisiae	1	BLAST2	TVYLEP ~~GH~~ AA	DQEQA ~~KL~~ IRS	AI ~~EH~~ TT ~~LK~~ SV	TIASEIYYDP
P04050	S. cerevisiae	1	MPsrch	TVYLEP ~~GH~~ AA	DQEQA ~~KL~~ IRS	AI ~~EH~~ TT ~~LK~~ SV	TIASEIYYDP
P04050	S. cerevisiae	1	Multa	TVYLEP ~~GH~~ AA	DQEQA ~~KL~~ IRS	AI ~~EH~~ TT ~~LK~~ SV	T-----
O35559	C. griseus		Multa	TVFLLGQ ~~SAR~~	DAERAKD ~~ILC~~	RLEHTTL ~~LRKV~~	T-----
P24928	H. sapiens	1	Multa	TVFLLGQ ~~SAR~~	DAERAKD ~~ILC~~	RLEHTTL ~~LRKV~~	T-----
P08775	M. musculus	1	Multa	TVFLLGQ ~~SAR~~	DAERAKD ~~ILC~~	RLEHTTL ~~LRKV~~	T-----
Q9AXN0	G. vacuolata	1	Multa	TVYLRGEAAR	DAERAKQVQA	ELQHTTLN ~~HV~~	T-----
Q9XZS1	D. gauchae		Multa	TVFLTGGAAAR	DAEKAKNVLC	RLEHTTL ~~LRKV~~	T-----
Q9XZS2	D. madeirensis		Multa	TVFLTGGAAAR	DAEKAKNVLC	RLEHTTL ~~LRKV~~	T-----
Q9XZU9	D. subobscura		Multa	TVFLTGGAAAR	DAEKAKNVLC	RLEHTTL ~~LRKV~~	T-----
P04052	D. melanogaster		Multa	TVFLTGGAAAR	DAEKAKNVLC	RLEHTTL ~~LRKV~~	T-----
Q9FRB0	O. sativa		Multa	SVHLKPEV ~~NK~~	KKELAKNVQC	ALEYTT ~~LRSV~~	T-----
P18616	A. thaliana	1	Multa	SVYLTPEA ~~SK~~	SKEGAKTVQC	ALEYTT ~~LRSV~~	T-----
P36594	S. pombe	1	Multa	TIYLM ~~PWIAA~~	NMDLAKNVQT	QIEHTTL ~~STV~~	T-----
O96446	V. necatrix	1	Multa	KIYLKDPF ~~NK~~	TLEMAKKIQS	ELEFSDI ~~KSL~~	C-----
P14248	P. falciparum		Multa	TIYLD ~~DMVSN~~	DQQA ~~KDILT~~	KLEYTT ~~LKQL~~	T-----
P16356	C. elegans	1	Multa	TVFLTGA ~~AAK~~	DPEKAKD ~~VLC~~	KLEHTTL ~~LKKV~~	T-----
Q25312	L. major		Multa	AVCLIREYQK	-RNKAQEAQQ	FIEYCT ~~LANI~~	T-----
Q9NJR9	L. donovani		Multa	AVCLIREYQK	-RNKAQEAQQ	FIEYCT ~~LANI~~	T-----
Q9BJ59	L. seymouri		Multa	AVCLTREYQK	-RQKAQEAQQ	YIEYCT ~~LANI~~	T-----
P17545	T. brucei brucei	1	Multa	TVSLFP ~~PPYDE~~	-KRNAQKAQH	LIEYCT ~~LESV~~	T-----
Q95W43	T. cruzi		Multa	NVSLYPP ~~WDK~~	-KRTAQQAQH	LIEYCT ~~LESV~~	T-----
O96452	N. locustae	1	Multa	KIYLK ~~P--GS~~	TIEQA ~~KKVAS~~	EIEYINI ~~EMV~~	C-----
Q9AVW8	G. theta	1	Multa	TI ~~FLHKKCSN~~	DPRK ~~VKKIQE~~	SIESV ~~RLNRI~~	S-----
Q9HGZ5	A. oryzae	A	Multa	TVYQEP ~~GRTH~~	DKEGA ~~KQLRS~~	VVEHT ~~SLRSV~~	T-----
Q27095	T. vaginalis		Multa	TVEFERD ~~VAK~~	DEEIV ~~KAVRA~~	QIESA ~~SFKKF~~	V-----

*Method:
 BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
 PSIPRED and PHD = Structure Prediction
 Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.5 Sequence alignment for *P. furiosus* RNAP subunit A (sheet 5)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	Sequence
Q8U0M5	P. furiosus	A	RMSD	--RSTTL-- --DLINFEFI
Q8U0M5	P. furiosus	A	PSIPRED	--RSTTL-- --DLINFEFI
Q8U0M5	P. furiosus	A	PHD	--RSTTL-- --DLINFEFI
P04050	S. cerevisiae	1	RMSD	DPRSTVI--- --PEDEEIIQ LHFSLLDEEA
P04050	S. cerevisiae	1	PSIPRED	DPRSTVI--- --PEDEEIIQ LHFSLLDEEA
P04050	S. cerevisiae	1	PHD	DPRSTVI--- --PEDEEIIQ LHFSLLDEEA
P04050	S. cerevisiae	1	MOE	DPRSTVIPED E----- --EIIQLHFS LLDEEAEQSF
P04050	S. cerevisiae	1	BLAST	DPRSTVIPED E----- --EIIQLHFS LLDEEAEQSF
P04050	S. cerevisiae	1	BLAST2	
P04050	S. cerevisiae	1	MPsrch	
P04050	S. cerevisiae	1	Multa	--IASEIYYD PDPSTVIPE DEEIIQLHFS LLDEEAEQSF
O35559	C. griseus		Multa	--ANTAIYYD PNPQSTVVAE DQEWVNVVYE MPDFD--VA-
P24928	H. sapiens	1	Multa	--ANTAIYYD PNPQSTVVAE DQEWVNVVYE MPDFD--VA-
P08775	M. musculus	1	Multa	--ANTAIYYD PNPQSTVVAE DQEWVNVVYE MPDFD--VA-
Q9AXN0	G. vacuolata	1	Multa	--QSTEIYYD PNPQDTVIEA DQELVRSYYE LPDDENSSA-
Q9XZS1	D. gauchae		Multa	--ANTAIYYD PDPQRTVISE DQEFVNVVYE MPDFD--PT-
Q9XZS2	D. madeirensis		Multa	--ANTAIYYD PDPQRTVISE DQEFVNVVYE MPDFD--PT-
Q9XZU9	D. subobscura		Multa	--ANTAIYYD PDPQRTVISE DQEFVNVVYE MPDFD--PT-
P04052	D. melanogaster		Multa	--ANTAIYYD PDPQRTVISE DQEFVNVVYE MPDFD--PT-
Q9FRB0	O. sativa		Multa	--HATEIYYD PDPLGTIEE DAEFVQSYFE MPDEDIDPD-
P18616	A. thaliana	1	Multa	--QATEVWYD PDPMTIIEE DFEFVRSYYE MPDEDVSPD-
P36594	S. pombe	1	Multa	--SATEIHYD PDPQDTVIEE DKDFVEAFFA IPDEEVEENL
O96446	V. necatrix	1	Multa	--EFSEIYYD FVIEDTSIKE DKDFVQYEFD FPDEHLDFS-
P14248	P. falciparum		Multa	--SHAQIIYD PNTTTTILEE DKSWVNEFFE FPDED--DTQ
P16356	C. elegans	1	Multa	--CNTAIYYD PDPKNTVIAE DEEWVSIFYE MPDHD--LS-
Q25312	L. major		Multa	--TTVQIIYD PDPRTTVVAE DEEMIRWEQA VMNADEEEDP
Q9NJR9	L. donovani		Multa	--TTVQIIYD PDPRTTVVAE DEEMIRWEQA VMNEDEEEDP
Q9BJ59	L. seymouri		Multa	--TTV----- PRPLNTVVVE DVNILEWEQA VMN-EDEGLD
P17545	T. brucei brucei	1	Multa	--RRIQFIYD PDPRTTVVEA DRDILELEWN VMDESDAELR
Q95W43	T. cruzi		Multa	--KRIQFIYD PDPRTTVVEA DRDILELEWN VMDESDAELR
O96452	N. locustae	1	Multa	--SEIKIIYD PKVDTTINE DRDFVEAYFE FPDEYVSLD-
Q9AVW8	G. theta	1	Multa	--NYFEIYYD PDFFFSISK NRILVKNYME IPDEDINFLQ
Q9HGZ5	A. oryzae	A	Multa	--EATEIYYD PDIQSTVIEN DRDMVESYFI IPEDVTDDS-
Q27095	T. vaginalis		Multa	--IKAEIYYD PNPEETIIEP DTKWINTRIR YGGEV-DREL

*Method:
 BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
 PSIPRED and PHD = Structure Prediction
 Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.6 Sequence alignment for *P. furiosus* RNAP subunit A (sheet 6)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	-----170	-----
Q8U0M5	P. furiosus	A	RMSD	-----VEIDPER	-----LERSGL-TME
Q8U0M5	P. furiosus	A	PSIPRED	-----VEIDPER	-----LERSGL-TME
Q8U0M5	P. furiosus	A	PHD	-----VEIDPER	-----LERSGL-TME
P04050	S. cerevisiae	1	RMSD	EQSFDQQSPW LLR-----	---LELDRAA MNDKDL-TMG
P04050	S. cerevisiae	1	PSIPRED	EQSFDQQSPW LLR-----	---LELDRAA MNDKDL-TMG
P04050	S. cerevisiae	1	PHD	EQSFDQQSPW LLR-----	---LELDRAA MNDKDL-TMG
P04050	S. cerevisiae	1	MOE	DQQSPWLLR- - - - -	---LELDRAA MNDKDL-TMG
P04050	S. cerevisiae	1	BLAST	DQQSPWLLR- - - - -	---LELDRAA MNDKDL-TMG
P04050	S. cerevisiae	1	BLAST2		
P04050	S. cerevisiae	1	MPsrch		
P04050	S. cerevisiae	1	Multa	DQQ----SPW LLR-----	---LELDRAA MNDKDL-TMG
O35559	C. griseus		Multa	-RI----SPW LLR-----	---VELDRKH MTDKRL-TME
P24928	H. sapiens	1	Multa	-RI----SPW LLR-----	---VELDRKH MTDKRL-TME
P08775	M. musculus	1	Multa	-RI----SPW LLR-----	---VELDRKH MTDKRL-TME
Q9AXN0	G. vacuolata	1	Multa	-NL----SPW LLR-----	---LNLSEKM MTDKRL-SMN
Q9XZS1	D. gauchae		Multa	-RI----SPW LLR-----	---IELDRKR MTDKRL-TME
Q9XZS2	D. madeirensis		Multa	-RI----SPW LLR-----	---IELDRKR MTDKRL-TME
Q9XZU9	D. subobscura		Multa	-RI----SPW LLR-----	---IELDRKR MTDKRL-TME
P04052	D. melanogaster		Multa	-RI----SPW LLR-----	---IELDRKR MTDKRL-TME
Q9FRB0	O. sativa		Multa	-KI----SPW LLR-----	---IELNREM MVDKRL-SMA
P18616	A. thaliana	1	Multa	-KI----SPW LLR-----	---IELNREM MVDKRL-SMA
P36594	S. pombe	1	Multa	YKQ----SPW LLR-----	---LELDRAK MLDKRL-SMS
O96446	V. necatrix	1	Multa	-KM----PKF IIR-----	---LKDRIK LVSKNL-KLE
P14248	P. falciparum		Multa	YSL----GEW VLR-----	---IQLTNIH VNEKKL-TMK
P16356	C. elegans	1	Multa	-RT----SPW LLR-----	---IELDRKR MVDKRL-TME
Q25312	L. major		Multa	AEQ--PPSPF IAR-----	---LILDNDL FNDKRL-NMK
Q9NJR9	L. donovani		Multa	AEQ--PPSPF IAR-----	---LILDNDL FNDKRL-NMK
Q9BJ59	L. seymouri		Multa	AAQ--P-SPF IAR-----	---LIINTEL FNNKQL-NMK
P17545	T. brucei brucei	1	Multa	IQEVVAGSPW VVR-----	---LELDVDM VTDKAL-DMK
Q95W43	T. cruzi		Multa	LQEVENGSPW IAR-----	---LELDADM FVDKAL-NMK
O96452	N. locustae	1	Multa	-TL----NPW LLR-----	---LVVARDR LISNNL-DFE
Q9AVW8	G. theta	1	Multa	A-----SSW ILK-----	---LKLLEE FVEKGL-FSN
Q9HGZ5	A. oryzae	A	Multa	SRQ----SKW LLR-----	---IILSRPK LLDKGL-TVQ
Q27095	T. vaginalis		Multa	QDL----APW VLR-----	---FEINREE MIASNIESIS

*Method:
 BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
 PSIPRED and PHD = Structure Prediction
 Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.7 Sequence alignment for *P. furiosus* RNAP subunit A (sheet 7)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	180-----	190-----	-----	-----	-----
Q8U0M5	P. furiosus	A	RMSD	KVVKKLESSF	KS-----A	EF-----EV--	---DG----	
Q8U0M5	P. furiosus	A	PSIPRED	KVVKKLESSF	KS-----A	EF-----EV--	---DG----	
Q8U0M5	P. furiosus	A	PHD	KVVKKLESSF	KS-----A	EF-----EV--	---DG----	
P04050	S. cerevisiae	1	RMSD	QVGERIKQTF	KN-----D	LF-----VI--	---WSEDNDE	
P04050	S. cerevisiae	1	PSIPRED	QVGERIKQTF	KN-----D	LF-----VI--	---WSEDNDE	
P04050	S. cerevisiae	1	PHD	QVGERIKQTF	KN-----D	LF-----VI--	---WSEDNDE	
P04050	S. cerevisiae	1	MOE	QVGERIKQTF	KN-----D	LF-----VIWS	EDNDE-----	
P04050	S. cerevisiae	1	BLAST	QVGERIKQTF	KN-----D	LFVIWSED--	---ND-----	
P04050	S. cerevisiae	1	BLAST2					
P04050	S. cerevisiae	1	MPsrch					
P04050	S. cerevisiae	1	Multa	QVGERIKQTF	KND-LFVIWS	ED----ND--	---EK-----	
O35559	C. griseus		Multa	QIAEKINAGF	GDD-LNCIFN	DD----NA--	---EK-----	
P24928	H. sapiens	1	Multa	QIAEKINAGF	GDD-LNCIFN	DD----NA--	---EK-----	
P08775	M. musculus	1	Multa	QIAEKINAGF	GDD-LNCIFN	DD----NA--	---EK-----	
Q9AXN0	G. vacuolata	1	Multa	HVKNKIHHD	GGD-VNVMAS	ED----ND--	---AN-----	
Q9XZS1	D. gauchae		Multa	QIAEKINVGF	GED-LNCIFN	DD----NA--	---DK-----	
Q9XZS2	D. madeirensis		Multa	QIAEKINVGF	GED-LNCIFN	DD----NA--	---DK-----	
Q9XZU9	D. subobscura		Multa	QIAEKINVGF	GED-LNCIFN	DD----NA--	---DK-----	
P04052	D. melanogaster		Multa	QIAEKINVGF	GED-LNCIFN	DD----NA--	---DK-----	
Q9FRB0	O. sativa		Multa	DIAEKINHEF	DDD-LSCIFN	DD----NA--	---DK-----	
P18616	A. thaliana	1	Multa	DIAEKINLEF	DDD-LTCIFN	DD----NA--	---QK-----	
P36594	S. pombe	1	Multa	DVAGKIAESF	ERD-LFTIWS	ED----NA--	---DK-----	
O96446	V. necatrix	1	Multa	NIVKSLHEAF	ENI-FHIIRS	DE----NS--	---QN-----	
P14248	P. falciparum		Multa	EIVYIIYSVF	SSDELDIITY	DD----NS--	---ED-----	
P16356	C. elegans	1	Multa	MIADRIHGGF	GND-VHTIYT	DD----NA--	---EK-----	
Q25312	L. major		Multa	DVKSARQ-V	DDTYMVQANM	EN----D---	---GQ-----	
Q9NJR9	L. donovani		Multa	DVKSARQ-V	DDTYMVQANM	EN----D---	---GQ-----	
Q9BJ59	L. seymouri		Multa	DVKSARQ-Q	DDSYLMQANL	KN----NI--	---RH-----	
P17545	T. brucei brucei	1	Multa	DVKQAILR-V	DESIIETGM	AN----NV--	---RQ-----	
Q95W43	T. cruzi		Multa	DVKQAILR-V	DDGYVIETGM	AN----DT--	---KE-----	
O96452	N. locustae	1	Multa	FIAQKIKKVF	HQD-IFLLHS	DA----NA--	---EE-----	
Q9AVW8	G. theta	1	Multa	YIIETIKKKI	KPIPLIIAS	DE----NS--	---GS-----	
Q9HGZ5	A. oryzae	A	Multa	DVATRIKQAY	PKD-IAVIFS	DN----NA--	---DE-----	
Q27095	T. vaginalis		Multa	DIVQIKKGM	GDK--FFVIG	DD----DA--	---RI-----	

*Method:
 BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
 PSIPRED and PHD = Structure Prediction
 Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.8 Sequence alignment for *P. furiosus* RNAP subunit A (sheet 8)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	200- - - - -	- - - - -	- - - - -	-210--	- - - - -
Q8U0M5	P. furiosus	A	RMSD	YTLI-----V	RPKK-----	-----ADKISD	-----L---	
Q8U0M5	P. furiosus	A	PSIPRED	YTLI-----V	RPKK-----	-----ADKISD	-----L---	
Q8U0M5	P. furiosus	A	PHD	YTLI-----V	RPKK-----	-----ADKISD	-----L---	
P04050	S. cerevisiae	1	RMSD	KLI-----R	CRVVRPKSLD	AETEAEEDHM	-----L---	
P04050	S. cerevisiae	1	PSIPRED	KLI-----R	CRVVRPKSLD	AETEAEEDHM	-----L---	
P04050	S. cerevisiae	1	PHD	KLI-----R	CRVVRPKSLD	AETEAEEDHM	-----L---	
P04050	S. cerevisiae	1	MOE	KLIIRCRV-V	-----RPKSLD	AETE-----	AEDDHML---	
P04050	S. cerevisiae	1	BLAST	EKLIIRCRVV	RPKS-----	-----LDAETE	AEDDHML---	
P04050	S. cerevisiae	1	BLAST2					
P04050	S. cerevisiae	1	MPsrch					
P04050	S. cerevisiae	1	Multa	LIIR-----C	RVV-----	-----RP---K	-----SLDA	
O35559	C. griseus		Multa	LVL-----I	RIM-----	-----NSDENK	-----MQEE	
P24928	H. sapiens	1	Multa	LVL-----I	RIM-----	-----NSDENK	-----MQEE	
P08775	M. musculus	1	Multa	LVL-----I	RIM-----	-----NSDENK	-----MQEE	
Q9AXN0	G. vacuolata	1	Multa	LVL-----I	RIAA-----	-----QKEPEK	-----MAEG	
Q9XZS1	D. gauchae		Multa	LVL-----I	RIM-----	-----NNEENK	-----FQDE	
Q9XZS2	D. madeirensis		Multa	LVL-----I	RIM-----	-----NNEENK	-----FQDE	
Q9XZU9	D. subobscura		Multa	LVL-----I	RIM-----	-----NNEENK	-----FQDE	
P04052	D. melanogaster		Multa	LVL-----I	RIM-----	-----NNEENK	-----FQDE	
Q9FRB0	O. sativa		Multa	LIIR-----V	RIT-----	-----NDEAQK	-----GEIQ	
P18616	A. thaliana	1	Multa	LIIR-----I	RIM-----	-----NDEGPK	-----GELQ	
P36594	S. pombe	1	Multa	LIIR-----C	RII-----	-----RDDRDK	-----AEDD	
O96446	V. necatrix	1	Multa	LIIR-----I	RCIS-----	-----SLNNNV	-----E---	
P14248	P. falciparum		Multa	LVL-----I	RVKY-----	-----LNGEYN	-----FMNY	
P16356	C. elegans	1	Multa	LVFR-----L	RIA-----	-----GEDKGE	-----AQEE	
Q25312	L. major		Multa	RIVR-----L	RPRK-----	-----CTGADS	-----XPA-	
Q9NJR9	L. donovani		Multa	RIVR-----L	RPRK-----	-----CTGADS	-----VPA-	
Q9BJ59	L. seymouri		Multa	RVVR-----L	RPRK-----	-----CAGADS	-----VPE-	
P17545	T. brucei brucei	1	Multa	RTIR-----M	RSRY-----	-----NEGADS	-----IPK-	
Q95W43	T. cruzi		Multa	RTIR-----M	RPKK-----	-----NDGVDS	-----IPR-	
O96452	N. locustae	1	Multa	LVIR-----I	RVLG-----	-----MSANE	-----E---	
Q9AVW8	G. theta	1	Multa	IEFL-----V	RFIW-----	-----PEFINN	-----YNSE	
Q9HGZ5	A. oryzae	A	Multa	QVIR-----I	RQI-----	-----QD--HK	-----EDED	
Q27095	T. vaginalis		Multa	IRMQ-----A	RLKD-----	-----EQAEFA	-----MDVK	

*Method:
 BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
 PSIPRED and PHD = Structure Prediction
 Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.9 Sequence alignment for *P. furiosus* RNAP subunit A (sheet 9)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	-----220--	--230	-----,,
Q8U0M5	P. furiosus	A	RMSD	-----RRFAEKIKK	H-RL-KGLSG	VGKTIVRK--
Q8U0M5	P. furiosus	A	PSIPRED	-----RRFAEKIKK	H-RL-KGLSG	VGKTIVRK--
Q8U0M5	P. furiosus	A	PHD	-----RRFAEKIKK	H-RL-KGLSG	VGKTIVRK--
P04050	S. cerevisiae	1	RMSD	-----KKIENTMLE	N-ITLRGVEN	IERVVMK--
P04050	S. cerevisiae	1	PSIPRED	-----KKIENTMLE	N-ITLRGVEN	IERVVMK--
P04050	S. cerevisiae	1	PHD	-----KKIENTMLE	N-ITLRGVEN	IERVVMK--
P04050	S. cerevisiae	1	MOE	-----KKIENTMLE	NITL-RGVEN	IERVVMKYD
P04050	S. cerevisiae	1	BLAST	-----KKIENTMLE	NITL-RGVEN	IERVVMKYD
P04050	S. cerevisiae	1	BLAST2		K Y-DR-KVPSP	TGEYV--K--
P04050	S. cerevisiae	1	MPsrch			
P04050	S. cerevisiae	1	Multa	-----HML	KKIENTMLEN	I-TL-RGVEN
O35559	C. griseus		Multa	-----VFL	RCIESNMLTD	M-TL-QGIEQ
P24928	H. sapiens	1	Multa	-----VFL	RCIESNMLTD	M-TL-QGIEQ
P08775	M. musculus	1	Multa	-----VFL	RCIESNMLTD	M-TL-QGIEQ
Q9AXN0	G. vacuolata	1	Multa	-----CLFL	RKVEGSLSE	M-NL-GGVPD
Q9XZS1	D. gauchae		Multa	-----MFL	RCIEANMLSD	M-TL-QGIEA
Q9XZS2	D. madeirensis		Multa	-----MFL	RCIEANMLSD	M-TL-QGIEA
Q9XZU9	D. subobscura		Multa	-----MFL	RCIEANMLSD	M-TL-QGIEA
P04052	D. melanogaster		Multa	-----MFL	RCIEANMLSD	M-TL-QGIEA
Q9FRB0	O. sativa		Multa	-----VFL	KKIESNMLTE	M-AL-RGIPG
P18616	A. thaliana	1	Multa	-----VFL	KKIESNMLTE	M-AL-RGIPD
P36594	S. pombe	1	Multa	-----VFL	KTIEGHMLES	I-SL-RGVPN
O96446	V. necatrix	1	Multa	-----YY	NLQYKNIL-N	L-KI-MGYNK
P14248	P. falciparum		Multa	YNNNTTNIFK	SKVKNNISSD	I-NT-KNEDS
P16356	C. elegans	1	Multa	-----VFL	RCIEANMLSD	L-TL-QGIPA
Q25312	L. major		Multa	-----L	TKAVAQLLQS	V-HL-RGIPG
Q9NJR9	L. donovani		Multa	-----L	TKAVAQLLQS	V-HH-RGIPG
Q9BJ59	L. seymouri		Multa	-----L	TKAVAQLLQS	V-HL-LGIPG
P17545	T. brucei brucei	1	Multa	-----L	KREIPALLAR	V-HL-RGIPG
Q95W43	T. cruzi		Multa	-----L	KHAPELLAR	V-HL-RGIPG
O96452	N. locustae	1	Multa	-----FY	RRLLNTIL-G	I-HL-KGITN
Q9AVW8	G. theta	1	Multa	NLFLEPNYFK	KIKHYFYNNY	L-KP-YGSQK
Q9HGZ5	A. oryzae	A	Multa	-----VTL	KKLEQHLLDT	L-TL-RGVNG
Q27095	T. vaginalis		Multa	-----FL	RETEQQMYKS	F-TL-KGIPG

*Method:
 BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
 PSIPRED and PHD = Structure Prediction
 Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.10 Sequence alignment for *P. furiosus* RNAP subunit A (sheet 10)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	-----220--	--230	-----,,
Q8U0M5	P. furiosus	A	RMSD	-----RRFAEKIKK	H-RL-KGLSG	VGKTIVRK--
Q8U0M5	P. furiosus	A	PSIPRED	-----RRFAEKIKK	H-RL-KGLSG	VGKTIVRK--
Q8U0M5	P. furiosus	A	PHD	-----RRFAEKIKK	H-RL-KGLSG	VGKTIVRK--
P04050	S. cerevisiae	1	RMSD	-----KKIENTMLE	N-ITLRGVEN	IERVVMK--
P04050	S. cerevisiae	1	PSIPRED	-----KKIENTMLE	N-ITLRGVEN	IERVVMK--
P04050	S. cerevisiae	1	PHD	-----KKIENTMLE	N-ITLRGVEN	IERVVMK--
P04050	S. cerevisiae	1	MOE	-----KKIENTMLE	NITL-RGVEN	IERVVMK--
P04050	S. cerevisiae	1	BLAST	-----KKIENTMLE	NITL-RGVEN	IERVVMK--
P04050	S. cerevisiae	1	BLAST2		K Y-DR-KVPSP	TGEYV--K--
P04050	S. cerevisiae	1	MPsrch			
P04050	S. cerevisiae	1	Multa	-----HML	KKIENTMLEN	I-TL-RGVEN
O35559	C. griseus		Multa	-----VFL	RCIESNMLTD	M-TL-QGIEQ
P24928	H. sapiens	1	Multa	-----VFL	RCIESNMLTD	M-TL-QGIEQ
P08775	M. musculus	1	Multa	-----VFL	RCIESNMLTD	M-TL-QGIEQ
Q9AXN0	G. vacuolata	1	Multa	-----CLFL	RKVEGSLSE	M-NL-GGVPD
Q9XZS1	D. gauchae		Multa	-----MFL	RCIEANMLSD	M-TL-QGIEA
Q9XZS2	D. madeirensis		Multa	-----MFL	RCIEANMLSD	M-TL-QGIEA
Q9XZU9	D. subobscura		Multa	-----MFL	RCIEANMLSD	M-TL-QGIEA
P04052	D. melanogaster		Multa	-----MFL	RCIEANMLSD	M-TL-QGIEA
Q9FRB0	O. sativa		Multa	-----VFL	KKIESNMLTE	M-AL-RGIPG
P18616	A. thaliana	1	Multa	-----VFL	KKIESNMLTE	M-AL-RGIPD
P36594	S. pombe	1	Multa	-----VFL	KTIEGHMLES	I-SL-RGVPN
O96446	V. necatrix	1	Multa	-----YY	NLQYKNIL-N	L-KI-MGYNK
P14248	P. falciparum		Multa	YNNNTTNIFK	SKVKNNISSD	I-NT-KNEDS
P16356	C. elegans	1	Multa	-----VFL	RCIEANMLSD	L-TL-QGIPA
Q25312	L. major		Multa	-----L	TKAVAQLLQS	V-HL-RGIPG
Q9NJR9	L. donovani		Multa	-----L	TKAVAQLLQS	V-HH-RGIPG
Q9BJ59	L. seymouri		Multa	-----L	TKAVAQLLQS	V-HL-LGIPG
P17545	T. brucei brucei	1	Multa	-----L	KREIPALLAR	V-HL-RGIPG
Q95W43	T. cruzi		Multa	-----L	KHAPELLAR	V-HL-RGIPG
O96452	N. locustae	1	Multa	-----FY	RRLINTIL-G	I-HL-KGITN
Q9AVW8	G. theta	1	Multa	NLFLEPNYFK	KIKHYFYNNY	L-KP-YGSQK
Q9HGZ5	A. oryzae	A	Multa	-----VTL	KKLEQHLLDT	L-TL-RGVNG
Q27095	T. vaginalis		Multa	-----FL	RETEQQMYKS	F-TL-KGIPG

*Method:
 BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
 PSIPRED and PHD = Structure Prediction
 Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.11 Sequence alignment for *P. furiosus* RNAP subunit A (sheet 11)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	240	240	240	240
Q8U0M5	P. furiosus	A	RMSD	-----	-EGD-----	-----	---E-----
Q8U0M5	P. furiosus	A	PSIPRED	-----	-EGD-----	-----	---E-----
Q8U0M5	P. furiosus	A	PHD	-----	-EGD-----	-----	---E-----
P04050	S. cerevisiae	1	RMSD	-----	-YDR-----	-----	---KVPSPTG
P04050	S. cerevisiae	1	PSIPRED	-----	-YDR-----	-----	---KVPSPTG
P04050	S. cerevisiae	1	PHD	-----	-YDR-----	-----	---KVPSPTG
P04050	S. cerevisiae	1	MOE	RKVPSP----	-TGEYVKEP-	-----	---E-----
P04050	S. cerevisiae	1	BLAST	RKVPSPPTGEY	VKEP-----	-----	---E-----
P04050	S. cerevisiae	1	BLAST2	-----	-E-P-----	-----	---E-----
P04050	S. cerevisiae	1	MPsrch	-----	-----	-----	E-----
P04050	S. cerevisiae	1	Multa	-----	-YDRKVPSPPT	GEYVKEP---	---E-----
O35559	C. griseus		Multa	-----	-PQTDNKKKI	IITEDGEFKA	L-QE-----
P24928	H. sapiens	1	Multa	-----	-PQTDNKKKI	IITEDGEFKA	L-QE-----
P08775	M. musculus	1	Multa	-----	-PQTDNKKKI	IITEDGEFKA	L-QE-----
Q9AXN0	G. vacuolata	1	Multa	-----	-AD-----RI	VPDPRTGGLS	VEKE-----
Q9XZS1	D. gauchae		Multa	-----	-PQTDSSKKRI	VITETGEFKA	I-GE-----
Q9XZS2	D. madeirensis		Multa	-----	-PQTDSSKKRI	VITETGEFKA	I-GE-----
Q9XZU9	D. subobscura		Multa	-----	-PQTDSSKKRI	VITETGEFKA	I-GE-----
P04052	D. melanogaster		Multa	-----	-PQTDSSKKRI	VITETGEFKA	I-GE-----
Q9FRB0	O. sativa		Multa	-----	-GNVVK----	-FEDNDGFKT	E-KG-----
P18616	A. thaliana	1	Multa	-----	-VRKSR----	-FDEEGGFKT	S-EE-----
P36594	S. pombe	1	Multa	-----	-HKIVRQIED	GTFERAD---	---E-----
O96446	V. necatrix	1	Multa	-----	-DKDKD----	-----	---E-----
P14248	P. falciparum		Multa	-----	-EQVKININSS	PVSNMHNHN	NNNN-----
P16356	C. elegans	1	Multa	-----	-PNTDDKKRI	IITPEGGFKS	V-AD-----
Q25312	L. major		Multa	-----	-GNTFR----	-VDPEAGGIK	NESS-----
Q9NJR9	L. donovani		Multa	-----	-GNTFR----	-VDPETGGIK	NESS-----
Q9BJ59	L. seymouri		Multa	-----	-GTTFN----	-VDPATGGIT	KESS-----
P17545	T. brucei brucei	1	Multa	-----	-TTEFT----	-VDQATGKMS	GNKI-----
Q95W43	T. cruzi		Multa	-----	-RTTFT----	-VDPATGKMT	GEKT-----
O96452	N. locustae	1	Multa	-----	--RERE-----	-----	---E-----
Q9AVW8	G. theta	1	Multa	-----	-IDIE-----	PEPNQSGKLG	NNFE-----
Q9HGZ5	A. oryzae	A	Multa	-----	-KSKVRVLED	GSLFTSKVDP	LCKE-----
Q27095	T. vaginalis		Multa	-----	-AAEKM----	FLNPETHGFD	KEKV-----

*Method:
 BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
 PSIPRED and PHD = Structure Prediction
 Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.12 Sequence alignment for *P. furiosus* RNAP subunit A (sheet 12)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	-----YVI	YTEGNSNFKQV	L-----KV	PGVDPTRTRT
Q8U0M5	P. furiosus	A	RMSD	-----YVI	YTEGNSNFKQV	L-----KV	PGVDPTRTRT
Q8U0M5	P. furiosus	A	PSIPRED	-----YVI	YTEGNSNFKQV	L-----KV	PGVDPTRTRT
Q8U0M5	P. furiosus	A	PHD	-----YVI	YTEGNSNFKQV	L-----KV	PGVDPTRTRT
P04050	S. cerevisiae	1	RMSD	EYVKEPEWVL	ETDGVNLSEV	M-----TV	PGIDPTRIYT
P04050	S. cerevisiae	1	PSIPRED	EYVKEPEWVL	ETDGVNLSEV	M-----TV	PGIDPTRIYT
P04050	S. cerevisiae	1	PHD	EYVKEPEWVL	ETDGVNLSEV	M-----TV	PGIDPTRIYT
P04050	S. cerevisiae	1	MOE	-----WVL	ETDGVNLSEV	M-----TV	PGIDPTRIYT
P04050	S. cerevisiae	1	BLAST	-----WVL	ETDGVNLSEV	M-----TV	PGIDPTRIYT
P04050	S. cerevisiae	1	BLAST2	-----WVL	ETDGVNLSEV	M-----TV	PGIDPTRIYT
P04050	S. cerevisiae	1	MPsrch	-----WVL	ETDGVNLSEV	M-----TV	PGIDPTRIYT
P04050	S. cerevisiae	1	Multa	-----WVL	ETDGVNLSEV	M-----TV	PGIDPTRIYT
O35559	C. griseus		Multa	-----WIL	ETDGVSLMRV	L-----SE	KDVPVVRTTS
P24928	H. sapiens	1	Multa	-----WIL	ETDGVSLMRV	L-----SE	KDVPVVRTTS
P08775	M. musculus	1	Multa	-----WIL	ETDGVSLMRV	L-----SE	KDVPVVRTTS
Q9AXN0	G. vacuolata	1	Multa	-----WVL	DTDGTNLLAV	M-----SH	KDVPVVRTTS
Q9XZS1	D. gauchae		Multa	-----WLL	ETDGTSMKVV	L-----SE	RDVPVVRTTS
Q9XZS2	D. madeirensis		Multa	-----WLL	ETDGTSMKVV	L-----SE	RDVPVVRTTS
Q9XZU9	D. subobscura		Multa	-----WLL	ETDGTSMKVV	L-----SE	RDVPVVRTTS
P04052	D. melanogaster		Multa	-----WLL	ETDGTSMKVV	L-----SE	RDVPVVRTTS
Q9FRB0	O. sativa		Multa	-----WML	DTEGVNLLAV	M-----CH	EDVDATRTTS
P18616	A. thaliana	1	Multa	-----WML	DTEGVNLLAV	M-----CH	EDVDATRTTS
P36594	S. pombe	1	Multa	-----WVL	ETDGINLTEA	M-----TV	EGVDATRTYS
O96446	V. necatrix	1	Multa	-----WYL	QTDGVCIREI	F-----SH	PNVEGHLVTS
P14248	P. falciparum		Multa	-----NDS	SNINDIKVKN	I-----KK	EDNEGALRG
P16356	C. elegans	1	Multa	-----WIL	ETDGTALLRV	L-----SE	RQIDPVRTTS
Q25312	L. major		Multa	-----WMV	DTEGTALQRI	FVGVNSEGG	NIIDFSKTSS
Q9NJR9	L. donovani		Multa	-----WMV	DTEGTALQRI	FVGVNSEGG	NIIDFSKTSS
Q9BJ59	L. seymouri		Multa	-----WMV	DTEGTALQRI	FVGVNQHG	NIVDFAKTSS
P17545	T. brucei brucei	1	Multa	-----WAI	DTDGTLARRA	FIGVVGEDGK	NIINAVKTSS
Q95W43	T. cruzi		Multa	-----WLI	DTDGTLARRA	FIGIVDEKGN	NIINATRTSS
O96452	N. locustae	1	Multa	-----WIL	QTDGKNLKQL	L-----LV	EDILANKVYC
Q9AVW8	G. theta	1	Multa	-----FVL	ETEGVDLRFV	M-----NY	NGIDYSRCTT
Q9HGZ5	A. oryzae	A	Multa	-----WVL	ETSGSALGEV	L-----AV	PGVDATRTYS
Q27095	T. vaginalis		Multa	-----SCL	YTEGTAFREI	L-----SL	DHVDVTHTIT

*Method:
 BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
 PSIPRED and PHD = Structure Prediction
 Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.13 Sequence alignment for *P. furiosus* RNAP subunit A (sheet 13)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	270-----	280-----	-290-----	-300-----
Q8U0M5	P. furiosus	A	RMSD	NNIHEIAEVL	GIEAARNA--	IIDEIVSTMQ	EQGLEVDIRH
Q8U0M5	P. furiosus	A	PSIPRED	NNIHEIAEVL	GIEAARNA--	IIDEIVSTMQ	EQGLEVDIRH
Q8U0M5	P. furiosus	A	PHD	NNIHEIAEVL	GIEAARNA--	IIDEIVSTMQ	EQGLEVDIRH
P04050	S. cerevisiae	1	RMSD	NSFIDIMEVL	GIEAGRAA--	LYKEVYNVIA	SDGSYVNYRH
P04050	S. cerevisiae	1	PSIPRED	NSFIDIMEVL	GIEAGRAA--	LYKEVYNVIA	SDGSYVNYRH
P04050	S. cerevisiae	1	PHD	NSFIDIMEVL	GIEAGRAA--	LYKEVYNVIA	SDGSYVNYRH
P04050	S. cerevisiae	1	MOE	NSFIDIMEVL	GIEAGRAA--	LYKEVYNVIA	SDGSYVNYRH
P04050	S. cerevisiae	1	BLAST	NSFIDIMEVL	GIEAGRAA--	LYKEVYNVIA	SDGSYVNYRH
P04050	S. cerevisiae	1	BLAST2	NSFIDIMEVL	GIEAGRAA--	LYKEVYNVIA	SDGSYVNYRH
P04050	S. cerevisiae	1	MPsrch	NSFIDIMEVL	GIEAGRAA--	LYKEVYNVIA	SDGSYVNYRH
P04050	S. cerevisiae	1	Multa	NSFIDIMEVL	GIEAGRAA--	LYKEVYNVIA	SDGSYVNYRH
O35559	C. griseus		Multa	NDIVEIFTVL	GIEAVRKA--	LERELYHVIS	FDGSYVNYRH
P24928	H. sapiens	1	Multa	NDIVEIFTVL	GIEAVRKA--	LERELYHVIS	FDGSYVNYRH
P08775	M. musculus	1	Multa	NDIVEIFTVL	GIEAVRKA--	LERELYHVIS	FDGSYVNYRH
Q9AXN0	P. vacuolata	1	Multa	NDVVMFVTL	GIEGVRAA--	LLNEIRGVIS	FDGAYVNYRH
Q9XZS1	D. gaunche		Multa	NDICEIFQVL	GIEAVRKS--	VEKEMNAVLO	FYGLYVNYRH
Q9XZS2	D. madeirensis		Multa	NDICEIFQVL	GIEAVRKS--	VEKEMNAVLO	FYGLYVNYRH
Q9XZU9	D. subobscura		Multa	NDICEIFQVL	GIEAVRKS--	VEKEMNAVLO	FYGLYVNYRH
P04052	D. melanogaster		Multa	NDICEIFQVL	GIEAVRKS--	VEKEMNAVLO	FYGLYVNYRH
Q9FRB0	O. sativa		Multa	NHLIEVIEVL	GIEAVRRA--	LLDELRVVIS	FDGSYVNYRH
P18616	A. thaliana	1	Multa	NHLIEVIEVL	GIEAVRRA--	LLDELRVVIS	FDGSYVNYRH
P36594	S. pombe	1	Multa	NSFVEILQIL	GIEATRSA--	LLKELRNVE	FDGSYVNYRH
O96446	V. necatrix	1	Multa	NDLNEIVEVL	GIEAARET--	ILNELTLVID	NGNSYVNNRH
P14248	P. falciparum		Multa	GGDSNTSALF	GNKNSQKEDN	IVNNNNNDND	DDDEEEEEED
P16356	C. elegans	1	Multa	NDICEIFEVL	GIEAVRKA--	IEREMDNVIS	FDGSYVNYRH
Q25312	L. major		Multa	NKIPVVTVL	GIEAARRK--	MLSELREAYL	AYGLNINRYH
Q9NJR9	L. donovani		Multa	NKIPVVTVL	GIEAARRK--	LLSELREAYL	AYGLNINRYH
Q9BJ59	L. seymouri		Multa	NKIPVVTVL	GIEAARRK--	LLSELREAYL	AYGLNINRYH
P17545	T. brucei brucei	1	Multa	NKVPEVCSLL	GIEAARSK--	MLTELREAYL	AYGLNINRYH
Q95W43	T. cruzi		Multa	NKVPEVCSLL	GIEAARYK--	MLHELREAYL	AYGLNINRYH
O96452	N. locustae	1	Multa	NDLLEVEVL	GIEAVREC--	ILRELNVIE	AEGSYVNNRH
Q9AVW8	G. theta	1	Multa	NDLLESLKVL	GIEATRKI--	LIQELRNLI	FDGSYVNNRH
Q9HGZ5	A. oryzae	A	Multa	NQFIEVFEVF	GIEAARTA--	VLRELTVLA	FDGSYVNNRH
Q27095	T. vaginalis		Multa	NDIYQVCDVL	GIEAARHS--	LSYEMMLIME	KAGAALNRRH

*Method:
 BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
 PSIPRED and PHD = Structure Prediction
 Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.14 Sequence alignment for *P. furiosus* RNAP subunit A (sheet 14)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	---,---,310	-----320--	-----330--	-,---340-
Q8U0M5	P. furiosus	A	RMSD	IML----VAD	MMTLDGIVRP	IGRHGVVGEK	S-SVLARAFA
Q8U0M5	P. furiosus	A	PSIPRED	IML----VAD	MMTLDGIVRP	IGRHGVVGEK	S-SVLARAFA
Q8U0M5	P. furiosus	A	PHD	IML----VAD	MMTLDGIVRP	IGRHGVVGEK	S-SVLARAFA
P04050	S. cerevisiae	1	RMSD	MAL----LVD	VMTTQGGGTS	VTRHGFNRSN	T-GALMRCSF
P04050	S. cerevisiae	1	PSIPRED	MAL----LVD	VMTTQGGGTS	VTRHGFNRSN	T-GALMRCSF
P04050	S. cerevisiae	1	PHD	MAL----LVD	VMTTQGGGTS	VTRHGFNRSN	T-GALMRCSF
P04050	S. cerevisiae	1	MOE	MAL----LVD	VMTTQGGGTS	VTRHGFNRSN	T-GALMRCSF
P04050	S. cerevisiae	1	BLAST	MAL----LVD	VMTTQGGGTS	VTRHGFNRSN	T-GALMRCSF
P04050	S. cerevisiae	1	BLAST2	MAL----LVD	VMTTQGGGTS	VTRHGFNRSN	T-GALMRCSF
P04050	S. cerevisiae	1	MPsrch	MAL----LVD	VMTTQGGGTS	VTRHGFNRSN	T-GALMRCSF
P04050	S. cerevisiae	1	Multa	MAL----LVD	VMTTQGGGTS	VTRHGFNRSN	T-GALMRCSF
O35559	C. griseus		Multa	LAL----LCD	TMTCRGHIMA	ITRHGVNRQD	T-GPLMKCSF
P24928	H. sapiens	1	Multa	LAL----LCD	TMTCRGHIMA	ITRHGVNRQD	T-GPLMKCSF
P08775	M. musculus	1	Multa	LAL----LCD	TMTCRGHIMA	ITRHGVNRQD	T-GPLMKCSF
Q9AXN0	G. vacuolata	1	Multa	LAI----LAD	IMTCRGHIMS	ITRHGINRVE	T-GALMRCSF
Q9XZS1	D. gauchae		Multa	LAL----LCD	VMTAKGHIMA	ITRHGINRQD	T-GALMRCSF
Q9XZS2	D. madeirensis		Multa	LAL----LCD	VMTAKGHIMA	ITRHGINRQD	T-GALMRCSF
Q9XZU9	D. subobscura		Multa	LAL----LCD	VMTAKGHIMA	ITRHGINRQD	T-GALMRCSF
P04052	D. melanogaster		Multa	LAL----LCD	VMTAKGHIMA	ITRHGINRQD	T-GALMRCSF
Q9FRB0	O. sativa		Multa	LAV----LCD	TMTYRGHIMA	ITRHGINRND	T-GPLMRCSF
P18616	A. thaliana	1	Multa	LAI----LCD	TMTYRGHIMA	ITRHGINRND	T-GPLMRCSF
P36594	S. pombe	1	Multa	LAL----LCD	VMTSRGHIMA	ITRHGINRAE	T-GALMRCSF
O96446	V. necatrix	1	Multa	ISL----LAD	VMTMKGYLTG	ITRHGVNKGK	F-GCTKRASF
P14248	P. falciparum		Multa	FLFGDHNVSF	KNTKDGKKNK	TNNKSNNNEN	K-NKKSGNNN
P16356	C. elegans	1	Multa	LAL----LCD	VMTAKGHIMA	ITRHGINRQE	V-GALMRCSF
Q25312	L. major		Multa	YTI----LVD	TMCQRGYIMA	VSRTGINRSE	TSGPLMRCSF
Q9NJR9	L. donovani		Multa	YTI----LVD	TMCQRGYIMA	VSRTGINRSE	TSGPLMRCSF
Q9BJ59	L. seymouri		Multa	YTI----LVD	TMCQRGYIMA	VSRTGINRSE	TSGPLMRCSF
P17545	T. brucei brucei	1	Multa	YTI----LVD	TICQHGYYIMA	VSRSGINRSD	TSGPLMRCSF
Q95W43	T. cruzi		Multa	YTI----LVD	TICQRGYIMA	VSRVGINRSE	MSGPLMRCSF
O96452	N. locustae	1	Multa	LYL----LCD	VMTLKG-LRG	ITRHGANRAA	T-GALKRCSF
Q9AVW8	G. theta	1	Multa	LCT----LVD	LITFRGSLMS	INRHGINKCN	S-GPIAKSTF
Q9HGZ5	A. oryzae	A	Multa	LAL----LVD	VMTVRGYLTP	VTRHGINRAD	N-GALMRCSF
Q27095	T. vaginalis		Multa	LDL----LAD	TMTQYGVLYP	VSRHGINKTP	T-GTLMRASV

*Method:
 BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
 PSIPRED and PHD = Structure Prediction
 Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.15 Sequence alignment for *P. furiosus* RNAP subunit A (sheet 15)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	-----350-	-----360-	-----,,	,,370-----
Q8U0M5	P. furiosus	A	RMSD	EITVQHLFEA	AEKGEVDNLN	GVIENVLI--	--GQVPVPGT
Q8U0M5	P. furiosus	A	PSIPRED	EITVQHLFEA	AEKGEVDNLN	GVIENVLI--	--GQVPVPGT
Q8U0M5	P. furiosus	A	PHD	EITVQHLFEA	AEKGEVDNLN	GVIENVLI--	--GQVPVPGT
P04050	S. cerevisiae	1	RMSD	EETVEILFEA	GASAELEDDCR	GVSENVIL--	--GQMAPIGT
P04050	S. cerevisiae	1	PSIPRED	EETVEILFEA	GASAELEDDCR	GVSENVIL--	--GQMAPIGT
P04050	S. cerevisiae	1	PHD	EETVEILFEA	GASAELEDDCR	GVSENVIL--	--GQMAPIGT
P04050	S. cerevisiae	1	MOE	EETVEILFEA	GASAELEDDCR	GVSENVIL--	--GQMAPIGT
P04050	S. cerevisiae	1	BLAST	EETVEILFEA	GASAELEDDCR	GVSENVIL--	--GQMAPIGT
P04050	S. cerevisiae	1	BLAST2	EETVEILFEA	GASAELEDDCR	GVSENVIL--	--GQMAPIGT
P04050	S. cerevisiae	1	MPsrch	EETVEILFEA	GASAELEDDCR	GVSENVIL--	--GQMAPIGT
P04050	S. cerevisiae	1	Multa	EETVEILFEA	GASAELEDDCR	GVSENVIL--	--GQMAPIGT
O35559	C. griseus		Multa	EETVDVLMFA	AAHGSDPMK	GVSENVIL--	--GQLAPAGT
P24928	H. sapiens	1	Multa	EETVDVLMFA	AAHGSDPMK	GVSENVIL--	--GQLAPAGT
P08775	M. musculus	1	Multa	EETVDVLMFA	AAHGSDPMK	GVSENVIL--	--GQLAPAGT
Q9AXN0	G. vacuolata	1	Multa	EETADILLEA	AMYGEVDDMK	GVAENIML--	--GQMAPVGT
Q9XZS1	D. gauchae		Multa	EETVDVLMFA	AAHAETDFMR	GVSENVIL--	--GQLPKMGT
Q9XZS2	D. madeirensis		Multa	EETVDVLMFA	AAHAETDFMR	GVSENVIL--	--GQLPKMGT
Q9XZU9	D. subobscura		Multa	EETVDVLMFA	AAHAETDFMR	GVSENVIL--	--GQLPKMGT
P04052	D. melanogaster		Multa	EETVDVLMFA	AAHAETDFMR	GVSENVIL--	--GQLPKMGT
Q9FRB0	O. sativa		Multa	EETVDILLEA	AVYAESDPLR	GVSENVIL--	--GQLAPIGT
P18616	A. thaliana	1	Multa	EETVDILLEA	AAHAETDCLR	GVSENVIL--	--GQLAPIGT
P36594	S. pombe	1	Multa	EETVEILMDA	AASGEKDDCK	GVSENVIL--	--GQLAPMGT
O96446	V. necatrix	1	Multa	EETVDILLEA	ALVAEKYVTK	GVSENVIL--	--GHLAPLGT
P14248	P. falciparum		Multa	SNNSNTYDDG	DVDNDNDDN	DDNKSDITIK	EDNDVAFMKT
P16356	C. elegans	1	Multa	EETVDILMEA	AVHAEEDPVK	GVSENVIL--	--GQLARCGT
Q25312	L. major		Multa	EETVKVLMFA	AAFGEKDPVR	GVSENVIL--	--GNQARIGT
Q9NJR9	L. donovani		Multa	EETVKVLMFA	AAFGEKDPVR	GVSENVIL--	--GNQARIGT
Q9BJ59	L. seymouri		Multa	EETVKVLMFA	AAFGEKDPVR	GVSENVIL--	--GNQARIGT
P17545	T. brucei brucei	1	Multa	EETVKVLMFA	ASFGECDPVR	GVSENVIL--	--GNQARVGT
Q95W43	T. cruzi		Multa	EETVKVLMFA	AAFGERDPVR	GVSENVIL--	--GNQARIGT
O96452	N. locustae	1	Multa	EETVEILLEA	SCFAERNYCR	GVSENVIL--	--GQLAPIGT
Q9AVW8	G. theta	1	Multa	EETVDIFYES	AVFGLNDNLR	GVSAQIMT--	--GMVPKIGT
Q9HGZ5	A. oryzae	A	Multa	EETVEILLEA	AAFGELEDDCR	GVSENVIL--	--GQMAPAGT
Q27095	T. vaginalis		Multa	EMTNDVFFDA	AVFSEIDYMR	DVSSNIIA--	--GKPSRAGT

*Method:

BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD

PSIPRED and PHD = Structure Prediction

Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.16 Sequence alignment for *P. furiosus* RNAP subunit A (sheet 16)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	380-----	390----397
Q8U0M5	P. furiosus	A	RMSD	GMVKLTMKLP	LRPQKEKEEV
Q8U0M5	P. furiosus	A	PSIPRED	GMVKLTMKLP	LRPQKEKEEV
Q8U0M5	P. furiosus	A	PHD	GMVKLTMKLP	LRPQKEKEEV
P04050	S. cerevisiae	1	RMSD	GAFDVMIDEE	SL
P04050	S. cerevisiae	1	PSIPRED	GAFDVMIDEE	SL
P04050	S. cerevisiae	1	PHD	GAFDVMIDEE	SL
P04050	S. cerevisiae	1	MOE	GAFDVMIDEE	SL
P04050	S. cerevisiae	1	BLAST	G	
P04050	S. cerevisiae	1	BLAST2	G	
P04050	S. cerevisiae	1	MPsrch	G	
P04050	S. cerevisiae	1	Multa	GAFDVMIDEE	-S-LVKYMPE
O35559	C. griseus		Multa	GCFDLLLLDAE	-K-CKYGMEI
P24928	H. sapiens	1	Multa	GCFDLLLLDAE	-K-CKYGMEI
P08775	M. musculus	1	Multa	GCFDLLLLDAE	-K-CKYGMEI
Q9AXN0	G. vacuolata	1	Multa	GEFGLLLNEK	-A-LKNLITA
Q9XZS1	D. gaunche		Multa	GCFDLLLLDAE	-K-CRFGIEI
Q9XZS2	D. madeirensis		Multa	GCFDLLLLDAE	-K-CRFGIEI
Q9XZU9	D. subobscura		Multa	GCFDLLLLDAE	-K-CRFGIEI
P04052	D. melanogaster		Multa	GCFDLLLLDAE	-K-CRFGIEI
Q9FRB0	O. sativa		Multa	GGCDLYLNDQ	-M-LKQAIEL
P18616	A. thaliana	1	Multa	GDCELYLNDE	-M-LKNAIEL
P36594	S. pombe	1	Multa	GAFDIYLDQD	-M-LMNYSLG
O96446	V. necatrix	1	Multa	GIGNLLLDVS	-K-LDKAIPL
P14248	P. falciparum		Multa	STKNAEEDLE	LK-NKNHIEH
P16356	C. elegans	1	Multa	GCFDLVLDVE	-K-CKYGMEI
Q25312	L. major		Multa	GLFDLLLDMS	-----KLQH
Q9NJR9	L. donovani		Multa	GLFDLLLDMS	-----KLQH
Q9BJ59	L. seymouri		Multa	GLFDLMLDMG	-----NCSN
P17545	T. brucei brucei	1	Multa	GLFDLVINMA	-----ALQQ
Q95W43	T. cruzi		Multa	GLFDLVINMA	-----ALQQ
O96452	N. locustae	1	Multa	GNTICILDKK	-A-LRSVIPV
Q9AVW8	G. theta	1	Multa	GKIDLFVNI	
Q9HGZ5	A. oryzae	A	Multa	GEFDIYLDQN	-M-LNTVVSN
Q27095	T. vaginalis		Multa	GIVEVLINED	LLPMTKGNNG

*Method:
 BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
 PSIPRED and PHD = Structure Prediction
 Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.17 Sequence alignment for *P. furiosus* RNAP subunit A (sheet 17)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*		-----10-----	-20-----	---30---
Q8U0M4	P. furiosus	A'	RMSD		MKKVIGS IEFGLSPQE	IRKMSAVE--	VTVPDTYDD-
Q8U0M4	P. furiosus	A'	PSIPRED		MKKVIGS IEFGLSPQE	IRKMSAVE--	VTVPDTYDD-
Q8U0M4	P. furiosus	A'	PHD		MKKVIGS IEFGLSPQE	IRKMSAVE--	VTVPDTYDD-
P04050	S. cerevisiae	1	RMSD	VGQQY SSAPLRTVKE	VQFGLFSPEE	VRAISVAK--	IRFPETMDE-
P04050	S. cerevisiae	1	PSIPRED	VGQQY SSAPLRTVKE	VQFGLFSPEE	VRAISVAK--	IRFPETMDE-
P04050	S. cerevisiae	1	PHD	VGQQY SSAPLRTVKE	VQFGLFSPEE	VRAISVAK--	IRFPETMDE-
P04050	S. cerevisiae	1	MOE	VGQQY SSAPLRTVKE	VQFGLFSPEE	VRAISVAK--	IRFPETMDE-
P04050	S. cerevisiae	1	BLAST	RTVKE VQFGLFSPEE		VRAISVAK--	IRFPETMDE-
P04050	S. cerevisiae	1	BLAST2	RTVKE VQFGLFSPEE		VRAISVAK--	IRFPETMDE-
P04050	S. cerevisiae	1	MPsrch		VQFGLFSPEE	VRAISVAK--	IRFPETMDE-
P04050	S. cerevisiae	1	Multa	MVGQQYS- -SAPLRTVKE	VQFGLFSPEE	VRAISVAK--	IRFPETMDET
P08775	M. musculus	1	Multa	MHGGGPPSGD	SACPLRTIKR VQFGLSPDE	LKRMSVTEGG	IKYPETTE--
P24928	H. sapiens	1	Multa	MHGGGPPSGD	SACPLRTIKR VQFGLSPDE	LKRMSVTEGG	IKYPETTE--
P36594	S. pombe	1	Multa	MSGIQFSP	SSVPLRRVEE VQFGLSPDE	IRMSVACE--	IKFETAFA--
O35559	C. griseus		Multa	MHGGGPPSGD	SACPLRTIKR VQFGLSPDE	LKRMSVTEGG	IKYPETTE--
Q9AVW8	G. theta	1	Multa		MVREEKLIK I QFGLSPDE	ILKLSVQV--	VDSMDTFE--
Q9AXN0	G. vacuolata	1	Multa	MSFVKDFLY	SSAEVKKVKA VQFGLSPDE	IRMSVACE--	IKFETAFA--
Q96452	N. locustae	1	Multa	MSC	QEIIKKRVKA I QFGLSPDE	IRASSVAL--	IRYPETLE--
P18616	A. thaliana	1	Multa	MDTRFPF	SPAENVSKVRV VQFGLSPDE	IRQMSVIH--	VEHSETTE--
Q25312	L. major		Multa	MSGGAALPP	SQMPLQKVHE VQFEVFKAQ	IKAYAKCI--	IEHAKSYA--
Q9XZS1	D. guanche		Multa	MSTTTD	SKAPLRQVQR VQFGLSPDE	IRMSVTEGG	VQFAETME--
Q9BJ59	L. seymouri		Multa	MSGGAPLPP	SQMPLQKVHD VQFEVFKERQ	IKAYATCR--	IEHAKSYE--
P16356	C. elegans		Multa	MALVGD	FQAPLRIVSR VQFGLSPDE	IKRMSVAH--	VEFPEVYE--
P17545	T. brucei brucei		Multa	MSGGAALPV	SQMLHKKVNE VQFEVFKERQ	IKSYAVCL--	VEHAKSYA--
P14248	P. falciparum		Multa	MTVDLNI PY	SACELKRVRK LELGVLDPDEI	IKKISVCE--	IVNVDIYK--
P04052	D. melanogaster		Multa	MSTPTD	SKAPLRQVQR VQFGLSPDE	IRMSVTEGG	VQFAETME--
Q980R2	S. solfataricus	A'	Multa		MSEKNIGK IKFGLSPDE	IRKMSVTA--	IITPDVYDE-
Q979F4	T. volcanium	A'	Multa		MMGISKRISS IKFALLSPDE	IRKLSQVK--	VITADTYDD-
P31813	T. celer	A'	Multa		MQSIKKVIGS IDFGI LSPQE	IRKMSAAE--	ITVPDTYDE-
Q9V114	P. abyssi	A'	Multa		MQSVKKVIGS IEFGLSPQE	IRKMSAVE--	VTVPDTYDD-
Q8ZYQ8	P. aerophilum	A'	Multa	MSLREEF	DAIPKKVIKS IKFGLSPDE	IRKYSVME--	VTTSEVYDE-
Q58445	M. jannaschii	A'	Multa	ME	RYEIPKEIGE IMFGLSPDY	IRQMSVAK--	IVTPDTYDE-
P11512	S. acidocaldarius	A'	Multa		MSEKIIRG VKFGLSPNE	IRQMSVTA--	IITSEVYDE-
O27125	M. thermoautotrophicum	A'	Multa		MRGILKKISQ INFGLMSPED	IRKMSVTQ--	IVTPDTYDE-
P09846	M. thermoautotrophicum (strain Winter)	A'	Multa		MRGILKKISQ INFGLMSPED	IRKMSVAK--	IVTPDTYDE-
O28391	A. fulgidus	A'	Multa		MVFKRISA IKFEVLSFQE	IRMSVVK--	IITPETYDD-
P15350	H. halobium	A'	Multa	MS	AGQAPKEIGE ISFGLMDPEE	YRDMSATK--	VITADTYDD-
P41556	M. vanielii	A'	Multa	MD	RFDVPKEIGD ITFGLLSPEQ	IRTMSVAK--	IVTADTYDD-
Q03585	T. acidophilum	A'	Multa		MMGISKRISS IKFALLSPDE	IRKLSQVK--	VITADTYDD-

*Method:

BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD

PSIPRED and PHD = Structure Prediction

Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.18 Sequence alignment for *P. furiosus* RNAP subunit A' (sheet 1)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	----40,--	----50--	-----60--	-----70--	-----80--
Q8U0M4	<i>P. furiosus</i>	A'	RMSD	DGYPIE-GG- VMDK-RMGVI	DPGLRCETCG	GRAGECPGHF	GHIELARPVI	
Q8U0M4	<i>P. furiosus</i>	A'	PSIPRED	DGYPIE-GG- VMDK-RMGVI	DPGLRCETCG	GRAGECPGHF	GHIELARPVI	
Q8U0M4	<i>P. furiosus</i>	A'	PHD	DGYPIE-GG- VMDK-RMGVI	DPGLRCETCG	GRAGECPGHF	GHIELARPVI	
P04050	<i>S. cerevisiae</i>	1	RMSD	TQTRAK-IGG LNDP-RLGSI	DRNLKCQTCQ	EGMNECPGHF	GHIIDLAKPVF	
P04050	<i>S. cerevisiae</i>	1	PSIPRED	TQTRAK-IGG LNDP-RLGSI	DRNLKCQTCQ	EGMNECPGHF	GHIIDLAKPVF	
P04050	<i>S. cerevisiae</i>	1	PHD	TQTRAK-IGG LNDP-RLGSI	DRNLKCQTCQ	EGMNECPGHF	GHIIDLAKPVF	
P04050	<i>S. cerevisiae</i>	1	MOE	TQTRAKIGG- LNDP-RLGSI	DRNLKCQTCQ	EGMNECPGHF	GHIIDLAKPVF	
P04050	<i>S. cerevisiae</i>	1	BLAST	TQTRAKIGG- LNDP-RLGSI	DRNLKCQTCQ	EGMNECPGHF	GHIIDLAKPVF	
P04050	<i>S. cerevisiae</i>	1	BLAST2	TQTRAKIGG- LNDP-RLGSI	DRNLKCQTCQ	EGMNECPGHF	GHIIDLAKPVF	
P04050	<i>S. cerevisiae</i>	1	MPsrch	TQTRAKIGG- LNDP-RLGSI	DRNLKCQTCQ	EGMNECPGHF	GHIIDLAKPVF	
P04050	<i>S. cerevisiae</i>	1	MPsrch	TQTRAKIGG- LNDP-RLGSI	DRNLKCQTCQ	EGMNECPGHF	GHIIDLAKPVF	
P08775	<i>M. musculus</i>	1	Multa	GGRPKL-GG- LMDP-RQGV	ERTGRCQTCA	GNMTECPGHF	GHIELAKPVF	
P24928	<i>H. sapiens</i>	1	Multa	GGRPKL-GG- LMDP-RQGV	ERTGRCQTCA	GNMTECPGHF	GHIELAKPVF	
P36594	<i>S. pombe</i>	1	Multa	GGRPRV-GG- LLDL-RLGT	DRQFKCQTCG	ETMADCPGHF	GHIELAKPVF	
O35559	<i>C. griseus</i>		Multa	GGRPKL-GG- LMDP-RQGV	ERTGRCQTCA	GNMTECPGHF	GHIELAKPVF	
Q9AVW8	<i>G. theta</i>	1	Multa	HNLKPL-GG- LLDL-RMGV	DKEFFCTTDM	GSFLESPGYF	GHIELAKPVF	
Q9AXN0	<i>G. vacuolata</i>	1	Multa	MGKPKT-EG- LMDP-RLGAI	GRDLPCKTCH	CDEKNCPGHF	GHIELAKPVF	
Q96452	<i>N. locustae</i>	1	Multa	NGVPKE-SG- LIDL-RMGV	ERDYLCACTN	QDNVVCAGHF	GHIELAKPVF	
P18616	<i>A. thaliana</i>	1	Multa	KGKPKV-GG- LSDT-RLGT	DRKVKCETCM	ANMAECPGHF	GHIELAKPVF	
Q25312	<i>L. major</i>		Multa	STASRC-GA- ASTTCAWARQ	TSSIACETCG	LRHPECPPHF	GHIELAKPVF	
Q9XZS1	<i>D. guanche</i>		Multa	GGRPKL-GG- LMDP-RQGV	DRTSRCQTCA	GNMTECPGHF	GHIIDLAKPVF	
Q9BJ59	<i>L. seymouri</i>		Multa	HGQPVV-GG- INDL-RMGV	DFDYSCETCG	LKHPECPPHF	GHIELAKPVF	
P16356	<i>C. elegans</i>		Multa	NGKPKL-GG- LMDP-RQGV	DRGRCMCTCA	GNLTDCPPHF	GHIELAKPVF	
P17545	<i>T. brucei brucei</i>		Multa	NAADQS-GE- ASMICVWVPL	TSNSACETCH	RKHPECPPHF	GHIELAKPVF	
P14248	<i>P. falciparum</i>		Multa	DGFPRE-GG- LNDI-RMGV	DYRTLCTGTCN	MNVKYCPGHF	GHIELAKPVF	
P04052	<i>D. melanogaster</i>		Multa	GGRPKL-GG- LMDP-RQGV	DRTSRCQTCA	GNMTECPGHF	GHIIDLAKPVF	
Q980R2	<i>S. solfataricus</i>	A'	Multa	DGTPIE-GG- VMDP-RLGVI	EPGQKCPCTG	NTLGNCPPHF	GHIELAKPVF	
Q979F4	<i>T. volcanium</i>	A'	Multa	DGYPIE-HG- LMDL-HMGVI	EPGLRCATCG	GKVDECPGHF	GHIELAMPV	
P31813	<i>T. celer</i>	A'	Multa	DGYPIE-GG- LMDK-RLGVI	DPGLRCETCG	ARAGECPGHF	GHIELAMPV	
Q9V114	<i>P. abyssi</i>	A'	Multa	DGYPIE-GG- VMDK-RMGVI	DPGLRCETCG	AKAGECPGHF	GHIELAMPV	
Q8ZYQ8	<i>P. aerophilum</i>	A'	Multa	GGLPVR-GG- IADR-RLGVA	EPGSRCECTG	QTHDVCPPHF	GHIELAKPVF	
Q58445	<i>M. jannaschii</i>	A'	Multa	DGYPID-GG- LMDT-RLGVI	DPGLVCKTCG	GRIGECPPHF	GHIELAKPVF	
P11512	<i>S. acidocaldarius</i>	A'	Multa	DGTPIE-GG- VMDP-KLGVI	EPGEKCPVCG	NTLAGCPGHF	GHIELAKPVF	
O27125	<i>M. thermoautotrophicum</i>	A'	Multa	DGYPIE-NG- LMDP-RLGVI	DPSLRCTCG	AKGGECPGHF	GSINLARPV	
P09846	<i>M. thermoautotrophicum</i> (strain Winter)	A'	Multa	DGYPIE-NE- LMDP-RLGVI	GSSLRCSRSG	AKGGECPGHF	GSINLARPV	
O28391	<i>A. fulgidus</i>	A'	Multa	DGFPID-GG- LMDT-RLGVI	DPGLRCKTCG	GKAGECPGHF	GHIELAKPVF	
P15350	<i>H. halobium</i>	A'	Multa	DGFPID-MG- LMDP-RLGVI	DPGLECKTCG	QRSGGCNGHF	GHIELAKPVF	
P41556	<i>M. vannielii</i>	A'	Multa	DGYPID-GG- LMDT-RLGVI	DPGLVCKSCS	GRVGTCPGHF	GHIELAKSVI	
Q03585	<i>T. acidophilum</i>	A'	Multa	DGYPIE-HG- LMDL-HMGVI	EPGLRCATCG	GKVDECPGHF	GHIELAMPV	

*Method:

BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD

PSIPRED and PHD = Structure Prediction

Red = Helix Green = Strand Blue = Turn Gray = No cord

Figure C.19 Sequence alignment for *P. furiosus* RNAP subunit A' (sheet 2)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	-----90-	-----100-	-----110-	-----120-----
Q8U0M4	P. furiosus	A'	RMSD	HVGFAKTIYR	IESTCRECG	RIKLTDEEIE	EYM---KKIE LA-RNRRSE-
Q8U0M4	P. furiosus	A'	PSIPRED	HVGFAKTIYR	IESTCRECG	RIKLTDEEIE	EYM---KKIE LA-RNRRSE-
Q8U0M4	P. furiosus	A'	PHD	HVGFAKTIYR	IESTCRECG	RIKLTDEEIE	EYM---KKIE LA-RNRRSE-
P04050	S. cerevisiae	1	RMSD	HVGFIKIKK	VCEVCVMHC-	GKLLLE-HN	ELM---RQAL AI-KDSKKRF
P04050	S. cerevisiae	1	PSIPRED	HVGFIKIKK	VCEVCVMHC-	GKLLLE-HN	ELM---RQAL AI-KDSKKRF
P04050	S. cerevisiae	1	PHD	HVGFIKIKK	VCEVCVMHC-	GKLLLE-HN	ELM---RQAL AI-KDSKKRF
P04050	S. cerevisiae	1	MOE	HVGFIKIKK	VCEVCVMHCG	KL-LLDEH-N	ELM---RQAL AI-KDSKKR-
P04050	S. cerevisiae	1	BLAST	HVGFIKIKK	VCEVCVMHCG	KLL--DEHN	ELM---RQAL AI-KDSKKR-
P04050	S. cerevisiae	1	BLAST2	HVGFIKIKK	VCEVCVMHCG	KL-LLDEH-N	ELM---RQAL AI-KDSKKR-
P04050	S. cerevisiae	1	MPsrch	HVGFIKIKK	VCEVCVMHCG	KL-LLDEH-N	ELM---RQ-A LAIKDSKKR-
P04050	S. cerevisiae	1	Multa	HVGFIKIKK	VCEVCVMHCG	KLLLDEHNEL	M-----RQAL AI---K--D-
P08775	M. musculus	1	Multa	HVGFLVKTMK	VLRCVCFPCS	KLLVDSNNPK	I-----KDIL AK-----SKGQ-
P24928	H. sapiens	1	Multa	HVGFLVKTMK	VLRCVCFPCS	KLLVDSNNPK	I-----KDIL AK-----SKGQ-
P36594	S. pombe	1	Multa	HIGFLSKIKK	ILECVCMWCG	KLKIDSSNPK	F-----NDTQ RY---R--D-
O35559	C. griseus		Multa	HVGFLVKTMK	VLRCVCFPCS	KLLVDSNNPK	I-----KDIL AK-----SKGQ-
Q9AVW8	G. theta	1	Multa	HEGFMDYVLR	ILRCIDHTTS	KLIINENSQK	I-----KNVL KI-----KN-
Q9AXN0	G. vacuolata	1	Multa	NIGFFGMTLR	VLRCVCFPCS	KLLIDPRDPK	A-----DFEV KK-VLRIKN-
Q96452	N. locustae	1	Multa	HVGFIKIKK	VLECVCFYCS	KVRINKNKT	I-----TK-- -----E-
P18616	A. thaliana	1	Multa	HVGFMKTVLS	IMRCVCFNCS	KILADEVCRS	L-----FRQA MK-----IK-N-
Q25312	L. major		Multa	NINVEDVLI	ALKCVCKYCG	ALLMDTNDPS	E-----MKKI AH-----LQ-
Q9XZS1	D. guanche		Multa	HIGFITKTIK	ILRCVCFYCS	KMLVSPHNPK	I-----KEIV VK-----SKGQ-
Q9BJ59	L. seymouri		Multa	NINVEDVLI	ALKCVCKYCG	ALLMDTNDPT	E-----MKKI TH-----VQ-
P16356	C. elegans		Multa	HIGFLTKTLK	ILRCVCFYCG	RLLIDKSAPR	V-----LEIL KK---TGNT-
P17545	T. brucei brucei		Multa	NIGVFDLVL	VLKVCCTCG	ALLNTREOD	V-----HKKL QH-----MT-
P14248	P. falciparum		Multa	HYGFMMVVLN	VLRCVCYHCG	RLLCNVNSSK	V-----KYIE KI-----KV-
P04052	D. melanogaster		Multa	HIGFITKTIK	ILRCVCFYCS	KMLVSPHNPK	I-----KEIV MK---SRGQ-
Q980R2	S. solfataricus	A'	Multa	HVGLVKHIYE	FLKATCRRCG	RVKISEDEIE	KYSRIYNAIK KR-W--PSA-
Q979F4	T. volcanium	A'	Multa	HVGfVKEIKM	FLDATCKSCG	RIKLTDEEIK	TYLPEVQKVD FE-TGDPED-
P31813	T. celer	A'	Multa	HVGFAKTIHR	VLESTCRECG	RIKLTDEEIE	EYF---HKFE VT-GNRKKA-
Q9V114	P. abyssi	A'	Multa	HVGFAKTIHR	IESTCRECG	RIKLTDEEIE	EYM---KKLE LA-KNRRSE-
Q8ZYQ8	P. aerophilum	A'	Multa	HVGFAKTIYD	VLATCPNCG	RIMLRDEEIE	RYRERLNKIK AR-W--RL-
Q58445	M. jannaschii	A'	Multa	HIGFAKTIYK	ILKAVCPHCG	RVAISETKRR	E-----ILEK ME-----K-
P11512	S. acidocaldarius	A'	Multa	HIGYVKHIYD	FLRSTCWRCG	RIKIKEQDLE	RYKRIYNAIK LR-W--PSA-
O27125	M. thermoautotrophicum	A'	Multa	HVGFAKTIHK	ILRSTCRKCG	RVLLTETEIE	EYR---QRIL DA-MEKEES-
P09846	M. thermoautotrophicum (strain Winter)	A'	Multa	HVGFAKTIHK	ILSSICRCKS	RILLTETEIQ	DYR---QRIL EA-MEREES-
Q28391	A. fulgidus	A'	Multa	HVGYAKMIAR	LLNGTCSECG	RILLKDERRD	KFI---AEIE RR-KELNQS-
P15350	H. halobium	A'	Multa	HVGFSKLIRR	LLRGTCRECA	SLLLTEEEKD	EYR---ENLD RT-RSLRQD-
P41556	M. vannielii	A'	Multa	HIGFAKDIYK	LLKAVCPHCG	KVTVTEIKRD	EYL---EKML KL-EEDGGD-
Q03585	T. acidophilum	A'	Multa	HVGfVKEIKM	FLDATCRSCG	RIKLTDEEIR	TYLPEIQKMD FE-TGDPED-

*Method:
 BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
 PSIPRED and PHD = Structure Prediction
 Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.20 Sequence alignment for *P. furiosus* RNAP subunit A' (sheet 3)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	130	140
Q8U0M4	<i>P. furiosus</i>	A'	RMSD	-----F--	-----NEI- -I-K-E---I -HKK-A-KE- --RMVC-P-
Q8U0M4	<i>P. furiosus</i>	A'	PSIPRED	-----F--	-----NEI- -I-K-E---I -HKK-A-KE- --RMVC-P-
Q8U0M4	<i>P. furiosus</i>	A'	PHD	-----F--	-----NEI- -I-K-E---I -HKK-A-KE- --RMVC-P-
P04050	<i>S. cerevisiae</i>	1	RMSD	AAIWTLC--	-----TKM- -V-C-E---T -DVP-SEDDP T--QLVS-RG
P04050	<i>S. cerevisiae</i>	1	PSIPRED	AAIWTLC--	-----TKM- -V-C-E---T -DVP-SEDDP T--QLVS-RG
P04050	<i>S. cerevisiae</i>	1	PHD	AAIWTLC--	-----TKM- -V-C-E---T -DVP-SEDDP T--QLVS-RG
P04050	<i>S. cerevisiae</i>	1	MOE	-----FAA IWTLC--	-----TKM- -V-C-E---T -DVP-SEDDP T--QLVS-RG
P04050	<i>S. cerevisiae</i>	1	BLAST	-----F--	-----AAIW TLCKTKM--V CETDVP-SED DPTQLVS-R-
P04050	<i>S. cerevisiae</i>	1	BLAST2	-----F--	-----AAIW TLCKTKM--V CETDVP-SED DPTQLVS-R-
P04050	<i>S. cerevisiae</i>	1	MPsrch	-----F--	-----AAIW TLCKTKM--V CETDVP-SED DPTQLVS-R-
P04050	<i>S. cerevisiae</i>	1	Multa	-----S--	-----KKR- -F-A-A---I -WTL-C-KT- --KMVC-E-
P08775	<i>M. musculus</i>	1	Multa	-----P--	-----KKR- -L-T-H---V -YDL-C-KG- --KNIC-E-
P24928	<i>H. sapiens</i>	1	Multa	-----P--	-----KKR- -L-T-H---V -YDL-C-KG- --KNIC-E-
P36594	<i>S. pombe</i>	1	Multa	-----P--	-----KNR- -L-N-A---V -WNV-C-KT- --KMVC-D-
O35559	<i>C. griseus</i>	1	Multa	-----P--	-----KKR- -L-T-H---V -YDL-C-KG- --KNIC-E-
Q9AVW8	<i>G. theta</i>	1	Multa	-----L--	-----SQR- -L-N-I---I -SKL-C-LT- --ACNL-K-
Q9AXN0	<i>G. vacuolata</i>	1	Multa	-----R--	-----INR- -L-K-K---V -QNL-C-SK- --VKIC-K-
Q96452	<i>N. locustae</i>	1	Multa	-----S--	-----KHK- -L-N-F---I -WGI-C-KN- --KMVC-E-
P18616	<i>A. thaliana</i>	1	Multa	-----P--	-----KNR- -L-K-K---I -LDA-C-KN- --KTKC-D-
Q25312	<i>L. major</i>	1	Multa	-----G--	-----LNR- -L-R-M---V -AKL-C-GS- --VCKR-S-
Q9XZS1	<i>D. guanche</i>	1	Multa	-----P--	-----RKR- -L-A-Y---V -YDL-C-KG- --KTIC-E-
Q9BJ59	<i>L. seymouri</i>	1	Multa	-----G--	-----LNR- -L-R-L---V -AKL-C-GS- --VCKR-S-
P16356	<i>C. elegans</i>	1	Multa	-----S--	-----KKR- -L-T-M---I -YDL-C-KA- --KSVK-E-
P17545	<i>T. brucei brucei</i>	1	Multa	-----G--	-----LNR- -L-R-Q---V -AKM-A-EA- --KCRV-S-
P14248	<i>P. falciparum</i>	1	Multa	-----N--	-----SLR- -L-R-K---L -AEL-C-LG- --IRAC-D-
P04052	<i>D. melanogaster</i>	1	Multa	-----P--	-----RKR- -L-A-Y---V -YDL-C-KG- --KTIC-E-
Q980R2	<i>S. solfataricus</i>	A'	Multa	-----A--	-----RRL- -T-E-Y---V -KKT-A-MK- --AQVC-P-
Q979F4	<i>T. volcanium</i>	A'	Multa	-----I--	-----ELM- -T-K-R---F -VDL-A-SQ- --RMVC-P-
P31813	<i>T. celer</i>	A'	Multa	-----K--	-----DRL- -I-K-E---I -HKK-A-KE- --RMVC-P-
Q9V114	<i>P. abyssi</i>	A'	Multa	-----V--	-----NKI- -L-K-E---I -HKK-A-RE- --RMVC-P-
Q8ZYQ8	<i>P. aerophilum</i>	A'	Multa	-----A--	-----LNL- -H-E-K---I -RKK-A-TE- --RMTC-P-
Q58445	<i>M. jannaschii</i>	A'	Multa	-----L--	-----ERD- -G-G-N---K -WEV-C-EE- --VYKE-A-
P11512	<i>S. acidocaldarius</i>	A'	Multa	-----A--	-----RRL- -V-E-Y---I -KKI-S-IK- --NLEC-P-
O27125	<i>M. thermoautotrophicum</i>	A'	Multa	-----L--	-----TPI- -I-K-E---I -YAE-A-R-- --RDKC-P-
P09846	<i>M. thermoautotrophicum (strain Winter)</i>	A'	Multa	-----L--	-----TPI- -I-K-E---I -YAE-A-R-- --RDRC-P-
Q28391	<i>A. fulgidus</i>	A'	Multa	-----Y--	-----EEV- -V-K-E---V -FRL-T-KA- --AKKC-P-
P15350	<i>H. halobium</i>	A'	Multa	-----V--	-----SDV- -M-T-A---A -IRE-A-RK- --KDHC-P-
P41556	<i>M. vannielii</i>	A'	Multa	-----P--	-----WTL- -S-D-D---L -LKE-A-AK- --SSVC-P-
Q03585	<i>T. acidophilum</i>	A'	Multa	-----I--	-----EIL- -T-K-K---Y -VDL-A-SQ- --RMVC-P-

*Method:
 BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
 PSIPRED and PHD = Structure Prediction
 Red = Helix Green = Strand Blue = Turn Gray = No cord

Figure C.21 Sequence alignment for *P. furiosus* RNAP subunit A' (sheet 4)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	170--	-----	-----	-----	-----	-----
Q8U0M4	P. furiosus	A'	RMSD	GNEYR	-----	-----	-----	-----	-----
Q8U0M4	P. furiosus	A'	PSIPRED	GNEYR	-----	-----	-----	-----	-----
Q8U0M4	P. furiosus	A'	PHD	GNEYR	-----	-----	-----	-----	-----
P04050	S. cerevisiae	1	RMSD	ADEPE	-----	-----	-----	-----	-----
P04050	S. cerevisiae	1	PSIPRED	ADEPE	-----	-----	-----	-----	-----
P04050	S. cerevisiae	1	PHD	ADEPE	-----	-----	-----	-----	-----
P04050	S. cerevisiae	1	MOE	ADEPE	-----	-----	-----	-----	-----
P04050	S. cerevisiae	1	BLAST	ADEPE	-----	-----	-----	-----	-----
P04050	S. cerevisiae	1	BLAST2	ADEPE	-----	-----	-----	-----	-----
P04050	S. cerevisiae	1	MPsrch	ADEPE	-----	-----	-----	-----	-----
P04050	S. cerevisiae	1	Multa	LVSRG	-----	-----	-----	CGNTQPTIRK	DGLKLVGSWK
P08775	M. musculus	1	Multa	EKGHG	-----	-----	-----	CGRYQPRIRR	SGLELYAEWK
P24928	H. sapiens	1	Multa	EKGHG	-----	-----	-----	CGRYQPRIRR	SGLELYAEWK
P36594	S. pombe	1	Multa	NMGHG	-----	-----	-----	CGAAQPTIRK	DGLRLWGSWK
O35559	C. griseus	1	Multa	EKGHG	-----	-----	-----	CGRYQPRIRR	SGLELYAEWK
Q9AVW8	G. theta	1	Multa	-----	-----	-----	-----	RISIQPKYSR	DGWCIVAKFD
Q9AXN0	G. vacuolata	1	Multa	-----	-----	-----	-----	CSNEQPKFAK	EGLQFRAEFK
Q96452	N. locustae	1	Multa	-----	-----	-----	-----	CNNRQPVIKK	EGLSLFAYMK
P18616	A. thaliana	1	Multa	KKSRG	-----	-----	-----	CGAQPKLTI	EGMKMIAEYK
Q25312	L. major	1	Multa	XGP	-----	-----	-----	PAARHP	F-VGIYPGLQ
Q9XZS1	D. guanche	1	Multa	KPGHG	-----	-----	-----	CGHYQPSIRR	TGLDLSAEWK
Q9BJ59	L. seymouri	1	Multa	KGR	-----	-----	-----	QPRIGR	F-LGIYPGLQ
P16356	C. elegans	1	Multa	VAG	-----	-----	-----	CGRYQPSYRR	VGIDINAEWK
P17545	T. brucei brucei	1	Multa	MGPATR	-----	-----	-----	CGASQPRVSR	F-YGIYPTLV
P14248	P. falciparum	1	Multa	LSNLSNMQQM	LLNKSNYTNI	FEMVSKEDVD	CGCVQPKYSR	EGPNMYIQFL	-----
P04052	D. melanogaster	1	Multa	KPGHG	-----	-----	-----	CGHYQPSIRR	TGLDLTAEWK
Q980R2	S. solfataricus	A'	Multa	V	-----	-----	-----	-----	-----
Q979F4	T. volcanium	A'	Multa	-EGTN	-----	-----	-----	-----	-----
P31813	T. celer	A'	Multa	GNEYK	-----	-----	-----	-----	-----
Q9V114	P. abyssi	A'	Multa	GNEYK	-----	-----	-----	-----	-----
Q8ZYQ8	P. aerophilum	A'	Multa	L	-----	-----	-----	-----	-----
Q58445	M. jannaschii	A'	Multa	-----	-----	-----	-----	H CGEIKYDIKF	EKPPTYRIRID
P11512	S. acidocaldarius	A'	Multa	I	-----	-----	-----	-----	-----
O27125	M. thermoautotrophicum	A'	Multa	---GD	-----	-----	-----	-----	-----
P09846	M. thermoautotrophicum (strain Winter)	A'	Multa	---GD	-----	-----	-----	-----	-----
O28391	A. fulgidus	A'	Multa	---GE	-----	-----	-----	-----	-----
P15350	H. halobium	A'	Multa	SDYSERIAAS	MQPDEDEDDA	GVSPQELAEQ	TDIDISRINE	ILSGEFPRR	-----
P41556	M. vannielii	A'	Multa	DGKSQ	-----	-----	-----	-----	-----
Q03585	T. acidophilum	A'	Multa	-EGTN	-----	-----	-----	-----	-----

*Method:
 BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
 PSIPRED and PHD = Structure Prediction
 Red = Helix Green = Strand Blue = Turn Gray = No coörd

Figure C.22 Sequence alignment for *P. furiosus* RNAP subunit A' (sheet 5)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	170																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														</
-------------	--------	----	---------	-----	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	----

*Method:
BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
PSIPRED and PHD = Structure Prediction
Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.23 Sequence alignment for *P. furiosus* RNAP subunit A' (sheet 6)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	-----180-----190-----200-----
Q8U0M4	<i>P. furiosus</i>	A'	RMSD	-----H RLMPT-E-VR DWLEKIPDKD LPLLGLDPEK
Q8U0M4	<i>P. furiosus</i>	A'	PSIPRED	-----H RLMPT-E-VR DWLEKIPDKD LPLLGLDPEK
Q8U0M4	<i>P. furiosus</i>	A'	PHD	-----H RLMPT-E-VR DWLEKIPDKD LPLLGLDPEK
P04050	<i>S. cerevisiae</i>	1	RMSD	-----L RVLSTEE-IL NIFKHISVKD FTSLGFNEVF
P04050	<i>S. cerevisiae</i>	1	PSIPRED	-----L RVLSTEE-IL NIFKHISVKD FTSLGFNEVF
P04050	<i>S. cerevisiae</i>	1	PHD	-----L RVLSTEE-IL NIFKHISVKD FTSLGFNEVF
P04050	<i>S. cerevisiae</i>	1	MOE	-----L RVLSTEE-IL NIFKHISVKD FTSLGFNEVF
P04050	<i>S. cerevisiae</i>	1	BLAST	-----L RVLSTEE-IL NIFKHISVKD FTSLGFNEVF
P04050	<i>S. cerevisiae</i>	1	BLAST2	-----L RVLSTEE-IL NIFKHISVKD FTSLGFNEVF
P04050	<i>S. cerevisiae</i>	1	MPsrch	-----L RVLSTEE-IL NIFKHISVKD FTSLGFNEVF
P04050	<i>S. cerevisiae</i>	1	Multa	KDRATGDADE-----PELR VLSTE-E-IL NIFKHISVKD FTSLGFNEVF
P08775	<i>M. musculus</i>	1	Multa	-HVNEDSQ-------EKKI LLSPE-R-VH EIFKRISDEE CFVLGMEPRY
P24928	<i>H. sapiens</i>	1	Multa	-HVNEDSQ-------EKKI LLSPE-R-VH EIFKRISDEE CFVLGMEPRY
P36594	<i>S. pombe</i>	1	Multa	--RGKDESDL-----PEKR LLSPL-E-VH TIFTHISSED LAHLGLNEQY
O35559	<i>C. griseus</i>		Multa	-HVNEDSQ-------EKKI LLSPE-R-VH EIFKRISDEE CFVLGMEPRY
Q9AVW8	<i>G. theta</i>	1	Multa	SRITS-----EPTR LVSAE-R-IH EIFVKITDEA CEKLGILNPKL
Q9AXN0	<i>G. vacuolata</i>	1	Multa	GSADDTM-----EKKQ IVSAE-K-VL SIFKHISDDD IEIMGLSPQH
Q96452	<i>N. locustae</i>	1	Multa	GEENN-----EGKV MNGE-R-VY NIFKKINNED LEFLGFINIKY
P18616	<i>A. thaliana</i>	1	Multa	IQRKKNDPEP QLPEPAERKQ TLGAD-R-VL SVLKRI SDAD CQLLGFPNPKF
Q25312	<i>L. major</i>		Multa	IKVTQE-----EQDY VWHAE-N-AR QVLDRVSDSD ALIMGFDRRF
Q9XZS1	<i>D. guanche</i>		Multa	-HPNEDSQ-------EKKI VVSAE-R-VW EILKHITDEE CFILGMDPKY
Q9BJ59	<i>L. seymouri</i>		Multa	IKVTQE-----EQDY VWHAD-N-AR QVLDRVSDSD ATIMGFDRRF
P16356	<i>C. elegans</i>		Multa	KNVNEDTQ-----ERKI MLTAE-R-VL EVFQITDED ILVIGMDPQF
P17545	<i>T. brucei brucei</i>		Multa	IKAVHE-----EQDA EWHAD-K-VR QVLDRVSDSD ARLMGFDPQR
P14248	<i>P. falciparum</i>		Multa	HSSEEDID-----ESKR KLSAE-E-AL EILKKIRKEE MSILGFNSDR
P04052	<i>D. melanogaster</i>		Multa	-HQNEDSQ-------EKKI VVSAE-R-VW EILKHITDEE CFILGMDPKY
Q980R2	<i>S. solfataricus</i>	A'	Multa	-----A KLTPS-E-IR ERLEKIPDSD VEILGYDPTT
Q979F4	<i>T. volcanium</i>	A'	Multa	-----V KITPK-E-IR ERLEKIPDDD LVFFGFNPKT
P31813	<i>T. celer</i>	A'	Multa	-----H RLMPT-E-VR DRLEKIPDRD LPLLGLDAEK
Q9V114	<i>P. abyssi</i>	A'	Multa	-----H RLMPT-E-IR DRLEKIPDKD LPLLGLHPEK
Q8ZYQ8	<i>P. aerophilum</i>	A'	Multa	-----V KLDPE-V-LR ERLSKVPSED LELLGINPSV
Q58445	<i>M. jannaschii</i>	A'	Multa	GN-----EEK TLTPS-D-VR EILEKIPDED CILLGINPEV
P11512	<i>S. acidocaldarius</i>	A'	Multa	-----V KLSPS-E-IR DRLEKIPDSD VELLGYDPKS
O27125	<i>M. thermoautotrophicum</i>	A'	Multa	-----Y KLTPS-E-VR ERLEKISDDD ALILGVNPEV
P09846	<i>M. thermoautotrophicum</i> (strain Winter)	A'	Multa	-----Y KLTPS-E-VR ERLEKIDND SLLLGVPNEV
Q28391	<i>A. fulgidus</i>	A'	Multa	-----H RLTPK-D-VR EWLEKIPDDD LPAFGINPKA
P15350	<i>H. halobium</i>	A'	Multa	EDREAIETAI GADLTTEDMN KLMPD-D-IR DWFEDIPGED LEALGVNSDR
P41556	<i>M. vannielii</i>	A'	Multa	-----K QLTSS-E-VR EILEKIPNED CKLLGINSRV
Q03585	<i>T. acidophilum</i>	A'	Multa	-----V KITPK-E-IR ERLEKIPDDD LIFFGFNPKT

*Method:
 BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
 PSIPRED and PHD = Structure Prediction
 Red = Helix Green = Strand Blue = Turn Gray = No cord

Figure C.24 Sequence alignment for *P. furiosus* RNAP subunit A' (sheet 7)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	---210---	---220---	---230---	---240---	---250---
Q8U0M4	<i>P. furiosus</i>	A'	RMSD	SRPEWMVLTV	LPVPPVTARP	SITLETGIRA	EDDLTHKLVD	IIRINNRLKQ
Q8U0M4	<i>P. furiosus</i>	A'	PSIPRED	SRPEWMVLTV	LPVPPVTARP	SITLETGIRA	EDDLTHKLVD	IIRINNRLKQ
Q8U0M4	<i>P. furiosus</i>	A'	PHD	SRPEWMVLTV	LPVPPVTARP	SITLETGIRA	EDDLTHKLVD	IIRINNRLKQ
P04050	<i>S. cerevisiae</i>	1	RMSD	SRPEWMI LTC	LPVPPPVVRP	SISFNESQRG	EDDLTFKLAD	ILKANISLET
P04050	<i>S. cerevisiae</i>	1	PSIPRED	SRPEWMI LTC	LPVPPPVVRP	SISFNESQRG	EDDLTFKLAD	ILKANISLET
P04050	<i>S. cerevisiae</i>	1	PHD	SRPEWMI LTC	LPVPPPVVRP	SISFNESQRG	EDDLTFKLAD	ILKANISLET
P04050	<i>S. cerevisiae</i>	1	MOE	SRPEWMI LTC	LPVPPPVVRP	SISFNESQRG	EDDLTFKLAD	ILKANISLET
P04050	<i>S. cerevisiae</i>	1	BLAST	SRPEWMI LTC	LPVPPPVVRP	SISFNESQRG	EDDLTFKLAD	ILKANISLET
P04050	<i>S. cerevisiae</i>	1	BLAST2	SRPEWMI LTC	LPVPPPVVRP	SISFNESQRG	EDDLTFKLAD	ILKANISLET
P04050	<i>S. cerevisiae</i>	1	MPsrch	SRPEWMI LTC	LPVPPPVVRP	SISFNESQRG	EDDLTFKLAD	ILKANISL-E
P04050	<i>S. cerevisiae</i>	1	Multa	SRPEWMI LTC	LPVPPPVVRP	SISFNESQRG	EDDLTFKLAD	ILKANISLET
P08775	<i>M. musculus</i>	1	Multa	ARPEWMI VTV	LPVPPPLSVRP	AVVMQGSARN	QDDLTHKLAD	IVKINNQLRR
P24928	<i>H. sapiens</i>	1	Multa	ARPEWMI VTV	LPVPPPLSVRP	AVVMQGSARN	QDDLTHKLAD	IVKINNQLRR
P36594	<i>S. pombe</i>	1	Multa	ARPDWMI LTV	LPVPPFVSVP	SISVDGTSRG	EDDLTHKLSD	IIKANANVR
O35559	<i>C. griseus</i>		Multa	ARPEWMI VTV	LPVPPPLSVRP	AVVMQGSARN	QDDLTHKLAD	IVKINNQLRR
Q9AVW8	<i>G. theta</i>	1	Multa	SRPDWMI LTV	LPVPPPTIRP	SVKFETLLRA	EDDLTYKLGD	IIRTNKI LRK
Q9AXN0	<i>G. vacuolata</i>	1	Multa	SRPEWFI LTV	FPVPPFHIRP	SVMMDASMRG	EDDLTYKLGD	IVRNNALRE
Q96452	<i>N. locustae</i>	1	Multa	SRPEYMI LSV	LLVAPPVVRP	SIVMEGTLRG	EDDLTYKYSD	IIKSNQNLKK
P18616	<i>A. thaliana</i>	1	Multa	ARPDWMI LEV	LPVPPFVSVP	SVMMDATSR	EDDLTHQLAM	IIRHNENLKR
Q25312	<i>L. major</i>		Multa	CHPRDLDLTV	LPVPPFQVRS	SIYL-GGLMQ	VDDLTHQIMS	IVKRNIQLRK
Q9XZS1	<i>D. guanche</i>		Multa	ARPDWMI VTV	LPVPPPLAVRP	AVVMFGAAKN	QDDLTHKLSD	IIKANNE LRK
Q9BJ59	<i>L. seymouri</i>		Multa	CHPRDLI LTV	VPVPPFQVVRP	AVAF-GSAKS	DDELTHQIMS	IVKREHQLRK
P16356	<i>C. elegans</i>		Multa	ARPEWMI CTV	LPVPPPLAVRP	AVVTFGSAKN	QDDLTHKLSD	IIKTNQQLQR
P17545	<i>T. brucei brucei</i>		Multa	CHPRDLV LTV	LPVPPFQVVRP	AISF-GGLRS	DDELTHQIMS	IVKRNQNLRR
P14248	<i>P. falciparum</i>		Multa	CVPASLI LTV	LPVPPFQVVRP	YV-QYGNQRS	EDDLTLKLDD	IVKTNQLKLR
P04052	<i>D. melanogaster</i>		Multa	ARPDWMI VTV	LPVPPPLAVRP	AVVMFGAAKN	QDDLTHKLSD	IIKANNE LRK
Q980R2	<i>S. solfataricus</i>	A'	Multa	SRPEWMI LTV	LPVPPITIRP	SIMIESGIRA	EDDLTHKLVD	IVRINER LKE
Q979F4	<i>T. volcanium</i>	A'	Multa	ARPEWMI LTV	LPVPPINVRP	SITLETGERS	EDDLTHKLVD	IIRISQRLRE
P31813	<i>T. celer</i>	A'	Multa	ARPEWMI LTV	LPVPPVTMRP	SITLESIGIRA	EDDLTHKLVD	IIRINNRLKQ
Q9V114	<i>P. abyssi</i>	A'	Multa	SRPEWMI LTV	LPVPPVTARP	SITLETGIRA	EDDLTHKLVD	IIRINNRLKQ
Q8ZYQ8	<i>P. aerophilum</i>	A'	Multa	ARPEWMI LTV	LPVPPPHVRP	SIQLETGIRS	EDDLTHKLVD	IIRNNEKLKI
Q58445	<i>M. jannaschii</i>	A'	Multa	ARPEWMI LTV	LPVPPVTVRP	SITLETGERS	EDDLTHKLVD	IIRINNRLKE
P11512	<i>S. acidocaldarius</i>	A'	Multa	SRPEWMI LTV	LPVPPITIRP	SITLESIGIRA	EDDLTHKLVD	IIRINER LKE
O27125	<i>M. thermoautotrophicum</i>	A'	Multa	ARPEWMI LTV	LPVPPVTVRP	SITLETGERS	EDDLTHKLVD	ILIRINQLKE
P09846	<i>M. thermoautotrophicum</i> (strain Winter)	A'	Multa	ARPEWMI LTV	LPVPPVTVRP	SITLETGERS	EDDLTHKLVD	ILIRINQLKE
O28391	<i>A. fulgidus</i>	A'	Multa	TRPEWFI LTV	LPVPPVTVRP	SITLETGERS	EDDLTHKLVD	IIRINQRFME
P15350	<i>H. halobium</i>	A'	Multa	SRPEWMI LTV	LPVPPVTARP	SITLDNGQRS	EDDLTHKLVD	IIRINQRFME
P41556	<i>M. vannielii</i>	A'	Multa	ARPEWMI LTV	LPVPPVTVRP	SITLESIGERS	EDDLTHKLVD	IIRINQRLKE
Q03585	<i>T. acidophilum</i>	A'	Multa	ARPEWMI LTV	LPVPPINVRP	SITLETGERS	EDDLTHKLVD	IIRISQRLRE

*Method:

BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD

PSIPRED and PHD = Structure Prediction

Red = Helix Green = Strand Blue = Turn Gray = No co-ord

Figure C.25 Sequence alignment for *P. furiosus* RNAP subunit A' (sheet 8)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	---,260---	---270---	---280---	---290---	---300---
Q8U0M4	P. furiosus	A'	RMSD	NIE-AGAPQL	IIEDLWDLQ	YHVTTYINNE	APGVPPAKHK	SGRPLKTLAQ
Q8U0M4	P. furiosus	A'	PSIPRED	NIE-AGAPQL	IIEDLWDLQ	YHVTTYINNE	APGVPPAKHK	SGRPLKTLAQ
Q8U0M4	P. furiosus	A'	PHD	NIE-AGAPQL	IIEDLWDLQ	YHVTTYINNE	APGVPPAKHK	SGRPLKTLAQ
P04050	S. cerevisiae	1	RMSD	LEH-NGAPHH	AIEEAESLLQ	FHVATYMDND	IAGQPQALQK	SGRPVKKSIRA
P04050	S. cerevisiae	1	PSIPRED	LEH-NGAPHH	AIEEAESLLQ	FHVATYMDND	IAGQPQALQK	SGRPVKKSIRA
P04050	S. cerevisiae	1	PHD	LEH-NGAPHH	AIEEAESLLQ	FHVATYMDND	IAGQPQALQK	SGRPVKKSIRA
P04050	S. cerevisiae	1	MOE	LEH-NGAPHH	AIEEAESLLQ	FHVATYMDND	IAGQPQALQK	SGRPVKKSIRA
P04050	S. cerevisiae	1	BLAST	LEH-NGAPHH	AIEEAESLLQ	FHVATYMDND	IAGQPQALQK	SGRPVKKSIRA
P04050	S. cerevisiae	1	BLAST2	LEH-NGAPHH	AIEEAESLLQ	FHVATYMDND	IAGQPQALQK	SGRPVKKSIRA
P04050	S. cerevisiae	1	MPsrch	TLEHNGAPHH	AIEEAESLLQ	FHVATYMDND	IAGQPQALQK	SGRPVKKSIRA
P04050	S. cerevisiae	1	Multa	LEH-NGAPHH	AIEEAESLLQ	FHVATYMDND	IAGQPQALQK	SGRPVKKSIRA
P08775	M. musculus	1	Multa	NEQ-NGAAAH	VIAEDVKLLQ	FHVATMVDNE	LPGLPRAMQK	SGRPLKSLKQ
P24928	H. sapiens	1	Multa	NEQ-NGAAAH	VIAEDVKLLQ	FHVATMVDNE	LPGLPRAMQK	SGRPLKSLKQ
P36594	S. pombe	1	Multa	CEQ-EGAPAH	IVSEYEQLLQ	FHVATYMDNE	IAGQPQALQK	SGRPLKKSIRA
O35559	C. griseus		Multa	NEQ-NGAAAH	VIAEDVKLLQ	FHVATMVDNE	LPGLPRAMQK	SGRPLKSLKQ
Q9AVW8	G. theta	1	Multa	AIF-SCLPSQ	LINEQINLLQ	YHIGSYMNS	VPSLAKSLQK	SGRPLKKSIRA
Q9AXN0	G. vacuolata	1	Multa	MER-TGAPAH	RINEQIHVLQ	AHIITIMND	LPGMFRAHQK	SGRPLKKSIRA
Q96452	N. locustae	1	Multa	YEE-EGAPAH	IIRDYEQLVQ	FHVATLIDNE	ISGQPQALQK	NGRPLKAISA
P18616	A. thaliana	1	Multa	QEK-NGAPAH	IISEFTQLLQ	FHIATYFDNE	LPQQPRATQK	SGRPLKKSIRA
Q25312	L. major		Multa	DKE-SGVKA-	AVDRSRALLQ	EHVATFFNNA	STYYKPAKVG	DTKKLKSLTE
Q9XZS1	D. guanche		Multa	NEA-SGAAAH	VIQENIKMLQ	FHVATLVDND	MPGMFRAHQK	SGKPLKAISA
Q9BJ59	L. seymouri		Multa	DKE-SGVKA-	AVDRSRALLQ	EHVATFFNNA	STYYKPAKVG	DTKKLKSLTE
P16356	C. elegans		Multa	NEA-NGAAAH	VLTDDVRLQ	FHVATLVDNC	IPGLPTATQK	GGRPLKKSIRA
P17545	T. brucei brucei		Multa	DKE-SDVQA-	AIDRSRALLQ	EHVATYFNNA	STYYKPTKVN	DTKKLKSLTE
P14248	P. falciparum		Multa	QTD-RGAKSH	VLQDLCSLLQ	FHITTLFDND	IPGMPIATTR	SKKPIKAIRT
P04052	D. melanogaster		Multa	NEA-SGAAAH	VIQENIKMLQ	FHVATLVDND	MPGMFRAHQK	SGKPLKAISA
Q980R2	S. solfataricus	A'	Multa	SID-AGAPQL	IIEDLWDLQ	YHVATYFDNE	IPGLPPSKHR	SGRPLKTLAQ
Q979F4	T. volcanium	A'	Multa	SRD-NGSPQL	IIEDLWDLQ	FHVTTYFDNQ	TPGIPPARHR	SGRPLKTLAQ
P31813	T. celer	A'	Multa	NIE-AGAPQL	IIEDLWDLQ	YHVTTYINNE	TSGIPPAKHK	SGRPLKTLAQ
Q9V114	P. abyssi	A'	Multa	NIE-AGAPQL	IIEDLWDLQ	YHVTTYINNE	TSGVPPAKHK	SGRPLKTLAQ
Q8ZYQ8	P. aerophilum	A'	Multa	AIE-TGAPTN	VVDNLWELLQ	YHVATYFDNE	LPGLPVAKHK	GGRPLKGIQA
Q58445	M. jannaschii	A'	Multa	NIE-GGAPNL	IIEDLWNLQ	YHVNTYFDNE	APGIPPAKHK	SGRPLKTLAQ
P11512	S. acidocaldarius	A'	Multa	SIE-AGAPQL	IIEDLWDLQ	YHVATYFDNE	IPGLPPAKHK	SGRPLKTLAQ
O27125	M. thermoautotrophicum	A'	Multa	NME-AGAPQL	IIVEDLWELLQ	YHVTTYFDNE	ASGVPPARHR	SGRPLKTLAQ
P09846	M. thermoautotrophicum (strain Winter)	A'	Multa	NME-AGAPQL	IIVEDLWELLQ	YHVTTYFDNE	ASGVPPARHR	SGRPLKTLAQ
O28391	A. fulgidus	A'	Multa	NKE-AGAPQQ	IIVEDLWELLQ	YHVTTYFDNE	VSGIPPARHR	SGRPLKTLAQ
P15350	H. halobium	A'	Multa	NRE-AGAPQL	IIVEDLWELLQ	YHVTTYFDNE	ISGTPPARHR	SGRPLKTLAQ
P41556	M. vannielii	A'	Multa	NIN-GGAPNL	IIEDLWDLQ	YHINTYFDNE	APGIPPARHR	SGRPLKTLAQ
Q03585	T. acidophilum	A'	Multa	SRD-NGSPQL	IIEDLWDLQ	FHVTTYFDNQ	TPGIPPARHR	SGRPLKTLAQ

*Method:

BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD

PSIPRED and PHD = Structure Prediction

Red = Helix Green = Strand Blue = Turn Gray = No co-ord

Figure C.26 Sequence alignment for *P. furiosus* RNAP subunit A' (sheet 9)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	----310----	----320----	----330----	----340----	----350----
Q8U0M4	P. furiosus	A'	RMSD	RLKGKEGRFR	GNLSGKRVNF	SARTVISFDP	MISINEVGVP	VEVAMELTVP
Q8U0M4	P. furiosus	A'	PSIPRED	RLKGKEGRFR	GNLSGKRVNF	SARTVISFDP	MISINEVGVP	VEVAMELTVP
Q8U0M4	P. furiosus	A'	PHD	RLKGKEGRFR	GNLSGKRVNF	SARTVISFDP	MISINEVGVP	VEVAMELTVP
P04050	S. cerevisiae	1	RMSD	RLKGKEGRIR	GNLMGKRVDF	SARTVISGDP	NLELDQVGVP	KSIAKTLTYP
P04050	S. cerevisiae	1	PSIPRED	RLKGKEGRIR	GNLMGKRVDF	SARTVISGDP	NLELDQVGVP	KSIAKTLTYP
P04050	S. cerevisiae	1	PHD	RLKGKEGRIR	GNLMGKRVDF	SARTVISGDP	NLELDQVGVP	KSIAKTLTYP
P04050	S. cerevisiae	1	MOE	RLKGKEGRIR	GNLMGKRVDF	SARTVISGDP	NLELDQVGVP	KSIAKTLTYP
P04050	S. cerevisiae	1	BLAST	RLKGKEGRIR	GNLMGKRVDF	SARTVISGDP	NLELDQVGVP	KSIAKTLTYP
P04050	S. cerevisiae	1	BLAST2	RLKGKEGRIR	GNLMGKRVDF	SARTVISGDP	NLELDQVGVP	KSIAKTLTYP
P04050	S. cerevisiae	1	MPsrch	RLKGKEGRIR	GNLMGKRVDF	SARTVISGDP	NLELDQVGVP	KSIAKTLTYP
P04050	S. cerevisiae	1	Multa	RLKGKEGRIR	GNLMGKRVDF	SARTVISGDP	NLELDQVGVP	KSIAKTLTYP
P08775	M. musculus	1	Multa	RLKGKEGRVR	GNLMGKRVDF	SARTVITPDP	NLSIDQVGVP	RSIAANMTFA
P24928	H. sapiens	1	Multa	RLKGKEGRVR	GNLMGKRVDF	SARTVITPDP	NLSIDQVGVP	RSIAANMTFA
P36594	S. pombe	1	Multa	RLKGKEGRLR	GNLMGKRVDF	SARTVITGDP	NLSLDELGVF	RSIAKTLTYP
O35559	C. griseus		Multa	RLKGKEGRVR	GNLMGKRVDF	SARTVITPDP	NLSIDQVGVP	RSIAANMTFA
Q9AVW8	G. theta	1	Multa	RLQGKEGR LR	GNLMGKRVDF	SARTVITPDP	NKLDELGVF	LSIAMNLTFF
Q9AXN0	G. vacuolata	1	Multa	RLKGKEGRVR	GNLMGKRVDF	SGRTVISFDP	NLRLDQVGVP	QTIAMNLTYP
Q96452	N. locustae	1	Multa	RLKGKEGR LR	GNLMGKRVDF	SARSVITADP	NISLEELGVF	MKVAMIHTFF
P18616	A. thaliana	1	Multa	RLKAKEGRIR	GNLMGKRVDF	SARTVITPDP	TINIDELGVF	WSIALNLTYP
Q25312	L. major		Multa	RLKGKYGR LR	GNLMGKRVDF	SARTVITGDP	NIDVDEGVF	FSVAMTLTFF
Q9XZS1	D. guanche		Multa	RLKGKEGRIR	GNLMGKRVDF	SARTVITPDP	NLRIDQVGVP	RSIAQNLTFF
Q9BJ59	L. seymouri		Multa	RLKGKYGR LR	GNLMGKRVDF	SARTVITGDP	NIDVDEGVF	FSIAMTLTFF
P16356	C. elegans		Multa	RLKGKEGRIR	GNLMGKRVDF	SARTVITADP	NLPIDTVGVF	RTIAQNLTFF
P17545	T. brucei brucei		Multa	RLKGKYGR LR	GNLMGKRVDF	SARTVITGDP	NIDVDEGVF	FSVAMTLTFF
P14248	P. falciparum		Multa	RLKGKEGR LR	GNLMGKRVDF	SARTVITGDP	NLNIDYIGVP	KSVAMTLTFC
P04052	D. melanogaster		Multa	RLKGKEGRIR	GNLMGKRVDF	SARTVITPDP	NLRIDQVGVP	RSIAQNLTFF
Q980R2	S. solfataricus	A'	Multa	RLKGKEGRFR	GNLSGKRVNF	SSRTVISFDP	NISIDEVGVP	EIIAKTLTVP
Q979F4	T. volcanium	A'	Multa	RLKGKEGRFR	GNLSGKRVNF	SSRTVISPEP	YLSVNEVGVP	EKAARELTVP
P31813	T. celer	A'	Multa	RLKGKEGRFR	GNLSGKRVNF	SARTVISFDP	MISINEVGVP	MAIAMELTVP
Q9V114	P. abyssi	A'	Multa	RLKGKEGRFR	GNLSGKRVNF	SARTVISFDP	MISINEVGVP	IQIAMELTVP
Q8ZYQ8	P. aerophilum	A'	Multa	RLKGKEGRFR	GNLSGKRVNF	SARTVISFDP	HISINEVGVP	YDIKAVLTVP
Q58445	M. jannaschii	A'	Multa	RLKGKEGRFR	GNLSGKRVNF	SSRTVISFDP	CLSINEVGVP	EVVAKELTVP
P11512	S. acidocaldarius	A'	Multa	RLKGKEGRFR	GNLSGKRVNF	SARTVISFDP	NLSIDEVGVP	YTIARMLTVP
O27125	M. thermoautotrophicum	A'	Multa	RLKGKEGRFR	GNLSGKRVNF	SARTVISFDP	NISINEVGVP	EIIAREVTVP
P09846	M. thermoautotrophicum (strain Winter)	A'	Multa	RLKGKEGRFR	GNLSGKRVNF	SARTVISFDP	NVSNEVGVP	ELIAKEVTVP
Q28391	A. fulgidus	A'	Multa	RLKGKEGRFR	GNLSGKRVNF	SARTVISFDP	CLSINEVGVP	REIAEELTVP
P15350	H. halobium	A'	Multa	RLKGKEGRFR	GNLSGKRVNF	SARTVISFDP	TLSLNEVGVP	DRVATEMTQT
P41556	M. vanielii	A'	Multa	RLKGKEGRFR	GNLSGKRVNF	SARTVISFDP	RLSINEVGVP	ELIAKELTVP
Q03585	T. acidophilum	A'	Multa	RLKGKEGRFR	GNLSGKRVNF	SSRTVISPEP	YLSVNEVGVP	ERAARELTVP

*Method:

BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD

PSIPRED and PHD = Structure Prediction

Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.27 Sequence alignment for *P. furiosus* RNAP subunit A' (sheet 10)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	-----360----	-----370----	-----380-----	-----390-----
Q8U0M4	<i>P. furiosus</i>	A'	RMSD	EKVTEFNIER	LKKMVLNGPD	KYP-----	GANYVIDP-- --EGR---R
Q8U0M4	<i>P. furiosus</i>	A'	PSIPRED	EKVTEFNIER	LKKMVLNGPD	KYP-----	GANYVIDP-- --EGR---R
Q8U0M4	<i>P. furiosus</i>	A'	PHD	EKVTEFNIER	LKKMVLNGPD	KYP-----	GANYVIDP-- --EGR---R
P04050	<i>S. cerevisiae</i>	1	RMSD	EVVTPYNIDR	LTQLVRNGPN	EHP-----	GAKYVIRD-- --SGD---R
P04050	<i>S. cerevisiae</i>	1	PSIPRED	EVVTPYNIDR	LTQLVRNGPN	EHP-----	GAKYVIRD-- --SGD---R
P04050	<i>S. cerevisiae</i>	1	PHD	EVVTPYNIDR	LTQLVRNGPN	EHP-----	GAKYVIRD-- --SGD---R
P04050	<i>S. cerevisiae</i>	1	MOE	EVVTPYNIDR	LTQLVRNGPN	EHP-----	GAKYVIRD-- --SGD---R
P04050	<i>S. cerevisiae</i>	1	BLAST	EVVTPYNIDR	LTQLVRNGPN	EHP-----	GAKYVIRD-- --SGD---R
P04050	<i>S. cerevisiae</i>	1	BLAST2	EVVTPYNIDR	LTQLVRNGPN	EHP-----	GAKYVIRD-- --SGD---R
P04050	<i>S. cerevisiae</i>	1	MPsrch	EVVTPYNIDR	LTQLVRNGPN	EHP-----	GAKYVIRD-- --SGD---R
P04050	<i>S. cerevisiae</i>	1	Multa	EVVTPYNIDR	LTQLVRNGPN	EHPG-----	A-KYVIRD-- --SGD---R
P08775	<i>M. musculus</i>	1	Multa	EIVTPFNIDR	LQELVRRGNS	QYPG-----	A-KYIIRD-- --NGD---R
P24928	<i>H. sapiens</i>	1	Multa	EIVTPFNIDR	LQELVRRGNS	QYPG-----	A-KYIIRD-- --NGD---R
P36594	<i>S. pombe</i>	1	Multa	ETVTPYNIYQ	LQELVRNGPD	EHPG-----	A-KYIIRD-- --TGE---R
O35559	<i>C. griseus</i>		Multa	EIVTPFNIDR	LQELVRRGNS	QYPG-----	A-KYIIRD-- --NGD---R
Q9AVW8	<i>G. theta</i>	1	Multa	EIVTSFNIEY	MKFLLYNGPY	KHPG-----	A-KYLIRN-- --DNN---R
Q9AXN0	<i>G. vacuolata</i>	1	Multa	EVVTPFNVEK	MRRLVMNGPN	EYPG-----	A-KYIERL-- --DGS---K
Q96452	<i>N. locustae</i>	1	Multa	ERVTSFNLNK	LRLVLSNGPN	KYPG-----	A-NYLIRR-- --DGQ---R
P18616	<i>A. thaliana</i>	1	Multa	ETVTPYNIER	LKELVDYGP	PPPGKT---G	A-KYIIRD-- --DGQ---R
Q25312	<i>L. major</i>		Multa	ERVNVINKKR	LTEFVQR--T	TYPG-----	A-NYIIRP-- --NGN---V
Q9XZS1	<i>D. guanche</i>		Multa	ELVTPFNIDR	MQELVRRGNS	QYPG-----	A-KYIVRD-- --NGE---R
Q9BJ59	<i>L. seymouri</i>		Multa	ERVNVINKKR	LTEFVQR--S	TYPG-----	A-NYIMRP-- --NGN---V
P16356	<i>C. elegans</i>		Multa	EIVTPFNVDK	LQELVNRGDT	QYPG-----	A-K-----E-- --NGA---R
P17545	<i>T. brucei brucei</i>		Multa	ERVNTINKKR	LTEFARR--T	VYPG-----	A-NYIHHP-- --NGT---I
P14248	<i>P. falciparum</i>		Multa	ETVTPYNIYD	LKKLVERGPY	EWPG-----	A-KYIIRD-- --NGT---K
P04052	<i>D. melanogaster</i>		Multa	ELVTPFNIDR	MQELVRRGNS	QYPG-----	A-KYIVRD-- --NGE---R
Q980R2	<i>S. solfataricus</i>	A'	Multa	ERITPWNIEK	LQFVINGPD	KWP-----	GANYVIRP-- --DGR---R
Q979F4	<i>T. volcanium</i>	A'	Multa	VIVNQFNIDE	MREFIKRGRN	PRDKYGYMA	GVNYVIRP-- --DGR---R
P31813	<i>T. celer</i>	A'	Multa	EKVTEFNIEK	LKKMVLNGPE	KYP-----	GANYVIDP-- --EGR---R
Q9V114	<i>P. abyssi</i>	A'	Multa	EKVTEFNIEK	LKKMVLNGPD	KYP-----	GANYVIDP-- --EGR---R
Q8ZYQ8	<i>P. aerophilum</i>	A'	Multa	ERVTAWNVDV	LREYVIRGPE	SWP-----	GANYVVP-- --EGR---R
Q58445	<i>M. jannaschii</i>	A'	Multa	EKVTKYNIER	IRQLLRNGSE	KHPGVN---Y	VIRKMIGR-- --DGT---E
P11512	<i>S. acidocaldarius</i>	A'	Multa	ERVTNINIER	IRQYIINGPD	KWP-----	GANYVIRP-- --DGR---R
O27125	<i>M. thermoautotrophicum</i>	A'	Multa	VYVTEWNIDR	MREYIENGPD	VHP-----	GANYVIRP-- --DGR---K
P09846	<i>M. thermoautotrophicum (strain Winter)</i>	A'	Multa	VYVTEWNIDR	MKEHIENGPD	VHP-----	GANYVIRP-- --DGR---K
O28391	<i>A. fulgidus</i>	A'	Multa	IYVTPQNIDM	AREFVLRS--	EHP-----	KANYVIRP-- --DGR---R
P15350	<i>H. halobium</i>	A'	Multa	MVVNEQNLER	ARRYVRNGPE	GHP-----	GANYVTRP-- --DGR---R
P41556	<i>M. vannielii</i>	A'	Multa	EKVTPYNIER	IRKLIENGSD	KHP-----	GVNYVKKVK TKDGKEEYK
Q03585	<i>T. acidophilum</i>	A'	Multa	VIVNQFNIDE	MRELIKGRN	PRDQFGRYVT	GVNYVIRP-- --DGR---R

*Method:
 BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
 PSIPRED and PHD = Structure Prediction
 Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.28 Sequence alignment for *P. furiosus* RNAP subunit A' (sheet 11)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	390,-----	--,,,400--	---410----	---420----	---430----
Q8U0M4	<i>P. furiosus</i>	A'	RMSD	RR1-MDSNKE	TL---ANQLD	IGWTVERHLM	DGDIVLFNRQ	PSLHRMSIMA
Q8U0M4	<i>P. furiosus</i>	A'	PSIPRED	RR1-MDSNKE	TL---ANQLD	IGWTVERHLM	DGDIVLFNRQ	PSLHRMSIMA
Q8U0M4	<i>P. furiosus</i>	A'	PHD	RR1-MDSNKE	TL---ANQLD	IGWTVERHLM	DGDIVLFNRQ	PSLHRMSIMA
P04050	<i>S. cerevisiae</i>	1	RMSD	IDL-RYS-KR	AG---DIQLQ	YGWVERHIM	DNDPVLFNRQ	PSLHKMSMMA
P04050	<i>S. cerevisiae</i>	1	PSIPRED	IDL-RYS-KR	AG---DIQLQ	YGWVERHIM	DNDPVLFNRQ	PSLHKMSMMA
P04050	<i>S. cerevisiae</i>	1	PHD	IDL-RYS-KR	AG---DIQLQ	YGWVERHIM	DNDPVLFNRQ	PSLHKMSMMA
P04050	<i>S. cerevisiae</i>	1	MOE	IDL-RYS-KR	AG---DIQLQ	YGWVERHIM	DNDPVLFNRQ	PSLHKMSMMA
P04050	<i>S. cerevisiae</i>	1	BLAST	IDL-RYS-KR	AG---DIQLQ	YGWVERHIM	DNDPVLFNRQ	PSLHKMSMMA
P04050	<i>S. cerevisiae</i>	1	BLAST2	IDL-RYS-KR	AG---DIQLQ	YGWVERHIM	DNDPVLFNRQ	PSLHKMSMMA
P04050	<i>S. cerevisiae</i>	1	MPsrch	IDL-RYSKRA	G----DIQLQ	YGWVERHIM	DNDPVLFNRQ	PSLHKMSMMA
P04050	<i>S. cerevisiae</i>	1	Multa	IDLRYSKRAG	DI-----QLQ	YGWVERHIM	DNDPVLFNRQ	PSLHKMSMMA
P08775	<i>M. musculus</i>	1	Multa	IDLRFPKPS	DL-----HLQ	TGYKVERHMC	DGDIVIFNRQ	PTLHKMSMMG
P24928	<i>H. sapiens</i>	1	Multa	IDLRFPKPS	DL-----HLQ	TGYKVERHMC	DGDIVIFNRQ	PTLHKMSMMG
P36594	<i>S. pombe</i>	1	Multa	IDLRYPKRA	DI-----PLR	YGRVVERHIR	DGDVVLFNQ	PSLHKMSMMG
O35559	<i>C. griseus</i>		Multa	IDLRFPKPS	DL-----HLQ	TGYKVERHMC	DGDIVIFNRQ	PTLHKMSMMG
Q9AVW8	<i>G. theta</i>	1	Multa	IDLRYPKLS	NI-----NLE	NGYCVVERHLQ	NGDLVLFNQ	PSLHKMSMMG
Q9AXN0	<i>G. vacuolata</i>	1	Multa	VNLAFVKNS	DI-----HLN	NGEKVIRHLL	DGDVVLFNQ	PSLHKMSMMG
Q96452	<i>N. locustae</i>	1	Multa	IDLNFSS--G	DV-----ILE	EGYVVERHML	NGDLVLFNQ	PSLHKMSMMA
P18616	<i>A. thaliana</i>	1	Multa	LDLRYLKKSS	DQ-----HLE	LGYKVERHLQ	DGDFVLFNQ	PSLHKMSMMG
Q25312	<i>L. major</i>		Multa	TKLALVKDRS	AV-----HLD	VGDVVERHVI	DGDVVLFNQ	PTLHRMSMMG
Q9XZS1	<i>D. guanche</i>		Multa	IDLRFPKSS	DL-----HLQ	CGYKVERHLR	DDDLVLFNQ	PTLHKMSMMG
Q9BJ59	<i>L. seymouri</i>		Multa	TKLALVKDRS	AV-----HLD	VGDVVERHVI	DGDVVLFNQ	PTLHRMSMMG
P16356	<i>C. elegans</i>		Multa	VDLRYPHRAA	DL-----HLQ	PGYRVERHMK	DGDIVIFNRQ	PTLHKMSMMG
P17545	<i>T. brucei brucei</i>		Multa	TKLALLRDRS	KV-----TLN	IGDVVERHVI	NGDVVLFNQ	PTLHRMSMMG
P14248	<i>P. falciparum</i>		Multa	YDLRHVRNS	EK-----ELE	YGYKVERHMT	DEDYILFNQ	PSLHKMSMMG
P04052	<i>D. melanogaster</i>		Multa	IDLRFPKSS	DL-----HLQ	CGYKVERHLR	DDDLVLFNQ	PTLHKMSMMG
Q980R2	<i>S. solfataricus</i>	A'	Multa	IDLRYPKDRK	EL---ASTLA	PGYIIRHLLI	DGDIVLFNRQ	PSLHRISMMMA
Q979F4	<i>T. volcanium</i>	A'	Multa	IKI-TDQNAE	EN---ANRIE	IGWTVERQLM	EGDIVLFNRQ	PSLHRMSMMG
P31813	<i>T. celer</i>	A'	Multa	IRL-MENNLE	TV---AEKLD	IGWTVERHLL	DGDVVLFNQ	PSLHRMSIMA
Q9V114	<i>P. abyssi</i>	A'	Multa	IRI-MESNRE	NL---ARMID	IGWTVERHLL	DGDIVLFNRQ	PSLHRMSIMA
Q8ZYQ8	<i>P. aerophilum</i>	A'	Multa	IDLRYPKDRK	AL---AERLA	PGWIVERHLLI	DGDIVLFNRQ	PSLHRVSMMA
Q58445	<i>M. jannaschii</i>	A'	Multa	QYKVKITES	NKDFWAENIR	EGDIVERHLM	DGDIVLYNRQ	PSLHRMSIMA
P11512	<i>S. acidocaldarius</i>	A'	Multa	IDLRYPKDRK	EL---ASSIT	AGYVVERHLV	DGDVVLFNQ	PSLHRISMMMA
O27125	<i>M. thermoautotrophicum</i>	A'	Multa	IRI-YNETKE	VV---LENLK	PGYIVERHLK	DGDIVLFNRQ	PSLHRMSMMA
P09846	<i>M. thermoautotrophicum</i> (strain Winter)	A'	Multa	IRA-YNETKD	VV---LENLK	PGYIVERHLK	DGDIVLFNRQ	PSLHRMSMMA
O28391	<i>A. fulgidus</i>	A'	Multa	IRV-IESNKE	EL---AEKLE	PGWIVERQLM	DGDIVLFNRQ	PSLHRMSIMA
P15350	<i>H. halobium</i>	A'	Multa	VRV-TEKVCE	EL---AERVE	PGWEVQRHLLI	DGDIIIFNRQ	PSLHRMSIMA
P41556	<i>M. vannielii</i>	A'	Multa	IKI-TDTNKK	MW---VENIV	DGMVVERHLG	DGDVVLVNRQ	PSLHRMSIMA
Q03585	<i>T. acidophilum</i>	A'	Multa	IKI-TDQNAE	EN---ADRID	IGWTVERQLM	EGDIVLFNRQ	PSLHRMSMMG

*Method:

BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD

PSIPRED and PHD = Structure Prediction

Red = Helix Green = Strand Blue = Turn Gray = No co-ord

Figure C.29 Sequence alignment for *P. furiosus* RNAP subunit A' (sheet 12)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	---440---	---450---	---460---	---470---	---480---
Q8U0M4	P. furiosus	A'	RMSD	HRVRVMPYRT	FRINLAVCPP	YNADFDGDEM	NLHVPQTTEE	QAEARI LMEV
Q8U0M4	P. furiosus	A'	PSIPRED	HRVRVMPYRT	FRINLAVCPP	YNADFDGDEM	NLHVPQTTEE	QAEARI LMEV
Q8U0M4	P. furiosus	A'	PHD	HRVRVMPYRT	FRINLAVCPP	YNADFDGDEM	NLHVPQTTEE	QAEARI LMEV
P04050	S. cerevisiae	1	RMSD	HRVKVIPYST	FRINLSVTSP	YNADFDGDEM	NLHVPQSEET	RAELSQLCAV
P04050	S. cerevisiae	1	PSIPRED	HRVKVIPYST	FRINLSVTSP	YNADFDGDEM	NLHVPQSEET	RAELSQLCAV
P04050	S. cerevisiae	1	PHD	HRVKVIPYST	FRINLSVTSP	YNADFDGDEM	NLHVPQSEET	RAELSQLCAV
P04050	S. cerevisiae	1	MOE	HRVKVIPYST	FRINLSVTSP	YNADFDGDEM	NLHVPQSEET	RAELSQLCAV
P04050	S. cerevisiae	1	BLAST	HRVKVIPYST	FRINLSVTSP	YNADFDGDEM	NLHVPQSEET	RAELSQLCAV
P04050	S. cerevisiae	1	BLAST2	HRVKVIPYST	FRINLSVTSP	YNADFDGDEM	NLHVPQSEET	RAELSQLCAV
P04050	S. cerevisiae	1	MPsrch	HRVKVIPYST	FRINLSVTSP	YNADFDGDEM	NLHVPQSEET	RAELSQLCAV
P04050	S. cerevisiae	1	Multa	HRVKVIPYST	FRINLSVTSP	YNADFDGDEM	NLHVPQSEET	RAELSQLCAV
P08775	M. musculus	1	Multa	HRVRI LPWST	FRINLSVTTP	YNADFDGDEM	NLHLPQSLET	RAEIQLAMV
P24928	H. sapiens	1	Multa	HRVRI LPWST	FRINLSVTTP	YNADFDGDEM	NLHLPQSLET	RAEIQLAMV
P36594	S. pombe	1	Multa	HRIRVMPYST	FRINLSVTSP	YNADFDGDEM	NMHVPQSEET	RAEQEITMV
O35559	C. griseus		Multa	HRVRI LPWST	FRINLSVTTP	YNADFDGDEM	NLHLPQSLET	RAEIQLAMV
Q9AVW8	G. theta	1	Multa	HIVKVMKFST	FRINLSATSP	YNADFDGDEM	NLHFPQNVES	KSELKNLMLV
Q9AXN0	G. vacuolata	1	Multa	HRVKVMRYSS	FRINLSCTSP	YNADFDGDEM	NLHVPQSPQA	RAEVMQLMMV
Q96452	N. locustae	1	Multa	HKVKVMDGLT	FRINLSATAP	YNADFDGDEM	NLHMPQSYNS	KAELESLMLV
P18616	A. thaliana	1	Multa	HRIRIMPYST	FRINLSVTSP	YNADFDGDEM	NMHVPQSFET	RAEVLELMMV
Q25312	L. major		Multa	HRVRVLNYNT	FRINLSCTTP	YNADFDGDEM	NLHVPQSLLT	KAEIEMMMV
Q9XZS1	D. guanche		Multa	HRVKVLPWST	FRINLSCTSP	YNADFDGDEM	NLHVPQSMET	RAEVENIHIT
Q9BJ59	L. seymouri		Multa	HRVRVLNYNT	FRINLSCTTP	YNADFDGDEM	NLHVPQSLLT	KAEIEMMMV
P16356	C. elegans		Multa	HRVKILPWST	FRINLSVTSP	YNADFDGDEM	NLHLPQSLET	RAEIEELAMV
P17545	T. brucei brucei		Multa	HRVRVLNYNT	FRINLSCTTP	YNADFDGDEM	NLHVPQSLLT	KAEIEMMMV
P14248	P. falciparum		Multa	HKAKILPYST	FRINLSVTSP	YNADFDGDEM	NLHLAQSHET	RSEIKHLMIV
P04052	D. melanogaster		Multa	HRVKVLPWST	FRINLSCTSP	YNADFDGDEM	NLHVPQSMET	RAEVENIHIT
Q980R2	S. solfataricus	A'	Multa	HRVRVLKGLT	FRINLLVCP	YNADFDGDEM	NLHVPQSEEA	IAEAKELMLV
Q979F4	T. volcanium	A'	Multa	HTVRVLPQGT	FRINLAVCTP	YNADFDGDEM	NLHVIQKEEA	RAEARIIMKV
P31813	T. celer	A'	Multa	HRVRVMPYRT	FRINLSVCP	YNADFDGDEM	NLHVPQTTEE	QAEAKI LMEV
Q9V114	P. abyssi	A'	Multa	HRVRVMPYRT	FRINLAVCPP	YNADFDGDEM	NLHVPQTTEE	QAEARI LMEV
Q8ZYQ8	P. aerophilum	A'	Multa	HLVKVLPGR	FRINLAVCPP	YNADFDGDEM	NLHVPQTTEE	RAEARALMLV
Q58445	M. jannaschii	A'	Multa	HRVRVLPYRT	FRINLCVCP	----VDGDTT	VLLDGK LIKI	K-DLEDKWKD
P11512	S. acidocaldarius	A'	Multa	HKVRVLPGR	FRINLLDCPP	YNADFDGDEM	NLHVPQSEEA	IAEARELMLV
O27125	M. thermoautotrophicum	A'	Multa	HQVRVLPYKT	FRINLCVCP	YNADFDGDEM	NMHVFQTEES	RAEAKT LMRV
P09846	M. thermoautotrophicum (strain Winter)	A'	Multa	HEVRVLPYKT	FRINLCVCP	YNADFDGDEM	NMHVFQTEES	RAEAKT LMRV
O28391	A. fulgidus	A'	Multa	HYVRVLPYKT	FRINPAVCP	YNADFDGDEM	NLHVPQSLEA	QAEAKI LMAV
P15350	H. halobium	A'	Multa	HEVVVMPYKT	FRINLVVCP	YNADFDGDEM	NMHALQNEEA	RAEARVLMRV
P41556	M. vannielii	A'	Multa	HKVKVLPYRT	FRINLCVCP	YNADFDGDEM	NVHVPQSEEA	RAEAET LMLV
Q03585	T. acidophilum	A'	Multa	HTVRI LPGQT	FRINLAVCTP	YNADFDGDEM	NLHVIQKEEA	RAEARIIMKV

*Method:
 BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
 PSIPRED and PHD = Structure Prediction
 Red = Helix Green = Strand Blue = Turn Gray = No co-ord

Figure C.30 Sequence alignment for *P. furiosus* RNAP subunit A' (sheet 13)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	---490---	---500---	---510---	---520---	---530---
Q8U0M4	P. furiosus	A'	RMSD	QNHIIISPRYG	GPIIGGIQDH	ISGGYLLTRE	GAYFTRDEVE	QMLMFAGVDI
Q8U0M4	P. furiosus	A'	PSIPRED	QNHIIISPRYG	GPIIGGIQDH	ISGGYLLTRE	GAYFTRDEVE	QMLMFAGVDI
Q8U0M4	P. furiosus	A'	PHD	QNHIIISPRYG	GPIIGGIQDH	ISGGYLLTRE	GAYFTRDEVE	QMLMFAGVDI
P04050	S. cerevisiae	1	RMSD	PLQIVSPQSN	KPCMGIQVDT	LCGIRKLTLR	DTFIELDQVL	NMLYWVPDWD
P04050	S. cerevisiae	1	PSIPRED	PLQIVSPQSN	KPCMGIQVDT	LCGIRKLTLR	DTFIELDQVL	NMLYWVPDWD
P04050	S. cerevisiae	1	PHD	PLQIVSPQSN	KPCMGIQVDT	LCGIRKLTLR	DTFIELDQVL	NMLYWVPDWD
P04050	S. cerevisiae	1	MOE	PLQIVSPQSN	KPCMGIQVDT	LCGIRKLTLR	DTFIELDQVL	NMLYWVPDWD
P04050	S. cerevisiae	1	BLAST	PLQIVSPQSN	KPCMGIQVDT	LCGIRKLTLR	DTFIELDQVL	NMLYWVPDWD
P04050	S. cerevisiae	1	BLAST2	PLQIVSPQSN	KPCMGIQVDT	LCGIRKLTLR	DTFIELDQVL	NMLYWVPDWD
P04050	S. cerevisiae	1	MPsrch	PLQIVSPQSN	KPCMGIQVDT	LCGIRKLTLR	DTFIELDQVL	NMLYWVPDWD
P04050	S. cerevisiae	1	Multa	PLQIVSPQSN	KPCMGIQVDT	LCGIRKLTLR	DTFIELDQVL	NMLYWVPDWD
P08775	M. musculus	1	Multa	PRMIVTPQSN	RPVMGIQVDT	LTAVRKFTKR	DVFLERGEVM	NLLMFLSTW-
P24928	H. sapiens	1	Multa	PRMIVTPQSN	RPVMGIQVDT	LTAVRKFTKR	DVFLERGEVM	NLLMFLSTW-
P36594	S. pombe	1	Multa	PKQIVSPQSN	KPVMGIQVDT	LAGVRKFSLR	DNFLTRNAV	NIMLWVPDWD
O35559	C. griseus		Multa	PRMIVTPQSN	RPVMGIQVDT	LTAVRKFTKR	DVFLERGEVM	NLLMFLSTW-
Q9AVW8	G. theta	1	Multa	SNCIVSPQGN	KPIMGIQVDS	LLGSMILTEK	GNFLSKYEIM	AYLNLIEGY-
Q9AXN0	G. vacuolata	1	Multa	PRCIVSPQGN	KPVMGIQVDT	LVTMLFTQR	DTFMEKDLVM	NLLHFRGW-
Q96452	N. locustae	1	Multa	SKHIIISQSN	KPVMGIQVDT	LLGVCRLTSK	GVFIKREEFC	NILVYASNLV-
P18616	A. thaliana	1	Multa	PKCIVSPQAN	RPVMGIQVDT	LLGCRKITKR	DTFIEKDVFM	NTLMWEDF-
Q25312	L. major		Multa	PKNFVSPNKS	APCMGIQVDS	VLGSKYKTEK	DTFLDKYFIA	ERPLWLDLW-
Q9XZS1	D. guanche		Multa	PRQIITPQAN	KPVMGIQVDT	LTAVRKMTKR	DVFIITREQVM	NLLMFLPTW-
Q9BJ59	L. seymouri		Multa	PKNFVSPNKS	APCMGIQVDS	LLGSYRLTDK	DTFVDKYFIQ	SVALWLDLW-
P16356	C. elegans		Multa	PRQLITPQAN	KPVMGIQVDT	LCAVRMMTKR	DVFLDWPFMM	DLMLYLPW-
P17545	T. brucei brucei		Multa	PKNFVSPNKS	APCMGIQVDS	LLGSYRLTDK	DTFLDKYFVQ	SVALWLDLW-
P14248	P. falciparum		Multa	QRQIVSPQGN	KPVMGIQVDS	LLAIRKFTTR	DNFLTKKEVM	SLLIWIPIYW-
P04052	D. melanogaster		Multa	PRQIITPQAN	KPVMGIQVDT	LTAVRKMTKR	DVFIITREQVM	NLLMFLPTW-
Q980R2	S. solfataricus	A'	Multa	HKNIITPRYG	GPIIGGAAQDY	ISGAYLLTVK	TTLLTKEEAQ	QILGVADV--
Q979F4	T. volcanium	A'	Multa	QEIQMSPRFG	RPIIGGIHGH	VTAMFLLTHN	NPLFTQEEMI	HIMCYVDPDL
P31813	T. celer	A'	Multa	QNHIIISPRYG	GPIIAGIQDH	ISGGYLLTRE	GAYFERTEVE	QMLMFAGVDV
Q9V114	P. abyssi	A'	Multa	QNHIIISPRYG	GPIIGGIQDH	ISGGYLLTRE	GAYFTREEVE	QMLMFAGVDI
Q8ZYQ8	P. aerophilum	A'	Multa	EKHIIITPRYG	GAIIGARQDY	IIGAYLLSHK	STFLTKKEVA	FLLGAGKS--
Q58445	M. jannaschii	A'	Multa	VKVLTSDDL	PKLTSLSKYW	KLNADEYGKK	IYKIKTELGR	EIIATED---
P11512	S. acidocaldarius	A'	Multa	HKNIITPRYG	GPIIGGGQDY	ISGAYLLSVK	TTLLTVEEVA	TI LGVTDF--
O27125	M. thermoautotrophicum	A'	Multa	QEHIILSPRFG	GPIIGGIHGH	ISGAYLLTRK	SAVFSEEKVF	QILKKAG---
P09846	M. thermoautotrophicum (strain Winter)	A'	Multa	QDHILSPRFR	D-LSSGVYTT	ISQEHTSSQG	KRLFSVRSRP	PDPEGR---
O28391	A. fulgidus	A'	Multa	QEHIILSPRFG	GPIIGGIHGH	ISGLYLLTRG	EKKFTREEAM	ELLR--SLDI
P15350	H. halobium	A'	Multa	QEIQILSPRFG	ENIIGAIQDH	ISGTYLLTND	NPRFNETQAS	DLRLQTRID-
P41556	M. vannielii	A'	Multa	EKHIIISPRYG	GPIIGAIHDF	VSGAYVLTSS	N--FIKDEAL	TLLKSSGFG-
Q03585	T. acidophilum	A'	Multa	QEIQMSPRFG	GPIIGGIHGH	VTALFLLTHN	NPRYTHEEMV	HIMAYLEPDL

*Method:
 BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
 PSIPRED and PHD = Structure Prediction
 Red = Helix Green = Strand Blue = Turn Gray = No co-ord

Figure C.31 Sequence alignment for *P. furiosus* RNAP subunit A' (sheet 14)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	---540---	---550---	---560---	-----	-----
Q8U0M4	<i>P. furiosus</i>	A'	RMSD	TELPEPDKYD	ENG ~~N~~ PLWSGK	TIFSLLL ~~P~~ --	-----	-----
Q8U0M4	<i>P. furiosus</i>	A'	PSIPRED	TELPEPDKYD	ENG ~~N~~ PLWSGK	TIFSLLL ~~P~~ --	-----	-----
Q8U0M4	<i>P. furiosus</i>	A'	PHD	TELPEPDKYD	ENG ~~N~~ PLWSGK	TIFSLLL ~~P~~ --	-----	-----
P04050	<i>S. cerevisiae</i>	1	RMSD	GVIPTPAII--	-KPKPLWSGK	QILSVAIP--	-----	-----
P04050	<i>S. cerevisiae</i>	1	PSIPRED	GVIPTPAII--	-KPKPLWSGK	QILSVAIP--	-----	-----
P04050	<i>S. cerevisiae</i>	1	PHD	GVIPTPAII--	-KPKPLWSGK	QILSVAIP--	-----	-----
P04050	<i>S. cerevisiae</i>	1	MOE	GVIPTPAII--	-KPKPLWSGK	QILSVAIP--	-----	-----
P04050	<i>S. cerevisiae</i>	1	BLAST	GVIPTPAIIK	P--KPLWSGK	QILSVAIP--	-----	-----
P04050	<i>S. cerevisiae</i>	1	BLAST2	GVIPTPAIIK	PK--PLWSGK	QILSVAIP--	-----	-----
P04050	<i>S. cerevisiae</i>	1	MPsrch	GVIPTPAII--	-KPKPLWSGK	QILSVAIP--	-----	-----
P04050	<i>S. cerevisiae</i>	1	Multa	-DGVIPTPAI	IKPKPLWSGK	QILSVAIP--	-----	-----
P08775	<i>M. musculus</i>	1	Multa	-DGKVPQPAI	LKPRPLWTGK	QIFSLIIP--	-----	-----
P24928	<i>H. sapiens</i>	1	Multa	-DGKVPQPAI	LKPRPLWTGK	QIFSLIIP--	-----	-----
P36594	<i>S. pombe</i>	1	Multa	-DGILPPPIV	LKPKVLWTGK	QILSLIIP--	-----	-----
O35559	<i>C. griseus</i>		Multa	-DGKVPQPAI	LKPRPLWTGK	QIFSLIIP--	-----	-----
Q9AVW8	<i>G. theta</i>	1	Multa	-SLNIPLPAI	YKPRILWTGK	QLISTIIP--	-----	-----
Q9AXN0	<i>G. vacuolata</i>	1	Multa	-DGVIPPPAI	LKPRQLWTGK	QLFSMILP--	-----	-----
Q96452	<i>N. locustae</i>	1	Multa	-ARANIKPAI	ARKILYTGK	QLFSLTLP--	-----	-----
P18616	<i>A. thaliana</i>	1	Multa	-DGKVPAPAI	LKPRPLWTGK	QVFNLIIP--	-----	-----
Q25312	<i>L. major</i>		Multa	-Q--LPIPTI	LKPRPLWTGK	QIFSLIIP--	-----	-----
Q9XZS1	<i>D. guanche</i>		Multa	-DGKMPQPCI	LKPRPLWTGK	QIFSLIIP--	-----	-----
Q9BJ59	<i>L. seymouri</i>		Multa	-E--LPIPAI	LKPRPLWTGK	QIFSLIIP--	-----	-----
P16356	<i>C. elegans</i>		Multa	-DGKVPQPAI	LKPKPLWTGK	QVFSLIIP--	-----	-----
P17545	<i>T. brucei brucei</i>		Multa	-Q--LPIPAI	LKPRPLWTGK	QVFSLIIP--	-----	-----
P14248	<i>P. falciparum</i>		Multa	-NHVIPTPAI	IKPRALWTGK	QIFSMLLQFD	DMNIEDDKND	TANNKVGRDV
P04052	<i>D. melanogaster</i>		Multa	-DAKMPQPCI	LKPRPLWTGK	QIFSLIIP--	-----	-----
Q980R2	<i>S. solfataricus</i>	A'	Multa	-KIDLGEPAI	LAPREYTGK	QVISAFILP--	-----	-----
Q979F4	<i>T. volcanium</i>	A'	Multa	---IPDATI	VNGKKYYSGR	NIFSILIP--	-----	-----
P31813	<i>T. celer</i>	A'	Multa	DRLEPEDKV-	ENGVELWSGK	TIFSLLL ~~P~~ --	-----	-----
Q9V114	<i>P. abyssi</i>	A'	Multa	KELPEPDKY-	ENGKPLWSGK	TIFSLLL ~~P~~ --	-----	-----
Q8ZYQ8	<i>P. aerophilum</i>	A'	Multa	-QEDPPEPAI	LHPVELWTGK	QIISHFLP--	-----	-----
Q58445	<i>M. jannaschii</i>	A'	Multa	-HPFYTTNGR	KRCGELKVG	EV--IIYP--	-----	-----
P11512	<i>S. acidocaldarius</i>	A'	Multa	-VGELGEPAI	LAPKPYTGK	QVISLFLP--	-----	-----
O27125	<i>M. thermoautotrophicum</i>	A'	Multa	--LPLPDSRG	RD---WTGK	EIFSMLVLP--	-----	-----
P09846	<i>M. thermoautotrophicum</i> (strain Winter)	A'	Multa	--APPPDREG	RE---WTVK	EIFSMLVLP--	-----	-----
O28391	<i>A. fulgidus</i>	A'	Multa	SEIPVKDEY-	-----TGK	EIFSILIP--	-----	-----
P15350	<i>H. halobium</i>	A'	Multa	-ELPAAAGTD	EDGDQYWTGH	QIFSELLP--	-----	-----
P41556	<i>M. vannielii</i>	A'	Multa	SELGEPDFV-	ENGIEYSGK	SLFSKTLF--	-----	-----
Q03585	<i>T. acidophilum</i>	A'	Multa	---LPEARI	ENGKYYGR	DIFSTILP--	-----	-----

*Method:
 BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
 PSIPRED and PHD = Structure Prediction
 Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.32 Sequence alignment for *P. furiosus* RNAP subunit A' (sheet 15)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	-----590-----	-----600-----	-----	-----	-----
Q8U0M4	P. furiosus	A'	RMSD	ALEKLI EE KL	MPD EE VRK-	-----	-----	-----
Q8U0M4	P. furiosus	A'	PSIPRED	ALEKLI EE KL	MPD EE VRK-	-----	-----	-----
Q8U0M4	P. furiosus	A'	PHD	ALEKLI EE KL	MPD EE VRK-	-----	-----	-----
P04050	S. cerevisiae	1	RMSD	HLQRFDEGTT	LLSP-----	-----	-----	-----
P04050	S. cerevisiae	1	PSIPRED	HLQRFDEGTT	LLSP-----	-----	-----	-----
P04050	S. cerevisiae	1	PHD	HLQRFDEGTT	LLSP-----	-----	-----	-----
P04050	S. cerevisiae	1	MOE	HLQRFDEGTT	LLSPKD----	-----	-----	-----
P04050	S. cerevisiae	1	BLAST	HLQRFDEGTT	LLSPK-----	-----	-----	-----
P04050	S. cerevisiae	1	BLAST2	HLQRFDEGTT	LLSPK-----	-----	-----	-----
P04050	S. cerevisiae	1	MPsrch	--E---GTTL	LS-P----K-	-----	-----	-----
P04050	S. cerevisiae	1	Multa	-----	-----	-----	-----	-----
P08775	M. musculus	1	Multa	-----	-----	-----	-----	-----
P24928	H. sapiens	1	Multa	-----	-----	-----	-----	-----
P36594	S. pombe	1	Multa	-----	-----	-----	-----	-----
O35559	C. griseus		Multa	-----	-----	-----	-----	-----
Q9AVV8	G. theta	1	Multa	-----	-----	-----	-----	-----
Q9AXN0	G. vacuolata	1	Multa	-----	-----	-----	-----	-----
Q96452	N. locustae	1	Multa	-----	-----	-----	-----	-----
P18616	A. thaliana	1	Multa	-----	-----	-----	-----	-----
Q25312	L. major		Multa	-----	-----	-----	-----	-----
Q9XZS1	D. guanche		Multa	-----	-----	-----	-----	-----
Q9BJ59	L. seymouri		Multa	-----	-----	-----	-----	-----
P16356	C. elegans		Multa	-----	-----	-----	-----	-----
P17545	T. brucei brucei		Multa	-----	-----	-----	-----	-----
P14248	P. falciparum		Multa	IGDNINVGNV	QQNDMSSPNN	NNNNNNNNNS	NNNNNNNIGG	GINSFKRFNM
P04052	D. melanogaster		Multa	-----	-----	-----	-----	-----
Q980R2	S. solfataricus	A'	Multa	NED-----	-----	-----	-----	-----
Q979F4	T. volcanium	A'	Multa	EYEK-----	-----	-----	-----	-----
P31813	T. celer	A'	Multa	ALEKLI EE KL	IPD EE EVRA-	-----	-----	-----
Q9V114	P. abyssi	A'	Multa	ALEKLI EE KL	MPD EE VRK-	-----	-----	-----
Q8ZYQ8	P. aerophilum	A'	Multa	AYT-----	-----	-----	-----	-----
Q58445	M. jannaschii	A'	Multa	-----	-----	-----	-----	-----
P11512	S. acidocaldarius	A'	Multa	DEI-----	-----	-----	-----	-----
O27125	M. thermoautotrophicum	A'	Multa	EME-----	-----	-----	-----	-----
P09846	M. thermoautotrophicum (strain Winter)	A'	Multa	EMD-----	-----	-----	-----	-----
O28391	A. fulgidus	A'	Multa	GAE-----	-----	-----	-----	-----
P15350	H. halobium	A'	Multa	-----	-----	-----	-----	-----
P41556	M. vannieli	A'	Multa	REE-----	-----	-----	-----	-----
Q03585	T. acidophilum	A'	Multa	EFED-----	-----	-----	-----	-----

*Method:
 BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
 PSIPRED and PHD = Structure Prediction
 Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.33 Sequence alignment for *P. furiosus* RNAP subunit A' (sheet 16)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	-----590-	-----600,	-----	-----	-----
Q8U0M4	P. furiosus	A'	RMDS	ALEKLI EE KL	MPD EE VRK-	-----	-----	-----
Q8U0M4	P. furiosus	A'	PSIPRED	ALEKLI EE KL	MPD EE VRK-	-----	-----	-----
Q8U0M4	P. furiosus	A'	PHD	ALEKLI EE KL	MPD EE VRK-	-----	-----	-----
P04050	S. cerevisiae	1	RMDS	HLQRFDEGTT	LLSP-----	-----	-----	-----
P04050	S. cerevisiae	1	PSIPRED	HLQRFDEGTT	LLSP-----	-----	-----	-----
P04050	S. cerevisiae	1	PHD	HLQRFDEGTT	LLSP-----	-----	-----	-----
P04050	S. cerevisiae	1	MOE	HLQRFDEGTT	LLSPKD----	-----	-----	-----
P04050	S. cerevisiae	1	BLAST	HLQRFDEGTT	LLSPK-----	-----	-----	-----
P04050	S. cerevisiae	1	BLAST2	HLQRFDEGTT	LLSPK-----	-----	-----	-----
P04050	S. cerevisiae	1	MPsrch	--E---GTTL	LS-P----K-	-----	-----	-----
P04050	S. cerevisiae	1	Multa	-----	-----	-----	-----	-----
P08775	M. musculus	1	Multa	-----	-----	-----	-----	-----
P24928	H. sapiens	1	Multa	-----	-----	-----	-----	-----
P36594	S. pombe	1	Multa	-----	-----	-----	-----	-----
O35559	C. griseus		Multa	-----	-----	-----	-----	-----
Q9AVV8	G. theta	1	Multa	-----	-----	-----	-----	-----
Q9AXN0	G. vacuolata	1	Multa	-----	-----	-----	-----	-----
Q96452	N. locustae	1	Multa	-----	-----	-----	-----	-----
P18616	A. thaliana	1	Multa	-----	-----	-----	-----	-----
Q25312	L. major		Multa	-----	-----	-----	-----	-----
Q9XZS1	D. guanche		Multa	-----	-----	-----	-----	-----
Q9BJ59	L. seymouri		Multa	-----	-----	-----	-----	-----
P16356	C. elegans		Multa	-----	-----	-----	-----	-----
P17545	T. brucei brucei		Multa	-----	-----	-----	-----	-----
P14248	P. falciparum		Multa	IGDNINVGNV	QQNDMSSPNN	NNNNNNNNNS	NNNNNNNIGG	GINSFKRFNM
P04052	D. melanogaster		Multa	-----	-----	-----	-----	-----
Q980R2	S. solfataricus	A'	Multa	NED-----	-----	-----	-----	-----
Q979F4	T. volcanium	A'	Multa	EYEK-----	-----	-----	-----	-----
P31813	T. celer	A'	Multa	ALEKLI EE KL	IPD EE VR-	-----	-----	-----
Q9V114	P. abyssi	A'	Multa	ALEKLI EE KL	MPD EE VRK-	-----	-----	-----
Q8ZYQ8	P. aerophilum	A'	Multa	AYT-----	-----	-----	-----	-----
Q58445	M. jannaschii	A'	Multa	-----	-----	-----	-----	-----
P11512	S. acidocaldarius	A'	Multa	DEI-----	-----	-----	-----	-----
O27125	M. thermoautotrophicum	A'	Multa	EME-----	-----	-----	-----	-----
P09846	M. thermoautotrophicum (strain Winter)	A'	Multa	EMD-----	-----	-----	-----	-----
O28391	A. fulgidus	A'	Multa	GAE-----	-----	-----	-----	-----
P15350	H. halobium	A'	Multa	-----	-----	-----	-----	-----
P41556	M. vannieli	A'	Multa	REE-----	-----	-----	-----	-----
Q03585	T. acidophilum	A'	Multa	EFED-----	-----	-----	-----	-----

*Method:
 BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMDS = MOE Minimized RMDS
 PSIPRED and PHD = Structure Prediction
 Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.34 Sequence alignment for *P. furiosus* RNAP subunit A' (sheet 17)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	-----610	-----620	-----630
Q8U0M4	P. furiosus	A'	RMSD	-----	LAYDGFVYIQ	NGKLLSGAID KKAYGRED-G
Q8U0M4	P. furiosus	A'	PSIPRED	-----	LAYDGFVYIQ	NGKLLSGAID KKAYGRED-G
Q8U0M4	P. furiosus	A'	PHD	-----	LAYDGFVYIQ	NGKLLSGAID KKAYGRED-G
P04050	S. cerevisiae	1	RMSD	-----	-KDNGMLII-	DGQIIFGVVE KKTVGSSN-G
P04050	S. cerevisiae	1	PSIPRED	-----	-KDNGMLII-	DGQIIFGVVE KKTVGSSN-G
P04050	S. cerevisiae	1	PHD	-----	-KDNGMLII-	DGQIIFGVVE KKTVGSSN-G
P04050	S. cerevisiae	1	MOE	-----	---NGMLII-	DGQIIFGVVE KKTVGSSN-G
P04050	S. cerevisiae	1	BLAST	-----	---DNGMLII	DGQIIFGVVE KKTVGSSN-G
P04050	S. cerevisiae	1	BLAST2	-----	DNGMLII	DGQIIFGVVE KKTVGSSN-G
P04050	S. cerevisiae	1	MPsrch	-----	-D-NGMLII-	DGQIIFGVVE KKTVGSSN-G
P04050	S. cerevisiae	1	Multa	-----	-GTT----LL	SPKDNMGLII DGQIIFGVVE KKTVGSSN-G
P08775	M. musculus	1	Multa	-----	-EDSGPYKHI	SPGDTKVVVE NGELIMGILC KKSIGTSA-G
P24928	H. sapiens	1	Multa	-----	-EDSGPYKHI	SPGDTKVVVE NGELIMGILC KKSIGTSA-G
P36594	S. pombe	1	Multa	-----	-KQS----LS	NPTDSGMLIE NGELIYGVDV KKTVGSAQ-G
O35559	C. griseus		Multa	-----	-EDSGPYKHI	SPGDTKVVVE NGELIMGILC KKSIGTSA-G
Q9AVW8	G. theta	1	Multa	-----	---EHEKAYV	TSGDTRVIIIF QGQLLSGMID KRTIGATS-G
Q9AXN0	G. vacuolata	1	Multa	-----	-DIE-----	NAGDTKVLIS RGEILSGTVD KRTVGSA-N
Q96452	N. locustae	1	Multa	-----	-----KL	NFFRDKVVIM DGVLLNGVIC KKSIVGTAQ-G
P18616	A. thaliana	1	Multa	-----	-ETG----FI	TPGDTQVRIE RGEILAGTLC KKTIGTSA-G
Q25312	L. major		Multa	-----	-ASPHDKPPF	PHNDSKIMIQ RGQLLGGAIT KGVVGAAP-G
Q9XZS1	D. guanche		Multa	-----	-EDDGPYKWI	SPGDTKVMVE HGDLMIGILC KKTIGTSA-G
Q9BJ59	L. seymouri		Multa	-----	-ASPYGKPPF	PHNDKKI-IQ RGLLLVGAIT KGVVGAAP-G
P16356	C. elegans		Multa	-----	-EDSGPYKWI	SPGDTKVIIE HGEILSGIVC SKTVGKSA-G
P17545	T. brucei brucei		Multa	-----	-ATPQDRPPF	PHNDSVVMIR RGQLLCPIT KSIIVGAAP-G
P14248	P. falciparum		Multa	VKINLMRDSS	TSSKDDNPYC	SINDGKVIK NNELLSGIIC KRTVGSSS-G
P04052	D. melanogaster		Multa	-----	-EDEGFPYKI	SPGDTKVMVE HGEILMGILC KKSIGTSA-G
Q980R2	S. solfataricus	A'	Multa	-----	-----	CPHDSYVVIK NGILLEGVFD KKAIGNQQPE
Q979F4	T. volcanium	A'	Multa	-----	-----	EKEDTYVTIIV DGKLIHGTID EAAISPFSS-G
P31813	T. celer	A'	Multa	-----	-----	LAYDGFYIIL NGKLLSGAID KTAYGRED-G
Q9V114	P. abyssi	A'	Multa	-----	-----	LAYDGFVYIQ NGKLLSGAID KKAYGRED-G
Q8ZYQ8	P. aerophilum	A'	Multa	-----	-----	CPTDEWIIVI NGNLVKGVL D KKSIGAEQVD
Q58445	M. jannaschii	A'	Multa	-----	-----	NDFPM FEDDNRIIVD EEKI----- KKVINNIG-G
P11512	S. acidocaldarius	A'	Multa	-----	-----	CPHDSFIVIK NGLLLEGVFD KKAIGNQQPE
O27125	M. thermoautotrophicum	A'	Multa	-----	-----	CENDAYVVIE NGQLISGVID EKAYGAFA-G
P09846	M. thermoautotrophicum (strain Winter)	A'	Multa	-----	-----	WENDAYVVIE NGQLITGVID EKAYGAFA-G
Q28391	A. fulgidus	A'	Multa	-----	-----	CEYDAYVIE NGKLLKGITID EKAYGAFA-G
P15350	H. halobium	A'	Multa	-----	-----	-----VVIE DGQLLEGITIA DDEVEGYSS-G
P41556	M. vannielii	A'	Multa	-----	-----	CENDAFVIR NGRVLQGV D KNGFGAET-G
Q03585	T. acidophilum	A'	Multa	-----	-----	DPSDTYVEIV DGKMIHGTID EAAVSFPSS-G

*Method:
 BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
 PSIPRED and PHD = Structure Prediction
 Red = Helix Green = Strand Blue = Turn Gray = No coörd

Figure C.35 Sequence alignment for *P. furiosus* RNAP subunit A' (sheet 18)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	630-----	640-,-----	650---,---	660----	670-----
Q8U0M4	P. furiosus	A'	RMSD	IILDLLVREY	GVER-ARQFL	DQVTKL-TIW	VI-THKGFTT	GIDDEDLPPE
Q8U0M4	P. furiosus	A'	PSIPRED	IILDLLVREY	GVER-ARQFL	DQVTKL-TIW	VI-THKGFTT	GIDDEDLPPE
Q8U0M4	P. furiosus	A'	PHD	IILDLLVREY	GVER-ARQFL	DQVTKL-TIW	VI-THKGFTT	GIDDEDLPPE
P04050	S. cerevisiae	1	RMSD	GLIHVVTTREK	GPQV-CAKLF	GNIQKV-VNF	WL-LHNGFST	GIGDTIADGP
P04050	S. cerevisiae	1	PSIPRED	GLIHVVTTREK	GPQV-CAKLF	GNIQKV-VNF	WL-LHNGFST	GIGDTIADGP
P04050	S. cerevisiae	1	PHD	GLIHVVTTREK	GPQV-CAKLF	GNIQKV-VNF	WL-LHNGFST	GIGDTIADGP
P04050	S. cerevisiae	1	MOE	GLIHVVTTREK	GPQV-CAKLF	GNIQKV-VNF	WL-LHNGFST	GIGDTIADGP
P04050	S. cerevisiae	1	BLAST	GLIHVVTTREK	GPQV-CAKLF	GNIQKV-VNF	WL-LHNGFST	GIGDTIADGP
P04050	S. cerevisiae	1	BLAST2	GLIHVVTTREK	GPQV-CAKLF	GNIQKV-VNF	WL-LHNGFST	GIGDTIADGP
P04050	S. cerevisiae	1	MPsrch	GLIHVVTTREK	GPQVCAKLF-	GNIQKV-VNF	LL--HNGFST	GIGD-TIADG
P04050	S. cerevisiae	1	Multa	GLIHVVTTREK	GPQV-CAKLF	GNIQKV-VNF	WL-LHNGFST	GIGDTIADGP
P08775	M. musculus	1	Multa	SLVHISYLEM	GHDV-TRLFY	SNIQTV-INN	WL-LIEGHTI	GIGDSIADSK
P24928	H. sapiens	1	Multa	SLVHISYLEM	GHDV-TRLFY	SNIQTV-INN	WL-LIEGHTI	GIGDSIADSK
P36594	S. pombe	1	Multa	GLVHTIWKEK	GPEI-CKGFF	NGIQRV-VNY	WL-LHNGFSI	GIGDTIADAD
O35559	C. griseus		Multa	SLVHISYLEM	GHDV-TRLFY	SNIQTV-INN	WL-LIEGHTI	GIGDSIADSK
Q9AVW8	G. theta	1	Multa	GLIHVWIKDI	GPSA-CSYFI	SNFQII-VNN	WL-LQNGFSV	GIGDCIPTKM
Q9AXN0	G. vacuolata	1	Multa	SLVHVTWKEH	GPEG-ACDLI	SNIQTV-INN	WL-LQNGFSI	GIGDTIADQ
Q96452	N. locustae	1	Multa	SLIHIIANDY	GHEE-ITRFI	DSLQKL-IST	YLTLISTFSV	GIGDTISSPE
P18616	A. thaliana	1	Multa	SLVHVWEEV	GPDA-ARKFL	GHTQWL-VNY	WL-LQNGFTI	GIGDTIADSS
Q25312	L. major		Multa	SLIHVIFNER	GSDE-VAKFI	NGVQRI-TTY	FN-YCFASFV	GVQDVTADAT
Q9XZS1	D. guanche		Multa	SLHICFLEL	GHEI-AGRFY	GNIQTV-INN	WL-LLEGHSI	GIGDTIADPO
Q9BJ59	L. seymouri		Multa	SLIHVIFNER	GSDE-VAKFI	NGVQRI-TTY	FN-YCFASFV	GVQDVTADAS
P16356	C. elegans		Multa	NLLHVVTLEL	GVEI-AANFY	SHIQTV-INA	WL-IREGHTI	GIGDTIADQA
P17545	T. brucei brucei		Multa	SLIHVIFNEH	GSDE-VARFI	NGVQRV-TTF	FL-LNFGFSV	GVQDVTADSD
P14248	P. falciparum		Multa	SLIHVWHEM	GPKK-TKDFL	SALQKV-TNN	WL-EYVGFTV	SCSDIIASNK
P04052	D. melanogaster		Multa	SLHICFLEL	GHDV-AGRFY	GNIQTV-INN	WL-LFEGHSI	GIGDTIADPO
Q980R2	S. solfataricus	A'	Multa	SILHWLIKEY	SDEY-GKWL	DNLFRV-FIR	FV-ELQGFTM	RLEDVSLGDD
Q979F4	T. volcanium	A'	Multa	AIIDKIFRKF	GPNE-AAKFI	DRMTRL-AVG	FI-TYGFST	GISDYDIPYS
P31813	T. celer	A'	Multa	KLLDIIVREY	GVER-ARQFL	DQVTKL-AIW	VI-THKGFTT	AIDDEDLPTE
Q9V114	P. abyssi	A'	Multa	IILDLLVREY	GVER-ARQFL	DQVTKL-TIW	VI-THKGFTT	GIDDEDLPPE
Q8ZYQ8	P. aerophilum	A'	Multa	SLWHRIARDY	SPEV-ARRWL	DSSLRL-FLR	FL-DLRGFTF	GMDSIYIPPE
Q58445	M. jannaschii	A'	Multa	TYKNKIINEL	KDRK-LIPLT	YNDQKA---S	I LARIVGHVM	GDGSLIINN
P11512	S. acidocaldarius	A'	Multa	SMLHWSIREY	GTEY-GKWL	DNVFKM-FIR	FL-EMRGFTM	TLEDITIPDE
O27125	M. thermoautotrophicum	A'	Multa	KILDHIVKEY	GTDA-AREFL	DSATKL-AIA	GI-MHAGFTT	STNDEEIPPE
P09846	M. thermoautotrophicum (strain Winter)	A'	Multa	KILDQIVKEY	GSDA-AKEFL	DSATKL-AIA	GI-MHAGFTT	STNDEEIPPE
O28391	A. fulgidus	A'	Multa	IIIDEIARKY	GKEM-AKKFI	DDMTQL-AIR	II-SKLGFTV	GIDDEDIPPE
P15350	H. halobium	A'	Multa	EIVDTITKVH	GNTR-ARIFI	NEVASL-AMR	SI-MHGFST	GIDDETIVSTE
P41556	M. vannielii	A'	Multa	VVLTNTIVKDF	GSED-ARTFL	DSATKM-SVK	SM-MLRGFTT	GIDDEDIPSE
Q03585	T. acidophilum	A'	Multa	AIIDKIFRKF	GSQE-AARFI	DRMTRL-AVG	FI-TYRGFTT	GISDYDIPES

*Method:
 BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
 PSIPRED and PHD = Structure Prediction
 Red = Helix Green = Strand Blue = Turn Gray = No co-ord

Figure C.36 Sequence alignment for *P. furiosus* RNAP subunit A' (sheet 19)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	-680,-----	--690-----	--700-,---	---710----	---720----
Q8U0M4	P. furiosus	A'	RMSD	ARDR-IREII	REAEERVNKL	IEAYKR-GEL	EPLPGKSLED	TLESIMAVL
Q8U0M4	P. furiosus	A'	PSIPRED	ARDR-IREII	REAEERVNKL	IEAYKR-GEL	EPLPGKSLED	TLESIMAVL
Q8U0M4	P. furiosus	A'	PHD	ARDR-IREII	REAEERVNKL	IEAYKR-GEL	EPLPGKSLED	TLESIMAVL
P04050	S. cerevisiae	1	RMSD	TMRE-ITETI	AEAKKKVLDV	TKEAQA-NLL	TAKHGMTLRE	SFEDNVVRF
P04050	S. cerevisiae	1	PSIPRED	TMRE-ITETI	AEAKKKVLDV	TKEAQA-NLL	TAKHGMTLRE	SFEDNVVRF
P04050	S. cerevisiae	1	PHD	TMRE-ITETI	AEAKKKVLDV	TKEAQA-NLL	TAKHGMTLRE	SFEDNVVRF
P04050	S. cerevisiae	1	MOE	TMRE-ITETI	AEAKKKVLDV	TKEAQA-NLL	TAKHGMTLRE	SFEDNVVRF
P04050	S. cerevisiae	1	BLAST	TMRE-ITETI	AEAKKKVLDV	TKEAQA-NLL	TAKHGMTLRE	SFEDNVVRF
P04050	S. cerevisiae	1	BLAST2	TMRE-ITETI	AEAKKKVLDV	TKEAQA-NLL	TAKHGMTLRE	SFEDNVVRF
P04050	S. cerevisiae	1	MPsrch	PTMREITETI	AEAKKKVLDV	TKEAQA-NLL	TAKHGMTLRE	SFEDNVVRF
P04050	S. cerevisiae	1	Multa	TMRE-ITETI	AEAKKKVLDV	TKEAQA-NLL	TAKHGMTLRE	SFEDNVVRF
P08775	M. musculus	1	Multa	TYQD-IQNTI	KKAKQDVIEV	IEKAHN-NEL	EPTPGNTLRQ	TFENQVNRIL
P24928	H. sapiens	1	Multa	TYQD-IQNTI	KKAKQDVIEV	IEKAHN-NEL	EPTPGNTLRQ	TFENQVNRIL
P36594	S. pombe	1	Multa	TMKE-VTRTV	KEARRQVAEC	IQDAQH-NRL	KPEPGMTLRE	SFEAKVSRFL
O35559	C. griseus		Multa	TYQD-IQNTI	KKAKQDVIEV	IEKAHN-NEL	EPTPGNTLRQ	TFENQVNRIL
Q9AVW8	G. theta	1	Multa	VKKD-VSLII	KTAKLFVKKS	IKRLLS-KKF	NGFFNKDSNY	NLEHNLKNVL
Q9AXN0	G. vacuolata	1	Multa	TRFN-VIEII	NKAKKEVKRL	VRKAQE-GEL	QLLPKGMTME	SFETGVNKNVL
Q96452	N. locustae	1	Multa	TMAH-ISRAI	GDAKAEVKQL	IYDSRRRGS	EKLPGMNLQE	TLESKINLAL
P18616	A. thaliana	1	Multa	TMEK-INETI	SNAKTAVKDL	IRQFQG-KEL	DPEPGRTMRD	TFENRVNQVL
Q25312	L. major		Multa	TLKE-MNNVL	HKTRQSVEKI	GAAANN-GKL	TRKAGMSLLQ	SFEADVNSAL
Q9XZS1	D. guanche		Multa	TYNE-IQMAI	KKAKDDVINV	IQKAHN-MEL	EPTPGNTLRQ	TFENKVNRI
Q9BJ59	L. seymouri		Multa	TLKE-MNNVL	HKTRQSVEKI	GAAANN-GKL	TRKAGMSLLQ	SFEADVNSAL
P16356	C. elegans		Multa	TYLD-IQNTI	KKAKQDVVDV	IEKAHN-DDL	EPTPGNTLRQ	TFENKVNRI
P17545	T. brucei brucei		Multa	TLRQ-MNDVL	VKTRRNVEKI	GAAANN-RTL	NRKAGMTLLQ	SFEADVNSAL
P14248	P. falciparum		Multa	VLKG-VREIL	DKSKSEVSKL	VEKAQK-GEL	ECQPGKSLYE	SFETRVNNEL
P04052	D. melanogaster		Multa	TYNE-IQQA	KKAKDDVINV	IQKAHN-MEL	EPTPGNTLRQ	TFENKVNRI
Q980R2	S. solfataricus	A'	Multa	VKKE-IYNEI	DRAKVEVDNL	IQKYKN-GEL	EPIPGRTLEE	SLENYILDTL
Q979F4	T. volcanium	A'	Multa	ATAR-IEELV	NQAEEDRINKL	IETYYK-GEL	QAPAGRSVED	TLEIEILSEA
P31813	T. celer	A'	Multa	ALDR-IREII	REAEERVNRL	IEAYRN-GEL	EPIPGKTLEE	TLESKIMAA
Q9V114	P. abyssi	A'	Multa	ARDR-IKEII	REAEERVQKL	IEAYKR-GEL	EPLPGKTLEE	TLESKIMAVL
Q8ZYQ8	P. aerophilum	A'	Multa	AYRE-IDNII	KASLEKVDKL	IDDFRS-GHL	EAMPGFVTEE	TFENKVTIEL
Q58445	M. jannaschii	A'	Multa	NSRV-V----	--FRGDIEDL	-----	-----K	TIKEDLKEG
P11512	S. acidocaldarius	A'	Multa	AQNE-ITTKI	KEGYSQVDEY	IRKFNE-GQL	EPIPGRTIEE	SLESYILDTL
O27125	M. thermoautotrophicum	A'	Multa	ARER-IEAHL	RNAEARVDQL	IEAYEN-GEL	EPLPGRSLEE	TLEMKIMQVL
P09846	M. thermoautotrophicum (strain Winter)	A'	Multa	AKER-IEAHL	RNAEARVDQL	IEAYEN-GEL	EPLPGRSLEE	TLEMKIMQVL
O28391	A. fulgidus	A'	Multa	AAEQ-IEEVL	KEAENEVNRL	IEAYRR-GDL	EPMPGRSIEE	TLEMKIMQVL
P15350	H. halobium	A'	Multa	ARER-IDEAI	QSAIDRVQEL	IETYYN-GDL	ESLPGRTVDE	TLEMKIMQTL
P41556	M. vannielii	A'	Multa	AIQE-IQDLL	DKAESDVQEI	VERYEG-GTL	DPLPGRGIEE	SREAYIMQIL
Q03585	T. acidophilum	A'	Multa	AVAR-IEELV	AQAEEDRINKL	IETFRR-GEL	QAPAGRSVED	TLEMEILSEA

*Method:

BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD

PSIPRED and PHD = Structure Prediction

Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.37 Sequence alignment for *P. furiosus* RNAP subunit A' (sheet 20)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	---730---	---740---	---750---	---760---	---770---
Q8U0M4	<i>P. furiosus</i>	A'	RMSD	AEARDNAGAV	AEKYLGM-D-N	HTVI-MAKTG	ARGKILNITQ	MAALLGQQSI
Q8U0M4	<i>P. furiosus</i>	A'	PSIPRED	AEARDNAGAV	AEKYLGM-D-N	HTVI-MAKTG	ARGKILNITQ	MAALLGQQSI
Q8U0M4	<i>P. furiosus</i>	A'	PHD	AEARDNAGAV	AEKYLGM-D-N	HTVI-MAKTG	ARGKILNITQ	MAALLGQQSI
P04050	<i>S. cerevisiae</i>	1	RMSD	NEARDKAGRL	AEVNLKDL-N	NVKQ-MVMAG	SKGSFINIAQ	MSACVGQQSV
P04050	<i>S. cerevisiae</i>	1	PSIPRED	NEARDKAGRL	AEVNLKDL-N	NVKQ-MVMAG	SKGSFINIAQ	MSACVGQQSV
P04050	<i>S. cerevisiae</i>	1	PHD	NEARDKAGRL	AEVNLKDL-N	NVKQ-MVMAG	SKGSFINIAQ	MSACVGQQSV
P04050	<i>S. cerevisiae</i>	1	MOE	NEARDKAGRL	AEVNLKDL-N	NVKQ-MVMAG	SKGSFINIAQ	MSACVGQQSV
P04050	<i>S. cerevisiae</i>	1	BLAST	NEARDKAGRL	AEVNLKDL-N	NVKQ-MVMAG	SKGSFINIAQ	MSACVGQQSV
P04050	<i>S. cerevisiae</i>	1	BLAST2	NEARDKAGRL	AEVNLKDL-N	NVKQ-MVMAG	SKGSFINIAQ	MSACVGQQSV
P04050	<i>S. cerevisiae</i>	1	MPsrch	NEARDKAGRL	AEVNL-KD-L	NNVKQMVMA	SKGSFINIAQ	MSACVGQQSV
P04050	<i>S. cerevisiae</i>	1	Multa	NEARDKAGRL	AEVNLKDL-N	NVKQ-MVMAG	SKGSFINIAQ	MSACVGQQSV
P08775	<i>M. musculus</i>	1	Multa	NDARDKTGSS	AQKSLSEY-N	NFKS-MVVS	ARGSKINISQ	VIADVQQNV
P24928	<i>H. sapiens</i>	1	Multa	NDARDKTGSS	AQKSLSEY-N	NFKS-MVVS	ARGSKINISQ	VIADVQQNV
P36594	<i>S. pombe</i>	1	Multa	NQARDNAGRS	AEHSLKDS-N	NVKQ-MVAG	SKGSFINISQ	MSACVGQQSV
O35559	<i>C. griseus</i>		Multa	NDARDKTGSS	AQKSLSEY-N	NFKS-MVVS	ARGSKINISQ	VIADVQQNV
Q9AVW8	<i>G. theta</i>	1	Multa	NNARDMAGSS	AQKCLTSY-N	NLKK-MIFSG	SKGSLINISQ	IVACVGQQNV
Q9AXN0	<i>G. vacuolata</i>	1	Multa	NGARDKSGSD	AQKSLKLT-N	NIKR-MVNAG	SKGSFINISQ	ICACVGQQNV
Q96452	<i>N. locustae</i>	1	Multa	NKARDISGTR	AVESLNHL-N	GLKQ-MLKAG	SKGSYINISQ	ITSCVGQQNI
P18616	<i>A. thaliana</i>	1	Multa	NKARDDAGSS	AQKSLAET-N	NLKA-MVTAG	SKGSFINISQ	MTACVGQQNV
Q25312	<i>L. major</i>		Multa	NKCREEAACK	ALSNVVRT-N	SFKV-MIEAG	SKGSDLNICQ	IADVFGQQNV
Q9XZS1	<i>D. guanche</i>		Multa	NDARDKTGGS	AKKSLTEY-N	NLKA-MVVS	SKGSNINISQ	VIADVQQNV
Q9BJ59	<i>L. seymouri</i>		Multa	NKCREEAACK	ALSNVVRT-N	SFKV-MIEAG	SKGSDLNICQ	IADVFGQQNV
P16356	<i>C. elegans</i>		Multa	NDARDKTGSS	AQKSLSEF-N	NFKS-MVVS	ARGSKINISQ	VIADVQQNV
P17545	<i>T. brucei brucei</i>		Multa	NKCREEAACK	ALSNVVRT-N	SFKV-MIEAG	SKGSDLNICQ	IADVFGQQNV
P14248	<i>P. falciparum</i>		Multa	NCAREMAGKV	ASESLDER-N	NIFS-MVASG	SKGSININISQ	IISCVGQQNV
P04052	<i>D. melanogaster</i>		Multa	NDARDKTGGS	AKKSLTEY-N	NLKA-MVVS	SKGSNINISQ	VIADVQQNV
Q980R2	<i>S. solfataricus</i>	A'	Multa	DKLRSTAGDI	ASKYLDPF-N	FAYV-MARTG	ARGSVLNITQ	MAALLGQQSV
Q979F4	<i>T. volcanium</i>	A'	Multa	GUVRDESCKI	ASSYLGLS-V	PSVI-MARSG	ARATMLNISE	VAGIVGQQSV
P31813	<i>T. celer</i>	A'	Multa	AEARDNAGKV	AERYLGMS-N	HAVI-MAKTG	ARGKILNITQ	MAALLGQQSI
Q9V114	<i>P. abyssi</i>	A'	Multa	AEARDNAGSI	AEKYLGM-D-N	HTVI-MAKTG	ARGKILNITQ	MAALLGQQSI
Q8ZYQ8	<i>P. aerophilum</i>	A'	Multa	SKVREDAVV	VEKYIDKN-S	EGYL-MAKTG	ARGSIVNIVQ	MVATLGQQTI
Q58445	<i>M. jannaschii</i>	A'	Multa	YDGEEIKLHE	GETEITDY-N	GKKR-IKIG-	-KGYSFEVRK	KSLLCILLKAL
P11512	<i>S. acidocaldarius</i>	A'	Multa	DKLRKVAGEI	ATKYLDPF-N	NVVI-MALTG	ARGSELNITQ	MTALLGQQSV
O27125	<i>M. thermoautotrophicum</i>	A'	Multa	GEARDKSGEI	AESYFDMEN	HAVI-MALTG	ARGSMLNLTQ	ITACVGQQSV
P09846	<i>M. thermoautotrophicum</i> (strain Winter)	A'	Multa	GEAKDKSGEI	AESYFDMEN	HAVI-MALTG	ARGAMLNLTQ	ITACVGQQSV
O28391	<i>A. fulgidus</i>	A'	Multa	GRARDRAGKI	AQRHLGMD-N	AAVI-MAVSG	ARGSMLNLTQ	MAACIGQQSV
P16350	<i>H. halobium</i>	A'	Multa	GKARDSAGDV	AENFEDEN-N	PAVV-MANSG	ARGSMLNLTQ	MAGCVGQQAV
P41556	<i>M. vanielii</i>	A'	Multa	GKARDDTGKV	AEKYLKSE-N	HAAL-MARTG	ARGSLNITM	MAASVGQQSV
Q03585	<i>T. acidophilum</i>	A'	Multa	GUVRDESCKI	ASSYLGLS-V	PSVI-MARSG	ARATMLNISE	VAGIVGQQSV

*Method:

BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD

PSIPRED and PHD = Structure Prediction

Red = Helix Green = Strand Blue = Turn Gray = No cord

Figure C.38 Sequence alignment for *P. furiosus* RNAP subunit A' (sheet 21)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	-----780--	-----790	-----800	-----810	-----820
Q8U0M4	<i>P. furiosus</i>	A'	RMSD	RGKRLYRGFR	--GRVLSHF	PGDLGARAKG	FVNSYKSG	TPQEYFFHAM
Q8U0M4	<i>P. furiosus</i>	A'	PSIPRED	RGKRLYRGFR	--GRVLSHF	PGDLGARAKG	FVNSYKSG	TPQEYFFHAM
Q8U0M4	<i>P. furiosus</i>	A'	PHD	RGKRLYRGFR	--GRVLSHF	PGDLGARAKG	FVNSYKSG	TPQEYFFHAM
P04050	<i>S. cerevisiae</i>	1	RMSD	EGKRIAFGFV	--DRTLPHFS	KDDYSPESKG	FVNSYLRGL	TPQEYFFHAM
P04050	<i>S. cerevisiae</i>	1	PSIPRED	EGKRIAFGFV	--DRTLPHFS	KDDYSPESKG	FVNSYLRGL	TPQEYFFHAM
P04050	<i>S. cerevisiae</i>	1	PHD	EGKRIAFGFV	--DRTLPHFS	KDDYSPESKG	FVNSYLRGL	TPQEYFFHAM
P04050	<i>S. cerevisiae</i>	1	MOE	EGKRIAFGFV	--DRTLPHFS	KDDYSPESKG	FVNSYLRGL	TPQEYFFHAM
P04050	<i>S. cerevisiae</i>	1	BLAST	EGKRIAFGFV	--DRTLPHFS	KDDYSPESKG	FVNSYLRGL	TPQEYFFHAM
P04050	<i>S. cerevisiae</i>	1	BLAST2	EGKRIAFGFV	--DRTLPHFS	KDDYSPESKG	FVNSYLRGL	TPQEYFFHAM
P04050	<i>S. cerevisiae</i>	1	MPsrch	EGKRIAFGFV	--DRTLPHFS	KDDYSPESKG	FVNSYLRGL	TPQEYFFHAM
P04050	<i>S. cerevisiae</i>	1	Multa	EGKRIAFGFV	--DRTLPHFS	KDDYSPESKG	FVNSYLRGL	TPQEYFFHAM
P08775	<i>M. musculus</i>	1	Multa	EGKRIPFGFK	--HRTLPHFI	KDDYGPESRG	FVNSYLAGL	TPTEFFFHAM
P24928	<i>H. sapiens</i>	1	Multa	EGKRIPFGFK	--HRTLPHFI	KDDYGPESRG	FVNSYLAGL	TPTEFFFHAM
P36594	<i>S. pombe</i>	1	Multa	EGKRIPFGFK	--YRTLPHFP	KDDYGPESRG	FVNSYLAGL	TPTEFFFHAM
O35559	<i>C. griseus</i>		Multa	EGKRIPFGFK	--HRTLPHFI	KDDYGPESRG	FVNSYLAGL	TPTEFFFHAM
Q9AVW8	<i>G. theta</i>	1	Multa	EGKRIPFGFR	--KRSPLPHY	YKNFEPETRG	FVNSYIEGL	FPDEFFFHSM
Q9AXN0	<i>G. vacuolata</i>	1	Multa	EGKRISYGF	--RRTLPHFV	LDDLGPESRG	FVNSYLRGL	TATEFFFHAM
Q96452	<i>N. locustae</i>	1	Multa	ESKRIPFGFR	--QRTLPHFV	KNDYGPESRG	FVNSYLSGL	NPEEFFFHAM
P18616	<i>A. thaliana</i>	1	Multa	EGKRIPFGFD	--GRTLPHFT	KDDYGPESRG	FVNSYLRGL	TPQEYFFHAM
Q25312	<i>L. major</i>		Multa	AGSRIPFGFR	--RRTLPHFM	LDDYGETSRG	MATRGYVEGL	QPHFYFFHTM
Q9XZS1	<i>D. guanche</i>		Multa	EGKRIPYGF	--KRTLPHFI	KDDYGPESRG	FVNSYLAGL	TPSEFYFFHAM
Q9BJ59	<i>L. seymouri</i>		Multa	AGSRIPFGFR	--RRTLPHFM	LDDYGETSRG	MATRGYVEGL	QPHFYFFHTM
P16356	<i>C. elegans</i>		Multa	EGKRIPFGFR	--HRTLPHFI	KDDYGPESRG	FVNSYLAGL	TPSEFFFHAM
P17545	<i>T. brucei brucei</i>		Multa	AGSRIPFGFR	--RRTLPHFM	LDDYGETSRG	MATRGYVEGL	KPEFFFHMTM
P14248	<i>P. falciparum</i>		Multa	EGKRIPFGFN	--HRSPLPHI	KFDYGPESRG	FVNSYLSGL	TPQEYFFHAM
P04052	<i>D. melanogaster</i>		Multa	EGKRIPYGF	--KRTLPHFI	KDDYGPESRG	FVNSYLAGL	TPSEFYFFHAM
Q980R2	<i>S. solfataricus</i>	A'	Multa	RGERIKRGM	--TRTLPHFK	PYDISPEARG	FIYSSFTGL	KPTLEFFHAA
Q979F4	<i>T. volcanium</i>	A'	Multa	RGGRLNRGYY	--NRTLPHFK	RGDIGADARG	FVKSSYMTGL	TPTEYFFHSI
P31813	<i>T. celer</i>	A'	Multa	RGKRLYRGYR	--ERVLTHFK	PGDLGARARG	FVNSYKSG	TPQEYFFHAM
Q9V114	<i>P. abyssi</i>	A'	Multa	RGKRLYRGFR	--GRVLSHF	PGDLGARAKG	FVNSYKSG	SPQEYFFHAM
Q8ZYQ8	<i>P. aerophilum</i>	A'	Multa	RGERIRRGFR	--TRTLPHFP	VGDIGAFSGG	FVRNCFRCGL	TPVEYFFHAA
Q58445	<i>M. jannaschii</i>	A'	Multa	GCVGGDKTKK	--MYGIPNWI	KTAPKYIKKE	FLSAYFGSEL	TTPKIRNHGT
P11512	<i>S. acidocaldarius</i>	A'	Multa	RGERIRRGYR	--ERTLSLFK	YGDIAPEARG	FVKNSFMRL	SPYEMFFHAA
O27125	<i>M. thermoautotrophicum</i>	A'	Multa	RGGRIISRGYD	--NRTLPHFK	KGELGAKSRG	FVHSSYKEGL	DPIEFFFHAM
P09846	<i>M. thermoautotrophicum</i> (strain Winter)	A'	Multa	HGGRIISRGYD	--NRTLPHFK	KGELGAKSRG	FVHSSYKEDS	ILLEF----
O28391	<i>A. fulgidus</i>	A'	Multa	RGERISRGYT	YTNRTLPHFK	PGDLGAEARG	FVRSSYKKG	SPVEFFFHAA
P15350	<i>H. halobium</i>	A'	Multa	RGERINRGYE	--DRTLPHFA	PNDLSSEAHG	FVNSYTSGL	TPKEFFFHAM
P41556	<i>M. vannielii</i>	A'	Multa	RGGRVFRGYR	--GRTLPHFG	EGSLDAKSHG	FVRSCYKKG	APTVEYFFHAM
Q03585	<i>T. acidophilum</i>	A'	Multa	RGGRLNRGYY	--NRTLPHFK	RGDIGADARG	FVRSSYMTGL	SPTEYFFHSI

*Method:
 BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
 PSIPRED and PHD = Structure Prediction
 Red = Helix Green = Strand Blue = Turn Gray = No coörd

Figure C.39 Sequence alignment for *P. furiosus* RNAP subunit A' (sheet 22)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	-----830	-----840	-----850	-----860	--,-----
Q8U0M4	P. furiosus	A'	RMSD	GGREGLVDTA	VRTAQSGYMQ	RRLINALQDL	KVEYDGTVRS	PE-GIIVQFK
Q8U0M4	P. furiosus	A'	PSIPRED	GGREGLVDTA	VRTAQSGYMQ	RRLINALQDL	KVEYDGTVRS	PE-GIIVQFK
Q8U0M4	P. furiosus	A'	PHD	GGREGLVDTA	VRTAQSGYMQ	RRLINALQDL	KVEYDGTVRS	PE-GIIVQFK
P04050	S. cerevisiae	1	RMSD	GGREGLIDTA	VKTAETGYIQ	RRLVKALEDI	MVHYDNTTRN	SL-GNVIQFI
P04050	S. cerevisiae	1	PSIPRED	GGREGLIDTA	VKTAETGYIQ	RRLVKALEDI	MVHYDNTTRN	SL-GNVIQFI
P04050	S. cerevisiae	1	PHD	GGREGLIDTA	VKTAETGYIQ	RRLVKALEDI	MVHYDNTTRN	SL-GNVIQFI
P04050	S. cerevisiae	1	MOE	GGREGLIDTA	VKTAETGYIQ	RRLVKALEDI	MVHYDNTTRN	SL-GNVIQFI
P04050	S. cerevisiae	1	BLAST	GGREGLIDTA	VKTAETGYIQ	RRLVKALEDI	MVHYDNTTRN	SL-GNVIQFI
P04050	S. cerevisiae	1	BLAST2	GGREGLIDTA	VKTAETGYIQ	RRLVKALEDI	MVHYDNTTRN	SL-GNVIQFI
P04050	S. cerevisiae	1	MPsrch	GGREGLIDTA	VKTAETGYIQ	RRLVKALEDI	MVHYDNTTRN	SL-GNVIQFI
P04050	S. cerevisiae	1	Multa	GGREGLIDTA	VKTAETGYIQ	RRLVKALEDI	MVHYDNTTRN	SL-GNVIQFI
P08775	M. musculus	1	Multa	GGREGLIDTA	VKTAETGYIQ	RRLIKSMESV	MVKYDATVRN	SI-NQVVQLR
P24928	H. sapiens	1	Multa	GGREGLIDTA	VKTAETGYIQ	RRLIKSMESV	MVKYDATVRN	SI-NQVVQLR
P36594	S. pombe	1	Multa	AGREGLIDTA	VKTAETGYIQ	RRLVKAMEDV	MVRYDGTVRN	AM-GDIIQFA
O35559	C. griseus		Multa	GGREGLIDTA	VKTAETGYIQ	RRLIKSMESV	MVKYDATVRN	SI-NQVVQLR
Q9AVW8	G. theta	1	Multa	GGREGLIDTA	IKTSETGYIQ	RRLSKSMEDI	VVEYDFSVRN	NQ-QEII SFL
Q9AXN0	G. vacuolata	1	Multa	GGREGLIDTA	VKTSETGYIQ	RRLIKALEDV	MVKYDGTVRN	ST-GLIVQFL
Q96452	N. locustae	1	Multa	GGREGLIDTA	CKTAETGYIQ	RRLVKAMEDA	RIEYDGSVRG	S--STVYQFM
P18616	A. thaliana	1	Multa	GGREGLIDTA	VKTSETGYIQ	RRLVKAMEDI	MVKYDGTVRN	SL-GDIVQFL
Q25312	L. major		Multa	AGREGLIDTA	VKTSDTGYLQ	RRLVKALEDV	HASYDGTVRN	AN-QELIQLA
Q9XZS1	D. guanche		Multa	GGREGLIDTA	VKTAETGYIQ	RRLIKAMESV	MVNYDGTVRN	SV-GQLIQLR
Q9BJ59	L. seymouri		Multa	AGREGLIDTA	VKTSDTGYLQ	RRLIKALEDV	HAAYDGTVRN	AN-QELIQLA
P16356	C. elegans		Multa	GGREGLIDTA	VKTAETGYIQ	RRLIKAMESV	MVNYDGTVRN	SL-AQMVQLR
P17545	T. brucei brucei		Multa	AGREGLIDTA	VKTSDTGYLQ	RRLIKALEDV	HAAYDGTVRN	AN-DELIQFM
P14248	P. falciparum		Multa	GGREGIIDTA	CKTSETGYIQ	RRLIKAMEDV	MVQYDRTVRN	SY-GDIIQFL
P04052	D. melanogaster		Multa	GGREGLIDTA	VKTAETGYIQ	RRLIKAMESV	MVNYDGTVRN	SV-GQLIQLR
Q980R2	S. solfataricus	A'	Multa	GGREGLVDTA	VRTSQSGYMQ	RRLINALSDL	RAEYDGTVRS	LY-GEVVQVA
Q979F4	T. volcanium	A'	Multa	GGREGLVDTA	VRTSRSGYMQ	RRLINAFEDL	KVDEERQVKD	TV-GSLIQIK
P31813	T. celer	A'	Multa	GGREGLVDTA	VRTAQSGYMQ	RRLINALQDL	KVEYDGTVRD	PT-GIIVQFR
Q9V114	P. abyssi	A'	Multa	GGREGLVDTA	VRTAQSGYMQ	RRLINALQDL	KVDYDGTVRD	PT-GVIVQFR
Q8ZYQ8	P. aerophilum	A'	Multa	AGRDGLIDTA	VRTAQSGYMQ	RRLINALQDV	YVAYDGTVRF	GG-AMLLQYL
Q58445	M. jannaschii	A'	Multa	SFKE--LSFK	IAKIEEIFDE	DRFIKDIKEM	LKEFGIELK-	-----VRVE
P11512	S. acidocaldarius	A'	Multa	GGREGLVDTA	VKTSQSGYMQ	RRLINALSDL	RIEYDGTVRS	LY-GDIVQVV
O27125	M. thermoautotrophicum	A'	Multa	GGREGLVDTA	IRTAQSGYMQ	RRLVNALQDL	TVDENGRVVD	NR-GVIIQNR
P09846	M. thermoautotrophicum (strain Winter)	A'	Multa	GGREGLVDTA	IRTAQSGYMQ	RRLVNALQDL	TVDENGRVVD	NR-GVIIQTR
O28391	A. fulgidus	A'	Multa	GGREGLVDTA	VRTSQSGYLQ	RRMINALQDL	KVEYDGTVRE	QTSGALVQFR
P15350	H. halobium	A'	Multa	GGREGLVDTA	VRTSKSGYLQ	RRLINALSEL	ETQYDGTVRD	TS-DTIVQFE
P41556	M. vannielii	A'	Multa	GGREGLVDQA	VRTAQSGYMQ	RRLVNALQDI	KAEDYDGTVRD	SR-GIIVQFN
Q03585	T. acidophilum	A'	Multa	GGREGLVDTA	VRTSRSGYMQ	RRLINAFEDL	KVDDSRREVVD	TV-GSLIQIR

*Method:

BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD

PSIPRED and PHD = Structure Prediction

Red = Helix Green = Strand Blue = Turn Gray = No co-ord

Figure C.40 Sequence alignment for *P. furiosus* RNAP subunit A' (sheet 23)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	870-----,, /-----,-----
-------------	--------	----	---------	---

*Method:
BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
PSIPRED and PHD = Structure Prediction
Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.41 Sequence alignment for *P. furiosus* RNAP subunit A' (sheet 24)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	-----,-- 890-----,-- --,900--- 907
Q8U0M4	P. furiosus	A'	RMSD	KTVDVD--RI IVRTL----- KM--RGNNKA KV
Q8U0M4	P. furiosus	A'	PSIPRED	KTVDVD--RI IVRTL----- KM--RGNNKA KV
Q8U0M4	P. furiosus	A'	PHD	KTVDVD--RI IVRTL----- KM--RGNNKA KV
P04050	S. cerevisiae	1	RMSD	DLNLT--HT LDP--LLES-- GSEILGDLKL QV
P04050	S. cerevisiae	1	PSIPRED	DLNLT--HT LDP--LLES-- GSEILGDLKL QV
P04050	S. cerevisiae	1	PHD	DLNLT--HT LDP--LLES-- GSEILGDLKL QV
P04050	S. cerevisiae	1	MOE	DLNLT--HT LDP--LLES-- GSEILGDLKL QV
P04050	S. cerevisiae	1	BLAST	
P04050	S. cerevisiae	1	BLAST2	
P04050	S. cerevisiae	1	MPsrch	
P04050	S. cerevisiae	1	Multa	QSLDTIGGSD AAFEKR---- YR--VDLLNT DH
P08775	M. musculus	1	Multa	QNLATLKPSN KAFEKK---- FR--FDYTNE RA
P24928	H. sapiens	1	Multa	QNLATLKPSN KAFEKK---- FR--FDYTNE RA
P36594	S. pombe	1	Multa	QVFDLSRLST KQFEKK---- YR--IDLME- DR
O35559	C. griseus		Multa	QNLATLKPSN KAFEKK---- FR--FDYTNE RA
Q9AVW8	G. theta	1	Multa	QKIFLGKMSD IEMNMV---- YK--VTQSSP CF
Q9AXN0	G. vacuolata	1	Multa	QNLRTLSMND AKLEQT---- FH--LDFFDL DF
Q96452	N. locustae	1	Multa	QAFPVESRCS EIKLSV---- CS--FKLHGM LE
P18616	A. thaliana	1	Multa	QKLDLSLKMKK SEFDRT---- FK--YEIDDE NW
Q25312	L. major		Multa	QAFPIPHMTN SEMADK---- YR--YEYNDE GS
Q9XZS1	D. guanche		Multa	QNMPTVKLSN KAFEKK---- FK--FDWSNE RY
Q9BJ59	L. seymouri		Multa	QAFPIPHMAN NEMIDK---- YR--YEYNEE GS
P16356	C. elegans		Multa	QNMPTMKPNN AVFERD---- FR--MDLTDN KF
P17545	T. brucei brucei		Multa	QLFPLPFRDD KEMEDT---- YK--YEYDND GT
P14248	P. falciparum		Multa	QIIDLKMLDN KEINKL---- YK--YNFDEE PF
P04052	D. melanogaster		Multa	QNMPTVKLSN KSEFKR---- FK--FDWSNE RL
Q980R2	S. solfataricus	A'	Multa	KTVDVN--RI FERVVG---- WK--A
Q979F4	T. volcanium	A'	Multa	RAIDIN--YI LFDENA---- GR
P31813	T. celer	A'	Multa	KTIDVN--RV VLRTLL----- KL--RGKG
Q9V114	P. abyssi	A'	Multa	KTVDVD--RV IMRTL----- KM--RANNKG
Q8ZYQ8	P. aerophilum	A'	Multa	KVVDIK--SL KMWIR
Q58445	M. jannaschii	A'	Multa	KVYVASIYNH KEFFGR---- IG--YTYANK KE
P11512	S. acidocaldarius	A'	Multa	KSVNVN--RV IERVIG---- WK--R
O27125	M. thermoautotrophicum	A'	Multa	KIVDL--KL VEEIRL----- KS--KG
P09846	M. thermoautotrophicum (strain Winter)	A'	Multa	KIVDL--KL VQEIRL----- KS--GK
O28391	A. fulgidus	A'	Multa	KPVVDV--RI IREVIG---- EV--KK
P15350	H. halobium	A'	Multa	EEVDIDVEHV ADRIIN----- SE--FDSDTQ KA
P41556	M. vannieli	A'	Multa	KGVDLD--KV FTKVIS----- KY--EN
Q03585	T. acidophilum	A'	Multa	KAIDMN--YI LFDEER----- R

*Method:

BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD

PSIPRED and PHD = Structure Prediction

Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.42 Sequence alignment for *P. furiosus* RNAP subunit A' (sheet 25)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	-----	~10~-----	~20~-----
P81412	P. furiosus	B	RMSD	MRGPTVV	DVT PDDLWIV MESYWQEKGL	
P81412	P. furiosus	B	PSIPRED	MRGPTVV	DVT PDDLWIV MESYWQEKGL	
P81412	P. furiosus	B	PHD	MRGPTVV	DVT PDDLWIV MESYWQEKGL	
P08518	S. cerevisiae	2	RMSD	FEDES-A	PITAEDSWAV ISAFFREKGL	
P08518	S. cerevisiae	2	PSIPRED	FEDES-A	PITAEDSWAV ISAFFREKGL	
P08518	S. cerevisiae	2	PHD	FEDES-A	PITAEDSWAV ISAFFREKGL	
P08518	S. cerevisiae	2	MOE	FEDES-A	PITAEDSWAV ISAFFREKGL	
P08518	S. cerevisiae	2	BLAST		ITAEDSWAV ISAFFREKGL	
P08518	S. cerevisiae	2	BLAST2		ITAEDSWAV ISAFFREKGL	
P08518	S. cerevisiae	2	MPsrch			
P08518	S. cerevisiae	2	Multa	MSDLANS EKYYDEDP-- --YGFEDESA	PITAEDSWAV ISAFFREKGL	
P38420	A. thaliana	2	Multa	MEYNEY EPEPQ----- --YVEDDDDE	EITQEDAWAV ISAYFEKGL	
Q42877	L. esculentum	2	Multa	MDMED EYEPQ----- --YNVDDDEE	EITQEDAWAV ISAYFEKGL	
Q02061	S. pombe	2	Multa	MS-- --YEDYQYNE	TLTQEDCWTV ISSFFEETSL	
Q10578	C. elegans	2	Multa	MYDDED EMVNDPMDGD	YIDDSDEISA EAWQEACWVV ISAYFDEKGL	
P08266	D. melanogaster	2	Multa	MMYDNEE ELYEE----- --ENAEIISH	ELWQEACWIV INAYFDEKGL	
P30876	H. sapiens	2	Multa	MYDADE DMQYD----- --EDDDEITP	DLWQEACWIV ISSYFDEKGL	
Q98RY5	G. theta	2	Multa		MNNS NFDIDI LWHI LKLFIIQKGL	
Q27826	E. octocarinatus	2	Multa	MKSGMSRISE GLGSGGRGDY	YNQHKMDIDD EITQEDAWAV IRAYFDQHGL	
Q9YAT8	A. pernix	B	Multa	MSN----- --KGGGQSRG	EIGVDELWTV YKSFIRETGL	
O59226	P. horikoshii	B	Multa		MRGPTVV EVTPDDLWIV MESYWEEKGL	
Q9V115	P. abyssi	B	Multa		MRGPTVV EVTPDDLWIV MESYWQEKGL	
P31814	T. celer	B	Multa	M----- --ASRGPTVV	DVTPDDLWLV MEAYWKEKGL	
Q979F3	T. volcanium	B	Multa	MR EIVDA----- --YFKKYGIV	NHQLDSMNSF YASSDNPNNSV	
Q03587	T. acidophilum	B	Multa	MK EIVDA----- --YFKKYGIV	NHQLDSMNSF YATPDNPNNSV	
P11513	S. acidocaldarius	B	Multa		MLDTERSWAI AESFFKTRGL	
Q980R1	S. solfataricus	B'	Multa			
O28392	A. fulgidus	B'	Multa			
Q60181	M. jannaschii	B'	Multa			
P41557	M. vannielii	B'	Multa			
O27124	M. thermoautotrophicum	B'	Multa			
P09845	M. thermoautotrophicum (st. Winter)	B'	Multa			
Q58444	M. jannaschii	B''	Multa		MAIMREL VDAYFKEHGL	
P41558	M. vannielii	B''	Multa		MESRVI VDAFFRENSL	
O27123	M. thermoautotrophicum	B''	Multa	MFWR KSMKSAWGL	VDAFFDKYDL	
P09844	M. thermoautotrophicum (str. Winter)	B''	Multa		MKKSAGWL VDAFFDKYDL	
O28393	A. fulgidus	B''	Multa		MLDTRVL ARAYFTHERL	
Q980R0	S. solfataricus	B''	Multa	MAS NLTIDERWRV	IEAYFKSKGL	
P15352	Halobacterium sp.	B''	Multa		MLQEDKREL SESYFSKERL	

*Method:
 BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
 PSIPRED and PHD = Structure Prediction
 Red = Helix Green = Strand Blue = Turn Gray = No cord

Figure C.43 Sequence alignment for *P. furiosus* RNAP subunit B (sheet 1)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	-30-----	-40-----	-50-----	-----	-----
P81412	P. furiosus	B	RMSD	VRQHLDSENA	FIERGLQEVV	NEFGGI----	-----KPD	-----
P81412	P. furiosus	B	PSIPRED	VRQHLDSENA	FIERGLQEVV	NEFGGI----	-----KPD	-----
P81412	P. furiosus	B	PHD	VRQHLDSENA	FIERGLQEVV	NEFGGI----	-----KPD	-----
P08518	S. cerevisiae	2	RMSD	VSQQLDSFNQ	FVDYTLQDII	CE-----	-----D	-----
P08518	S. cerevisiae	2	PSIPRED	VSQQLDSFNQ	FVDYTLQDII	CE-----	-----D	-----
P08518	S. cerevisiae	2	PHD	VSQQLDSFNQ	FVDYTLQDII	CE-----	-----D	-----
P08518	S. cerevisiae	2	MOE	VSQQLDSFNQ	FVDYTLQDII	CE-----	-----D	-----
P08518	S. cerevisiae	2	BLAST	VSQQLDSFNQ	FVDYTLQDII	CEDSTLILEQ	LAQHTTESDN	-----
P08518	S. cerevisiae	2	BLAST2	VSQQLDSFNQ	FVDYTLQDII	CEDSTLILEQ	LAQHTTESDN	-----
P08518	S. cerevisiae	2	MPsrch	VSQQLDSFNQ	FVDYTLQDII	CEDSTLILEQ	LAQHTTESDN	-----
P08518	S. cerevisiae	2	Multa	VSQQLDSFNQ	FVDYTLQDII	CEDSTL----	-----ILE	-----QLAQ
P38420	A. thaliana	2	Multa	VRQQLDSFDE	FIQNTMQEIV	DESADI----	-----EIR	-----PESQ
Q42877	L. esculentum	2	Multa	VRQQLDSFDE	FIQNTMQEIV	DESADI----	-----EIR	-----PESQ
Q02061	S. pombe	2	Multa	ARQQLFSFDE	FVQNTMQEIV	DDSTL----	-----TLD	-----QYAQ
Q10578	C. elegans	2	Multa	VRQQLDSFDE	FVQNTMQEIV	EDSPPV----	-----ELQ	-----SENQ
P08266	D. melanogaster	2	Multa	VRQQLDSFDE	FIQMSVQRIV	EDSPA----	-----ELQ	-----AEAQ
P30876	H. sapiens	2	Multa	VRQQLDSFDE	FIQMSVQRIV	EDAPPI----	-----DLQ	-----AEAQ
Q98RY5	G. theta	2	Multa	VDQQLINSFNN	FIQNTMQDII	DDMPPI----	-----TVY	-----IDND
Q27826	E. octocarinatus	2	Multa	VSQQLSSFDR	FLSYTVQDIV	TENNTI----	-----SIQ	-----PELQ
Q9YAT8	A. pernix	B	Multa	ARQHLDSENR	LVTQEIQKII	DSIREV----	-----RPQ	WRIYAGRRRY
O59226	P. horikoshii	B	Multa	VRQHLDSENA	FIERGLQEVV	DEFGGI----	-----KPD	-----
Q9V115	P. abyssi	B	Multa	VRQHLDSENA	FIERGLQEVV	DEFGGI----	-----KPD	-----
P31814	T. celer	B	Multa	VRQHLDSENA	FIERGLQEVV	DEFGGI----	-----KPD	-----
Q979F3	T. volcanium	B	Multa	MQQIVDETKV	SDDADPGFV	LDPAKT----	-----GGH	-----
Q03587	T. acidophilum	B	Multa	MQQIVDETKV	SDDADPGFV	LDPAKT----	-----GGH	-----
P11513	S. acidocaldarius	B	Multa	VRQHLDSEND	FLRNKQQVI	YEQGEI----	-----VTE	-----
Q980R1	S. solfataricus	B'	Multa					
O28392	A. fulgidus	B'	Multa					
Q60181	M. jannaschii	B'	Multa					
P41557	M. vannielii	B'	Multa					
O27124	M. thermoautotrophicum	B'	Multa					
P09845	M. thermoautotrophicum (str. Winter)	B'	Multa					
Q58444	M. jannaschii	B''	Multa	IDHQIESYND	FVENRIQKII	DEVGVI----	-----ETE	-----
P41558	M. vannielii	B''	Multa	VKHHISYND	FVENRIQKII	DEVTGV----	-----ETE	-----
O27123	M. thermoautotrophicum	B''	Multa	VDHHISYND	FVSNRIQKII	DTSEPI----	-----ELE	-----
P09844	M. thermoautotrophicum (str. Winter)	B''	Multa	VDHHISYND	FVSNRIQKII	DTSEPI----	-----ELE	-----
O28393	A. fulgidus	B''	Multa	VKHQIDSFNR	FIDEGLQRVI	DEQGVI----	-----ELD	-----
Q980R0	S. solfataricus	B''	Multa	VRQHLDSEND	FVRNKQKII	DEQGEI----	-----PTE	-----
P15352	Halobacterium sp.	B''	Multa	AEHHFRSYNA	FLEHGMQDVV	TEKERI----	-----ETD	-----

*Method:
 BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
 PSIPRED and PHD = Structure Prediction
 Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.44 Sequence alignment for *P. furiosus* RNAP subunit B (sheet 2)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	-----60-----	-----70-----
P81412	P. furiosus	B	RMSD	-----IP-DF-----	-----EVKFGK IRVGEPEFQE
P81412	P. furiosus	B	PSIPRED	-----IP-DF-----	-----EVKFGK IRVGEPEFQE
P81412	P. furiosus	B	PHD	-----IP-DF-----	-----EVKFGK IRVGEPEFQE
P08518	S. cerevisiae	2	RMSD	-----ST-LILEQL AQHTTESDNI SRKYEISFGK	IYVTKPMVNE
P08518	S. cerevisiae	2	PSIPRED	-----ST-LILEQL AQHTTESDNI SRKYEISFGK	IYVTKPMVNE
P08518	S. cerevisiae	2	PHD	-----ST-LILEQL AQHTTESDNI SRKYEISFGK	IYVTKPMVNE
P08518	S. cerevisiae	2	MOE	-----ST-LILEQL AQHTTESDNI SRKYEISFGK	IYVTKPMVNE
P08518	S. cerevisiae	2	BLAST	-----ISRKY-----	-----EISFGK IYVTKPMVNE
P08518	S. cerevisiae	2	BLAST2	-----ISRKY-----	-----EISFGK IYVTKPMVNE
P08518	S. cerevisiae	2	MPsrch		
P08518	S. cerevisiae	2	Multa	HTTESDNISR K---Y----	-----EISFGK IYVT-KPMVN
P38420	A. thaliana	2	Multa	HNPQHQSDFE ETI--Y----	-----KISFGQ IYLS-KPMT
Q42877	L. esculentum	2	Multa	HNPQHQSDFE ETI--Y----	-----KINFGQ IYLS-KPMT
Q02061	S. pombe	2	Multa	HTGAQGDVTR R---Y----	-----EINFGQ IYLS-RPTMT
Q10578	C. elegans	2	Multa	HLGTDMENPA K---F----	-----SLKFNQ IYLS-KPTHW
P08266	D. melanogaster	2	Multa	HTSGEVETPP R---F----	-----SLKFEQ IYLS-KPTHW
P30876	H. sapiens	2	Multa	HASGEVEEPP R---Y----	-----LLKFEQ IYLS-KPTHW
Q98RY5	G. theta	2	Multa	PGNDLLLSKI R---M----	-----VIRLQG LHLS-KPTYI
Q27826	E. octocarinatus	2	Multa	FAPGSMNESR ENI-KY----	-----DII LGQ VRVNEKPRFK
Q9YAT8	A. pernix	B	Multa	RRGRFEEHEI HLP-DV----	-----KIVFGK ARFKEPMARE
O59226	P. horikoshii	B	Multa	-----IP-DF-----	-----EVKFGK IRVGEPEFQE
Q9V115	P. abyssi	B	Multa	-----IP-DF-----	-----EVKFGK IRVGEPEFQE
P31814	T. celer	B	Multa	-----IP-DF-----	-----EVKFGK VRLGEPEFQE
Q979F3	T. volcanium	B	Multa	-DIRIYYGRV REN-GH----	-----YVGEQT IFVGKPEIKE
Q03587	T. acidophilum	B	Multa	-DIRIYYGRV REN-GH----	-----YVGEQT IFIGKPEIKE
P11513	S. acidocaldarius	B	Multa	-----VP-GL-----	-----KIKLGK IRYEKPSIRE
Q980R1	S. solfataricus	B'	Multa		
O28392	A. fulgidus	B'	Multa		
Q60181	M. jannaschii	B'	Multa		
P41557	M. vannielii	B'	Multa		
O27124	M. thermoautotrophicum	B'	Multa		
P09845	M. thermoautotrophicum (str. Winter)	B'	Multa		
Q58444	M. jannaschii	B''	Multa	-----ITGG-----Y----	-----KVKLGK IKVG-KPVIK
P41558	M. vannielii	B''	Multa	-----IKGG-----Y----	-----KVSFGK VRVT-KPINK
O27123	M. thermoautotrophicum	B''	Multa	-----QGK-----Y----	-----RVETGK VSIE-KPFIK
P09844	M. thermoautotrophicum (str. Winter)	B''	Multa	-----QGQ-----Y----	-----TVETGK VTIE-KPFIK
O28393	A. fulgidus	B''	Multa	-----IPD-----T----	-----YVELGR IKVG-TPIIK
Q980R0	S. solfataricus	B''	Multa	-----IPG-----L----	-----KVR LGK IRIK-KPRVR
P15352	Halobacterium sp.	B''	Multa	-----IGDKE DEE-PV----	-----WVELGD VRAV-TPRVR

*Method:
BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
PSIPRED and PHD = Structure Prediction
Red = Helix Green = Strand Blue = Turn Gray = No cord

Figure C.45 Sequence alignment for *P. furiosus* RNAP subunit B (sheet 3)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	~80~	~90~	~100~	~110~
P81412	P. furiosus	B	RMSD	P-HGQRKPLY	PMDARIRNLT	YAAPLYLEM-	-----IP-----VIK
P81412	P. furiosus	B	PSIPRED	P-HGQRKPLY	PMDARIRNLT	YAAPLYLEM-	-----IP-----VIK
P81412	P. furiosus	B	PHD	P-HGQRKPLY	PMDARIRNLT	YAAPLYLEM-	-----IP-----VIK
P08518	S. cerevisiae	2	RMSD	S-DGVTHALY	PQEARLRNLT	YSSGLFVD--	-----
P08518	S. cerevisiae	2	PSIPRED	S-DGVTHALY	PQEARLRNLT	YSSGLFVD--	-----
P08518	S. cerevisiae	2	PHD	S-DGVTHALY	PQEARLRNLT	YSSGLFVD--	-----
P08518	S. cerevisiae	2	MOE	S-DGVTHALY	PQEARLRNLT	YSSGLFVD--	-----VK
P08518	S. cerevisiae	2	BLAST	S-DGVTHALY	PQEARLRNLT	YSSGLFVDVK	KRTYEADVP GRELKYELIA
P08518	S. cerevisiae	2	BLAST2	S-DGVTHALY	PQEARLRNLT	YSSGLFVD--	-----VK
P08518	S. cerevisiae	2	MPsrch			EL-----	IA-----EES
P08518	S. cerevisiae	2	Multa	ESDGVTHALY	PQEARLRNLT	YSSGLFVDV-	-----KK-----RTY
P38420	A. thaliana	2	Multa	ESDGETATLF	PKAARLRNLT	YSAPLYVDV-	-----TK-----RVI
Q42877	L. esculentum	2	Multa	ESDGETATLF	PKAARLRNLT	YSAPLYVDV-	-----TK-----RVI
Q02061	S. pombe	2	Multa	EADGSTTTF	PQEARLRNLT	YSSPLYVDM-	-----RK-----KVM
Q10578	C. elegans	2	Multa	EKDGPMPMM	PNEARLRNLT	YSAPLYVDI-	-----TK-----VVT
P08266	D. melanogaster	2	Multa	EKDGPSPMM	PNEARLRNLT	YSAPLYVDI-	-----TK-----TKN
P30876	H. sapiens	2	Multa	ERDGPSPMM	PNEARLRNLT	YSAPLYVDI-	-----TK-----TVI
Q98RY5	G. theta	2	Multa	EDDGLVHNLV	PNEARLRNLT	YSSPLYCEI-	-----ST-----SII
Q27826	E. octocarinatus	2	Multa	EFDDKYNVIF	PNEARLRNLT	YQTSLYVEV-	-----KK-----QDK
Q9YAT8	A. pernix	B	Multa	VIESFDRRVT	PFECRVRGIT	YSAPLMVEM-	-----TI-----VQD
O59226	P. horikoshii	B	Multa	P-HGQRKPLY	PMDARIRNLT	YAAPLYLEM-	-----IP-----VIR
Q9V115	P. abyssi	B	Multa	P-HGQRKPLY	PMDARIRNLT	YAAPLYLEM-	-----IP-----VIR
P31814	T. celer	B	Multa	A-QGQRKPLY	PMDARIRNLT	YSAPIYLEL-	-----IP-----VVN
Q979F3	T. volcanium	B	Multa	A-SGASNQIT	PNEARLRDLN	YMAPVTLKL-	-----RI-----VED
Q03587	T. acidophilum	B	Multa	A-SGASNQIT	PNEARLRDLN	YMAPVTLKL-	-----RI-----VED
P11513	S. acidocaldarius	B	Multa	TDKGPMPREIT	PMEARLRNLT	YSSPIFLSM-	-----IP-----VEN
Q980R1	S. solfataricus	B'	Multa				
O28392	A. fulgidus	B'	Multa				
Q60181	M. jannaschii	B'	Multa				
P41557	M. vannielii	B'	Multa				
O27124	M. thermoautotrophicum	B'	Multa				
P09845	M. thermoautotrophicum (str. Winter)	B'	Multa				
Q58444	M. jannaschii	B''	Multa	EADGSIRPIT	PMEARIRDLT	YSVPLYLEM-	-----TP-----IIG
P41558	M. vannielii	B''	Multa	EADGSVKEIT	PMEARIRNLA	YSAPLYLEM-	-----IP-----LIG
O27123	M. thermoautotrophicum	B''	Multa	EADGSKSIY	PTEARLRNLT	YSAHMSLEM-	-----RL-----L--
P09844	M. thermoautotrophicum (str. Winter)	B''	Multa	EADGSKSIY	PTEARLRNLT	YSAHMSLEM-	-----RL-----L--
O28393	A. fulgidus	B''	Multa	EAEGTMPLL	PADARLRNLN	YAAPIYLEC-	-----TV-----Y--
Q980R0	S. solfataricus	B''	Multa	ESDRGEREIS	PMEARLRNLT	YAAPLWLTM-	-----IP-----V--
P15352	Halobacterium sp.	B''	Multa	EADGSEELLY	PQEARLRNLT	YSAPVFMEM-	-----SI-----VRG

*Method:
 BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
 PSIPRED and PHD = Structure Prediction
 Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.46 Sequence alignment for *P. furiosus* RNAP subunit B (sheet 4)

RNAP Subunit Alignments for *Pyrococcus furiosus*

[illegible]

*Method:
BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
PSIPRED and PHD = Structure Prediction
Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.47 Sequence alignment for *P. furiosus* RNAP subunit B (sheet 5)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	150	160	170	180
P81412	<i>P. furiosus</i>	B	RMSD	--GEDPKDPG	GYFIINGSER	VIVSIEDIAP	NRTLVE--K DE-R-----
P81412	<i>P. furiosus</i>	B	PSIPRED	--GEDPKDPG	GYFIINGSER	VIVSIEDIAP	NRTLVE--K DE-R-----
P81412	<i>P. furiosus</i>	B	PHD	--GEDPKDPG	GYFIINGSER	VIVSIEDIAP	NRTLVE--K DE-R-----
P08518	<i>S. cerevisiae</i>	2	RMSD	--KECPFDMG	GYFIINGSEK	VLIAQERSAG	NIVQVF--K KA-A-----
P08518	<i>S. cerevisiae</i>	2	PSIPRED	--KECPFDMG	GYFIINGSEK	VLIAQERSAG	NIVQVF--K KA-A-----
P08518	<i>S. cerevisiae</i>	2	PHD	--KECPFDMG	GYFIINGSEK	VLIAQERSAG	NIVQVF--K KA-A-----
P08518	<i>S. cerevisiae</i>	2	MOE	--KECPFDMG	GYFIINGSEK	VLIAQERSAG	NIVQVF--K KA-A-----
P08518	<i>S. cerevisiae</i>	2	BLAST	--KECPFDMG	GYFIINGSEK	VLIAQERSAG	NIVQVF--K KA-A-----
P08518	<i>S. cerevisiae</i>	2	BLAST2	--SEDDSESG	KVFI--GRLP	IMLRSKNCYL	SEA-T-E--S DL-Y-----
P08518	<i>S. cerevisiae</i>	2	MPsrch	--KECPFDMG	GYFIINGSEK	VLIAQERSAG	NIVQVFK--K AAPS-----
P08518	<i>S. cerevisiae</i>	2	Multa	--KECPFDMG	GYFIINGSEK	VLIAQERSAG	NIVQV-FKKA AP-SP-----
P38420	<i>A. thaliana</i>	2	Multa	--GECPYDQG	GYFIINGSEK	VLIAQEKMS	NHVVV-FKKR QP-NK-----
Q42877	<i>L. esculentum</i>	2	Multa	--GECPLDQG	GYFIINGSEK	VLIAQEKMS	NHVVV-FKKR QP-NK-----
Q02061	<i>S. pombe</i>	2	Multa	--NECPYDQG	GYFIINGSEK	VIAQERSAA	NIVQV-FKKA AP-SP-----
Q10578	<i>C. elegans</i>	2	Multa	--NECPDLP	GYFVINGSEK	VLIAQEKMAT	NTVYV-FSMK D--GK-----
P08266	<i>D. melanogaster</i>	2	Multa	--NECPDLP	GYFVINGSEK	VLIAQEKMAT	NTVYV-FSMK D--GK-----
P30876	<i>H. sapiens</i>	2	Multa	--NECPDLP	GYFIINGSEK	VLIAQEKMAT	NTVYV-FAKK D--SK-----
Q98RY5	<i>G. theta</i>	2	Multa	--GECRDDPG	GYFVINGSEK	VIVAQEQLSW	NKVYI-FQKN EN-LRNTIKI
Q27826	<i>E. octocarinatus</i>	2	Multa	--KECFHDQG	GYFIINGSEK	VIVAHERRMAN	NFVVV-FHKK QP-SK-----
Q9YAT8	<i>A. pernix</i>	B	Multa	--GEDPGDPG	GYFIINGSEK	VIVAQEDLAV	NRIIV-NKST AA-T-----
O59226	<i>P. horikoshii</i>	B	Multa	--GEDPKDPG	GYFIINGSER	VIVSIEDIAP	NRTLVE--K DE-R-----
Q9V115	<i>P. abyssi</i>	B	Multa	--GEDPKDPG	GYFIINGSER	VIVSIEDIAP	NRTLVE--K DE-R-----
P31814	<i>T. celer</i>	B	Multa	--GEDPKDPG	GYFIVNGSER	VIVSIEDLAP	NKTLVE--R DE-R-----
Q979F3	<i>T. volcanium</i>	B	Multa	--GEDPDDPG	GYFIIGGSEK	VIVSLEDLAP	NKIMV-EW-E DR-Y-----
Q03587	<i>T. acidophilum</i>	B	Multa	--GEDPDDPG	GYFIIGGSEK	VIVSLEDLAP	NKIMV-EW-E DR-Y-----
P11513	<i>S. acidocaldarius</i>	B	Multa	--GEDPKDPG	GYFIVNGSEK	MIAQEDLAT	NRVLV-DYKG SG-S-----
Q980R1	<i>S. solfataricus</i>	B'	Multa				
O28392	<i>A. fulgidus</i>	B'	Multa				
Q60181	<i>M. jannaschii</i>	B'	Multa				
P41557	<i>M. vannielii</i>	B'	Multa				
O27124	<i>M. thermoautotrophicum</i>	B'	Multa				
P09845	<i>M. thermoautotrophicum</i> (str. Winter)	B'	Multa				
Q58444	<i>M. jannaschii</i>	B''	Multa	DLGEDPEDPF	GYFIINGTEK	VLITQEDLIP	NRILC-EKAE RS-G-----
P41558	<i>M. vannielii</i>	B''	Multa	NYGEDPKDPL	GYFIVNGSEK	AVVAQEDLIP	NRILC-EKVE KN-N-----
O27123	<i>M. thermoautotrophicum</i>	B''	Multa	EKGEDPADPG	GYFIVNGSER	SIVTMEEIAP	NKII L-ERIG EE-D-----
P09844	<i>M. thermoautotrophicum</i> (str. Winter)	B''	Multa	EKGEDPADLG	GYFIVNASER	SIVTMEEIAP	NKII L-ERIG EE-D-----
O28393	<i>A. fulgidus</i>	B'	Multa	IAGEDPYDPG	GYFIINGTER	ALVTLEDLSP	NKILL-EREE KY-G-----
Q980R0	<i>S. solfataricus</i>	B''	Multa	EIGEDPKDPG	GYFIVNGSER	VIVTQEDLAP	NRVLV-DTGK TG-S-----
P15352	<i>Halobacterium</i> sp.	B''	Multa	EIGEDPADPG	GYFIINGSER	VIMTSEDLAP	NKILA-EYDS KY-G-----

*Method:
 BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
 PSIPRED and PHD = Structure Prediction
 Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.48 Sequence alignment for *P. furiosus* RNAP subunit B (sheet 6)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	190	200	210
P81412	P. furiosus	B	RMSD	-----Q E--KY----I AKVFS-YRH- GYRALVA-VE RR---KDG--		
P81412	P. furiosus	B	PSIPRED	-----Q E--KY----I AKVFS-YRH- GYRALVA-VE RR---KDG--		
P81412	P. furiosus	B	PHD	-----Q E--KY----I AKVFS-YRH- GYRALVA-VE RR---KDG--		
P08518	S. cerevisiae	2	RMSD	-----PSP I--SH----V AEIRS-ALEK GSRFISTLQV KL---YGR--		
P08518	S. cerevisiae	2	PSIPRED	-----PSP I--SH----V AEIRS-ALEK GSRFISTLQV KL---YGR--		
P08518	S. cerevisiae	2	PHD	-----PSP I--SH----V AEIRS-ALEK GSRFISTLQV KL---YGR--		
P08518	S. cerevisiae	2	MOE	-----P SPISH----V AEIRSALEK- GSRFIST-LQ VKLYGREG--		
P08518	S. cerevisiae	2	BLAST	-----P SPISH----V AEIRSALEK- GSRFIST-LQ VK---LYG--		
P08518	S. cerevisiae	2	BLAST2	-----K L--KECPFDM GGYFI-ING- SEKVLIA-QE RS---AGN--		
P08518	S. cerevisiae	2	MPsrch	-----P I--SH----V AEIRSALEK- GSRFIST-LQ VKLYGREG--		
P08518	S. cerevisiae	2	Multa	-----ISH V--AE----I RSALE-KGS- RFISTLQ-VK LY---GR--		
P38420	A. thaliana	2	Multa	-----YAY V--GE----V RSMAE-NQN- RPPSTMF-VR ML---ARAS-		
Q42877	L. esculentum	2	Multa	-----YAF V--AE----V RSMAD-TQN- RPPSTMF-VR ML---SRTS-		
Q02061	S. pombe	2	Multa	-----IAY V--AE----I RSALE-RGS- RLISSMQ-IK LM---AR--		
Q10578	C. elegans	2	Multa	-----YAF K--TE----C RSCLE-NSS- RPTSTMW-VN ML---ARGGG		
P08266	D. melanogaster	2	Multa	-----YAF K--TE----I RSCLE-HSS- RPTSTLW-VN MM---ARGSQ		
P30876	H. sapiens	2	Multa	-----YAY T--GE----C RSCLE-NSS- RPTSTIW-VS ML---ARGGQ		
Q98RY5	G. theta	2	Multa	KNYTVGEGLY F--AE----C RSVAD-YSK- WSPSLLT-VK VC---LCPYK		
Q27826	E. octocarinatus	2	Multa	-----FSW V--AE----V RSQVD-ASN- RPPNQFS-IK MY---SKGQR		
Q9YAT8	A. pernix	B	Multa	-----AR I--TH----T AKAVS-TVL- GLRRQAI-VD RM---SDG--		
O59226	P. horikoshii	B	Multa	-----Q E--KY----I AKVFS-YRH- GYRALVT-VE RK---KDG--		
Q9V115	P. abyssi	B	Multa	-----Q E--KY----I AKVFS-YRH- GYRALVT-VE RK---KDG--		
P31814	T. celer	B	Multa	-----Q N--RI----I AKCFS-YRH- GYRALIT-VE RR---KDG--		
Q979F3	T. volcanium	B	Multa	-----ES K--VE----V AKVFS-QRG- GFRAITS-ME KG---SDG--		
Q03587	T. acidophilum	B	Multa	-----ES K--VE----V SKVFS-QRG- GFRAITS-ME KG---TDG--		
P11513	S. acidocaldarius	B	Multa	-----N I--TH----V AKVTS-SAA- GYRVQVM-IE RL---KDS--		
Q980R1	S. solfataricus	B'	Multa			
O28392	A. fulgidus	B'	Multa			
Q60181	M. jannaschii	B'	Multa			
P41557	M. vannielii	B'	Multa			
O27124	M. thermoautotrophicum	B'	Multa			
P09845	M. thermoautotrophicum (str. Winter)	B'	Multa			
Q58444	M. jannaschii	B''	Multa	-----K I--VD----V AKVFS-TRH- GFRAICT-VE -----RHP		
P41558	M. vannielii	B''	Multa	-----K I--VD----I AKVFS-TRH- GFRAICT-VE -----RSP		
O27123	M. thermoautotrophicum	B''	Multa	-----E N--RA----R AIVTS-IRS- GFRARIS-LE YR---KPRKT		
P09844	M. thermoautotrophicum (str. Winter)	B''	Multa	-----E K--RA----R AIVTS-IRS- GFRARIS-LE YR---KPRKT		
O28393	A. fulgidus	B''	Multa	-----E M--VE----L AKCFS-QRA- GYRALIV-VE -----RGR		
Q980R0	S. solfataricus	B''	Multa	-----N I--TH----T AKIIS-STA- GYRVPTV-IE -----RLK		
P15352	Halobacterium sp.	B''	Multa	-----D E--IQ----V AKTFS-QRR- GYRALVL-VE -----RNR		

*Method:
BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
PSIPRED and PHD = Structure Prediction
Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.49 Sequence alignment for *P. furiosus* RNAP subunit B (sheet 7)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	190--	--200--	--210--
P81412	<i>P. furiosus</i>	B	RMSD	-----Q E--KY----I AKVFS-YRH- GYRALVA-VE RR---KDG--		
P81412	<i>P. furiosus</i>	B	PSIPRED	-----Q E--KY----I AKVFS-YRH- GYRALVA-VE RR---KDG--		
P81412	<i>P. furiosus</i>	B	PHD	-----Q E--KY----I AKVFS-YRH- GYRALVA-VE RR---KDG--		
P08518	<i>S. cerevisiae</i>	2	RMSD	-----PSP I--SH---V AEIRS-ALEK GSRFISTLQV KL---YGR--		
P08518	<i>S. cerevisiae</i>	2	PSIPRED	-----PSP I--SH---V AEIRS-ALEK GSRFISTLQV KL---YGR--		
P08518	<i>S. cerevisiae</i>	2	PHD	-----PSP I--SH---V AEIRS-ALEK GSRFISTLQV KL---YGR--		
P08518	<i>S. cerevisiae</i>	2	MOE	-----P SPISH---V AEIRSALEK- GSRFIST-LQ VKLYGREG--		
P08518	<i>S. cerevisiae</i>	2	BLAST	-----P SPISH---V AEIRSALEK- GSRFIST-LQ VK---LYG--		
P08518	<i>S. cerevisiae</i>	2	BLAST2	-----K L--KECPFDM GGYFI-ING- SEKVLIA-QE RS---AGN--		
P08518	<i>S. cerevisiae</i>	2	MPsrch	-----P I--SH---V AEIRSALEK- GSRFIST-LQ VKLYGREG--		
P08518	<i>S. cerevisiae</i>	2	Multa	-----ISH V--AE----I RSALE-KGS- RFISTLQ-VK LY---GR--		
P38420	<i>A. thaliana</i>	2	Multa	-----YAY V--GE---V RSMAE-NQN- RPPSTMF-VR ML---ARAS-		
Q42877	<i>L. esculentum</i>	2	Multa	-----YAF V--AE---V RSMAD-TQN- RPPSTMF-VR ML---SRTS-		
Q02061	<i>S. pombe</i>	2	Multa	-----IAY V--AE---I RSALE-RGS- RLISSMQ-IK LM---AR--		
Q10578	<i>C. elegans</i>	2	Multa	-----YAF K--TE---C RSCLE-NSS- RPTSTMW-VN ML---ARGGG		
P08266	<i>D. melanogaster</i>	2	Multa	-----YAF K--TE---I RSCLE-HSS- RPTSTLW-VN MM---ARGSQ		
P30876	<i>H. sapiens</i>	2	Multa	-----YAY T--GE---C RSCLE-NSS- RPTSTIW-VS ML---ARGGQ		
Q98RY5	<i>G. theta</i>	2	Multa	KNYTVGEGLY F--AE---C RSVAD-YSK- WSPSLLT-VK VC---LCPYK		
Q27826	<i>E. octocarinatus</i>	2	Multa	-----FSW V--AE---V RSQVD-ASN- RPPNQFS-IK MY---SKGQR		
Q9YAT8	<i>A. pernix</i>	B	Multa	-----AR I--TH---T AKAVS-TVL- GLRRQAI-VD RM---SDG--		
O59226	<i>P. horikoshii</i>	B	Multa	-----Q E--KY----I AKVFS-YRH- GYRALVT-VE RK---KDG--		
Q9V115	<i>P. abyssi</i>	B	Multa	-----Q E--KY----I AKVFS-YRH- GYRALVT-VE RK---KDG--		
P31814	<i>T. celer</i>	B	Multa	-----Q N--RI---I AKCFS-YRH- GYRALIT-VE RR---KDG--		
Q979F3	<i>T. volcanium</i>	B	Multa	-----ES K--VE---V AKVFS-QRG- GFRAITS-ME KG---SDG--		
Q03587	<i>T. acidophilum</i>	B	Multa	-----ES K--VE---V SKVFS-QRG- GFRAITS-ME KG---TDG--		
P11513	<i>S. acidocaldarius</i>	B	Multa	-----N I--TH---V AKVTS-SAA- GYRVQVM-IE RL---KDS--		
Q980R1	<i>S. solfataricus</i>	B'	Multa			
O28392	<i>A. fulgidus</i>	B'	Multa			
Q60181	<i>M. jannaschii</i>	B'	Multa			
P41557	<i>M. vannielii</i>	B'	Multa			
O27124	<i>M. thermoautotrophicum</i>	B'	Multa			
P09845	<i>M. thermoautotrophicum</i> (str. Winter)	B'	Multa			
Q58444	<i>M. jannaschii</i>	B''	Multa	-----K I--VD---V AKVFS-TRH- GFRAICT-VE -----RHP		
P41558	<i>M. vannielii</i>	B''	Multa	-----K I--VD---I AKVFS-TRH- GFRAICT-VE -----RSP		
O27123	<i>M. thermoautotrophicum</i>	B''	Multa	-----E N--RA---R AIVTS-IRS- GFRARIS-LE YR---KPRKT		
P09844	<i>M. thermoautotrophicum</i> (str. Winter)	B''	Multa	-----E K--RA---R AIVTS-IRS- GFRARIS-LE YR---KPRKT		
O28393	<i>A. fulgidus</i>	B''	Multa	-----E M--VE---L AKCFS-QRA- GYRALIV-VE -----RGR		
Q980R0	<i>S. solfataricus</i>	B''	Multa	-----N I--TH---T AKIIS-STA- GYRVPTV-IE -----RLK		
P15352	<i>Halobacterium</i> sp.	B''	Multa	-----D E--IQ---V AKTFS-QRR- GYRALVL-VE -----RNR		

*Method:

BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD

PSIPRED and PHD = Structure Prediction

Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.50 Sequence alignment for *P. furiosus* RNAP subunit B (sheet 8)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	-----220-----	-----230-----
P81412	<i>P. furiosus</i>	B	RMSD	-----ILYVDI	-PNVP--KP- VKFVYVMRAL
P81412	<i>P. furiosus</i>	B	PSIPRED	-----ILYVDI	-PNVP--KP- VKFVYVMRAL
P81412	<i>P. furiosus</i>	B	PHD	-----ILYVDI	-PNVP--KP- VKFVYVMRAL
P08518	<i>S. cerevisiae</i>	2	RMSD	-----EG SSARTIKATL	-PYIK--QD- IPIVIFRAL
P08518	<i>S. cerevisiae</i>	2	PSIPRED	-----EG SSARTIKATL	-PYIK--QD- IPIVIFRAL
P08518	<i>S. cerevisiae</i>	2	PHD	-----EG SSARTIKATL	-PYIK--QD- IPIVIFRAL
P08518	<i>S. cerevisiae</i>	2	MOE	-----SSARTIKATL	-PYIK--QD- IPIVIFRAL
P08518	<i>S. cerevisiae</i>	2	BLAST	-----REG SSARTIKATL	-PYIK--QD- IPIVIFRAL
P08518	<i>S. cerevisiae</i>	2	BLAST2	-----IVQVFK	-KAAP--SP- ISHVAEIRS-
P08518	<i>S. cerevisiae</i>	2	MPsrch	-----SSARTI	KATLPYIKQD IPIVIFRAL
P08518	<i>S. cerevisiae</i>	2	Multa	--EG-SSA-- --R-----	---TIKATL -PYIK--QD- IPIVIFRAL
P38420	<i>A. thaliana</i>	2	Multa	-AKGSSG-- --Q-----	---YIRCTL -PYIR--TE- IPIIIVFRAL
Q42877	<i>L. esculentum</i>	2	Multa	-AKGSSG-- --Q-----	---YIRATL -PYIR--TE- IPIIIVFRAL
Q02061	<i>S. pombe</i>	2	Multa	--NTENSG-- --Q-----	---TIRATL -PYIR--SD- IPIVIVFRAL
Q10578	<i>C. elegans</i>	2	Multa	GGKKTAMG-- --Q-----	---RIIGIL -PYIK--QE- IPIMIVFRAL
P08266	<i>D. melanogaster</i>	2	Multa	NIKKAIG-- --Q-----	---RIIAIL -PYIK--QE- IPIMIVFRAL
P30876	<i>H. sapiens</i>	2	Multa	GAKKSAIG-- --Q-----	---RIVATL -PYIK--QE- VPIIIVFRAL
Q98RY5	<i>G. theta</i>	2	Multa	DQKFSIFKKN LKFNSPEIFL	SGGFYLRMML -PYFK--KD- IPITWIFKAL
Q27826	<i>E. octocarinatus</i>	2	Multa	RAGGSFSGGK IFGQ-----	---TLCASI -PYIN--AD- IPVAILFRAL
Q9YAT8	<i>A. pernix</i>	B	Multa	-----TLEASL	-SRLR--YR- IPLVLLFRAL
O59226	<i>P. horikoshii</i>	B	Multa	-----ILYVGI	-PNVP--RP- VKFVYVMRAL
Q9V115	<i>P. abyssi</i>	B	Multa	-----ILYVGI	-PNVP--KP- VKFVYVMRAL
P31814	<i>T. celer</i>	B	Multa	-----ILYVKL	-PNVP--RP- VKFVYVMRAL
Q979F3	<i>T. volcanium</i>	B	Multa	-----TINVSI	-PSVA--GT- VPLVILMKAL
Q03587	<i>T. acidophilum</i>	B	Multa	-----TINVSI	-PSVA--GT- VPLVILMKAL
P11513	<i>S. acidocaldarius</i>	B	Multa	-----TIQISF	-ATVP--GR- IPFAIIMRAL
Q980R1	<i>S. solfataricus</i>	B'	Multa		
O28392	<i>A. fulgidus</i>	B'	Multa		
Q60181	<i>M. jannaschii</i>	B'	Multa		
P41557	<i>M. vannielii</i>	B'	Multa		
O27124	<i>M. thermoautotrophicum</i>	B'	Multa		
P09845	<i>M. thermoautotrophicum</i> (str. Winter)	B'	Multa		
Q58444	<i>M. jannaschii</i>	B''	Multa	DG-----	----LLYATF -PGMP--GQ- IPLVILMKAL
P41558	<i>M. vannielii</i>	B''	Multa	DG-----	----LLNVSF -PGMP--ST- IPLVILMRAL
O27123	<i>M. thermoautotrophicum</i>	B''	Multa	GV-----	----FLRISF -PYVP--GE- IPLVILLRAL
P09844	<i>M. thermoautotrophicum</i> (str. Winter)	B''	Multa	GV-----	----FLRISF -PYVP--GE- LPLVILLRAL
O28393	<i>A. fulgidus</i>	B''	Multa	NN-----	----ILEVTF -PQLP--KP- IKFVILMRAL
Q980R0	<i>S. solfataricus</i>	B''	Multa	DG-----	----TFHVSF -PAVP--GK- IPFVILMRAL
P15352	<i>Halobacterium</i> sp.	B''	Multa	EG-----	----LLEVSE -PSVS--GS- ISFVILVRAL

*Method:
 BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
 PSIPRED and PHD = Structure Prediction
 Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.51 Sequence alignment for *P. furiosus* RNAP subunit B (sheet 9)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	-----240--	-----250--	-----260
P81412	<i>P. furiosus</i>	B	RMSD	GLERDKDIVE ---AV---G- --NDPEV-QQ I-MFDN-LE- -----DAS		
P81412	<i>P. furiosus</i>	B	PSIPRED	GLERDKDIVE ---AV---G- --NDPEV-QQ I-MFDN-LE- -----DAS		
P81412	<i>P. furiosus</i>	B	PHD	GLERDKDIVE ---AV---G- --NDPEV-QQ I-MFDN-LE- -----DAS		
P08518	<i>S. cerevisiae</i>	2	RMSD	GIIPDGEILE HICYD---V- --NDWQM-LE M-LKPC-VE- -----DGF		
P08518	<i>S. cerevisiae</i>	2	PSIPRED	GIIPDGEILE HICYD---V- --NDWQM-LE M-LKPC-VE- -----DGF		
P08518	<i>S. cerevisiae</i>	2	PHD	GIIPDGEILE HICYD---V- --NDWQM-LE M-LKPC-VE- -----DGF		
P08518	<i>S. cerevisiae</i>	2	MOE	GIIPDGEILE ---HICYDV- --NDWQM-LE M-LKPC-VE- -----DGF		
P08518	<i>S. cerevisiae</i>	2	BLAST	GIIPDGEILE ---HI---CY DVNDWQM-LE M-LKPC-VE- -----DGF		
P08518	<i>S. cerevisiae</i>	2	BLAST2	ALEKSGRFIS ---TL		
P08518	<i>S. cerevisiae</i>	2	MPsrch	GIIPDGEILE ---HI---C- --YDVNDWQM LEMLKPCVE- -----DGF		
P08518	<i>S. cerevisiae</i>	2	Multa	GIIPDGEILE ---HICYDV- --NDWQM-LE M-LKPC-VE- -----D-		
P38420	<i>A. thaliana</i>	2	Multa	GFVADKDILE ---HICYDF- --ADTQM-ME L-LRPS-LE- -----E--		
Q42877	<i>L. esculentum</i>	2	Multa	GFVADKDILE ---HICYDF- --NDTQM-ME L-LRPS-LE- -----E--		
Q02061	<i>S. pombe</i>	2	Multa	GVVPRDRILE ---HICYDF- --NDFQM-LE M-MKPC-IE- -----E--		
Q10578	<i>C. elegans</i>	2	Multa	GFVSDRDILG ---HIIYDF- --NDPEM-ME M-VKPS-LD- -----E--		
P08266	<i>D. melanogaster</i>	2	Multa	GFVADRDILE ---HIIYDF- --DDPEM-ME M-VKPS-LD- -----E--		
P30876	<i>H. sapiens</i>	2	Multa	GFVSDRDILE ---HIIYDF- --EDPEM-ME M-VKPS-LD- -----E--		
Q98RY5	<i>G. theta</i>	2	Multa	GFENEIEILD ---YICYDS- --QDEDL-IS I-VRYI-VE- -----DDK		
Q27826	<i>E. octocarinatus</i>	2	Multa	NCISDKETILD ---RICYDS- --TDSHM-ME A-MRAS-LE- -----E--		
Q9YAT8	<i>A. pernix</i>	B	Multa	GLEKDMIEIML ---AV---S- --ADPDV-QR E-MLPS-LE- -----KGR		
O59226	<i>P. horikoshii</i>	B	Multa	GLERDKDIVD ---AV---G- --NDPEV-QQ I-LFDN-LE- -----DAS		
Q9V115	<i>P. abyssi</i>	B	Multa	GLERDKDIVD ---AV---G- --NDPEV-QQ I-LFDN-LE- -----DAS		
P31814	<i>T. celer</i>	B	Multa	GLLSDRIVE ---AV---S- --DDPRI-QH V-LFDN-LE- -----DAS		
Q979F3	<i>T. volcanium</i>	B	Multa	GLERDVVDVHD ---AI---A- --SVPEM-EP I-IYSN-IE- -----DSK		
Q03587	<i>T. acidophilum</i>	B	Multa	GLERDVVDVHD ---AI---A- --SVPEM-EP I-IYSN-IE- -----DSK		
P11513	<i>S. acidocaldarius</i>	B	Multa	GFVTRDRIVY ---AV---S- --LDPQI-QN E-LLPS-LE- -----QAS		
Q980R1	<i>S. solfataricus</i>	B'	Multa			
O28392	<i>A. fulgidus</i>	B'	Multa			
Q60181	<i>M. jannaschii</i>	B'	Multa			
P41557	<i>M. vannielii</i>	B'	Multa			
O27124	<i>M. thermoautotrophicum</i>	B'	Multa			
P09845	<i>M. thermoautotrophicum</i> (str. Winter)	B'	Multa			
Q58444	<i>M. jannaschii</i>	B''	Multa	GAETDKDIE ---SI---D- --DE-RF-FM E-IVIN-IQE IREEHNINSP		
P41558	<i>M. vannielii</i>	B''	Multa	GAESDREIME ---LI---S- --DEPTV-VM Q-LVAN-LQE AREEHGINTT		
O27123	<i>M. thermoautotrophicum</i>	B''	Multa	GLATDQEIIIT ---SI---S- --DDFNY-QM I-AADD-IQV SLDRNLNKK		
P09844	<i>M. thermoautotrophicum</i> (str. Winter)	B''	Multa	GLATDQEIIIT ---SI---S- --DDFNY-QM I-AADD-IQV SLDKLKLSD		
O28393	<i>A. fulgidus</i>	B''	Multa	GVETDQEIVE ---MV---S- --TNPEI-MK Y-MLQN-I-- --EDSDVGSQ		
Q980R0	<i>S. solfataricus</i>	B''	Multa	GILTDRDIVY ---AV---S- --LDPEV-QN E-LFPS-LE- --QASSIANV		
P15352	<i>Halobacterium</i> sp.	B''	Multa	GLSEDEEIVH ---RV---S- --EDPEI-VK F-MLEN-L-- --EADVQTQ		

*Method:
 BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
 PSIPRED and PHD = Structure Prediction
 Red = Helix Green = Strand Blue = Turn Gray = No coörd

Figure C.52 Sequence alignment for *P. furiosus* RNAP subunit B (sheet 10)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	-----270	-----280	-----290
P81412	<i>P. furiosus</i>	B	RMSD	DI-----T	----TQQEAL EFIGK-L-VA	-PGQAREYRL KRAEYVIDNN
P81412	<i>P. furiosus</i>	B	PSIPRED	DI-----T	----TQQEAL EFIGK-L-VA	-PGQAREYRL KRAEYVIDNN
P81412	<i>P. furiosus</i>	B	PHD	DI-----T	----TQQEAL EFIGK-L-VA	-PGQAREYRL KRAEYVIDNN
P08518	<i>S. cerevisiae</i>	2	RMSD	VI-----Q	----DRETAL DFIGRRG-TA	-LGIKKEKRI QYAKDILQKE
P08518	<i>S. cerevisiae</i>	2	PSIPRED	VI-----Q	----DRETAL DFIGRRG-TA	-LGIKKEKRI QYAKDILQKE
P08518	<i>S. cerevisiae</i>	2	PHD	VI-----Q	----DRETAL DFIGRRG-TA	-LGIKKEKRI QYAKDILQKE
P08518	<i>S. cerevisiae</i>	2	MOE	VI-----Q	----DRETAL DFIGR-RGTA	-LGIKKEKRI QYAKDILQKE
P08518	<i>S. cerevisiae</i>	2	BLAST	VI-----Q	----DRETAL DFIGR-RGTA	-LGIKKEKRI QYAKDILQKE
P08518	<i>S. cerevisiae</i>	2	BLAST2	VI-----Q	----DRETAL DFIGR-RGTA	-LGIKKEKRI QYAKDILQKE
P08518	<i>S. cerevisiae</i>	2	MPsrch	VI-----Q	----DRETAL DFIGR-RGTA	-LGIKKEKRI QYAKDILQKE
P08518	<i>S. cerevisiae</i>	2	Multa	-----G	FVIQDRETAL DFIGR-R-GT	ALGIKKEKRI QYAKDILQKE
P38420	<i>A. thaliana</i>	2	Multa	-----A	FVIQNQLVAL DYIGK-R-GA	TVGVTKKRI KYARDILQKE
Q42877	<i>L. esculentum</i>	2	Multa	-----A	FVIQNQQVAL DYIGK-R-GA	TVGVTKKRI KYAKEILQKE
Q02061	<i>S. pombe</i>	2	Multa	-----A	FVIQDKDIAL DYIGK-R-GS	TTGVTKKRI RYAHDILQKE
Q10578	<i>C. elegans</i>	2	Multa	-----A	FVIQEQNVAL NFIGA-R-GA	KPGVTREKRI KYAREILQKE
P08266	<i>D. melanogaster</i>	2	Multa	-----A	FVVQEQNVAL NFIGA-R-GA	RPGVTKDKRI KYAKEILQKE
P30876	<i>H. sapiens</i>	2	Multa	-----A	FVIQEQNVAL NFIGS-R-GA	KPGVTKEKRI KYAKEVLQKE
Q98RY5	<i>G. theta</i>	2	Multa	ISTQNSLSIS	CPILDQETAL AAIQK-H-VS	KG--PQGIRI KFAYEVLQKE
Q27826	<i>E. octocarinatus</i>	2	Multa	-----G	TPFLTQEDAL SYIGS-R-GT	AGGVGREERI RHAREILKRH
Q9YAT8	<i>A. pernix</i>	B	Multa	AL-----A	---PTKEDAL KYLGN-R-IA	-PGQPLDVRI RRAEFDLDQ
O59226	<i>P. horikoshii</i>	B	Multa	DI-----T	----TQQEAL EYIGK-L-VA	-PGQAREYRL KRAEYVIDNN
Q9V115	<i>P. abyssi</i>	B	Multa	DI-----T	----TQQEAL EYIGK-L-VA	-PGQAREYRL KRAEYVIDNN
P31814	<i>T. celer</i>	B	Multa	DV-----T	----TQEEAL DYIGK-L-SL	-PGQPKRYRL RRAQNIIDNN
Q979F3	<i>T. volcanium</i>	B	Multa	NPKILPPNGV	---NTTEDAI SYLEK-R-FA	-AGQAKEFRD KKISQMLDHS
Q03587	<i>T. acidophilum</i>	B	Multa	NPKVLPNGV	---NTTEDAI SYLEK-R-FA	-AGQAKEFRD KKISQMLDHS
P11513	<i>S. acidocaldarius</i>	B	Multa	SI-----T	----SAEEAL DFIGN-R-VA	-IGQKRENRI QKAEQVIDKY
Q980R1	<i>S. solfataricus</i>	B'	Multa			
O28392	<i>A. fulgidus</i>	B'	Multa			
Q60181	<i>M. jannaschii</i>	B'	Multa			
P41557	<i>M. vannielii</i>	B'	Multa			
O27124	<i>M. thermoautotrophicum</i>	B'	Multa			
P09845	<i>M. thermoautotrophicum</i> (str. Winter)	B'	Multa			
Q58444	<i>M. jannaschii</i>	B''	Multa	E-----	-----DAL EFIGK-R-VA	-PGQAKDYRL KRAETVLCNY
P41558	<i>M. vannielii</i>	B''	Multa	E-----	-----DAL EFIGK-R-VA	-PGQAKDYRL KRAETVLCNY
O27123	<i>M. thermoautotrophicum</i>	B''	Multa	EMKE-LDEED	RREYLIRSAI KYIGN-R-VA	-KGMTEDYRI KRAEDVIDRY
P09844	<i>M. thermoautotrophicum</i> (str. Winter)	B''	Multa	KMEEMDEEE	RREYLIRSAI KYIGN-R-VA	-KGMTEDYRI KRAEDVIDRY
O28393	<i>A. fulgidus</i>	B''	Multa	E-----	-----EAI EYIGK-R-VA	-PGQSREYRY QRAEQVLDQY
Q980R0	<i>S. solfataricus</i>	B''	Multa	D-----	-----DAL DFIGS-R-VA	-IGQKRENRI EKAQQIIDKY
P15352	<i>Halobacterium</i> sp.	B''	Multa	E-----	-----EAI EDLQK-R-VA	-SGQKKNYQL KRANYVIDRY

*Method:

BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD

PSIPRED and PHD = Structure Prediction

Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.53 Sequence alignment for *P. furiosus* RNAP subunit B (sheet 11)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	300-----, , , , , 310- ----320--- ----330--- ----340---
P81412	<i>P. furiosus</i>	B	RMSD	LLPHMGVS-- --PED-RIKK AYYLGMMALK VIELSLGRRP EDDKDHYANK
P81412	<i>P. furiosus</i>	B	PSIPRED	LLPHMGVS-- --PED-RIKK AYYLGMMALK VIELSLGRRP EDDKDHYANK
P81412	<i>P. furiosus</i>	B	PHD	LLPHMGVS-- --PED-RIKK AYYLGMMALK VIELSLGRRP EDDKDHYANK
P08518	<i>S. cerevisiae</i>	2	RMSD	FLPHITQL-- --EGF-ESRK AFFLGMINR LLLCALDRKD QDDRDFHGKK
P08518	<i>S. cerevisiae</i>	2	PSIPRED	FLPHITQL-- --EGF-ESRK AFFLGMINR LLLCALDRKD QDDRDFHGKK
P08518	<i>S. cerevisiae</i>	2	PHD	FLPHITQL-- --EGF-ESRK AFFLGMINR LLLCALDRKD QDDRDFHGKK
P08518	<i>S. cerevisiae</i>	2	MOE	FLPHITQL-- --EGF-ESRK AFFLGMINR LLLCALDRKD QDDRDFHGKK
P08518	<i>S. cerevisiae</i>	2	BLAST	FLPHITQL-- --EGF-ESRK AFFLGMINR LLLCALDRKD QDDRDFHGKK
P08518	<i>S. cerevisiae</i>	2	BLAST2	FLPHITQL-- --EGF-ESRK AFFLGMINR LLLCALDRKD QDDRDFHGKK
P08518	<i>S. cerevisiae</i>	2	MPsrch	FLPHITQL-- --EGF-ESRK AFFLGMINR LLLCALDRKD QDDRDFHGKK
P08518	<i>S. cerevisiae</i>	2	Multa	FLPHITQL-- --EGF-ESRK AFFLGMINR LLLCALDRKD QDDRDFHGKK
P38420	<i>A. thaliana</i>	2	Multa	MLPHVGIG-- --EHC-ETKK AYYFGYIHR LLLCALGRRP EDDRDHYGNK
Q42877	<i>L. esculentum</i>	2	Multa	MLPHVGIG-- --EHC-ETKK AYYFGYIHR LLLCALGRRP EDDRDHYGNK
Q02061	<i>S. pombe</i>	2	Multa	LLPHITTM-- --EGF-ETRK AFFLGMIHR MLLCALERRE PDDRDFHGKK
Q10578	<i>C. elegans</i>	2	Multa	LLPHVGVS-- --EHC-ETKK AYYFGYIHR LLLCALGRRP EDDRDHYGNK
P08266	<i>D. melanogaster</i>	2	Multa	MLPHVGVS-- --DFC-ETKK AYYFGYIHR LLLCALGRRP EDDRDHYGNK
P30876	<i>H. sapiens</i>	2	Multa	MLPHVGVS-- --DFC-ETKK AYYFGYIHR LLLCALGRRP EDDRDHYGNK
Q98RY5	<i>G. theta</i>	2	Multa	FLPHIGIG-- --QGF-ELRK GYFFGYIINK LISTQIGRRK MDDRDFHYGK
Q27826	<i>E. octocarinatus</i>	2	Multa	FLPHVSTK-- --DSC-ESQK SYFFGYIHR LLLCALGRRP EDDRDHYGK
Q9YAT8	<i>A. pernix</i>	B	Multa	LLPHLGNR-- --PED-RIKK AYYLAEMANR VLQYLGRGK EDDKDFLGK
O59226	<i>P. horikoshii</i>	B	Multa	LLPHMGVT-- --PED-RIRK AYYLGMMALK VIELSLGRRP EDDKDHYANK
Q9V115	<i>P. abyssi</i>	B	Multa	LLPHIGVS-- --PED-RIRK AYYLGMMALK VIELSLGRRP EDDKDHYANK
P31814	<i>T. celer</i>	B	Multa	LLPHMGVE-- --EKD-RKAK AYYLGMMALR VIELSLGRRP EDDKDHYANK
Q979F3	<i>T. volcanium</i>	B	Multa	LLPHLGDS-- --ASD-RIRK AYYLGMMALR VIELSLGRRP EDDKDHYANK
Q03587	<i>T. acidophilum</i>	B	Multa	LLPHLGDS-- --PSD-RIRK AYYLGMMALR VIELSLGRRP EDDKDHYANK
P11513	<i>S. acidocaldarius</i>	B	Multa	FLPHLGTS-- --PED-RKKK GYYLASAVNK ILELYLGRRE PDDKDHYANK
Q980R1	<i>S. solfataricus</i>	B'	Multa	FLPHLGTS-- --PED-RKKK GYYLASAVNK ILELYLGRRE PDDKDHYANK
O28392	<i>A. fulgidus</i>	B'	Multa	FLPHLGTS-- --PED-RKKK GYYLASAVNK ILELYLGRRE PDDKDHYANK
Q60181	<i>M. jannaschii</i>	B'	Multa	FLPHLGTS-- --PED-RKKK GYYLASAVNK ILELYLGRRE PDDKDHYANK
P41557	<i>M. vannielii</i>	B'	Multa	FLPHLGTS-- --PED-RKKK GYYLASAVNK ILELYLGRRE PDDKDHYANK
O27124	<i>M. thermoautotrophicum</i>	B'	Multa	FLPHLGTS-- --PED-RKKK GYYLASAVNK ILELYLGRRE PDDKDHYANK
P09845	<i>M. thermoautotrophicum</i> (str. Winter)	B'	Multa	FLPHLGTS-- --PED-RKKK GYYLASAVNK ILELYLGRRE PDDKDHYANK
Q58444	<i>M. jannaschii</i>	B''	Multa	LLPHLGVT-- --KED-FPKK IRFLGIMARN ALELYFGYRG EDDKDHYANK
P41558	<i>M. vannielii</i>	B''	Multa	LLPHMGIE-- --SEK-LGAK KYLGRMAKN SIELYLGSRV EDDKDHYANK
O27123	<i>M. thermoautotrophicum</i>	B''	Multa	LLPHIGTE-- --PEK-RLEK ATYLAEMTEM LLQVIGERR PHDKDHYTNK
P09844	<i>M. thermoautotrophicum</i> (str. Winter)	B''	Multa	LLPHIGTE-- --PDK-RLEK AVYLAEMTEM LLQVIGERR PHDKDHYTNK
O28393	<i>A. fulgidus</i>	B''	Multa	FLPHLGTD-- --KES-RAAK AFFLARMMAE VIELYLYGRR EDDKDHYANK
Q980R0	<i>S. solfataricus</i>	B''	Multa	FLPHLGTS-- --AED-RKKK AYYLAYAI SK VIELYLYGRR EDDKDHYANK
P15352	<i>Halobacterium</i> sp.	B''	Multa	LLPHLHEDGV EEEET-RINK AYYLCRMAEA CFELALGRRE ADDKDHYANK

*Method:
 BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
 PSIPRED and PHD = Structure Prediction
 Red = Helix Green = Strand Blue = Turn Gray = No coörd

Figure C.54 Sequence alignment for *P. furiosus* RNAP subunit B (sheet 12)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	----350----	----360----	----,370--	----380--	----,390
P81412	P. furiosus	B	RMSD	RLRLAGDLLK	DLFRVAFSQL	VKDI-QYQMT	KTYQRRGDKY	TFG-NIH-RF
P81412	P. furiosus	B	PSIPRED	RLRLAGDLLK	DLFRVAFSQL	VKDI-QYQMT	KTYQRRGDKY	TFG-NIH-RF
P81412	P. furiosus	B	PHD	RLRLAGDLLK	DLFRVAFSQL	VKDI-QYQMT	KTYQRRGDKY	TFG-NIH-RF
P08518	S. cerevisiae	2	RMSD	RDLAAGPLLA	QLFKTLFKKL	TKDI-F----	-----	--R-YMQ-RT
P08518	S. cerevisiae	2	PSIPRED	RDLAAGPLLA	QLFKTLFKKL	TKDI-F----	-----	--R-YMQ-RT
P08518	S. cerevisiae	2	PHD	RDLAAGPLLA	QLFKTLFKKL	TKDI-F----	-----	--R-YMQ-RT
P08518	S. cerevisiae	2	MOE	RDLAAGPLLA	QLFKTLFKKL	TKDI-----	-----	-FR-YMQ-RT
P08518	S. cerevisiae	2	BLAST	RDLAAGPLLA	QLFKTLFKKL	TKDI-FRYMQ	RTVEEA----	-----H-DF
P08518	S. cerevisiae	2	BLAST2					
P08518	S. cerevisiae	2	MPsrch	RDLAAGPLLA	QLFKTLFKKL	TKDIFRY-MQ	RTVE-EAHDF	NMKLAINAKT
P08518	S. cerevisiae	2	Multa	RDLAAGPLLA	QLFKTLFKKL	TKDI-FRYMQ	RTVEE-AHDF	NMK-LA----
P38420	A. thaliana	2	Multa	RDLAAGPLLG	GLFRMLFRKL	TRDV-RSYVQ	KCVDN-GKEV	NLQ-FA----
Q42877	L. esculentum	2	Multa	RDLAAGPLLG	GLFRMLFRKL	TRDV-RGYVQ	KCVDN-GKDV	NLQ-FA----
Q02061	S. pombe	2	Multa	RDLAAGPLLA	SLFRMLFRKM	TRDV-YKYMQ	KCVET-NREF	NLT-LA----
Q10578	C. elegans	2	Multa	RDLAAGPLLA	FLFRSLFRNL	LKEM-RMTAQ	KYINK-NDDF	ALD-VC----
P08266	D. melanogaster	2	Multa	RDLAAGPLLA	FLFRGLFRNL	MKEV-RMYTQ	KFIDR-GKDF	NLE-LA----
P30876	H. sapiens	2	Multa	RDLAAGPLLA	FLFRGMFKNL	LKEV-RIYAQ	KFIDR-GKDF	NLE-LA----
Q98RY5	G. theta	2	Multa	RLDTVGPLLS	QLFRQLLGKV	LKEF-KTNVQ	KKIHLRDKKF	ELK-DF----
Q27826	E. octocarinatus	2	Multa	RLDMAGSLLG	MLFRNLFRRF	VKEA-GRYLK	AQADK-GDSL	NIT-KA----
Q9YAT8	A. pernix	B	Multa	RLRLAGDLIA	ELFRSLEAL	ANTM-AQKLE	EQLAER-RQI	RDL-FL----
O59226	P. horikoshii	B	Multa	RLRLAGDLLK	DLFRVAFSQL	IKDI-QYQMT	KTYQRRGEKY	VFS-NIH-RF
Q9V115	P. abyssi	B	Multa	RLRLAGDLLK	DLFRVAFSQL	VKDI-QYQMT	KTYQRRGEKY	VFG-NIH-RF
P31814	T. celer	B	Multa	RLKLAGDLLM	DLFRVAFSQL	VKDM-QYQMT	KTYQRRGERY	TFE-NIQ-RF
Q979F3	T. volcanium	B	Multa	RIKLAGDLMD	ELFRSAFQSV	MKDL-KYQLE	KTYNRK-RGI	KIR-PA----
Q03587	T. acidophilum	B	Multa	RIKLAGDLMD	ELFRSAFQSV	MKDL-KYQLE	KTYNRK-RGI	KIR-PA----
P11513	S. acidocaldarius	B	Multa	RVRLAGDLFT	SLFRVAFKAF	VKDL-VYQLE	KS-KVRGRRL	SLT-AL----
Q980R1	S. solfataricus	B'	Multa					
O28392	A. fulgidus	B'	Multa					
Q60181	M. jannaschii	B'	Multa					
P41557	M. vannielii	B'	Multa					
O27124	M. thermoautotrophicum	B'	Multa					
P09845	M. thermoautotrophicum (str. Winter)	B'	Multa					
Q58444	M. jannaschii	B''	Multa	RAKLAGDLME	DLFRYAFSQL	VKDI-KYQLE	RQTLR-NKTP	SIQ-AA----
P41558	M. vannielii	B''	Multa	RLKLAGDLME	DLFRHSFNQL	IKDI-KYQLE	RQAIR-NKEP	SIQ-AA----
O27123	M. thermoautotrophicum	B''	Multa	RLRVSGDLME	DLFRVAFSQL	SRDM-SYQLE	RSLAR-GKEP	SVK-QA----
P09844	M. thermoautotrophicum (str. Winter)	B''	Multa	RLRVSGDLME	DLFRVAFSQL	TRDM-SYQLE	RSLAR-GKEP	SVK-QA----
O28393	A. fulgidus	B''	Multa	RLKLAGDLME	ELFRVAFSQL	IKDV-KYQLE	RAKIR-GRPL	KMS-TA----
Q980R0	S. solfataricus	B''	Multa	RLRLAGDLFA	SLFRVAFKAF	VKDL-TYQLE	KSKVR-GRKL	ALK-AL----
P15352	Halobacterium sp.	B''	Multa	RLKVSGDLMK	DLFRTALNKL	ARDV-KYQLE	RANMR-NREL	TVN-TV----

*Method:
BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
PSIPRED and PHD = Structure Prediction
Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.55 Sequence alignment for *P. furiosus* RNAP subunit B (sheet 13)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	-----400-----410-----	
P81412	P. furiosus	B	RMSD	-IR----- -NSIRPDVLT ERIEHALATG AWP-----	-----GG
P81412	P. furiosus	B	PSIPRED	-IR----- -NSIRPDVLT ERIEHALATG AWP-----	-----GG
P81412	P. furiosus	B	PHD	-IR----- -NSIRPDVLT ERIEHALATG AWP-----	-----GG
P08518	S. cerevisiae	2	RMSD	-VEEAHDFNM KLAINAKTIT SGLKYALATG NWGEQKKAMS SR-----	
P08518	S. cerevisiae	2	PSIPRED	-VEEAHDFNM KLAINAKTIT SGLKYALATG NWGEQKKAMS SR-----	
P08518	S. cerevisiae	2	PHD	-VEEAHDFNM KLAINAKTIT SGLKYALATG NWGEQKKAMS SR-----	
P08518	S. cerevisiae	2	MOE	-VEEAHDFNM KLAINAKTIT SGLKYALATG NWGEQKKAMS SR-----	
P08518	S. cerevisiae	2	BLAST	NMK----- -LAINAKTIT SGLKYALATG NWGEQKKAM-	-----SS
P08518	S. cerevisiae	2	BLAST2		
P08518	S. cerevisiae	2	MPsrch	-IT----- -SGLK-YALA T-GNWGEQKK A-M-----	-----SS
P08518	S. cerevisiae	2	Multa	----- -INAKTIT SGLKYALATG NWGEQKKAM-	-----SS
P38420	A. thaliana	2	Multa	----- -IKAKTIT SGLKYSLATG NWG-QANAA-	-----GT
Q42877	L. esculentum	2	Multa	----- -IKAKTIT SGLKYSLATG NWG-QANAA-	-----GT
Q02061	S. pombe	2	Multa	----- -VKSNIIT NGLRYSLATG NWGDQKRSM-	-----VN
Q10578	C. elegans	2	Multa	----- -VKTSTIT RGLTYSLATG NWGDQKKAH-	-----QS
P08266	D. melanogaster	2	Multa	----- -IKTNIIT DGLRYSLATG NWGDQKKAH-	-----QA
P30876	H. sapiens	2	Multa	----- -IKTRIIIS DGLKYSLATG NWGDQKKAH-	-----QA
Q98RY5	G. theta	2	Multa	----- -FKSNYVT TGIQYALSTG NWGTDKQSL-	-----
Q27826	E. octocarinatus	2	Multa	----- -FKKDIIS DGLKYALATG NWGTNAAGE-	-----VS
Q9YAT8	A. pernix	B	Multa	----- -VRPTIIS DRLKGAIATG NWG-----	-----QD
O59226	P. horikoshii	B	Multa	-IR----- -NSIRPDVLT ERIEHALATG AWP-----	-----GG
Q9V115	P. abyssi	B	Multa	-IR----- -NSIRPDVLT ERIEHALATG AWP-----	-----GG
P31814	T. celer	B	Multa	-VR----- -NSIRPDVLS ERIEHALATG SWP-----	-----GG
Q979F3	T. volcanium	B	Multa	----- -VRQDLLT QRI LHAMSTG NWI-----	-----GG
Q03587	T. acidophilum	B	Multa	----- -VRQDLLT QRV LHAMSTG NWI-----	-----GG
P11513	S. acidocaldarius	B	Multa	----- -VRADIIT ERIRHALATG NWV-----	-----GG
Q980R1	S. solfataricus	B'	Multa		
O28392	A. fulgidus	B'	Multa		
Q60181	M. jannaschii	B'	Multa		
P41557	M. vannielii	B'	Multa		
O27124	M. thermoautotrophicum	B'	Multa		
P09845	M. thermoautotrophicum (str. Winter)	B'	Multa		
Q58444	M. jannaschii	B''	Multa	----- -VRSDILT ERIKHAMATG TWV-----	-----GG
P41558	M. vannielii	B''	Multa	----- -VRSDVLT ERMHRAMATG NWV-----	-----GG
O27123	M. thermoautotrophicum	B''	Multa	----- -VRSDVLT ENLKHAATG NWV-----	-----GG
P09844	M. thermoautotrophicum (str. Winter)	B''	Multa	----- -VRSDVLS ENLKHAATG NWV-----	-----GG
O28393	A. fulgidus	B''	Multa	----- -VRSDVLT DRLMHPMATG NWV-----	-----GG
Q980R0	S. solfataricus	B''	Multa	----- -VRPDIVT ERIRHALATG NWV-----	-----GG
P15352	Halobacterium sp.	B''	Multa	----- -VRSDVLT ERLEHPIATG NWV-----	-----GG

*Method:
BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
PSIPRED and PHD = Structure Prediction
Red = Helix Green = Strand Blue = Turn Gray = No cord

Figure C.56 Sequence alignment for *P. furiosus* RNAP subunit B (sheet 14)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	-420-----	-430-----	-440---,, , , ,---450	-----460
P81412	<i>P. furiosus</i>	B	RMSD	RTGVSQLLDR	TNYMSTLSHL	RRVTSPL---	---DREQPHF EARDLHGTHW
P81412	<i>P. furiosus</i>	B	PSIPRED	RTGVSQLLDR	TNYMSTLSHL	RRVTSPL---	---DREQPHF EARDLHGTHW
P81412	<i>P. furiosus</i>	B	PHD	RTGVSQLLDR	TNYMSTLSHL	RRVTSPL---	---DREQPHF EARDLHGTHW
P08518	<i>S. cerevisiae</i>	2	RMSD	-AGVSQVLNR	YTYSTLSHL	RRNTNPIGRD	GKL--AKPRQ L----HNTHW
P08518	<i>S. cerevisiae</i>	2	PSIPRED	-AGVSQVLNR	YTYSTLSHL	RRNTNPIGRD	GKL--AKPRQ L----HNTHW
P08518	<i>S. cerevisiae</i>	2	PHD	-AGVSQVLNR	YTYSTLSHL	RRNTNPIGRD	GKL--AKPRQ L----HNTHW
P08518	<i>S. cerevisiae</i>	2	MOE	A-GVSQVLNR	YTYSTLSHL	RRNTNPIGRD	GKL--AKP-- --RQLHNTHW
P08518	<i>S. cerevisiae</i>	2	BLAST	RAGVSQVLNR	YTYSTLSHL	RRNTNPI---	---GRDGKLA KPRQLHNTHW
P08518	<i>S. cerevisiae</i>	2	BLAST2	RAGVSQVLNR	YTYSTLSHL	RRNTNPI---	---GRDGKLA KPRQLHNTHW
P08518	<i>S. cerevisiae</i>	2	MPsrch	RAGVSQVLNR	YTYSTLSHL	RRNTNPI---	---GRDGKLA KPRQLHNTHW
P08518	<i>S. cerevisiae</i>	2	Multa	RAGVSQVLNR	YTYSTLSHL	RRNTNPI---	---GRDGKLA KPRQLHNTHW
P38420	<i>A. thaliana</i>	2	Multa	RAGVSQVLNR	LYASTLSHL	RRLNSPI---	---GREGKLA KPRQLHNSQW
Q42877	<i>L. esculentum</i>	2	Multa	RAGVSQVLNR	LYASTLSHL	RRLNSPI---	---GREGKLA KPRQLHNSQW
Q02061	<i>S. pombe</i>	2	Multa	RVGSQVLNR	YTFASTLSHL	RRNTNPI---	---GRDGKLA KPRQLHNTHW
Q10578	<i>C. elegans</i>	2	Multa	RAGVSQVLNR	LYTATLSHL	RRANSPI---	---GREGKLA KPRQLHNTQW
P08266	<i>D. melanogaster</i>	2	Multa	RAGVSQVLNR	LYFASTLSHL	RRVNSPI---	---GRDGKLA KPRQLHNTLW
P30876	<i>H. sapiens</i>	2	Multa	RAGVSQVLNR	LYFASTLSHL	RRLNSPI---	---GRDGKLA KPRQLHNTLW
Q98RY5	<i>G. theta</i>	2	Multa	RTGISQVLNR	LSYSTLSHL	RRVNSPS---	---AKGNKIT KPRHLHNSHW
Q27826	<i>E. octocarinatus</i>	2	Multa	RTGVSQVLNR	LYFASTLSHL	RRLNTPL---	---QKSGKLT KPRQLHNTHW
Q9YAT8	<i>A. pernix</i>	B	Multa	RTGISQQLDR	TNWISLLSHL	RRVISPL---	---SRGQAHF EARDLHGSHW
O59226	<i>P. horikoshii</i>	B	Multa	RTGVSQLLDR	TNYVSTLSHL	RRVTSPL---	---DREQPHF EARDLHGTHW
Q9V115	<i>P. abyssi</i>	B	Multa	RTGVSQLLDR	TNYVSTLSHL	RRVTSPL---	---DREQPHF EARDLHGTHW
P31814	<i>T. celer</i>	B	Multa	RTGVSQLLDR	TNYISTLSHL	RRVTSPL---	---SREQPHF EARDLHGTHW
Q979F3	<i>T. volcanium</i>	B	Multa	RTGVSQLLDR	VSNLSTISHL	RRIISPL---	---TRTQPHF EARDLHPTQW
Q03587	<i>T. acidophilum</i>	B	Multa	RTGVSQLLDR	VSNLSTISHL	RRIISPL---	---TRTQPHF EARDLHPTQW
P11513	<i>S. acidocaldarius</i>	B	Multa	RTGVSQLLDR	TNWLMSLSHL	RRVSSSL---	---ARGQPNF EARDLHGTHW
Q980R1	<i>S. solfataricus</i>	B'	Multa		MLYGIN ERIVEKTYE	-MGVVPV---	---EEVIRRV TEGGEDQNEY
O28392	<i>A. fulgidus</i>	B'	Multa				
Q60181	<i>M. jannaschii</i>	B'	Multa				MKNLA
P41557	<i>M. vannielii</i>	B'	Multa				MDTL
O27124	<i>M. thermoautotrophicum</i>	B'	Multa				
P09845	<i>M. thermoautotrophicum</i> (str. Winter)	B'	Multa				
Q58444	<i>M. jannaschii</i>	B''	Multa	KTGVSQLLDR	TSYLATNSQL	RRIVSPL---	---SRSQPHF EARELHGTHW
P41558	<i>M. vannielii</i>	B''	Multa	RTGVSQLLDR	TSYLATVSQL	RRVVSPL---	---SRSQPHF EARDLHATQW
O27123	<i>M. thermoautotrophicum</i>	B''	Multa	RAGVSQLLDR	TSYMGTLSHM	RRVVSPL---	---SRSQPHF EARDLHPTQF
P09844	<i>M. thermoautotrophicum</i> (str. Winter)	B''	Multa	RAGVSQLLDR	TSYMGTLSHM	RRVVSPL---	---SRSQPHF EARDLHPTQF
O28393	<i>A. fulgidus</i>	B''	Multa	RTGVSQLLDR	HNYSVISHL	RRVVSPL---	---SRSQPHF EARDLHPTQW
Q980R0	<i>S. solfataricus</i>	B''	Multa	RTGVSQLLDR	TNWLMSLSHL	RRVSSSL---	---ARGQPNF EARDLHGTHW
P15352	<i>Halobacterium</i> sp.	B''	Multa	RSGVSQVLDR	TDFMGVLSHL	RRLRSPL---	---SRSQPHF EARDLHATQW

*Method:
BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
PSIPRED and PHD = Structure Prediction
Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.57 Sequence alignment for *P. furiosus* RNAP subunit B (sheet 15)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	-----470	-----480	-----490	-----500	-,-----
P81412	P. furiosus	B	RMSD	GRICPTETPE	GPNCGLVKNL	ALMAQITTAV	PEEEIREYLM	S-LGIVPIE-
P81412	P. furiosus	B	PSIPRED	GRICPTETPE	GPNCGLVKNL	ALMAQITTAV	PEEEIREYLM	S-LGIVPIE-
P81412	P. furiosus	B	PHD	GRICPTETPE	GPNCGLVKNL	ALMAQITTAV	PEEEIREYLM	S-LGIVPIE-
P08518	S. cerevisiae	2	RMSD	GLVCPAETPE	GQACGLVKNL	SLMSCISVGT	DPMPITFSL	E-WGMEPLE-
P08518	S. cerevisiae	2	PSIPRED	GLVCPAETPE	GQACGLVKNL	SLMSCISVGT	DPMPITFSL	E-WGMEPLE-
P08518	S. cerevisiae	2	PHD	GLVCPAETPE	GQACGLVKNL	SLMSCISVGT	DPMPITFSL	E-WGMEPLE-
P08518	S. cerevisiae	2	MOE	GLVCPAETPE	GQACGLVKNL	SLMSCISVGT	DPMPITFSL	E-WGMEPLE-
P08518	S. cerevisiae	2	BLAST	GLVCPAETPE	GQACGLVKNL	SLMSCISVGT	DPMPITFSL	E-WGMEPLE-
P08518	S. cerevisiae	2	BLAST2	GLVCPAETPE	GQACGLVKNL	SLMSCISVGT	DPMPITFSL	E-WGMEPLE-
P08518	S. cerevisiae	2	MPsrch	GLVCPAETPE	GQACGLVKNL	SLMSCISVGT	DPMPITFSL	E-WGMEPLE-
P08518	S. cerevisiae	2	Multa	GLVCPAETPE	GQACGLVKNL	SLMSCISVGT	DPMPITFSL	E-WGMEPLE-
P38420	A. thaliana	2	Multa	GMMCPAETPE	GQACGLVKNL	ALMVYITVGS	AAYPILEFLE	E-WGTENFEE
Q42877	L. esculentum	2	Multa	GMMCPAETPE	GQACGLVKNL	ALMVYITVGS	AAYPILEFLE	E-WGTENFEE
Q02061	S. pombe	2	Multa	GMVCPAETPE	GQACGLVKNL	SLMSYVSVGS	PSAPIIEFLE	E-WGLETLED
Q10578	C. elegans	2	Multa	GMVCPAETPE	GQAVGLVKNL	ALMAYISVGS	LPEPILEFLE	E-WSMENLEE
P08266	D. melanogaster	2	Multa	GMLCPAETPE	GAAVGLVKNL	ALMAYISVGS	QPSPILEFLE	E-WSMENLEE
P30876	H. sapiens	2	Multa	GMVCPAETPE	GHAUGLVKNL	ALMAYISVGS	QPSPILEFLE	E-WSMENLEE
Q98RY5	G. theta	2	Multa	GYICPVETPE	GHSCGLVKNL	ALTAMITIGT	SPLFIEKLE	N-LGLESLEE
Q27826	E. octocarinatus	2	Multa	GMVCPAETPE	GQACGLVKNL	SLMAFVSVGS	AAERLREILE	N-YGTENLVLS
Q9YAT8	A. pernix	B	Multa	GRYCPFETPE	GINCGLVKNL	ALSAYISVGV	DESUVVEILV	KDLGVAPLER
O59226	P. horikoshii	B	Multa	GRICPTETPE	GPNCGLVKNL	ALMAQITTAV	PEEEVREYLM	S-LGIVPIE-
Q9V115	P. abyssi	B	Multa	GRICPTETPE	GPNCGLVKNL	ALMAQITTAV	PEEEVKEYLM	S-LGIVPIE-
P31814	T. celer	B	Multa	GRICPTETPE	GPNCGLVKNL	ALMSQITTVG	PEEEVREYLE	R-LGVVPIE-
Q979F3	T. volcanium	B	Multa	GRICPNETPE	GQNCGLVKNL	ALLINVTQGI	DSETVMEILK	---GMDVKEV
Q03587	T. acidophilum	B	Multa	GRICPNETPE	GQNCGLVKNL	ALLINVTQGI	DPDSVMEILK	---GMDVREV
P11513	S. acidocaldarius	B	Multa	GRMCPFETPE	GPNSGLVKNL	ALLAQVSVGI	NESVVERVAY	E-LGVVSVED
Q980R1	S. solfataricus	B'	Multa	LKWSKVI LNG	RLIGYYQDGG	ELANKIRERR	RKGEISDE--	--VNVGHIVT
O28392	A. fulgidus	B'	Multa	MRKCRVYING	ALVGFHDNGE	ELANKIRELR	RKGRISNQ--	--VNVVYYPD
Q60181	M. jannaschii	B'	Multa	SREVNI YVNG	KLVTGTDKPE	ELVNFIREKR	RKGEIPQY--	--TTVAYNEE
P41557	M. vannielii	B'	Multa	EKQANVYVNG	KLIDTSKDPE	NLVKSLRIQR	RSGLSPN--	--TSISFNEE
Q27124	M. thermoautotrophicum	B'	Multa	MSKTKIYING	KLIGTCENPE	EFVEEMRKKR	RSGEVSPE--	--MNITHYPE
P09845	M. thermoautotrophicum (str. Winter)	B'	Multa	MNKTKIYING	KLIGTCENPE	EFVEEIRAKR	RSGEVSHE--	--MNITHYPE
Q58444	M. jannaschii	B''	Multa	GKICPSETPE	GPNCGLVKNL	ALMCKVTTREE	DDSKVI-ELL	KSFGINVS
P41558	M. vannielii	B''	Multa	GKICPSETPE	GPNCGLVKNL	AVMCKVTTDE	EDEGII-QLI	KEIGLSKDI
Q27123	M. thermoautotrophicum	B''	Multa	GKICPNETPE	GPNCGLVKNL	ALLAKISEGS	D-ASEVEEVI	K-MGVVN
P09844	M. thermoautotrophicum (str. Winter)	B''	Multa	GKICPNETPE	GPNCGLVKNL	ALMAKISEGS	D-PDEIEEVI	KKMGVIN
O28393	A. fulgidus	B''	Multa	GRICPSETPE	GPNCGLVKNL	AEYAEVSVGT	D-ENEIKKVV	FRLGVEPIRG
Q980R0	S. solfataricus	B''	Multa	GRMCPFETPE	GPNSGLVKNL	AVNGSNCTE		
P15352	Halobacterium sp.	B''	Multa	GRICPSETPE	GPNCGLVKNL	AQAMELSQDV	DDERDLKQEL	SSMGVEGIPG

*Method:
 BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
 PSIPRED and PHD = Structure Prediction
 Red = Helix Green = Strand Blue = Turn Gray = No coörd

Figure C.58 Sequence alignment for *P. furiosus* RNAP subunit B (sheet 16)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	,510,-----,520-----,530-----,540-----
P81412	<i>P. furiosus</i>	B	RMSD	-VRR--PDPT -----LW -RVYLNGLVLI GTIEDGKALV QRIRQDRRSG
P81412	<i>P. furiosus</i>	B	PSIPRED	-VRR--PDPT -----LW -RVYLNGLVLI GTIEDGKALV QRIRQDRRSG
P81412	<i>P. furiosus</i>	B	PHD	-VRR--PDPT -----LW -RVYLNGLVLI GTIEDGKALV QRIRQDRRSG
P08518	<i>S. cerevisiae</i>	2	RMSD	-DYVPHQSPD -----AT -RVFVNGVWH GVHRNPARLM ETLRTLRRKG
P08518	<i>S. cerevisiae</i>	2	PSIPRED	-DYVPHQSPD -----AT -RVFVNGVWH GVHRNPARLM ETLRTLRRKG
P08518	<i>S. cerevisiae</i>	2	PHD	-DYVPHQSPD -----AT -RVFVNGVWH GVHRNPARLM ETLRTLRRKG
P08518	<i>S. cerevisiae</i>	2	MOE	YVPH--QSPD -----AT -RVFVNGVWH GVHRNPARLM ETLRTLRRKG
P08518	<i>S. cerevisiae</i>	2	BLAST	YVPH--QSPD -----AT -RVFVNGVWH GVHRNPARLM ETLRTLRRKG
P08518	<i>S. cerevisiae</i>	2	BLAST2	
P08518	<i>S. cerevisiae</i>	2	MPsrch	YVPH--QSPD -----AT -RVFVNGVWH GVHRNPARLM ETLRTLRRKG
P08518	<i>S. cerevisiae</i>	2	Multa	YVPH--QSPD -----A TRVFNGLVLI GVHRNPARLM ETLRTLRRKG
P38420	<i>A. thaliana</i>	2	Multa	ISPS--VIPQ -----A TKIFVNGMWV GVHRDPDMLV KTLRLRRRV
Q42877	<i>L. esculentum</i>	2	Multa	ISPA--VIPQ -----A TKIFVNGTWV GIHRDPDMLV RTLRLRRRV
Q02061	<i>S. pombe</i>	2	Multa	YNPS--ASPN -----A TKVFNGLVLI GVHRDPDMLV ETLRTLRRKG
Q10578	<i>C. elegans</i>	2	Multa	VSPS--AIAD -----A TKIFVNGAWV GIHRDPDMLV TTKLRLRRQ
P08266	<i>D. melanogaster</i>	2	Multa	IAPS--AIAD -----A TKIFVNGCWV GIHRDPDMLV ATLRLRRRV
P30876	<i>H. sapiens</i>	2	Multa	ISPA--AIAD -----A TKIFVNGCWV GIHRDPDMLV NTLRLRRQ
Q98RY5	<i>G. theta</i>	2	Multa	YNSD--KLKS -----S FKVFVNGCWI GTHKNPFALI SKMKIYRRR--
Q27826	<i>E. octocarinatus</i>	2	Multa	ISPS--DIHE -----S TKVFVNGCWI GIHNESEDLI NNKLMRRAL
Q9YAT8	<i>A. pernix</i>	B	Multa	ILEE--AKQA GEPPEIMLKG TKVFINGRPL GYHPNGAELA RKIRELRRSD
O59226	<i>P. horikoshii</i>	B	Multa	-VRR--PSPE -----LW -RVYLNGLVLI GTIEDGKALV ERIKKDRRSG
Q9V115	<i>P. abyssi</i>	B	Multa	-VRR--PSPE -----LW -RVYLNGLVLI GTIEDGKALV ERIKKDRRSG
P31814	<i>T. celer</i>	B	Multa	-ERR--PNPD -----LW -RLYLNGVLV GTVEDGGEFV NRIETDRRSG
Q979F3	<i>T. volcanium</i>	B	Multa	LEES--PNKG -----S -RVYLNGLVLI GYHDDPRYL V RLREERRAG
Q03587	<i>T. acidophilum</i>	B	Multa	LEES--PKKG -----S -RVYLNGLVLI GYHDDPRYL V SRIREERRSG
P11513	<i>S. acidocaldarius</i>	B	Multa	VIRR--ISEQ NEDVEKYSW SKVYLNGLL GYEDGKELA KKIRESRQ
Q980R1	<i>S. solfataricus</i>	B'	Multa	DFIN--EVHV NCDSGRVRP LIIVSNG--- ---NPLVTI EDIEKLES--
O28392	<i>A. fulgidus</i>	B'	Multa	S--N--EVRI NTDAGRARRP LIVVKDG--- ---KPLVTP EHIEKIKK--
Q60181	<i>M. jannaschii</i>	B'	Multa	S--N--DIHI NTDAGRIVRP LIVVENG--- ---KPLTK EHIEKLK--
P41557	<i>M. vannieli</i>	B'	Multa	S--N--DIHI STDGGRVRP LIVVENG--- ---FSKLTN ELLEKVN--
O27124	<i>M. thermoautotrophicum</i>	B'	Multa	N--H--EIYI FTDGGRARRP LIIVEDG--- ---EPLLTE EHLEKLET--
P09845	<i>M. thermoautotrophicum</i> (st. Winter)	B'	Multa	N--H--EIYI FTDGGRARRP LIIVEDG--- ---EPLLKE EHLEKLSS--
Q58444	<i>M. jannaschii</i>	B''	Multa	
P41558	<i>M. vannieli</i>	B''	Multa	
O27123	<i>M. thermoautotrophicum</i>	B''	Multa	
P09844	<i>M. thermoautotrophicum</i> (str. Winter)	B''	Multa	
O28393	<i>A. fulgidus</i>	B''	Multa	E
Q980R0	<i>S. solfataricus</i>	B''	Multa	
P15352	<i>Halobacterium</i> sp.	B''	Multa	ISME--TTST TSADD

*Method:
 BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
 PSIPRED and PHD = Structure Prediction
 Red = Helix Green = Strand Blue = Turn Gray = No cord

Figure C.59 Sequence alignment for *P. furiosus* RNAP subunit B (sheet 17)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	550,--	560	570	580	--
P81412	P. furiosus	B	RMSD	KIS-DV----	INVAYY--ED	VKEVYVNSDD	GRVRRPLIIV	EN--G-----
P81412	P. furiosus	B	PSIPRED	KIS-DV----	INVAYY--ED	VKEVYVNSDD	GRVRRPLIIV	EN--G-----
P81412	P. furiosus	B	PHD	KIS-DV----	INVAYY--ED	VKEVYVNSDD	GRVRRPLIIV	EN--G-----
P08518	S. cerevisiae	2	RMSD	DIN-PE----	VSMIRD--IR	EKELKIFTDA	GRVYRPLFIV	ED--D-----
P08518	S. cerevisiae	2	PSIPRED	DIN-PE----	VSMIRD--IR	EKELKIFTDA	GRVYRPLFIV	ED--D-----
P08518	S. cerevisiae	2	PHD	DIN-PE----	VSMIRD--IR	EKELKIFTDA	GRVYRPLFIV	ED--D-----
P08518	S. cerevisiae	2	MOE	DIN-PE----	VSMIRD--IR	EKELKIFTDA	GRVYRPLFIV	ED--D-----
P08518	S. cerevisiae	2	BLAST	DIN-PE----	VSMIRD--IR	EKELKIFTDA	GRVYRPLFIV	ED--DESLGH
P08518	S. cerevisiae	2	BLAST2					
P08518	S. cerevisiae	2	MPsrch	DINFEV----	SMIRDI--RE	-KELKIFTDA	GRVYRPLFIV	EDDES-----
P08518	S. cerevisiae	2	Multa	DIN--P----	-EVSMIRDIR	EKELKIFTDA	GRVYRPLFIV	ED--DESLGH
P38420	A. thaliana	2	Multa	DVN--T----	-EVGVVRDIR	LKELRIYTDY	GRCSRPLFIV	DN--Q-----
Q42877	L. esculentum	2	Multa	DVN--T----	-EVGVVRDIR	LKELRIYTDY	GRCSRPLFIV	EK--Q-----
Q02061	S. pombe	2	Multa	DIS--A----	-EVSIVRDIR	EKELRLFTDA	GRICRPLFIV	DN--NPNSE
Q10578	C. elegans	2	Multa	DII--VS----	-EVSMVRDIR	DREIRIYTDY	GRVCRPLIIV	EN--Q-----
P08266	D. melanogaster	2	Multa	DII--VS----	-EVSMIRDIR	DREIRIYTDY	GRICRPLIIV	EN--G-----
P30876	H. sapiens	2	Multa	DII--VS----	-EVSMIRDIR	EREIRIYTDY	GRICRPLIIV	EK--Q-----
Q98RY5	G. theta	2	Multa	NIF--SE----	-EISIIIDDY	DKEIKIFS	GRVIRPLIIV	NE--ENI--
Q27826	E. octocarinatus	2	Multa	LIP--K----	-EISIVRDIR	NREVVFYTDH	GRVQRPLFIV	DD--N-----
Q9YAT8	A. pernix	B	Multa	KLE--VERPWE	VSVYHLREGE	IDELYVNTDS	GRLMRPVVVV	EN--G-----
O59226	P. horikoshii	B	Multa	KIS--DI----	INVAYY--ED	VKEVYVNSDD	GRVRRPLIIV	EN--G-----
Q9V115	P. abyssi	B	Multa	KIS--DI----	INVAYY--ED	VKEVYVNSDD	GRVRRPLIIV	EN--G-----
P31814	T. celer	B	Multa	KIS--DI----	INVAYY--ED	VKEVYVNSDD	GRVRRPLIIV	EN--G-----
Q979F3	T. volcanium	B	Multa	RIS--DE----	INVRY--DDN	TNEVIVNCDR	GRLRPLIIV	KD--G-----
Q03587	T. acidophilum	B	Multa	RMS--DE----	VNVRY--DDN	TNEVIVNSDR	GRLRPLIIV	KD--G-----
P11513	S. acidocaldarius	B	Multa	KLS--DE----	VNVYIATDY	LNEVHINCDA	GRVRRPLIIV	NN--G-----
Q980R1	S. solfataricus	B'	Multa	-----	--GAITFDDL	VRQGKIEYLD	AEEEEENAYVA	LE--FND---
O28392	A. fulgidus	B'	Multa	-----	--GELTIEDL	VRQGLIEYLD	AEEEEENAYVA	VY--EHE---
Q60181	M. jannaschii	B'	Multa	-----	--GEITFSDL	VKEGVIEYLD	AEEEEENAYIA	LS--EEE---
P41557	M. vannielii	B'	Multa	-----	--NELTFEYL	VKTGVIEFLD	AEEEEENARIA	MY--NDE---
O27124	M. thermoautotrophicum	B'	Multa	-----	--GEMSWDDL	IAAGIIEYLD	AEEEEENAYIA	MS--PDE---
P09845	M. thermoautotrophicum (st. Winter)	B'	Multa	-----	--GEMEWDDL	ISQGIIEYLD	AEEEEENTYIA	MS--PEE---
Q58444	M. jannaschii	B''	Multa					
P41558	M. vannielii	B''	Multa					
O27123	M. thermoautotrophicum	B''	Multa					
P09844	M. thermoautotrophicum (str. Winter)	B''	Multa					
O28393	A. fulgidus	B''	Multa					
Q980R0	S. solfataricus	B''	Multa					
P15352	Halobacterium sp.	B''	Multa					

*Method:
 BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
 PSIPRED and PHD = Structure Prediction
 Red = Helix Green = Strand Blue = Turn Gray = No cord

Figure C.60 Sequence alignment for *P. furiosus* RNAP subunit B (sheet 18)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	-----590-----
P81412	<i>P. furiosus</i>	B	RMSD	-----V-----P---K-LT-K EHV---Q--AV--K--E--
P81412	<i>P. furiosus</i>	B	PSIPRED	-----V-----P---K-LT-K EHV---Q--AV--K--E--
P81412	<i>P. furiosus</i>	B	PHD	-----V-----P---K-LT-K EHV---Q--AV--K--E--
P08518	<i>S. cerevisiae</i>	2	RMSD	-----E SLGHKEL---K-VR-K GHIAKLMA--TE--Y--Q--
P08518	<i>S. cerevisiae</i>	2	PSIPRED	-----E SLGHKEL---K-VR-K GHIAKLMA--TE--Y--Q--
P08518	<i>S. cerevisiae</i>	2	PHD	-----E SLGHKEL---K-VR-K GHIAKLMA--TE--Y--Q--
P08518	<i>S. cerevisiae</i>	2	MOE	-----E SLGH KELK-VR-K GHIAKL-M--AT--EY-Q--
P08518	<i>S. cerevisiae</i>	2	BLAST	KELKVRKGHI-----A---K-LM-A TEY---Q--DI--E--G--
P08518	<i>S. cerevisiae</i>	2	BLAST2	
P08518	<i>S. cerevisiae</i>	2	MPsrch	-----L G-----H---KELKVRK GHIA---KLM ATEYQDIE--
P08518	<i>S. cerevisiae</i>	2	Multa	K-----E-----L---K-VR-K GHI---A--KL--M--ATE
P38420	<i>A. thaliana</i>	2	Multa	-----K-----L---L-IK-K RDI---Y--AL--Q--QR-
Q42877	<i>L. esculentum</i>	2	Multa	-----R-----L---M-IK-K KDI---Q--TL--Q--QR-
Q02061	<i>S. pombe</i>	2	Multa	RG-----L---C-IR-K EHI---Q--QL--I--EDK
Q10578	<i>C. elegans</i>	2	Multa	-----K-----L---A-LK-K RHI---D--QL--K--EAA
P08266	<i>D. melanogaster</i>	2	Multa	-----S-----L---L-LK-K THV---E--ML--K--ER-
P30876	<i>H. sapiens</i>	2	Multa	-----K-----L---L-LK-K RHI---D--QL--K--ER-
Q98RY5	<i>G. theta</i>	2	Multa	-----K-----I---N-NK-Q KYL---H--NL--T--D--
Q27826	<i>E. octocarinatus</i>	2	Multa	-----R-----L---I-LK-K THV---V--RL--Q--DH-
Q9YAT8	<i>A. pernix</i>	B	Multa	-----E-----P---K-LT-R EHV---E--KL--S--R-
O59226	<i>P. horikoshii</i>	B	Multa	-----V-----P---K-LT-K EHV---Q--AI--K--E--
Q9V115	<i>P. abyssi</i>	B	Multa	-----K-----P---K-LT-K EHV---Q--AI--K--E--
P31814	<i>T. celer</i>	B	Multa	-----R-----P---K-LT-R EHV---E--AI--K--N-
Q979F3	<i>T. volcanium</i>	B	Multa	-----R-----T---V-IT-E DMI---E--KL--R--K-
Q03587	<i>T. acidophilum</i>	B	Multa	-----K-----T---V-LD-R TMI---E--RL--K--H-
P11513	<i>S. acidocaldarius</i>	B	Multa	-----T-----P---L-VD-T EDI---K--KL--K--N-
Q980R1	<i>S. solfataricus</i>	B'	Multa	-----
O28392	<i>A. fulgidus</i>	B'	Multa	-----
Q60181	<i>M. jannaschii</i>	B'	Multa	-----
P41557	<i>M. vannielii</i>	B'	Multa	-----
O27124	<i>M. thermoautotrophicum</i>	B'	Multa	-----
P09845	<i>M. thermoautotrophicum</i> (str. Winter)	B'	Multa	-----
Q58444	<i>M. jannaschii</i>	B''	Multa	-----
P41558	<i>M. vannielii</i>	B''	Multa	-----
O27123	<i>M. thermoautotrophicum</i>	B''	Multa	-----
P09844	<i>M. thermoautotrophicum</i> (str. Winter)	B''	Multa	-----
O28393	<i>A. fulgidus</i>	B''	Multa	-----
Q980R0	<i>S. solfataricus</i>	B''	Multa	-----
P15352	<i>Halobacterium</i> sp.	B''	Multa	-----

*Method:
 BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
 PSIPRED and PHD = Structure Prediction
 Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.61 Sequence alignment for *P. furiosus* RNAP subunit B (sheet 19)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	600-----	610-----
P81412	P. furiosus	B	RMSD	-----G-----R-L- --KWSDLIKM	GIIEYLDAAE
P81412	P. furiosus	B	PSIPRED	-----G-----R-L- --KWSDLIKM	GIIEYLDAAE
P81412	P. furiosus	B	PHD	-----G-----R-L- --KWSDLIKM	GIIEYLDAAE
P08518	S. cerevisiae	2	RMSD	-----DIEGGFEDV E-----E-Y- --TWSSLLNE	GLVEYIDAAE
P08518	S. cerevisiae	2	PSIPRED	-----DIEGGFEDV E-----E-Y- --TWSSLLNE	GLVEYIDAAE
P08518	S. cerevisiae	2	PHD	-----DIEGGFEDV E-----E-Y- --TWSSLLNE	GLVEYIDAAE
P08518	S. cerevisiae	2	MOE	-----DIEGGFEDV E-----E-Y- --TWSSLLNE	GLVEYIDAAE
P08518	S. cerevisiae	2	BLAST	-----GFEDVE--- -----E-Y- --TWSSLLNE	GLVEYIDAAE
P08518	S. cerevisiae	2	BLAST2		
P08518	S. cerevisiae	2	MPsrch	-----GGF-----EDVE EYTWSSLLNE	GLVEYIDAAE
P08518	S. cerevisiae	2	Multa	YQDIEGGFED VE-----E-Y- --TWSSLLNE	GLVEYIDAAE
P38420	A. thaliana	2	Multa	-----ES AE-----E-D- --GWHHLVAK	GFIEYIDTEE
Q42877	L. esculentum	2	Multa	-----ES PD-----E-G- --GWHDLVAK	GYIEYIDTEE
Q02061	S. pombe	2	Multa	DRY---DIDP EQ-----R-F- --GWTALVSS	GLIEYLDAAE
Q10578	C. elegans	2	Multa	-----DE AN-----K-Y- --TWSDLVGG	GVVELIDSME
P08266	D. melanogaster	2	Multa	-----D YN-----N-Y- --SWQVLVAS	GVVEYIDTLE
P30876	H. sapiens	2	Multa	-----E YN-----N-Y- --SWQDLVAS	GVVEYIDTLE
Q98RY5	G. theta	2	Multa	-----DQ SI-----Y-S- --NFQQLLEE	GLIEYVDPPEE
Q27826	E. octocarinatus	2	Multa	-----ER SD-----S-M- --NFDLTKE	GLVEFLDVPEE
Q9YAT8	A. pernix	B	Multa	-----G-----E-I- --TFWDLVKQ	GVVEMLDPPEE
O59226	P. horikoshii	B	Multa	-----G-----R-L- --KWSDLVRM	GVIEYLDAAE
Q9V115	P. abyssi	B	Multa	-----G-----R-L- --KWSDLVRM	GVIEYLDAAE
P31814	T. celer	B	Multa	-----G-----S-L- --TWSDLVKM	GVIEYLDAAE
Q979F3	T. volcanium	B	Multa	-----D-----E-Y- --SFEDLVKV	GALEWLDAAE
Q03587	T. acidophilum	B	Multa	-----G-----E-I- --SFEDLVKV	GAIEWLDAAE
P11513	S. acidocaldarius	B	Multa	-----G-----E-I- --TFDDLKVQ	GKIEFIDAAE
Q980R1	S. solfataricus	B'	Multa		LTPEHTHLEI
O28392	A. fulgidus	B'	Multa		LTPEHTHLEL
Q60181	M. jannaschii	B'	Multa		LTEKHTHLEI
P41557	M. vannielii	B'	Multa		ITFENTHVEI
O27124	M. thermoautotrophicum	B'	Multa		INDEHHTHLEI
P09845	M. thermoautotrophicum (st. Winter)	B'	Multa		VTEETHHLEI
Q58444	M. jannaschii	B''	Multa		
P41558	M. vannielii	B''	Multa		
O27123	M. thermoautotrophicum	B''	Multa		
P09844	M. thermoautotrophicum (str. Winter)	B''	Multa		
O28393	A. fulgidus	B''	Multa		
Q980R0	S. solfataricus	B''	Multa		
P15352	Halobacterium sp.	B''	Multa		

*Method:
 BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
 PSIPRED and PHD = Structure Prediction
 Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.62 Sequence alignment for *P. furiosus* RNAP subunit B (sheet 20)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	620-----	-----	-----	-----	-----	-----
P81412	P. furiosus	B	RMSD	EENAYVAMW-	-----	-----	-----	-----	-----
P81412	P. furiosus	B	PSIPRED	EENAYVAMW-	-----	-----	-----	-----	-----
P81412	P. furiosus	B	PHD	EENAYVAMW-	-----	-----	-----	-----	-----
P08518	S. cerevisiae	2	RMSD	EESILIAMQ-	-----	-----	-----	-----	-----
P08518	S. cerevisiae	2	PSIPRED	EESILIAMQ-	-----	-----	-----	-----	-----
P08518	S. cerevisiae	2	PHD	EESILIAMQ-	-----	-----	-----	-----	-----
P08518	S. cerevisiae	2	MOE	EESILIAMQ-	-----	-----	-----	-----	-----
P08518	S. cerevisiae	2	BLAST	EESILIAMQP	EDLE-----	-----	-----	-----	-----
P08518	S. cerevisiae	2	BLAST2	EESILIAMQP	EDLE-----	-----	-----	-----	-----
P08518	S. cerevisiae	2	MPsrch	EESILIAMQ-	-----	-----	-----	-----	-----
P08518	S. cerevisiae	2	Multa	EESILIAMQP	EDLEPA---E	ANEENDLDVD	PAKRIRVSH-	-----	-----
P38420	A. thaliana	2	Multa	EETTMISMTI	SDLVQARLRP	EEA-----	-----	-----	-----
Q42877	L. esculentum	2	Multa	EETTMISMTI	NDLVQARLNP	GDA-----	-----	-----	-----
Q02061	S. pombe	2	Multa	EETVMIAMSP	EDLEASRQMQ	AGYEVKEELD	PAQRVKPAPN	P-----	-----
Q10578	C. elegans	2	Multa	EETSMIAMPP	EDL--R---	SGG-----	-----	-----	-----
P08266	D. melanogaster	2	Multa	EETVMIAMSP	YDLKQDK---	DYA-----	-----	-----	-----
P30876	H. sapiens	2	Multa	EETVMLAMTP	DDL-QEK---	EVA-----	-----	-----	-----
Q98RY5	G. theta	2	Multa	EETCLI SMFY	NFKYLNKLNHT	-----	-----	-----	-----
Q27826	E. octocarinatus	2	Multa	EETSMIAMNV	GDLENRN---	-----	-----	-----	-----
Q9YAT8	A. permix	B	Multa	EQNALIAEL-	-----	-----	-----	-----	-----
O59226	P. horikoshii	B	Multa	EENAYVAMW-	-----	-----	-----	-----	-----
Q9V115	P. abyssii	B	Multa	EENAYVAMW-	-----	-----	-----	-----	-----
P31814	T. celer	B	Multa	EENALVATW-	-----	-----	-----	-----	-----
Q979F3	T. volcanium	B	Multa	EEDAYVAVYA	YNIPEKCPHC	SSYLRSMDV	WSNPGSNII I	LECGFCYNKF	-----
Q03587	T. acidophilum	B	Multa	EEDTYVAVYA	YDIPEKCPHC	NSYLRSMDV	WVNPGESEIT	LECGFCQHRF	-----
P11513	S. acidocaldarius	B	Multa	EENAYVALN-	-----	-----	-----	-----	-----
Q980R1	S. solfataricus	B'	Multa	WSPAILGITA	SIIPYPEHNQ	SPRNTYQSAM	AKQALGLYAA	NYQLRTDTRA	-----
O28392	A. fulgidus	B'	Multa	DPSSLIVGICT	GSIPYAEHNA	SPRNTMGAAM	IKQSLGIPYS	NLAWRTDTRG	-----
Q60181	M. jannaschii	B'	Multa	DPLTILGIGA	GVAPYPEHNS	APRITMAAAM	GKQSLGIPMS	NIKWRLDTRG	-----
P41557	M. vannielii	B'	Multa	DPVLILGIGA	GVAPYPEHNS	APRITMAAAM	GKQSLGIPMA	NIKWRMDTRG	-----
Q27124	M. thermoautotrophicum	B'	Multa	DPSTMLGIGA	GIIPFANHNS	SPRNTMEAGM	TKQALGLYAS	NYDLRTDTRA	-----
P09845	M. thermoautotrophicum (st. Winter)	B'	Multa	DPSTMLGIGA	GIIPFANHNS	SPRNTMEAGM	TKQALGLYAS	NYNLRDTRA	-----
Q58444	M. jannaschii	B''	Multa	-----	-----	-----	-----	-----	-----
P41558	M. vannielii	B''	Multa	-----	-----	-----	-----	-----	-----
Q27123	M. thermoautotrophicum	B''	Multa	-----	-----	-----	-----	-----	-----
P09844	M. thermoautotrophicum (str. Winter)	B''	Multa	-----	-----	-----	-----	-----	-----
O28393	A. fulgidus	B''	Multa	-----	-----	-----	-----	-----	-----
Q980R0	S. solfataricus	B''	Multa	-----	-----	-----	-----	-----	-----
P15352	Halobacterium sp.	B''	Multa	-----	-----	-----	-----	-----	-----

*Method:
BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
PSIPRED and PHD = Structure Prediction
Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.63 Sequence alignment for *P. furiosus* RNAP subunit B (sheet 21)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	630, -	-	-
P81412	P. furiosus	B	RMSD	--P--WE-V-	-----T	-----
P81412	P. furiosus	B	PSIPRED	--P--WE-V-	-----T	-----
P81412	P. furiosus	B	PHD	--P--WE-V-	-----T	-----
P08518	S. cerevisiae	2	RMSD	--P--ED-L-	-----E	PAEANEENDL DVDPAKRIRV
P08518	S. cerevisiae	2	PSIPRED	--P--ED-L-	-----E	PAEANEENDL DVDPAKRIRV
P08518	S. cerevisiae	2	PHD	--P--ED-L-	-----E	PAEANEENDL DVDPAKRIRV
P08518	S. cerevisiae	2	MOE	--P--ED-LE	PAEANEENDL	DVDPAKRIRV
P08518	S. cerevisiae	2	BLAST	--P--AE-A-	-----N	-----
P08518	S. cerevisiae	2	BLAST2	-----	-----	-----
P08518	S. cerevisiae	2	MPsrch	--PEDLEPAE	-----A	N-----
P08518	S. cerevisiae	2	Multa	-----H-	-----A	-----
P38420	A. thaliana	2	Multa	-----Y-	-----T	-----
Q42877	L. esculentum	2	Multa	-----Y-	-----S	-----
Q02061	S. pombe	2	Multa	-----H-	-----V	-----
Q10578	C. elegans	2	Multa	-----Y-	-----C	-----
P08266	D. melanogaster	2	Multa	-----Y-	-----C	-----
P30876	H. sapiens	2	Multa	-----Y-	-----C	-----
Q98RY5	G. theta	2	Multa	-----Q-	-----K	-----
Q27826	E. octocarinatus	2	Multa	-----Q-	-----C	-----
Q9YAT8	A. pernix	B	Multa	--P--WE-V-	-----T	-----
O59226	P. horikoshii	B	Multa	--P--WE-V-	-----T	-----
Q9V115	P. abyssi	B	Multa	--P--WE-V-	-----T	-----
P31814	T. celer	B	Multa	--P--WE-V-	-----T	-----
Q979F3	T. volcanium	B	Multa	EVP--SL-I-	-----T	-----
Q03587	T. acidophilum	B	Multa	NVP--SK-L-	-----S	-----
P11513	S. acidocaldarius	B	Multa	--P--QD-L-	-----T	-----
Q980R1	S. solfataricus	B'	Multa	HLL--HY-P-	-----Q	-----
O28392	A. fulgidus	B'	Multa	HLM--QH-P-	-----Q	-----
Q60181	M. jannaschii	B'	Multa	HYL--HY-P-	-----Q	-----
P41557	M. vannielii	B'	Multa	HLL--HY-P-	-----Q	-----
Q27124	M. thermoautotrophicum	B'	Multa	HLL--HH-P-	-----Q	-----
P09845	M. thermoautotrophicum (st. Winter)	B'	Multa	HLL--HH-P-	-----Q	-----
Q58444	M. jannaschii	B''	Multa	-----	-----	-----
P41558	M. vannielii	B''	Multa	-----	-----	-----
O27123	M. thermoautotrophicum	B''	Multa	-----	-----	-----
P09844	M. thermoautotrophicum (str. Winter)	B''	Multa	-----	-----	-----
O28393	A. fulgidus	B''	Multa	-----	-----	-----
Q980R0	S. solfataricus	B''	Multa	-----	-----	-----
P15352	Halobacterium sp.	B''	Multa	-----	-----	-----

*Method:
BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
PSIPRED and PHD = Structure Prediction
Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.64 Sequence alignment for *P. furiosus* RNAP subunit B (sheet 22)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	-----640-----
P81412	P. furiosus	B	RMSD	---EE--- -H----- -THLELMPAA
P81412	P. furiosus	B	PSIPRED	---EE--- -H----- -THLELMPAA
P81412	P. furiosus	B	PHD	---EE--- -H----- -THLELMPAA
P08518	S. cerevisiae	2	RMSD	SHHATT--- -F----- -THCEIHPSM
P08518	S. cerevisiae	2	PSIPRED	SHHATT--- -F----- -THCEIHPSM
P08518	S. cerevisiae	2	PHD	SHHATT--- -F----- -THCEIHPSM
P08518	S. cerevisiae	2	MOE	---SH--- -HATTF--- -THCEIHPSM
P08518	S. cerevisiae	2	BLAST	---EE--- -NDLDVDPK RIRVSHHATT FTHCEIHPSM
P08518	S. cerevisiae	2	BLAST2	THCEIHPSM
P08518	S. cerevisiae	2	MPsrch	---EENDLD VDKAKRIRVS HHATTF--- -THCEIHPSM
P08518	S. cerevisiae	2	Multa	---TT--- -F----- -THCEIHPSM
P38420	A. thaliana	2	Multa	---EN--- -Y----- -THCEIHPSL
Q42877	L. esculentum	2	Multa	---DT--- -Y----- -THCEIHPSL
Q02061	S. pombe	2	Multa	---HA--- -W----- -THCEIHPAM
Q10578	C. elegans	2	Multa	---DT--- -H----- -THCEIHPAM
P08266	D. melanogaster	2	Multa	---TT--- -Y----- -THCEIHPAM
P30876	H. sapiens	2	Multa	---ST--- -Y----- -THCEIHPSM
Q98RY5	G. theta	2	Multa	---IK--- -Y----- -THCELHPSV
Q27826	E. octocarinatus	2	Multa	---ST--- -Y----- -THCEIHPSM
Q9YAT8	A. pernix	B	Multa	---KE--- -H----- -THMELWVPA
O59226	P. horikoshii	B	Multa	---KE--- -H----- -THLELMPAA
Q9V115	P. abyssi	B	Multa	---EE--- -H----- -THLELMPAA
P31814	T. celer	B	Multa	---EE--- -H----- -THLELMPAA
Q979F3	T. volcanium	B	Multa	---PE--- -H----- -THLEIDPAM
Q03587	T. acidophilum	B	Multa	---KE--- -N----- -THLEIDPAM
P11513	S. acidocaldarius	B	Multa	---PD--- -H----- -THLEIWPSA
Q980R1	S. solfataricus	B'	Multa	---RP--- -L----- -VQTRALDII
O28392	A. fulgidus	B'	Multa	---VP--- -I----- -VTTDTQVEI
Q60181	M. jannaschii	B'	Multa	---VP--- -I----- -VRTKHQEIL
P41557	M. vannielii	B'	Multa	---VP--- -L----- -VRTKHQEIL
O27124	M. thermoautotrophicum	B'	Multa	---VP--- -I----- -VKTRIIDVT
P09845	M. thermoautotrophicum (str. Winter)	B'	Multa	---VP--- -I----- -VKTRIIDVT
Q58444	M. jannaschii	B''	Multa	
P41558	M. vannielii	B''	Multa	
O27123	M. thermoautotrophicum	B''	Multa	
P09844	M. thermoautotrophicum (str. Winter)	B''	Multa	
O28393	A. fulgidus	B''	Multa	
Q980R0	S. solfataricus	B''	Multa	
P15352	Halobacterium sp.	B''	Multa	

*Method:
 BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
 PSIPRED and PHD = Structure Prediction
 Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.65 Sequence alignment for *P. furiosus* RNAP subunit B (sheet 23)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	-----650----	-----660----	-----670----	-----680----	-----690----
P81412	<i>P. furiosus</i>	B	RMSD	I L G L P A S L V P	Y P E H N A A P R N	T Y G A G M A K Q S	L G L G W A N F R I	R V D T R G H L L H
P81412	<i>P. furiosus</i>	B	PSIPRED	I L G L P A S L V P	Y P E H N A A P R N	T Y G A G M A K Q S	L G L G W A N F R I	R V D T R G H L L H
P81412	<i>P. furiosus</i>	B	PHD	I L G L P A S L V P	Y P E H N A A P R N	T Y G A G M A K Q S	L G L G W A N F R I	R V D T R G H L L H
P08518	<i>S. cerevisiae</i>	2	RMSD	I L G V A A S I I P	F P D H N Q S P R N	T Y Q S A M G K Q A	M G V F L T N Y N V	R M D T M A N I L Y
P08518	<i>S. cerevisiae</i>	2	PSIPRED	I L G V A A S I I P	F P D H N Q S P R N	T Y Q S A M G K Q A	M G V F L T N Y N V	R M D T M A N I L Y
P08518	<i>S. cerevisiae</i>	2	PHD	I L G V A A S I I P	F P D H N Q S P R N	T Y Q S A M G K Q A	M G V F L T N Y N V	R M D T M A N I L Y
P08518	<i>S. cerevisiae</i>	2	MOE	I L G V A A S I I P	F P D H N Q S P R N	T Y Q S A M G K Q A	M G V F L T N Y N V	R M D T M A N I L Y
P08518	<i>S. cerevisiae</i>	2	BLAST	I L G V A A S I I P	F P D H N Q S P R N	T Y Q S A M G K Q A	M G V F L T N Y N V	R M D T M A N I L Y
P08518	<i>S. cerevisiae</i>	2	BLAST2	I L G V A A S I I P	F P D H N Q S P R N	T Y Q S A M G K Q A	M G V F L T N Y N V	R M D T M A N I L Y
P08518	<i>S. cerevisiae</i>	2	MPsrch	I L G V A A S I I P	F P D H N Q S P R N	T Y Q S A M G K Q A	M G V F L T N Y N V	R M D T M A N I L Y
P08518	<i>S. cerevisiae</i>	2	Multa	I L G V A A S I I P	F P D H N Q S P R N	T Y Q S A M G K Q A	M G V F L T N Y N V	R M D T M A N I L Y
P38420	<i>A. thaliana</i>	2	Multa	I L G V C A S I I P	F P D H N Q S P R N	T Y Q S A M G K Q A	M G I Y V T N Y Q F	R M D T L A V V L Y
Q42877	<i>L. esculentum</i>	2	Multa	I L G V C A S I I P	F P D H N Q S P R N	T Y Q S A M G K Q A	M G I Y V T N Y Q F	R M D T L A V V L Y
Q02061	<i>S. pombe</i>	2	Multa	I L G I L A S I I P	F P D H N Q S P R N	T Y Q S A M G K Q A	M G V Y L T N Y Q F	R M D T M A N I L Y
Q10578	<i>C. elegans</i>	2	Multa	I L G V C A S I I P	F P D H N Q S P R N	T Y Q S A M G K Q A	M G V Y T T N F H V	R M D T L A H V L Y
P08266	<i>D. melanogaster</i>	2	Multa	I L G V C A S I I P	F P D H N Q S P R N	T Y Q S A M G K Q A	M G V Y I T N F H V	R M D T L A H V L Y
P30876	<i>H. sapiens</i>	2	Multa	I L G V C A S I I P	F P D H N Q S P R N	T Y Q S A M G K Q A	M G V Y I T N F H V	R M D T L A H V L Y
Q98RY5	<i>G. theta</i>	2	Multa	I L G V C A S I I P	F C D H N Q S P R N	T Y Q A A M G K Q A	I G I N V S N Y D F	R M D T M C H I L Y
Q27826	<i>E. octocarinatus</i>	2	Multa	I L G V C A S I I P	F P D H N Q S P R N	C Y Q S A M G K Q A	M G V Y C S N Y Q I	R M D T L A H V L Y
Q9YAT8	<i>A. pernix</i>	B	Multa	I V G V A A T I G	Y L E H N Q S P R N	T Y Q A A M A K Q A	L G V Y A L N Y R I	R L D S H S Y L L Q
O59226	<i>P. horikoshii</i>	B	Multa	I L G L P A S L I P	Y P E H N A A P R N	T Y G A G M A K Q S	L G L G W A N F R I	R V D T R G H L L H
Q9V115	<i>P. abyssi</i>	B	Multa	I L G L P A S L I P	Y P E H N A A P R N	T Y G A G M A K Q S	L G L G W A N F R I	R V D T R G H L L H
P31814	<i>T. celer</i>	B	Multa	I L G I P A S L V P	Y P E H N A R P R N	T Y G A G M A K Q S	L G L G W A N F R I	R V D T R G H L M H
Q979F3	<i>T. volcanium</i>	B	Multa	I L G V V A S I I P	Y P E H N S S P R I	T I A S A M A K Q S	L G F A Q S N V R I	R P D T R G H L L H
Q03587	<i>T. acidophilum</i>	B	Multa	I L G V V A S I I P	Y P E H N S S P R I	T I A S A M A K Q S	L G F A Q S N V R I	R P D T R G H L L H
P11513	<i>S. acidocaldarius</i>	B	Multa	I L G I I A S I I P	Y P E H N Q S P R N	T Y Q S A M A K Q S	L G L Y A S N Y Q I	R T D T R A H L L H
Q980R1	<i>S. solfataricus</i>	B'	Multa	G Y T N R P A G N N	A I L A V M S F T G	Y N M E D S I I M N	R S S V E R G M Y R	S T F F R L Y S T E
O28392	<i>A. fulgidus</i>	B'	Multa	K F D E R P A G Q N	F V V A V L S Y E G	Y N I E D A L I M N	K G S V E R G L G R	S F F Y R T Y E S E
Q60181	<i>M. jannaschii</i>	B'	Multa	G F D K R P A G Q N	F V V A I M S Y E G	Y N M E D A I V F N	K S A I D R G L G R	S T F F R T Y D A C
P41557	<i>M. vannielii</i>	B'	Multa	G F D K R P A G Q N	F V V A V M S Y E G	Y N M E D A F V I N	K A S L E R G L G R	S T F F R S Y E S F
O27124	<i>M. thermoautotrophicum</i>	B'	Multa	G Y D E R P S G Q N	F V V A V M S Y E G	Y N M E D A L I L N	K A S L E R G L A R	S S F F R S Y E A A
P09845	<i>M. thermoautotrophicum</i> (st. Winter)	B'	Multa	G Y D E R P S G Q N	F V V A V M S Y E G	Y N M E D A L I L N	K A S L E R G L A R	S S F F R S Y E A T
Q58444	<i>M. jannaschii</i>	B''	Multa					
P41558	<i>M. vannielii</i>	B''	Multa					
O27123	<i>M. thermoautotrophicum</i>	B''	Multa					
P09844	<i>M. thermoautotrophicum</i> (str. Winter)	B''	Multa					
O28393	<i>A. fulgidus</i>	B''	Multa					
Q980R0	<i>S. solfataricus</i>	B''	Multa					
P15352	<i>Halobacterium</i> sp.	B''	Multa					

*Method:
 BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
 PSIPRED and PHD = Structure Prediction
 Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.66 Sequence alignment for *P. furiosus* RNAP subunit B (sheet 24)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	-----700----	-----710----	-----720----	-----730----	-----740----
P81412	<i>P. furiosus</i>	B	RMSD	YFQIPLVNSR	IMKAVGYEDR	PAGQNFVVAV	LAYQGYNMED	AIINKASIE
P81412	<i>P. furiosus</i>	B	PSIPRED	YFQIPLVNSR	IMKAVGYEDR	PAGQNFVVAV	LAYQGYNMED	AIINKASIE
P81412	<i>P. furiosus</i>	B	PHD	YFQIPLVNSR	IMKAVGYEDR	PAGQNFVVAV	LAYQGYNMED	AIINKASIE
P08518	<i>S. cerevisiae</i>	2	RMSD	YFQKPLGTTR	AMEYLFKFREL	PAGQNAIVAI	ACYSGYNQED	SMIMNQSSID
P08518	<i>S. cerevisiae</i>	2	PSIPRED	YFQKPLGTTR	AMEYLFKFREL	PAGQNAIVAI	ACYSGYNQED	SMIMNQSSID
P08518	<i>S. cerevisiae</i>	2	PHD	YFQKPLGTTR	AMEYLFKFREL	PAGQNAIVAI	ACYSGYNQED	SMIMNQSSID
P08518	<i>S. cerevisiae</i>	2	MOE	YFQKPLGTTR	AMEYLFKFREL	PAGQNAIVAI	ACYSGYNQED	SMIMNQSSID
P08518	<i>S. cerevisiae</i>	2	BLAST	YFQKPLGTTR	AMEYLFKFREL	PAGQNAIVAI	ACYSGYNQED	SMIMNQSSID
P08518	<i>S. cerevisiae</i>	2	BLAST2	YFQKPLGTTR	AMEYLFKFREL	PAGQNAIVAI	ACYSGYNQED	SMIMNQSSID
P08518	<i>S. cerevisiae</i>	2	MPsrch	YFQKPLGTTR	AMEYLFKFREL	PAGQNAIVAI	ACYSGYNQED	SMIMNQSSID
P08518	<i>S. cerevisiae</i>	2	Multa	YFQKPLGTTR	AMEYLFKFREL	PAGQNAIVAI	ACYSGYNQED	SMIMNQSSID
P38420	<i>A. thaliana</i>	2	Multa	YFQKPLVTTR	AMEHLHFRQL	PAGINAIVAI	SCYSGYNQED	SVIMNQSSID
Q42877	<i>L. esculentum</i>	2	Multa	YFQKPLVTTR	AMEHLHFRQL	PAGINAIVAI	SCYSGYNQED	SVIMNQSSID
Q02061	<i>S. pombe</i>	2	Multa	YFQKPLATTR	SMEYLFKFREL	PAGQNAIVAI	LCYSGYNQED	SIIMNQASID
Q10578	<i>C. elegans</i>	2	Multa	YFQKPLVTTR	SMEYLFKFREL	PAGINAIVAI	LSYSGYNQED	SVIMNNSAID
P08266	<i>D. melanogaster</i>	2	Multa	YPMKPLVTTR	SMEYLFKFREL	PAGINSIVAI	LCYTGYNQED	SVILNASAVE
P30876	<i>H. sapiens</i>	2	Multa	YFQKPLVTTR	SMEYLFKFREL	PAGINSIVAI	ASYTGYNQED	SVIMNNSAID
Q98RY5	<i>G. theta</i>	2	Multa	YFQKPLVCTK	SMNFIFKDF	PNGINAIVAI	ACYGGYNQED	SIIMSQDSID
Q27826	<i>E. octocarinatus</i>	2	Multa	YFQKPLVVAR	SLDFPHVKEL	PPGCDIVAI	TCYTGYNQED	SVIMSQAPID
Q9YAT8	<i>A. pernix</i>	B	Multa	YFQVPIVQTR	MLDIMGYNER	PAGQNMVVAI	MSYTGYNIED	ALIFNKGSID
O59226	<i>P. horikoshii</i>	B	Multa	YFQIPLVNSR	IMKAVGYEDR	PAGQNFVVAV	LAYHGYNMED	AIINKASIE
Q9V115	<i>P. abyssi</i>	B	Multa	YFQIPLVNSR	IMKAVGYEDR	PAGQNFVVAV	LAYHGYNMED	AIINKASIE
P31814	<i>T. celer</i>	B	Multa	YFQVPLVNSR	IMKAVGFEEER	PAGQNFVVAV	LSYAGYNMED	AIIMNKASIE
Q979F3	<i>T. volcanium</i>	B	Multa	YFQVPLVRTR	VMEYTHYDRR	PAGQNFVVAV	LSYEGYNIQD	ALVINKAAID
Q03587	<i>T. acidophilum</i>	B	Multa	YFQVPLVRTR	VMDYIHYDRR	PAGQNFVVAV	LSYEGYNIQD	ALVINKAAIE
P11513	<i>S. acidocaldarius</i>	B	Multa	YFQVPLVQTR	MLGVIGYNDR	PAGANALAI	MSYTGYNMED	SIIMNKSSIE
Q980R1	<i>S. solfataricus</i>	B'	Multa	EVKYPGGQED	KIVMPEAG-V	RGYKGEYR	LLEDNGVVSF	EVEVKGQDVL
O28392	<i>A. fulgidus</i>	B'	Multa	EMRYPGGQED	KFEIPGAD-I	SGFRGAAYA	HLDEDGLIFP	ETEVEGPDVFL
Q60181	<i>M. jannaschii</i>	B'	Multa	ERRYPGGQMD	RFEIPEKG-V	RGYRSEECYR	YLEEDGIVAV	ESHVKGGDVI
P41557	<i>M. vannielii</i>	B'	Multa	EKKYPGGQLD	KFEVPEKG-V	RGYRAEEAYR	NLDDGLIDL	ESEVRSGDVI
O27124	<i>M. thermoautotrophicum</i>	B'	Multa	ERRYPGGQED	RFEIPEKG-V	RGYRSESDYR	HLDEDGIINP	EAEVSSGDVFL
P09845	<i>M. thermoautotrophicum</i> (str. Winter)	B'	Multa	ERRYPGGQED	RFEIPEKG-V	RGYRSESDYR	HLDEDGIINP	ETEVSQDVL
Q58444	<i>M. jannaschii</i>	B''	Multa					
P41558	<i>M. vannielii</i>	B''	Multa					
O27123	<i>M. thermoautotrophicum</i>	B''	Multa					
P09844	<i>M. thermoautotrophicum</i> (str. Winter)	B''	Multa					
O28393	<i>A. fulgidus</i>	B''	Multa					
Q980R0	<i>S. solfataricus</i>	B''	Multa					
P15352	<i>Halobacterium</i> sp.	B''	Multa					

*Method:

BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD

PSIPRED and PHD = Structure Prediction

Red = Helix Green = Strand Blue = Turn Gray = No cord

Figure C.67 Sequence alignment for *P. furiosus* RNAP subunit B (sheet 25)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	---750---	---760---	---770---	---780---	---790---
P81412	<i>P. furiosus</i>	B	RMSD	RGLARSTFFR	TYEAEKKYL	GGQTDKFEIP	DPT--VRGYR	GEKYYRNLDL
P81412	<i>P. furiosus</i>	B	PSIPRED	RGLARSTFFR	TYEAEKKYL	GGQTDKFEIP	DPT--VRGYR	GEKYYRNLDL
P81412	<i>P. furiosus</i>	B	PHD	RGLARSTFFR	TYEAEKKYL	GGQTDKFEIP	DPT--VRGYR	GEKYYRNLDL
P08518	<i>S. cerevisiae</i>	2	RMSD	RGLFRSLFFR	SYMDQEKKYG	MSITETFEKP	QRT--NTLRM	KHGTYDKLDD
P08518	<i>S. cerevisiae</i>	2	PSIPRED	RGLFRSLFFR	SYMDQEKKYG	MSITETFEKP	QRT--NTLRM	KHGTYDKLDD
P08518	<i>S. cerevisiae</i>	2	PHD	RGLFRSLFFR	SYMDQEKKYG	MSITETFEKP	QRT--NTLRM	KHGTYDKLDD
P08518	<i>S. cerevisiae</i>	2	MOE	RGLFRSLFFR	SYMDQEKKYG	MSITETFEKP	QRT--NTLRM	KHGTYDKLDD
P08518	<i>S. cerevisiae</i>	2	BLAST	RGLFRSLFFR	SYMDQEKKYG	MSITETFEKP	QRT--NTLRM	KHGTYDKLDD
P08518	<i>S. cerevisiae</i>	2	BLAST2	RGLFRSLFFR	SYMDQEKKYG	MSITETFEKP	QRTNTLRMKH	GT--YDKLDD
P08518	<i>S. cerevisiae</i>	2	MPsrch	RGLFRSLFFR	SYMDQEKKYG	MSITETFEKP	QRT--NTLRM	KHGTYDKLDD
P08518	<i>S. cerevisiae</i>	2	Multa	RGLFRSLFFR	SYMDQEKKYG	MSITETFEKP	QRTN-TLRMK	-HGTYDKLDD
P38420	<i>A. thaliana</i>	2	Multa	RGFFRSVFFR	SYRDEEKKMG	TLVKEDFGRP	DRGS-TMGMR	-HGSYDKLDD
Q42877	<i>L. esculentum</i>	2	Multa	RGFFRSVFFR	SYRDEEKKMG	TLVKEDFGRP	DRAS-TMGMR	-HGSYDKLDD
Q02061	<i>S. pombe</i>	2	Multa	RGLFRSIFVR	TYTDQEKKIG	MTVMEEFERP	VRST-TLRMK	-HGTYDKLDD
Q10578	<i>C. elegans</i>	2	Multa	RGLFRSVFVR	SYRDEEKKMG	NANEELIEKP	TREK-CQGM	-HSLYDKLDD
P08266	<i>D. melanogaster</i>	2	Multa	RGFFRSVFFR	SYRDEEKKMG	GDQEEVFEP	HRGT-CQGM	-NAHYDKLDD
P30876	<i>H. sapiens</i>	2	Multa	RGFFRSVFFR	SYRDEEKKMG	FDQEEVFEP	TRET-CQGM	-HAIYDKLDD
Q98RY5	<i>G. theta</i>	2	Multa	RGLFRSFFFR	NYKDVEKIKF	GGKKTTFEIP	NWKS-CNGIK	-IHSYDKLDD
Q27826	<i>E. octocarinatus</i>	2	Multa	RGCFRSVFFR	NI FEIPVRV-	---ERLLRY	PDST-VVGR	-NGDYTKLDD
Q9YAT8	<i>A. pernix</i>	B	Multa	MGLARAYFFR	VYTGIEENYYP	GGEDRITLP	KPDE-VYDVR	GAAYAKLEA
O59226	<i>P. horikoshii</i>	B	Multa	RGLARSTFFR	TYEAEKKYL	GGQTDKFEIP	APT--VRGYR	GEKYYRNLDL
Q9V115	<i>P. abyssi</i>	B	Multa	RGLARSTFFR	TYEAEKKYL	GGQTDKFEIP	SPT--VRGYR	GEKYYRNLDL
P31814	<i>T. celer</i>	B	Multa	RGLARSTFFR	TYEAEKKYL	GGQTDKFEIP	DPT--IQGYL	GERYYRNLDL
Q979F3	<i>T. volcanium</i>	B	Multa	RGLARSTFFR	TYEAEKKYL	GGQTDKFEIP	THD--VIGAR	AEYYKNLDD
Q03587	<i>T. acidophilum</i>	B	Multa	RGLARSTFFR	TYEAEKKYL	GGQTDKFEIP	THD--IIGAR	AEYYKNLDD
P11513	<i>S. acidocaldarius</i>	B	Multa	RGMYRSTFFR	LYSTEEVKYP	GGQEDKIVTP	EAG--VKGYK	GKDYRLLED
Q980R1	<i>S. solfataricus</i>	B'	Multa	IGKVSPPRFL	EEFEIT----	-SPEQAKRDT	SVTV-RHGE	GIVDLVLTE
O28392	<i>A. fulgidus</i>	B'	Multa	IGKVSPPRFL	EEFEIT----	GISPQKRRET	SVTM-RSNEK	GVVDAVFLMQ
Q60181	<i>M. jannaschii</i>	B'	Multa	VGKTSPPRFL	EEFEIT----	IQVKPQRDS	SVTV-RHGE	GVIDKVLTE
P41557	<i>M. vannielii</i>	B'	Multa	LGKTSPPRFL	EEFEIT----	LQTKSQRDT	SVTI-RHGE	GVVDLVLTE
O27124	<i>M. thermoautotrophicum</i>	B'	Multa	IGKTSPPRFL	EEFEIT----	GTVAERRRET	SVTV-RHGE	GIVDAVLLTE
P09845	<i>M. thermoautotrophicum</i> (st. Winter)	B'	Multa	IGKTSPPRFL	EEFEIT----	GTVAERRRET	SVTV-RHGE	GIVDAVLLTE
Q58444	<i>M. jannaschii</i>	B''	Multa					
P41558	<i>M. vannielii</i>	B''	Multa					
O27123	<i>M. thermoautotrophicum</i>	B''	Multa					
P09844	<i>M. thermoautotrophicum</i> (str. Winter)	B''	Multa					
O28393	<i>A. fulgidus</i>	B''	Multa					
Q980R0	<i>S. solfataricus</i>	B''	Multa					
P15352	<i>Halobacterium</i> sp.	B''	Multa					

*Method:
 BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
 PSIPRED and PHD = Structure Prediction
 Red = Helix Green = Strand Blue = Turn Gray = No cord

Figure C.68 Sequence alignment for *P. furiosus* RNAP subunit B (sheet 26)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	-----800--	-----810--	-----820----
P81412	P. furiosus	B	RMSD	DGLIFPESKV	EGKDVLVGRT	SPPR-----FLE ER-GLG---T
P81412	P. furiosus	B	PSIPRED	DGLIFPESKV	EGKDVLVGRT	SPPR-----FLE ER-GLG---T
P81412	P. furiosus	B	PHD	DGLIFPESKV	EGKDVLVGRT	SPPR-----FLE ER-GLG---T
P08518	S. cerevisiae	2	RMSD	DGLIAPGVRV	SGEDVIIIGKT	TPISPDDEEL QQRTAYH---
P08518	S. cerevisiae	2	PSIPRED	DGLIAPGVRV	SGEDVIIIGKT	TPISPDDEEL QQRTAYH---
P08518	S. cerevisiae	2	PHD	DGLIAPGVRV	SGEDVIIIGKT	TPISPDDEEL QQRTAYH---
P08518	S. cerevisiae	2	MOE	DGLIAPGVRV	SGEDVIIIGKT	TP-----PDE EE-LGQ---R
P08518	S. cerevisiae	2	BLAST	DGLIAPGVRV	SGEDVIIIGKT	TPIS-----PDE EE-LGQ---R
P08518	S. cerevisiae	2	BLAST2	DGLIAPGVRV	SGEDVIIIGKT	TPIS-----PDE EE-LGQ---R
P08518	S. cerevisiae	2	MPsrch	DGLIAPGVRV	SGEDVIIIGKT	TPIS-----PD EE-ELG---Q
P08518	S. cerevisiae	2	Multa	DGLIAPGVRV	SGEDVIIIGKT	TPIS-----PDE EELGQ---R
P38420	A. thaliana	2	Multa	DGLAPPGTRV	SGEDVIIIGKT	T--P-----ISQ DEAQG---Q
Q42877	L. esculentum	2	Multa	DGLAPPGTRV	SGEDVIIIGKT	T--P-----ISQ DDAQG---Q
Q02061	S. pombe	2	Multa	DGLIAPGTRV	SGEDIIIGKT	APIP-----LDH EELGQ---R
Q10578	C. elegans	2	Multa	DGIISPGMRV	SGDDVIIIGKT	VALP-----DID DDLDA---S
P08266	D. melanogaster	2	Multa	DGLIAPGIRV	SGDDVIIIGKT	ITLP-----END DELDS---N
P30876	H. sapiens	2	Multa	DGLIAPGVRV	SGDDVIIIGKT	VTLP-----ENE DELES---T
Q98RY5	G. theta	2	Multa	DGLIGEGKKV	NSNDIIIGKT	CPTN-----LFD NNNEDLRLG
Q27826	E. octocarinatus	2	Multa	DGLIFPGKNV	LGDDIIIGKT	ALPR-----NLF NDMD---D
Q9YAT8	A. pernix	B	Multa	DGIVAPETEV	EGGEILIGRV	SPPR-----FLG EGRVIS---M
O59226	P. horikoshii	B	Multa	DGLIFPESKV	EGKDVLIIGRT	SPPR-----FLE ER-GLG---A
Q9V115	P. abyssi	B	Multa	DGLIFPESKV	EGKDVLIIGRT	SPPR-----FLE ER-GLG---A
P31814	T. celer	B	Multa	DGLIFPESKV	NGKDVLVGRT	SPPR-----FLE EQSGLG---G
Q979F3	T. volcanium	B	Multa	SGIIFPEAYV	EGSDVLIGKT	SPPR-----FLE EGGERL---G
Q03587	T. acidophilum	B	Multa	SGIIFPEAYV	EGSDVLIGKT	SPPR-----FLE EGGERL---G
P11513	S. acidocaldarius	B	Multa	NGVVSPEVEV	KGGDVLIIGKT	SPPR-----FLQ EFKELS---P
Q980R1	S. solfataricus	B'	Multa	TAEGNKLKVK	RVRDLRIPE-	IGDK-----FAS RHGQKGVIGM
O28392	A. fulgidus	B'	Multa	SEGGSKLAKV	RVRDLRIPE-	LGDK-----FAS RHGQKGVVGL
Q60181	M. jannaschii	B'	Multa	TKEGNRLKVK	KVRDLRIPE-	LGDK-----FAS RHGQKGVVGL
P41557	M. vannielii	B'	Multa	TKEGNRLGKV	RVRDLRVPE-	FGDK-----FAS RHGQKGVIGL
O27124	M. thermoautotrophicum	B'	Multa	TVEGSR LAKI	RVREQRQPE-	LGDK-----FAS RHGQKGVVGL
P09845	M. thermoautotrophicum (str. Winter)	B'	Multa	TVEGSR LAKI	RVREQRQPEF	IGDK-----FAS RHGQKGVVGL
Q58444	M. jannaschii	B''	Multa			
P41558	M. vannielii	B''	Multa			
O27123	M. thermoautotrophicum	B''	Multa			
P09844	M. thermoautotrophicum (str. Winter)	B''	Multa			
O28393	A. fulgidus	B''	Multa			
Q980R0	S. solfataricus	B''	Multa			
P15352	Halobacterium sp.	B''	Multa			

*Method:
 BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
 PSIPRED and PHD = Structure Prediction
 Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.69 Sequence alignment for *P. furiosus* RNAP subunit B (sheet 27)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	830	840	850	860
P81412	<i>P. furiosus</i>	B	RMSD	-VSLQ-----ER	RETSVAVRPS	EKGVDVKVII	TETGDGTLV
P81412	<i>P. furiosus</i>	B	PSIPRED	-VSLQ-----ER	RETSVAVRPS	EKGVDVKVII	TETGDGTLV
P81412	<i>P. furiosus</i>	B	PHD	-VSLQ-----ER	RETSVAVRPS	EKGVDVKVII	TETGDGTLV
P08518	<i>S. cerevisiae</i>	2	RMSD	-----SK	RDASTPLRST	ENGIVDQVLV	TTNQDGLKFF
P08518	<i>S. cerevisiae</i>	2	PSIPRED	-----SK	RDASTPLRST	ENGIVDQVLV	TTNQDGLKFF
P08518	<i>S. cerevisiae</i>	2	PHD	-----SK	RDASTPLRST	ENGIVDQVLV	TTNQDGLKFF
P08518	<i>S. cerevisiae</i>	2	MOE	---ISPDEEE LGQRTAYHSK	RDASTPLRST	ENGIVDQVLV	TTNQDGLKFF
P08518	<i>S. cerevisiae</i>	2	BLAST	-TAYH-----SK	RDASTPLRST	ENGIVDQVLV	TTNQDGLKFF
P08518	<i>S. cerevisiae</i>	2	BLAST2	RTAYH-----SK	RDASTPLRST	ENGIVDQVLV	TTNQDGLKFF
P08518	<i>S. cerevisiae</i>	2	MPsrch	RTAYH-----SK	RDASTPLRST	ENGIVDQVLV	TTNQDGLKFF
P08518	<i>S. cerevisiae</i>	2	Multa	-TAYH-----SK	RDASTPLRST	ENGIVDQVLV	TTNQDGLKFF
P38420	<i>A. thaliana</i>	2	Multa	-SSRY-----TR	RDHSISLRHS	ETGMVDQVLL	TTNADGLRFV
Q42877	<i>L. esculentum</i>	2	Multa	-ASRY-----TR	KDHSTSLRHS	ETGMVDQVLL	TTNADGLRFV
Q02061	<i>S. pombe</i>	2	Multa	-TQLH-----AK	RDVSTPLRST	ESGIVDQVMV	TTNQEGKLFV
Q10578	<i>C. elegans</i>	2	Multa	-GKKY-----PK	RDASTFLRSS	ETGIVDQVML	SLNSDGNKFF
P08266	<i>D. melanogaster</i>	2	Multa	-TKRF-----SK	RDASTFLRNS	ETGIVDQVML	TLNSEGYKFC
P30876	<i>H. sapiens</i>	2	Multa	-NRRY-----TK	RDCSTFLRST	ETGIVDQVMV	TLNQEGYKFC
Q98RY5	<i>G. theta</i>	2	Multa	-KNLK-----IK	TDCSSSIKNF	ETGVVDKVII	AKSDEGFRV
Q27826	<i>E. octocarinatus</i>	2	Multa	-QTNL-----IK	KDDSTSLRHS	EKGIDSIVL	TSNEEGDSFT
Q9YAT8	<i>A. pernix</i>	B	Multa	-ATVR-----R	RDTISPMRLG	EKGVDVMVI	STTVRNKL
O59226	<i>P. horikoshii</i>	B	Multa	-VSLQ-----ER	RETSVAVRPS	EKGVDKVII	TETGDGTLV
Q9V115	<i>P. abyssi</i>	B	Multa	-VSLQ-----ER	RETSVAVRPS	ERGVDKVII	TETGDGTLV
P31814	<i>T. celer</i>	B	Multa	-IILQ-----ER	RETSVTVRPS	ETGVVDKVII	TETGDGTLV
Q979F3	<i>T. volcanium</i>	B	Multa	-P--Q-----RR	RESSVTMRPN	ESGYVDNVFL	TVSESNSRVV
Q03587	<i>T. acidophilum</i>	B	Multa	-P--Q-----RR	RESSVTMRPN	ESGYVDNVFL	TVSESNSRVV
P11513	<i>S. acidocaldarius</i>	B	Multa	-E--Q-----AK	RDTISVTRHG	ENGIVDLVLI	TETLEGKLV
Q980R1	<i>S. solfataricus</i>	B'	Multa	-LIPQ-----VD	MPYTVKGVVP	DIILNPHALP	SRMTLGQIME
O28392	<i>A. fulgidus</i>	B'	Multa	-LVPE-----ED	MPFTESGIVP	DLIINPHAI	SRMTVGHVLE
Q60181	<i>M. jannaschii</i>	B'	Multa	-TVFQ-----ED	LPFTESGIVP	DLIINPHAI	SRMTVGQLE
P41557	<i>M. vanielii</i>	B'	Multa	-VVFQ-----ED	LPFTEDGVIP	DLIINPHAI	SRMTIGQVLE
O27124	<i>M. thermoautotrophicum</i>	B'	Multa	-IVSQ-----ED	MPFTEDGVIP	DLIINPHAI	SRMSVGQVLE
P09845	<i>M. thermoautotrophicum</i> (str. Winter)	B'	Multa	-IVSQ-----ED	MPFTEDGVVP	DLIINPHAI	SRMSVGQVLE
Q58444	<i>M. jannaschii</i>	B''	Multa				
P41558	<i>M. vanielii</i>	B''	Multa				
O27123	<i>M. thermoautotrophicum</i>	B''	Multa				
P09844	<i>M. thermoautotrophicum</i> (str. Winter)	B''	Multa				
O28393	<i>A. fulgidus</i>	B''	Multa				
Q980R0	<i>S. solfataricus</i>	B''	Multa				
P15352	<i>Halobacterium</i> sp.	B''	Multa				

*Method:
BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
PSIPRED and PHD = Structure Prediction
Red = Helix Green = Strand Blue = Turn Gray = No cord

Figure C.70 Sequence alignment for *P. furiosus* RNAP subunit B (sheet 28)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	-----870	-----880	-----890	-----900	-----910
P81412	<i>P. furiosus</i>	B	RMSD	KVTVRDLRIP	E-LGDKFASR	HGQKGVIGLI	VPQEDMPWTE	SGIVPDLIVN
P81412	<i>P. furiosus</i>	B	PSIPRED	KVTVRDLRIP	E-LGDKFASR	HGQKGVIGLI	VPQEDMPWTE	SGIVPDLIVN
P81412	<i>P. furiosus</i>	B	PHD	KVTVRDLRIP	E-LGDKFASR	HGQKGVIGLI	VPQEDMPWTE	SGIVPDLIVN
P08518	<i>S. cerevisiae</i>	2	RMSD	KVRVRTTKIP	Q-IGDKFASR	HGQKGTIGIT	YRREDMPFTA	EGIVPDLIIN
P08518	<i>S. cerevisiae</i>	2	PSIPRED	KVRVRTTKIP	Q-IGDKFASR	HGQKGTIGIT	YRREDMPFTA	EGIVPDLIIN
P08518	<i>S. cerevisiae</i>	2	PHD	KVRVRTTKIP	Q-IGDKFASR	HGQKGTIGIT	YRREDMPFTA	EGIVPDLIIN
P08518	<i>S. cerevisiae</i>	2	MOE	KVRVRTTKIP	Q-IGDKFASR	HGQKGTIGIT	YRREDMPFTA	EGIVPDLIIN
P08518	<i>S. cerevisiae</i>	2	BLAST	KVRVRTTKIP	Q-IGDKFASR	HGQKGTIGIT	YRREDMPFTA	EGIVPDLIIN
P08518	<i>S. cerevisiae</i>	2	BLAST2	KVRVRTTKIP	Q-IGDKFASR	HGQKGTIGIT	YRREDMPFTA	EGIVPDLIIN
P08518	<i>S. cerevisiae</i>	2	MPsrch	KVRVRTTKIP	Q-IGDKFASR	HGQKGTIGIT	YRREDMPFTA	EGIVPDLIIN
P08518	<i>S. cerevisiae</i>	2	Multa	KVRVRTTKIP	Q-IGDKFASR	HGQKGTIGIT	YRREDMPFTA	EGIVPDLIIN
P38420	<i>A. thaliana</i>	2	Multa	KVRVRSVRIP	Q-IGDKFSSR	HGQKGTVGMT	YTQEDMPWTI	EGVTPDIIIN
Q42877	<i>L. esculentum</i>	2	Multa	KVRVRSVRIP	Q-IGDKFSSR	HGQKGTVGMT	YTQEDMPWTI	EGVTPDIIIN
Q02061	<i>S. pombe</i>	2	Multa	KVRMRSTRIP	Q-IGDKFASR	HGQKGTIGMT	YRHEDMPFSA	QGIVPDLIIN
Q10578	<i>C. elegans</i>	2	Multa	KIRMRSTRIP	Q-IGDKFASR	HGQKGTIGMT	YRHEDMPFSA	QGIVPDLIIN
P08266	<i>D. melanogaster</i>	2	Multa	KIRVRSVRIP	Q-IGDKFASR	HGQKGTIGMT	YRHEDMPFSA	QGIVPDLIIN
P30876	<i>H. sapiens</i>	2	Multa	KIRVRSVRIP	Q-IGDKFASR	HGQKGTIGMT	YRHEDMPFSA	QGIVPDLIIN
Q98RY5	<i>G. theta</i>	2	Multa	KIRIRSVRLP	Q-IGDKFASR	HGQKGTIGMT	YRHEDMPFSA	QGIVPDLIIN
Q27826	<i>E. octocarinatus</i>	2	Multa	KVRMRSTRIP	Q-IGDKFASR	HGQKGTIGMT	YRHEDMPFSA	QGIVPDLIIN
Q9YAT8	<i>A. pernix</i>	B	Multa	KVRVRNLRIP	E-IGDKFASR	HGQKGVIGMI	FPRYDMPYTE	EGIEPDVILN
O59226	<i>P. horikoshii</i>	B	Multa	KVTVRDLRIP	E-IGDKFASR	HGQKGVIGLI	VPQEDMPWTE	SGIVPDLIVN
Q9V115	<i>P. abyssi</i>	B	Multa	KVTVRDLRIP	E-IGDKFASR	HGQKGVIGLI	VPQEDMPWTE	SGIVPDLIVN
P31814	<i>T. celer</i>	B	Multa	KVTTRDLRIP	E-FGDKFASR	HGQKGVIGLI	VPQEDMPWTE	SGIVPDLIVN
Q979F3	<i>T. volcanium</i>	B	Multa	KIKVRSERIP	E-LGDKFASR	HGQKGVIGLV	VPQEDMPFTE	EGIVPDLIIN
Q03587	<i>T. acidophilum</i>	B	Multa	KIKVRSERIP	E-LGDKFASR	HGQKGVIGLV	VPQEDMPFTE	EGIVPDLIIN
P11513	<i>S. acidocaldarius</i>	B	Multa	KVRVRDLRIP	E-IGDKFATR	HGQKGVVGLI	IDQVDMPYTA	KGIVPDLIIN
Q980R1	<i>S. solfataricus</i>	B'	Multa	GIAGKYAALS	GNIVDATPFY	KTPV-EQLQN	EILRYGYLPD	ATEVVYDGRT
O28392	<i>A. fulgidus</i>	B'	Multa	MIGGKVGSL	GRRVDGSLFH	GESE-EDLRA	TLKKYGFST	GKEVMYDGIT
Q60181	<i>M. jannaschii</i>	B'	Multa	MLGGKVGSL	GRRVDGTIFS	GEKE-WDLRK	ALEALGFKHH	GKEVMYDGKT
P41557	<i>M. vannielii</i>	B'	Multa	MIGGKVGSL	CRRVDGTIFS	GEKE-WALRH	ALENYGFTHS	GKETMYDGKT
O27124	<i>M. thermoautotrophicum</i>	B'	Multa	MLAGKAACME	GRRVDGTFFT	GEDE-DDIKE	ALRANGFETA	GVESLYNGIT
P09845	<i>M. thermoautotrophicum</i> (st. Winter)	B'	Multa	MLAGKAACME	GRRVDGTFFT	GEDE-DDIKE	ALRANGFETA	GVESLYNGIT
Q58444	<i>M. jannaschii</i>	B''	Multa					
P41558	<i>M. vannielii</i>	B''	Multa					
O27123	<i>M. thermoautotrophicum</i>	B''	Multa					
P09844	<i>M. thermoautotrophicum</i> (str. Winter)	B''	Multa					
O28393	<i>A. fulgidus</i>	B''	Multa					
Q980R0	<i>S. solfataricus</i>	B''	Multa					
P15352	<i>Halobacterium</i> sp.	B''	Multa					

*Method:
 BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
 PSIPRED and PHD = Structure Prediction
 Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.71 Sequence alignment for *P. furiosus* RNAP subunit B (sheet 29)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	-----920	-----930	-----940	-----950-----
P81412	<i>P. furiosus</i>	B	RMSD	PHGIPSRMTV	GQLIEAIGGK	VASLKGRRV	GTAFIGEPE- EKLKLEEL
P81412	<i>P. furiosus</i>	B	PSIPRED	PHGIPSRMTV	GQLIEAIGGK	VASLKGRRV	GTAFIGEPE- EKLKLEEL
P81412	<i>P. furiosus</i>	B	PHD	PHGIPSRMTV	GQLIEAIGGK	VASLKGRRV	GTAFIGEPE- EKLKLEEL
P08518	<i>S. cerevisiae</i>	2	RMSD	PHAIPSRMTV	AHLIECLLSK	VAALSGNEGD	ASPFTDITV- EGISKLLREH
P08518	<i>S. cerevisiae</i>	2	PSIPRED	PHAIPSRMTV	AHLIECLLSK	VAALSGNEGD	ASPFTDITV- EGISKLLREH
P08518	<i>S. cerevisiae</i>	2	PHD	PHAIPSRMTV	AHLIECLLSK	VAALSGNEGD	ASPFTDITV- EGISKLLREH
P08518	<i>S. cerevisiae</i>	2	MOE	PHAIPSRMTV	AHLIECLLSK	VAALSGNEGD	ASPFTDITV- EGISKLLREH
P08518	<i>S. cerevisiae</i>	2	BLAST	PHAIPSRMTV	AHLIECLLSK	VAALSGNEGD	ASPFTDITV- EGISKLLREH
P08518	<i>S. cerevisiae</i>	2	BLAST2	PHAIPSRMTV	AHLIECLLSK	VAALSGNEGD	ASPFTDITV- EGISKLLREH
P08518	<i>S. cerevisiae</i>	2	MPsrch	PHAIPSRMTV	AHLIECLLSK	VAALSGNEGD	ASPFTDITV- EGISKLLREH
P08518	<i>S. cerevisiae</i>	2	Multa	PHAIPSRMTV	AHLIECLLSK	VAALSGNEGD	ASPFTD-ITV EGISKLLREH
P38420	<i>A. thaliana</i>	2	Multa	PHAIPSRMTI	GQLIECIMGK	VAAHMGKEGD	ATPFTD-VTV DNISKALHHC
Q42877	<i>L. esculentum</i>	2	Multa	PHAIPSRMTI	GQLIECIMGK	VAAHMGKEGD	ATPFTD-VTV DNISKALHHC
Q02061	<i>S. pombe</i>	2	Multa	PHAIPSRMTV	AHLVECLLSK	VSAISGFEGD	ATPFTD-VTV EAVSKLLRSH
Q10578	<i>C. elegans</i>	2	Multa	PHAVPSRMTI	GHLECLQSK	LSANKGEIGD	ATPFNDTVNV QKISGLLCEY
P08266	<i>D. melanogaster</i>	2	Multa	PHAIPSRMTI	GHLECLQSK	LGSNKGEIGD	ATPFNDVNV QKISTFLQY
P30876	<i>H. sapiens</i>	2	Multa	PHAIPSRMTI	GHLECLQSK	VSANKGEIGD	ATPFNDVNV QKISNLLSDY
Q98RY5	<i>G. theta</i>	2	Multa	PHAIPSRMTI	GHLECLLSK	AVSMKGIEGD	GTPFER-QNM EDIMNLFLEKN
Q27826	<i>E. octocarinatus</i>	2	Multa	PHAIPSRMTI	GHLECLASK	VACFKGEEGD	ATPFQD-VTV EHSNDLHRL
Q9YAT8	<i>A. pernix</i>	B	Multa	PHAIPSRMTV	GQLEEIAGK	LGAIKGRFVD	ATPFKE-DI DGLRIELLKE
O59226	<i>P. horikoshii</i>	B	Multa	PHGIPSRMTV	GQLEAIGGK	VAAKGRRV	GTAFIGEPE- EKLKLEEL
Q9V115	<i>P. abyssi</i>	B	Multa	PHGIPSRMTV	GQLEAIGGK	VAAKGRRV	GTAFIGEPE- EKLKLEEL
P31814	<i>T. celer</i>	B	Multa	PHGIPSRMTV	GQLIEAIGGK	VASLKGRRV	GTAFIGEPE- EKLKLEEL
Q979F3	<i>T. volcanium</i>	B	Multa	PHSIPSRMTV	GHILEMIGK	IASRTGKFID	GTIFSGEPE- KSLRDLVKY
Q03587	<i>T. acidophilum</i>	B	Multa	PHSIPSRMTV	GHILEMIGK	IASRTGKFID	GTIFSGEPE- KSLRDLVKY
P11513	<i>S. acidocaldarius</i>	B	Multa	PHALPSRMTI	GQIMEAIGGK	YAALSGKPV	ATPFLETPL QEMQKEILKL
Q980R1	<i>S. solfataricus</i>	B'	Multa	GQKIKSRIYF	GVVYQKLHH	MVADKLHARA	RGPVQILTRQ PTEGRAREGG
O28392	<i>A. fulgidus</i>	B'	Multa	GRRYLVDIVF	GVVYQKLHH	MVSSKIHARS	RGPVQVLTTRQ PTEGRARKGG
Q60181	<i>M. jannaschii</i>	B'	Multa	GKKFEVEIYI	GIAYYQKLHH	LVAGKIHARS	RGPVQVLTTRQ PTEGRAREGG
P41557	<i>M. vannieli</i>	B'	Multa	GRKLECEIFV	GVAYYQKLHH	LVAGKIHARS	RGPIQVLTTRQ PTEGRAREGG
O27124	<i>M. thermoautotrophicum</i>	B'	Multa	GERIEAEIFI	GVAYYQKLHH	MTTDKVYARS	RGPVQVLTTRQ PTEGRAREGG
P09845	<i>M. thermoautotrophicum</i> (st. Winter)	B'	Multa	GERIEAEIFI	GVAYYQKLHH	MTTDRIYARS	RGPVQVLTTRQ PTEGRAREGG
Q58444	<i>M. jannaschii</i>	B''	Multa				
P41558	<i>M. vannieli</i>	B''	Multa				
O27123	<i>M. thermoautotrophicum</i>	B''	Multa				
P09844	<i>M. thermoautotrophicum</i> (str. Winter)	B''	Multa				
O28393	<i>A. fulgidus</i>	B''	Multa				
Q980R0	<i>S. solfataricus</i>	B''	Multa				
P15352	<i>Halobacterium</i> sp.	B''	Multa				

*Method:

BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD

PSIPRED and PHD = Structure Prediction

Red = Helix Green = Strand Blue = Turn Gray = No cord

Figure C.72 Sequence alignment for *P. furiosus* RNAP subunit B (sheet 30)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	960-----	970-----	980-----	990-----	1000-----
P81412	P. furiosus	B	RMSD	GFKHTGREIM	YDGTGKRLE	ADIFIGVIYY	QRLHHMVADK	IHARSRGP--
P81412	P. furiosus	B	PSIPRED	GFKHTGREIM	YDGTGKRLE	ADIFIGVIYY	QRLHHMVADK	IHARSRGP--
P81412	P. furiosus	B	PHD	GFKHTGREIM	YDGTGKRLE	ADIFIGVIYY	QRLHHMVADK	IHARSRGP--
P08518	S. cerevisiae	2	RMSD	GYQSRGFEVM	YNGHTGKKLM	AQIFFGPTY	QRLHHMVDDK	IHARARGPMQ
P08518	S. cerevisiae	2	PSIPRED	GYQSRGFEVM	YNGHTGKKLM	AQIFFGPTY	QRLHHMVDDK	IHARARGPMQ
P08518	S. cerevisiae	2	PHD	GYQSRGFEVM	YNGHTGKKLM	AQIFFGPTY	QRLHHMVDDK	IHARARGPMQ
P08518	S. cerevisiae	2	MOE	GYQSRGFEVM	YNGHTGKKLM	AQIFFGPTY	QRLHHMVDDK	IHARARGPMQ
P08518	S. cerevisiae	2	BLAST	GYQSRGFEVM	YNGHTGKKLM	AQIFFGPTY	QRLHHMVDDK	IHARARGP--
P08518	S. cerevisiae	2	BLAST2	GYQSRGFEVM	YNGHTGKKLM	AQIFFGPTY	QRLHHMVDDK	IHARARGP--
P08518	S. cerevisiae	2	MPsrch	GYQSRGFEVM	YNGHTGKKLM	AQIFFGPTY	QRLHHMVDDK	IHARARGP--
P08518	S. cerevisiae	2	Multa	GYQSRGFEVM	YNGHTGKKLM	AQIFFGPTY	QRLHHMVDDK	IHARARGP--
P38420	A. thaliana	2	Multa	GYQMRGFERM	YNGHTGRPLT	AMIFLGPTY	QRLHHMVDDK	IHSRGRGP--
Q42877	L. esculentum	2	Multa	GYQMRGFETM	YNGHTGRRLS	AMIFLGPTY	QRLHHMVDDK	IHSRGRGP--
Q02061	S. pombe	2	Multa	GFKSRGFEVM	YHGHTGRKLV	AQVFLGPTY	QRLKHLVDDK	IHARARGP--
Q10578	C. elegans	2	Multa	GYHLRGNEVM	YNGHTGKKLT	TQIFFGPTY	QRLHHMVDDK	IHSRARGP--
P08266	D. melanogaster	2	Multa	GYHLRGNEVM	YNGHTGRKIN	AQVFLGPTY	QRLHHMVDDK	IHSRARGP--
P30876	H. sapiens	2	Multa	GYHLRGNEVL	YNGFTGRKIT	SQIFIGPTY	QRLHHMVDDK	IHSRARGP--
Q98RY5	G. theta	2	Multa	NFDKNGWEIL	YNGFNGKRIK	SLIFMGPTY	QRLHHMVDDK	IHSRARGP--
Q27826	E. octocarinatus	2	Multa	GYQKRGNEVM	YDGTGGRKMD	TMIFLGPTY	QRLHHMVDDK	IHSRARGP--
Q9YAT8	A. pernix	B	Multa	GYDPNAREVM	YDGTGQIIE	APITIGVAY	QRLHHMVADK	IHARATGR--
O59226	P. horikoshii	B	Multa	GFKHTGREIM	YDGTGKRFE	ADIFIGIYY	QRLHHMVADK	IHARSRGP--
Q9V115	P. abyssi	B	Multa	GFKHTGREIM	YDGTGKRFE	ADIFVGVIYY	QRLHHMVADK	IHARSRGP--
P31814	T. celer	B	Multa	GFKHSGREVM	YDGTGRRLE	ADVFGVIYY	QRLHHMVADK	MHARSRGP--
Q979F3	T. volcanium	B	Multa	GFRKSGTEVM	YDGTGKRFR	ADIFVGVIYY	QRLHHMVAGK	FHARSRGP--
Q03587	T. acidophilum	B	Multa	GFRKSSTEVM	YDGTGRRFK	ADIFVGVIYY	QRLHHMVAGK	FHARSRGP--
P11513	S. acidocaldarius	B	Multa	GHLPDSTEVM	YDGTGQKLE	SRLFGIYY	QRLHHMVADK	MHARSRGP--
Q980R1	S. solfataricus	B'	Multa	LRFGEMERDC	LIGFGTAMLL	KDRLLDSDR	TMIYVCDQ--	CGYIG-WY--
O28392	A. fulgidus	B'	Multa	LRFGEMERDV	LIGHGAALLL	KDRLLDESDK	VEVWVCGN--	CGHVA-TY--
Q60181	M. jannaschii	B'	Multa	LRFGEMERDV	LIGHGAALLL	KERLMDESDP	YDICICSK--	CGDFA-IL--
P41557	M. vannielii	B'	Multa	LRFGEMERDV	LVAHGAALLL	KERLLDESDP	HEDYVCAK--	CGEIA-IF--
O27124	M. thermoautotrophicum	B'	Multa	LRFGEMERDC	LIAHGAALLL	KERLLDESDK	YEALVCA--	CGMIA-VY--
P09845	M. thermoautotrophicum (st. Winter)	B'	Multa	LRFGEMERDC	LIAHGAALLL	KERLLDESDK	YEALVCAE--	CGMIA-IY--
Q58444	M. jannaschii	B''	Multa					
P41558	M. vannielii	B''	Multa					
O27123	M. thermoautotrophicum	B''	Multa					
P09844	M. thermoautotrophicum (str. Winter)	B''	Multa					
O28393	A. fulgidus	B''	Multa					
Q980R0	S. solfataricus	B''	Multa					
P15352	Halobacterium sp.	B''	Multa					

*Method:
 BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
 PSIPRED and PHD = Structure Prediction
 Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.73 Sequence alignment for *P. furiosus* RNAP subunit B (sheet 31)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	1010--	1020--	1030--	1040--
P81412	<i>P. furiosus</i>	B	RMSD	-----VQVLTK	QPTTEGRAREG	GLRFGEMERD	VLVGHGAAML
P81412	<i>P. furiosus</i>	B	PSIPRED	-----VQVLTK	QPTTEGRAREG	GLRFGEMERD	VLVGHGAAML
P81412	<i>P. furiosus</i>	B	PHD	-----VQVLTK	QPTTEGRAREG	GLRFGEMERD	VLVGHGAAML
P08518	<i>S. cerevisiae</i>	2	RMSD	VLTRQPVVEGR	SRDG-----	GLRFGEMERD	CMIAHGAASF
P08518	<i>S. cerevisiae</i>	2	PSIPRED	VLTRQPVVEGR	SRDG-----	GLRFGEMERD	CMIAHGAASF
P08518	<i>S. cerevisiae</i>	2	PHD	VLTRQPVVEGR	SRDG-----	GLRFGEMERD	CMIAHGAASF
P08518	<i>S. cerevisiae</i>	2	MOE	VLTRQPVVEGR	SRDG-----	GLRFGEMERD	CMIAHGAASF
P08518	<i>S. cerevisiae</i>	2	BLAST	-----MQVLTR	QPVEGRSRDG	GLRFGEMERD	CMIAHGAASF
P08518	<i>S. cerevisiae</i>	2	BLAST2	-----MQVLTR	QPVEGRSRDG	GLRFGEMERD	CMIAHGAASF
P08518	<i>S. cerevisiae</i>	2	MPsrch	-----MQVLTR	QPVEGRSRDG	GLRFGEMERD	CMIAHGAASF
P08518	<i>S. cerevisiae</i>	2	Multa	-----MQVLTR	QPVEGRSRDG	GLRFGEMERD	CMIAHGAASF
P38420	<i>A. thaliana</i>	2	Multa	-----VQI LTR	QPAEGRSRDG	GLRFGEMERD	CMIAHGAASF
Q42877	<i>L. esculentum</i>	2	Multa	-----VQI LTR	QPAEGRSRDG	GLRFGEMERD	CMIAHGAASF
Q02061	<i>S. pombe</i>	2	Multa	-----VQI LTR	QPVEGRSRDG	GLRFGEMERD	CQISHGCSSV
Q10578	<i>C. elegans</i>	2	Multa	-----IQMMNR	QPMGRARDG	GLRFGEMERD	CQISHGATQF
P08266	<i>D. melanogaster</i>	2	Multa	-----VQI LTR	QPMGRARDG	GLRFGEMERD	CQISHGAAQF
P30876	<i>H. sapiens</i>	2	Multa	-----IQI LNR	QPMGRSRDG	GLRFGEMERD	CQIAHGAASF
Q98RY5	<i>G. theta</i>	2	Multa	-----IQI LTR	QPVEGRSRDG	GLRFGEMERD	CMISHGAASF
Q27826	<i>E. octocarinatus</i>	2	Multa	-----LQI LTR	QPTGRSRHG	GLRFGEMERD	CMVSHGAASF
Q9YAT8	<i>A. pernix</i>	B	Multa	-----VHLLTR	QPTGKARQG	GLRFGEMERD	CLVGHGASML
O59226	<i>P. horikoshii</i>	B	Multa	-----VQVLTK	QPTTEGRAREG	GLRFGEMERD	VLVGHGAAML
Q9V115	<i>P. abyssi</i>	B	Multa	-----VQVLTK	QPTTEGRAREG	GLRFGEMERD	VLVGHGAAML
P31814	<i>T. celer</i>	B	Multa	-----VQVLTK	QPTTEGRAREG	GLRFGEMERD	VLIGHGAAML
Q979F3	<i>T. volcanium</i>	B	Multa	-----VQI LTR	QPTGRSRQG	GLRFGEMERD	TLIAHGAAMV
Q03587	<i>T. acidophilum</i>	B	Multa	-----VQI LTR	QPTGRSRQG	GLRFGEMERD	TLIAHGAAMV
P11513	<i>S. acidocaldarius</i>	B	Multa	-----VQI LTR	QPTTEGRAREG	GLRFGEMERD	CLIGFGTAML
Q980R1	<i>S. solfataricus</i>	B'	Multa	-----DKNKNK	YVCPIHGDKS	NLF-----P	VTVSAYAFKLL
O28392	<i>A. fulgidus</i>	B'	Multa	-----DYRRNT	AYCHICDDSE	NIH-----K	VEMSYAFKLL
Q60181	<i>M. jannaschii</i>	B'	Multa	-----DYKRGL	KYCPCIGIE	NLYSKKIPF	VRIPIYAFKLL
P41557	<i>M. vannielii</i>	B'	Multa	-----DYKRGM	KFCPVCGESE	DIQDNRKIPP	VKIAYAFKLL
O27124	<i>M. thermoautotrophicum</i>	B'	Multa	-----DKIRDK	KYCPCICEDSE	SF-----P	VEVSAYAFKLL
P09845	<i>M. thermoautotrophicum</i> (st. Winter)	B'	Multa	-----DKIRDK	KYCPCICEDSE	SF-----P	VEISYAFKLL
Q58444	<i>M. jannaschii</i>	B''	Multa				
P41558	<i>M. vannielii</i>	B''	Multa				
O27123	<i>M. thermoautotrophicum</i>	B''	Multa				
P09844	<i>M. thermoautotrophicum</i> (str. Winter)	B''	Multa				
O28393	<i>A. fulgidus</i>	B''	Multa				
Q980R0	<i>S. solfataricus</i>	B''	Multa				
P15352	<i>Halobacterium</i> sp.	B''	Multa				

*Method:
 BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
 PSIPRED and PHD = Structure Prediction
 Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.74 Sequence alignment for *P. furiosus* RNAP subunit B (sheet 32)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	-----1050--	-----1060--	,-----,---	1070--,-,-	1080-----
P81412	<i>P. furiosus</i>	B	RMSD	LIERLLEESD	KTEVWVCEN-	-CGHIA-LED	KRRRRV-Y-C	PVCGEEERIS
P81412	<i>P. furiosus</i>	B	PSIPRED	LIERLLEESD	KTEVWVCEN-	-CGHIA-LED	KRRRRV-Y-C	PVCGEEERIS
P81412	<i>P. furiosus</i>	B	PHD	LIERLLEESD	KTEVWVCEN-	-CGHIA-LED	KRRRRV-Y-C	PVCGEEERIS
P08518	<i>S. cerevisiae</i>	2	RMSD	LKERLMEASD	AFRVHICGI-	-CGLMT-VIA	KLNNHQFE-C	KGCDNKIDIY
P08518	<i>S. cerevisiae</i>	2	PSIPRED	LKERLMEASD	AFRVHICGI-	-CGLMT-VIA	KLNNHQFE-C	KGCDNKIDIY
P08518	<i>S. cerevisiae</i>	2	PHD	LKERLMEASD	AFRVHICGI-	-CGLMT-VIA	KLNNHQFE-C	KGCDNKIDIY
P08518	<i>S. cerevisiae</i>	2	MOE	LKERLMEASD	AFRVHICGI-	-CGLMT-VIA	KLNNHQ-FEC	KGCDNKIDIY
P08518	<i>S. cerevisiae</i>	2	BLAST	LKERLMEASD	AFRVHICGI-	-CGLMT-VIA	KLNNHQ-FEC	KGCDNKIDIY
P08518	<i>S. cerevisiae</i>	2	BLAST2	LKERLMEASD	AFRVHICGI-	-CGLMT-VIA	KLNNHQ-FEC	KGCDNKIDIY
P08518	<i>S. cerevisiae</i>	2	MPsrch	LKERLMEASD	AFRVHICGI-	-CGLMT-VIA	KLNNHQ-FEC	KGCDNKIDIY
P08518	<i>S. cerevisiae</i>	2	Multa	LKERLMEASD	AFRVHICGI-	-CGLMTVIK	LNHNQF-E-C	KGCDNKIDIY
P38420	<i>A. thaliana</i>	2	Multa	LKERLFDQSD	AYRVHVCEV-	-CGLIA-IAN	LKKNSE-E-C	RGCKNKTDIV
Q42877	<i>L. esculentum</i>	2	Multa	LKERLFDQSD	AYRVHVCEV-	-CGLIA-IAN	LKKNSE-E-C	RGCKNKTDIV
Q02061	<i>S. pombe</i>	2	Multa	LKERLFDQSD	AYRVHVCDI-	-CGLIA-IAS	YKDSY-E-C	RSCQNRTRFS
Q10578	<i>C. elegans</i>	2	Multa	LKERLFEVSD	PYHVYVCNN-	-CGLIV-VAN	LRTNSF-E-C	KACRNKTQVS
P08266	<i>D. melanogaster</i>	2	Multa	LKERLFEVSD	PYRVHICNF-	-CGLIA-IAN	LRTNTF-E-C	KGCKNKTDIV
P30876	<i>H. sapiens</i>	2	Multa	LKERLFEASD	PYQVHVCNL-	-CGIMA-IAN	TRTHTY-E-C	RGCRNKTDIV
Q98RY5	<i>G. theta</i>	2	Multa	LKERLFDQSD	PFGIYVCDI-	-CGFIV-VSN	KQNNF-E-C	RNCNRSMSIS
Q27826	<i>E. octocarinatus</i>	2	Multa	LKERLFDVSD	CYTVHVCRI-	-CGLIC-EAN	LQQKY-L-C	RGCKNSTNIR
Q9YAT8	<i>A. pernix</i>	B	Multa	LKERLFDQSD	AYVMYVCNT-	-CGNIA-WFN	RKKVL-E-C	PIHGEEDIR
O59226	<i>P. horikoshii</i>	B	Multa	LIERLLEESD	KTEVWVCEN-	-CGHIA-IED	KRRHKV-Y-C	PVCGEEERIS
Q9V115	<i>P. abyssi</i>	B	Multa	LIERLLEESD	KTEVWVCEN-	-CGHIA-LED	KRRHKV-Y-C	PVCGEEERIS
P31814	<i>T. celer</i>	B	Multa	LIERLLEESD	KTEVWVCEN-	-CGHIA-LED	KRRGV-Y-C	PVCGEEERIS
Q979F3	<i>T. volcanium</i>	B	Multa	LKERLFDQSD	GTVLYVCGNP	SCGHIA-IYD	RKGT-L-R-C	PVCGNTGNIY
Q03587	<i>T. acidophilum</i>	B	Multa	LKERLFDQSD	GTVLYVCGNP	SCGHIA-IYD	RKGT-L-R-C	PVCGNTGNIY
P11513	<i>S. acidocaldarius</i>	B	Multa	LKERLFDQSD	KAVVYICDQ-	-CGYVG-WYD	RSKNRY-V-C	PVHGDKSVLH
Q980R1	<i>S. solfataricus</i>	B'	Multa	LKERLFDQSD	PRLVLEKVG	LSGG		
O28392	<i>A. fulgidus</i>	B'	Multa	LDELKSMIA	PRILGDKA			
Q60181	<i>M. jannaschii</i>	B'	Multa	LDELKSMIL	PRIKVRDKVE	LEDFFEFVEK	LEKEKV-KQK	Q
P41557	<i>M. vannielii</i>	B'	Multa	LDELKSMGID	PKLKLKRA			
O27124	<i>M. thermoautotrophicum</i>	B'	Multa	LDELKSLCIF	PKLILEDKA			
P09845	<i>M. thermoautotrophicum</i> (str. Winter)	B'	Multa	LDELKSLCIF	PKLILEDKA			
Q58444	<i>M. jannaschii</i>	B''	Multa					
P41558	<i>M. vannielii</i>	B''	Multa					
O27123	<i>M. thermoautotrophicum</i>	B''	Multa					
P09844	<i>M. thermoautotrophicum</i> (str. Winter)	B''	Multa					
O28393	<i>A. fulgidus</i>	B''	Multa					
Q980R0	<i>S. solfataricus</i>	B''	Multa					
P15352	<i>Halobacterium</i> sp.	B''	Multa					

*Method:

BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
PSIPRED and PHD = Structure Prediction

Red = Helix Green = Strand Blue = Turn Gray = No cord

Figure C.75 Sequence alignment for *P. furiosus* RNAP subunit B (sheet 33)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	1090-----	1100-----	1110,,,-	1117
P81412	P. furiosus	B	RMSD	KVEMSYAFKL	LLDELKAMVI	RPKL--N--L	SERV
P81412	P. furiosus	B	PSIPRED	KVEMSYAFKL	LLDELKAMVI	RPKL--N--L	SERV
P81412	P. furiosus	B	PHD	KVEMSYAFKL	LLDELKAMVI	RPKL--N--L	SERV
P08518	S. cerevisiae	2	RMSD	QIHIPYAAKL	LFQELMAMNI	TPRL--YTDR	SRDF
P08518	S. cerevisiae	2	PSIPRED	QIHIPYAAKL	LFQELMAMNI	TPRL--YTDR	SRDF
P08518	S. cerevisiae	2	PHD	QIHIPYAAKL	LFQELMAMNI	TPRL--YTDR	SRDF
P08518	S. cerevisiae	2	MOE	QIHIPYAAKL	LFQELMAMNI	TPRLYTDR--R	SRDF
P08518	S. cerevisiae	2	BLAST	QIHIPYAAKL	LFQELMAMNI	TPRL	
P08518	S. cerevisiae	2	BLAST2	QIHIPYAAKL	LFQELMAMNI	TPRL	
P08518	S. cerevisiae	2	MPsrch	QIHIPYAAKL	LFQELMAMNI	TPRL	
P08518	S. cerevisiae	2	Multa	QIHIPYAAKL	LFQELMAMNI	TPRL--Y--T	DRSRDF
P38420	A. thaliana	2	Multa	QVYIPYACKL	LFQELMSMAI	APRM--L--T	KHLKSAKGRQ
Q42877	L. esculentum	2	Multa	QVHIPYACKL	LFQELMAMAI	APRM--L--T	KDVKLAKDQK KKGA
Q02061	S. pombe	2	Multa	QVYLPYAAKL	LFQELMSMNI	APRL--F--T	KNHK
Q10578	C. elegans	2	Multa	AVRIPYACKL	LFQELMSMSI	APRL--M--V	KPRQSKRSKH QSEA
P08266	D. melanogaster	2	Multa	QVRLPYAAKL	LFQELMSMNI	APRL--M--V	T
P30876	H. sapiens	2	Multa	LVRMPYACKL	LFQELMSMSI	APRM--M--S	V
Q98RY5	G. theta	2	Multa	FIRIPYAFKL	LIQELMSLSV	GVRLL--I--P	GEF
Q27826	E. octocarinatus	2	Multa	QVGLPYACKL	LFQELMAMQI	APRM--I--T	SVS
Q9YAT8	A. pernix	B	Multa	PVKVPYAFKL	LLQELTSMNI	KPEL--K--V	ADKISVIRKV VGDYT
O59226	P. horikoshii	B	Multa	KVEMSYAFKL	LLDELKAMVI	RPKL--N--L	TERV
Q9V115	P. abyssi	B	Multa	KVEMSYAFKL	LLDELKAMVI	RPKL--N--L	TERV
P31814	T. celer	B	Multa	KVEMSYAFKL	LLDELKAMGI	RPSL--K--L	VDRV
Q979F3	T. volcanium	B	Multa	PIETSYAFKL	MRDELISLGV	VMRL--M--L	GDMK
Q03587	T. acidophilum	B	Multa	PIETSYAFKL	MRDELISLGV	VMRL--M--L	GDMK
P11513	S. acidocaldarius	B	Multa	PVTVSYAFKL	LIQELMSMVI	SPRL--I--L	GEKVNIGGAS NE
Q980R1	S. solfataricus	B'	Multa				
O28392	A. fulgidus	B'	Multa				
Q60181	M. jannaschii	B'	Multa				
P41557	M. vannielii	B'	Multa				
O27124	M. thermoautotrophicum	B'	Multa				
P09845	M. thermoautotrophicum (str. Winter)	B'	Multa				
Q58444	M. jannaschii	B''	Multa				
P41558	M. vannielii	B''	Multa				
O27123	M. thermoautotrophicum	B''	Multa				
P09844	M. thermoautotrophicum (str. Winter)	B''	Multa				
O28393	A. fulgidus	B''	Multa				
Q980R0	S. solfataricus	B''	Multa				
P15352	Halobacterium sp.	B''	Multa				

*Method:

BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
 PSIPRED and PHD = Structure Prediction
 Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.76 Sequence alignment for *P. furiosus* RNAP subunit B (sheet 34)

RNAP Subunit Alignments for *Pyrococcus furiosus*

1

Accession #	Source	SU	Method*	----	-----10-----	-----20-----	-----30-----
Q8U0E4	P. furiosus	D	RMSD	MAGI	EVQILEKKED	SIKFFVLKGVH	VSFANALRRRT
Q8U0E4	P. furiosus	D	PSIPred	MAGI	EVQILEKKED	SIKFFVLKGVH	VSFANALRRRT
Q8U0E4	P. furiosus	D	PHD	MAGI	EVQILEKKED	SIKFFVLKGVH	VSFANALRRRT
P16370	S. cerevisiae	3	RMSD	EEGP	QVKIREASKD	NVDFILSNVD	LAMANSLLRRV
P16370	S. cerevisiae	3	PSIPred	EEGP	QVKIREASKD	NVDFILSNVD	LAMANSLLRRV
P16370	S. cerevisiae	3	PHD	EEGP	QVKIREASKD	NVDFILSNVD	LAMANSLLRRV
P16370	S. cerevisiae	3	MOE	EEGP	QVKIREASKD	NVDFILSNVD	LAMANSLLRRV
P16370	S. cerevisiae	3	BLAST	GP	QVKIREASKD	NVDFILSNVD	LAMANSLLRRV
P16370	S. cerevisiae	3	BLAST2	GP	QVKIREASKD	NVDFILSNVD	LAMANSLLRRV
P16370	S. cerevisiae	3	MPsrch	GP	QVKIREASKD	NVDFILSNVD	LAMANSLLRRV
P16370	S. cerevisiae	3	Multa	MSEEGP	QVKIREASKD	NVDFILSNVD	LAMANSLLRRV
O97183	D. melanogaster	3	Multa	MPYANQP	SVQITELTDD	NVKFVLEDE	LSVANSLRRV
P19387	H. sapiens	3	Multa	MPYANQP	TVRITELTDE	NVKFIIENTD	LAVANSIRRV
P97760	M. musculus	3	Multa	MPYANQP	TVRITELTEE	NVKFIIENTD	LAVANSIRRV
Q39211	A. thaliana	3A	Multa	MDGATYQRF	KIKIRELKDD	YAKFELRETD	VSMANALRRV
Q9AW23	G. theta	3	Multa		MKVEKISDF	EIKINFNNFP	IVYANSIRRI
P37382	S. pombe	3	Multa	MDSET	HITIRNISK	SVDFVLTNTS	LAVANSLLRV
O26144	M. thermoautotrophicum	D	Multa	M	DIALKEKTD	EMVFFVSGVT	VFFINAIRRI
Q8ZYQ3	P. aerophilum	D	Multa	MA	RATVVERGEY	FLKAVVEGVP	PSLVNSLLRV

*Method:
BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
PSIPRED and PHD = Structure Prediction
Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.77 Sequence alignment for *P. furiosus* RNAP subunit D (sheet 1)

RNAP Subunit Alignments for *Pyrococcus furiosus*

1

Accession #	Source	SU	Method*	----40----	----50----	----60----	----70----
Q8U0E4	P. furiosus	D	RMSD	ILGEVPTFAV	DEVEFYENDS	ALFDEIIAHR	LAMIPL-TT-
Q8U0E4	P. furiosus	D	PSIPred	ILGEVPTFAV	DEVEFYENDS	ALFDEIIAHR	LAMIPL-TT-
Q8U0E4	P. furiosus	D	PHD	ILGEVPTFAV	DEVEFYENDS	ALFDEIIAHR	LAMIPL-TT-
P16370	S. cerevisiae	3	RMSD	MIAEIPTLAI	DSVEVETNTT	VLADEFIAHR	LGLIPLQSM-
P16370	S. cerevisiae	3	PSIPred	MIAEIPTLAI	DSVEVETNTT	VLADEFIAHR	LGLIPLQSM-
P16370	S. cerevisiae	3	PHD	MIAEIPTLAI	DSVEVETNTT	VLADEFIAHR	LGLIPLQSM-
P16370	S. cerevisiae	3	MOE	MIAEIPTLAI	DSVEVETNTT	VLADEFIAHR	LGLIPLQSM-
P16370	S. cerevisiae	3	BLAST	MIAEIPTLAI	DSVEVETNTT	VLADEFIAHR	LGLIPL-QS-
P16370	S. cerevisiae	3	BLAST2	MIAEIPTLAI	DSVEVETNTT	VLADEFIAHR	LGLIPL-QSM
P16370	S. cerevisiae	3	MPsrch	MIAEIPTLAI	DSVEVETNTT	VLADEFIAHR	LGLIPLQSM-
P16370	S. cerevisiae	3	Multa	MIAEIPTLAI	DSVEVETNTT	VLADEFIAHR	LGLIPL-QS-
O97183	D. melanogaster	3	Multa	FIAETPTLAI	DWVQLEANST	VLSDEFLAHR	IGLIPL-IS-
P19387	H.sapiens	3	Multa	FIAEVPIIAI	DWQIDANSS	VLHDEFIAHR	LGLIPL-IS-
P97760	M. musculus	3	Multa	FIAEVPIIAI	DWQIDGNSS	VLHDEFIAHR	LGLIPL-TS-
Q39211	A. thaliana	3A	Multa	MISEVPTVAI	DLVEIEVNSS	VLNDEFIAHR	LGLIPL-TS-
Q9AW23	G. theta	3	Multa	MMNEIPTLAI	DKVYIEKNTS	SFKDEFIAHR	LGLIPI-FS-
P37382	S. pombe	3	Multa	VLAEIPTVAI	DLVEINVNTS	VMPDEFIAHR	LGLIPL-DS-
O26144	M. thermoautotrophicum	D	Multa	CMMEVPKLAI	EYVNMRYNDA	KMFDEVLAHR	LGLVPL-AA-
Q8ZYQ3	P. aerophilum	D	Multa	IISELPVMAI	DNVVVVNNTS	VMYDEMLAHR	LGLIPL-TT-

*Method:

BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD

PSIPRED and PHD = Structure Prediction

Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.78 Sequence alignment for *P. furiosus* RNAP subunit D (sheet 2)

RNAP Subunit Alignments for *Pyrococcus furiosus*

1

Accession #	Source	SU	Method*	-----, , , , , 80- - - - -
Q8U0E4	P. furiosus	D	RMSD	PVDRFEL--- ---DAL--- E---L--- ---DDY--
Q8U0E4	P. furiosus	D	PSIPred	PVDRFEL--- ---DAL--- E---L--- ---DDY--
Q8U0E4	P. furiosus	D	PHD	PVDRFEL--- ---DAL--- E---L--- ---DDY--
P16370	S. cerevisiae	3	RMSD	DIEQLEYSR- ---DCF--- CEDHC-----DKC--
P16370	S. cerevisiae	3	PSIPred	DIEQLEYSR- ---DCF--- CEDHC-----DKC--
P16370	S. cerevisiae	3	PHD	DIEQLEYSR- ---DCF--- CEDHC-----DKC--
P16370	S. cerevisiae	3	MOE	DIEQLEYSR- ---DCFCEH H---C-----DKC--
P16370	S. cerevisiae	3	BLAST	---MDI--- ---EQL--- E---YSRDCF CEDHCDKC--
P16370	S. cerevisiae	3	BLAST2	DIEQLEYSR- ---DCFCEH H---C-----DKC--
P16370	S. cerevisiae	3	MPsrch	DIEQLE
P16370	S. cerevisiae	3	Multa	MDIEQ----- LEYSRDC--- F---C-----EDHCD
O97183	D. melanogaster	3	Multa	DDVVER--- LQYTRDC--- I---C-----LDFCP
P19387	H.sapiens	3	Multa	DDIVDK--- LQYSRDC--- T---C-----EEFCP
P97760	M. musculus	3	Multa	DDIVDK--- LQYSRDC--- T---C-----EEFCP
Q39211	A. thaliana	3A	Multa	ERAMSMR--F SRDCDAC--- D---G-----DGQCE
Q9AW23	G. theta	3	Multa	ENADL----- IKYSREC--- D---C-----ENECE
P37382	S. pombe	3	Multa	SNIDEPPVVG LEYTRNC--- D---C-----DQYCP
O26144	M. thermoautotrophicum	D	Multa	DPEFAES--- LKMPREC--- D---C-----EEYCS
Q8ZYQ3	P. aerophilum	D	Multa	PLQSLPP--- YEDCVSG--- L---A-----DPSEC

*Method:
BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
PSIPRED and PHD = Structure Prediction
Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.79 Sequence alignment for *P. furiosus* RNAP subunit D (sheet 3)

RNAP Subunit Alignments for *Pyrococcus furiosus*

1

Accession #	Source	SU	Method*	90	100
Q8U0E4	P. furiosus	D	RMSD	--TVTLSTL-- --EA----E- ---GPGI--- -VYSGDL---	
Q8U0E4	P. furiosus	D	PSIPred	--TVTLSTL-- --EA----E- ---GPGI--- -VYSGDL---	
Q8U0E4	P. furiosus	D	PHD	--TVTLSTL-- --EA----E- ---GPGI--- -VYSGDL---	
P16370	S. cerevisiae	3	RMSD	--SVVLTLL-- --QA----FG ESESTTN--- -VYSKDL---	
P16370	S. cerevisiae	3	PSIPred	--SVVLTLL-- --QA----FG ESESTTN--- -VYSKDL---	
P16370	S. cerevisiae	3	PHD	--SVVLTLL-- --QA----FG ESESTTN--- -VYSKDL---	
P16370	S. cerevisiae	3	MOE	--SVVLTLLQA FGES----E- ---STTN--- -VYSKDLVIV	
P16370	S. cerevisiae	3	BLAST	--SVVLTLLQA FGES----E- ---STTN--- -VYSKDLVIV	
P16370	S. cerevisiae	3	BLAST2	--SVVLTLL-- --QA----F- ---GESESTT NVYSKDL---	
P16370	S. cerevisiae	3	MPsrch		
P16370	S. cerevisiae	3	Multa	KCSVVLTLL-- --QAFGESE- ---STTN--- -VYSKDL---	
O97183	D. melanogaster	3	Multa	ECSVEFTLL-- --DVKCSEE- ---QTRH--- -VTTADL---	
P19387	H.sapiens	3	Multa	ECSVEFTLL-- --DVRCNED- ---QTRH--- -VTSRDL---	
P97760	M. musculus	3	Multa	ECSVEFTLL-- --DVRCNED- ---QTRH--- -VTSRDL---	
Q39211	A. thaliana	3A	Multa	FCSVEFRL-- --SSKCVTD- ---QTLT--- -VTSRDL---	
Q9AW23	G. theta	3	Multa	KCSILFTLL-- --DVHAN-E- ---TEKT--- -ITSSMV---	
P37382	S. pombe	3	Multa	KCSVELFL-- --NAKCTGE- ---GTME--- -IYARDL---	
O26144	M. thermoautotrophicum	D	Multa	ECSVSLTLL-- --RK----K- ---GPGV--- -VYSGDL---	
Q8ZYQ3	P. aerophilum	D	Multa	STRLTLLQV-- --TA----E- ---GDTT--- -IYSGDL---	

*Method:

BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD

PSIPRED and PHD = Structure Prediction

Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.80 Sequence alignment for *P. furiosus* RNAP subunit D (sheet 4)

RNAP Subunit Alignments for *Pyrococcus furiosus*

1

Accession #	Source	SU	Method*	110
Q8U0E4	P. furiosus	D	RMSD	-----KS-- --DDP DVKPVN-PN- -----
Q8U0E4	P. furiosus	D	PSIPred	-----KS-- --DDP DVKPVN-PN- -----
Q8U0E4	P. furiosus	D	PHD	-----KS-- --DDP DVKPVN-PN- -----
P16370	S. cerevisiae	3	RMSD	-----VIVS NLMGRNIGHP IIQDKEGNG- -----
P16370	S. cerevisiae	3	PSIPred	-----VIVS NLMGRNIGHP IIQDKEGNG- -----
P16370	S. cerevisiae	3	PHD	-----VIVS NLMGRNIGHP IIQDKEGNG- -----
P16370	S. cerevisiae	3	MOE	SNLMGRNI-- --GHP IIQDKE-GNG -----
P16370	S. cerevisiae	3	BLAST	SNLMGRNI-- --GHP IIQDKE-GNG -----
P16370	S. cerevisiae	3	BLAST2	-----VI-- --VS- NLMGRN-IG- -----
P16370	S. cerevisiae	3	MPsrch	-----VI-- --VS- NLMGRN-IG- -----
P16370	S. cerevisiae	3	Multa	-----VI-- --VSN LMGRNI-GHP IIQDKEGNG- -----
O97183	D. melanogaster	3	Multa	-----KS-- --SNA KVLPT-SRN QGEEDNEYGE
P19387	H.sapiens	3	Multa	-----IS-- --NSP RVIPVT-SRN RDNDPNDYVE
P97760	M. musculus	3	Multa	-----IS-- --NSP RVIPVT-SRN RDNDPNDYVE
Q39211	A. thaliana	3A	Multa	-----YS-- --ADP TVTPVD-FTI DSSVSDSSEH
Q9AW23	G. theta	3	Multa	-----KC-- --STA CYPLKT-FFI INNSGNSRNY
P37382	S. pombe	3	Multa	-----VV-- --SSN ---SSL-GHP ILADPKSRG-
O26144	M. thermoautotrophicum	D	Multa	-----VS-- --ETP AVKPVY-PD- -----
Q8ZYQ3	P. aerophilum	D	Multa	-----AS-- --DRP DVVPVY-KD- -----

*Method:

BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD

PSIPRED and PHD = Structure Prediction

Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.81 Sequence alignment for *P. furiosus* RNAP subunit D (sheet 5)

RNAP Subunit Alignments for *Pyrococcus furiosus*

1

Accession #	Source	SU	Method*	120---130---140---150---
Q8U0E4	P. furiosus	D	RMSD	----IPIVKL AEGQRLVFNA YAKLGRGKDH AKWQPG----
Q8U0E4	P. furiosus	D	PSIPred	----IPIVKL AEGQRLVFNA YAKLGRGKDH AKWQPG----
Q8U0E4	P. furiosus	D	PHD	----IPIVKL AEGQRLVFNA YAKLGRGKDH AKWQPG----
P16370	S. cerevisiae	3	RMSD	----VLICKL RKQQLKLTTC VAKKGIKEH AKWGPA----
P16370	S. cerevisiae	3	PSIPred	----VLICKL RKQQLKLTTC VAKKGIKEH AKWGPA----
P16370	S. cerevisiae	3	PHD	----VLICKL RKQQLKLTTC VAKKGIKEH AKWGPA----
P16370	S. cerevisiae	3	MOE	----VLICKL RKQQLKLTTC VAKKGIKEH AKWGPA----
P16370	S. cerevisiae	3	BLAST	----VLICKL RKQQLKLTTC VAKKGIKEH AKWGPA--
P16370	S. cerevisiae	3	BLAST2	----HPIIQD KEGNGVLI-- -CKLRKGQE
P16370	S. cerevisiae	3	MPsrch	----HPIIQD KEGNGVLI-- -CKLRKGQE
P16370	S. cerevisiae	3	Multa	----VLICKL RKQQLKLTTC VAKKGIKEH AKWGPA----
O97183	D. melanogaster	3	Multa	SNDEILIKL RKQQLKLTTC VAKKGIKEH AKWGPA----
P19387	H.sapiens	3	Multa	Q-DDILIVKL RKQQLKLTTC VAKKGIKEH AKWGPA----
P97760	M. musculus	3	Multa	Q-DDILIVKL RKQQLKLTTC VAKKGIKEH AKWGPA----
Q39211	A. thaliana	3A	Multa	K--GIIIVKL RKQQLKLTTC VAKKGIKEH AKWGPA----
Q9AW23	G. theta	3	Multa	DDYAI IIAKL NPGQNIKLA YVKKGIKEH AKWGPA----
P37382	S. pombe	3	Multa	----PLICKL RKEQEISLRC IAKKGIKEH AKWGPA----
O26144	M. thermoautotrophicum	D	Multa	----IPLVKL GDDDELELEA VAQLGVGREH AKWGPA----
Q8ZYQ3	P. aerophilum	D	Multa	----IPIVKL VKGQSIVLEA YAKLGVAKDH AKWGPA----

*Method:
BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
PSIPRED and PHD = Structure Prediction
Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.82 Sequence alignment for *P. furiosus* RNAP subunit D (sheet 6)

RNAP Subunit Alignments for *Pyrococcus furiosus*

1

Accession #	Source	SU	Method*	---,---160 ---,---,--- --170,--- ---180-
Q8U0E4	P. furiosus	D	RMSD	FVY-YKYYTI VHI-----SK SIPEWK--EL KKLAKKRGLP
Q8U0E4	P. furiosus	D	PSIPred	FVY-YKYYTI VHI-----SK SIPEWK--EL KKLAKKRGLP
Q8U0E4	P. furiosus	D	PHD	FVY-YKYYTI VHI-----SK SIPEWK--EL KKLAKKRGLP
P16370	S. cerevisiae	3	RMSD	AAI-EFEYDP WNK-----LK HTDYWY---E QDSAK--EWP
P16370	S. cerevisiae	3	PSIPred	AAI-EFEYDP WNK-----LK HTDYWY---E QDSAK--EWP
P16370	S. cerevisiae	3	PHD	AAI-EFEYDP WNK-----LK HTDYWY---E QDSAK--EWP
P16370	S. cerevisiae	3	MOE	AAI-EFEYDP WNK-----LK HTDYWY---E QDSAK--EWP
P16370	S. cerevisiae	3	BLAST	FEY-DPWNKL KHTDYWYEQD SAKWEP--QS KNCEYED--P
P16370	S. cerevisiae	3	BLAST2	
P16370	S. cerevisiae	3	MPsrch	
P16370	S. cerevisiae	3	Multa	AAIEFEYDPW NKL-----KH TDYWEQDSA KEWPQSKNCE
O97183	D. melanogaster	3	Multa	AGVCFEYDPD NSM-----RH TLY----PKP DEWPKSEHTE
P19387	H.sapiens	3	Multa	AGVAFEYDPD NAL-----RH TVY----PKP EEWPKSEYSE
P97760	M. musculus	3	Multa	AGVAFEYDPD NGL-----RH TVY----PKP EEGPKSEYSE
Q39211	A. thaliana	3A	Multa	ATVTFMYEPD III-----NE DMMDTSLDEE KIDLIESPT
Q9AW23	G. theta	3	Multa	SSIKFRFYPI YQV-----N- -IFEMKKNLR PDILNKLENF
P37382	S. pombe	3	Multa	SAVAFEYDPW NKL-----QH TDYWFENDAD AEWPQSKNAD
O26144	M. thermoautotrophicum	D	Multa	TACAYKYYPR IEF-----SE DCDECL--EC IE-ACPRDVL
Q8ZYQ3	P. aerophilum	D	Multa	TASYYYYPKV VVK-----SE NCREMC--KE I-----

*Method:

BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD

PSIPRED and PHD = Structure Prediction

Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.83 Sequence alignment for *P. furiosus* RNAP subunit D (sheet 7)

RNAP Subunit Alignments for *Pyrococcus furiosus*

1

Accession #	Source	SU	Method*	-----190- -----200- -----210- -----220-
Q8U0E4	P. furiosus	D	RMSD	VEETEEVLV TTIKPFYIPK DFEEYEGKEI WEEIVPNTYI
Q8U0E4	P. furiosus	D	PSIPred	VEETEEVLV TTIKPFYIPK DFEEYEGKEI WEEIVPNTYI
Q8U0E4	P. furiosus	D	PHD	VEETEEVLV TTIKPFYIPK DFEEYEGKEI WEEIVPNTYI
P16370	S. cerevisiae	3	RMSD	QSKNCEY--- -----EDFP N---EGDPF DYKAQADTFY
P16370	S. cerevisiae	3	PSIPred	QSKNCEY--- -----EDFP N---EGDPF DYKAQADTFY
P16370	S. cerevisiae	3	PHD	QSKNCEY--- -----EDFP N---EGDPF DYKAQADTFY
P16370	S. cerevisiae	3	MOE	QSKNCE--- -----YEDFP N---EGDPF DYKAQADTFY
P16370	S. cerevisiae	3	BLAST	PNEGD----- -----PF DYKAQ----- -----ADTFY
P16370	S. cerevisiae	3	BLAST2	
P16370	S. cerevisiae	3	MPsrch	
P16370	S. cerevisiae	3	Multa	YEDPPNEGDP FDYKAQADTF YMNVESVSGSI PVDQVVVRGI
O97183	D. melanogaster	3	Multa	LEDDQYEA-P YNWEAKPNKF FFNVESAGAL KPENIVVMGV
P19387	H.sapiens	3	Multa	LDEDESQA-P YDPNGKPERF YYNVESCGSL RPETIVLSAL
P97760	M. musculus	3	Multa	LDEDESQA-P YDPNGKPERF YYNVESCGSL RPETIGLSAL
Q39211	A. thaliana	3A	Multa	KVFGMDPVTR QVVVDPEAY TYDEEVIK-- KAEAMGKPGI
Q9AW23	G. theta	3	Multa	IKDGIKNSDL CYNEKNSDYK CINLDNERKI ELLSIFEKNI
P37382	S. pombe	3	Multa	WEPPREGEP FNFQEEPRF YMDVESVSGSI PFNEIMVQGL
O26144	M. thermoautotrophicum	D	Multa	GEESGKPVVV DLENCMCKS CVRACDKRAI DVGEEGKFI
Q8ZYQ3	P. aerophilum	D	Multa	CRQLENPIEC SFNKAWTCKD LCKE---GL EVTWDKNKYV

*Method:

BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD

PSIPRED and PHD = Structure Prediction

Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.84 Sequence alignment for *P. furiosus* RNAP subunit D (sheet 8)

RNAP Subunit Alignments for *Pyrococcus furiosus*

1

Accession #	Source	SU	Method*	-----230-	-----240-	-----250-	-----261
Q8U0E4	P. furiosus	D	RMSD	FTVETNGELP	VEEIVSIALK	ILMRKADRFI	SELQKLTS
Q8U0E4	P. furiosus	D	PSIPred	FTVETNGELP	VEEIVSIALK	ILMRKADRFI	SELQKLTS
Q8U0E4	P. furiosus	D	PHD	FTVETNGELP	VEEIVSIALK	ILMRKADRFI	SELQKLTS
P16370	S. cerevisiae	3	RMSD	MNVESVGSIP	VDQVVVRGID	TLQKKVASIL	LALTQMDQD
P16370	S. cerevisiae	3	PSIPred	MNVESVGSIP	VDQVVVRGID	TLQKKVASIL	LALTQMDQD
P16370	S. cerevisiae	3	PHD	MNVESVGSIP	VDQVVVRGID	TLQKKVASIL	LALTQMDQD
P16370	S. cerevisiae	3	MOE	MNVESVGSIP	VDQVVVRGID	TLQKKVASIL	LALTQMDQD
P16370	S. cerevisiae	3	BLAST	MNVESVGSIP	VDQVVVRGID	TLQKKVASIL	LALTQMDQD
P16370	S. cerevisiae	3	BLAST2	MNVESVGSIP	VDQVVVRGID	TLQKKVASIL	LALTQMDQD
P16370	S. cerevisiae	3	MPsrch	MNVESVGSIP	VDQVVVRGID	TLQKKVASIL	LALTQMDQD
P16370	S. cerevisiae	3	Multa	DTLQKKV-AS	ILLALTQMDQ	DKVNFASGDN	NTASNMLGSN
O97183	D. melanogaster	3	Multa	QVLKKNKL-SN	LQ---TQLSH	ESQNDALAV	
P19387	H.sapiens	3	Multa	SGLKKKL-SD	LQ---TQLSH	EIQSDVLTIN	
P97760	M. musculus	3	Multa	SGLKKKL-SD	LQ---TQLSH	EIQSDVLTIN	
Q39211	A. thaliana	3A	Multa	IEISPKD-DS	FI-----FTV	ESTGAVKASQ	LVLNAIDLLK
Q9AW23	G. theta	3	Multa	LNLEEIYIYIS	ENDANVEFFI	ETTGVINPIN	LINKSLIYLK
P37382	S. pombe	3	Multa	RILQEKL-AV	LV---RDLDE	EQPTQLSANE	LNMEENAEMN
O26144	M. thermoautotrophicum	D	Multa	FRIETDGSVD	PKDVLLKACD	ILRDKAEQVI	TFCEGG
Q8ZYQ3	P. aerophilum	D	Multa	FWVESFGNYD	VNTALREAFR	ILKRKFAVFA	EELFRRASVE

*Method:

BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD

PSIPRED and PHD = Structure Prediction

Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.85 Sequence alignment for *P. furiosus* RNAP subunit D (sheet 9)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	-----	--10-----	--20-----	--30-----	--40-----
Q8U0M2	P. furiosus	H	RMSD	MAGKKE	FSIFDHVLP	EHRVLSSEK	KALLEKYKIT	LAQLPQIKAS
Q8U0M2	P. furiosus	H	PSIPRED	MAGKKE	FSIFDHVLP	EHRVLSSEK	KALLEKYKIT	LAQLPQIKAS
Q8U0M2	P. furiosus	H	PHD	MAGKKE	FSIFDHVLP	EHRVLSSEK	KALLEKYKIT	LAQLPQIKAS
P20434	S. cerevisiae	5	RMSD	IETFNEAALV	VNITHHELVP	KHIRLSSEK	RELLKRYRLK	ESQLPRIQRA
P20434	S. cerevisiae	5	PSIPRED	IETFNEAALV	VNITHHELVP	KHIRLSSEK	RELLKRYRLK	ESQLPRIQRA
P20434	S. cerevisiae	5	PHD	IETFNEAALV	VNITHHELVP	KHIRLSSEK	RELLKRYRLK	ESQLPRIQRA
P20434	S. cerevisiae	5	MOE	IETFNEAALV	VNITHHELVP	KHIRLSSEK	RELLKRYRLK	ESQLPRIQRA
P20434	S. cerevisiae	5	BLAST		NITHHELVP	KHIRLSSEK	RELLKRYRLK	ESQLPRIQRA
P20434	S. cerevisiae	5	BLAST2		NITHHELVP	KHIRLSSEK	RELLKRYRLK	ESQLPRIQRA
P20434	S. cerevisiae	5	MPsrch		NITHHELVP	KHIRLSSEK	RELLKRYRLK	ESQLPRIQRA
P20434	S. cerevisiae	5	Multa	IETFNEAALV	VNITHHELVP	KHIRLSSEK	RELLKRYRLK	ESQLPRIQRA
Q9V116	P. abyssi	H	Multa	MAGKSE	FNIFKHVLP	EHRILSEEEK	KALLEKYRIT	PAQLPQIKAS
O74019	P. horikoshii	H	Multa	MSGKNE	FNIFKHVLP	EHRILSEEEK	KALLEKYKIT	PAQLPQIRAS
P31815	T. celer	H	Multa	MATKKE	FDIFAHVLP	EHRILSEET	KELLRRYRIQ	ISQLPQIKAS
Q9YAT3	A. permix	H	Multa	MSSKTRKSS	RVILEHEYVP	EHRLLSIEEA	VQVLKMLGK	PWQLPKISVN
P11521	S. acidocaldarius	H	Multa	MRSSSKKK	IDISNHELVP	KHEILQLEEA	YKLVKELGK	PEQLPWIRAS
Q976A3	S. tokodaii	H	Multa	MRSSSKIK	IDPSQHELVP	KHEVLSIEEA	YKILKELGK	PEQLPWIRAS
Q980Q9	S. solfataricus	H	Multa	MRGSSNKK	IDPRIHYLVP	KHEVLNIDEA	YKILKELGIR	PEQLPWIRAS
O28394	A. fulgidus	H	Multa	MK	FKLQDHMLVP	KHEVLSKEEA	EELLKILGIG	KEQLPKIKAD
O27122	M. thermoautotrophicum	H	Multa	MK	REILKHQLVP	EHVILNESEA	KRVLKELDAH	PEQLPKIKTT
Q58443	M. jannaschii	H	Multa		MKVTDHILVP	KHEIVPKEEV	EEILKRYNIK	IQQLPKIYED
P41559	M. vannieli	H	Multa		MKVSSHEMVP	EHEIIPGDEI	PFLLEKYGIK	IQQFPKLLIS
Q03588	T. acidophilum	H	Multa	MAK	FNVLDHNLVP	EHHIVPEEEE	KNILKELNIE	KEFLPKISPN
Q979F2	T. volcanium	H	Multa	MAK	FNVLDHNLVP	EHHIVSEEEE	KTILKELGIE	KEFLPRISPN
Q9P4B9	K. marxianus	5	Multa	IETFPEASLV	VNITHHELVP	KHIRLSDAEK	KELLKRYRLK	ESQLPRIQRM
Q9ZVB9	A. thaliana	Multa		IETFQESDLI	VNITHHELVP	KHILLSPDEK	KELLDYKRLR	ETQLPRIQLA
Q09191	S. pombe	5	Multa	LEVFEQAEML	VNIKEHVLVP	EHQVLTTEEK	KTLLERYTVK	ETQLPRIQVT
O81098	A. thaliana	Multa		LEQFLEQELL	INITEHELVP	EHVMTKEEV	TELLARYKLR	ENQLPRIQAG
P19388	H. sapiens	5	Multa	LEQFLESELL	INITEHELVP	EHVMTVEEK	QELLSRYKLR	ENMLMRIQAG
Q9V5S0	D. melanogaster	Multa		LEHFLEAELM	VNITEHELVP	EHVMTAEK	AELLARYKLR	DSQLPRIQQC
Q9N5K2	C. elegans	Multa		IEYFSENELI	VNITHMLVP	KHEILSKNES	KSLLERYSIK	FDHLPKILKN
Q98RN7	G. theta	Multa		VELFQLTELL	VNITKHVLRP	KHHVLNEQEK	ESLFFKFSIQ	EQQLPKLLKK
Q9LKX4	B. napus	Multa			MSRISLGVK	EVTVIPREEA	KELLKRMKLR	PWQLPWIRAS

*Method:

BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD

PSIPRED and PHD = Structure Prediction

Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.86 Sequence alignment for *P. furiosus* RNAP subunit H (sheet 1)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	--50----	----60--	-----70-	-----80	82
Q8U0M2	P. furiosus	H	RMSD	DPVAVKALGA-	----KPGDVI	EIKRKSPTA-	GVYYYYRVVV	ED
Q8U0M2	P. furiosus	H	PSIPRED	DPVAVKALGA-	----KPGDVI	EIKRKSPTA-	GVYYYYRVVV	ED
Q8U0M2	P. furiosus	H	PHD	DPVAVKALGA-	----KPGDVI	EIKRKSPTA-	GVYYYYRVVV	ED
P20434	S. cerevisiae	5	RMSD	DPVALYLGL-	----KRGEVV	KIIRKSETS-	GRYASYRICM	
P20434	S. cerevisiae	5	PSIPRED	DPVALYLGL-	----KRGEVV	KIIRKSETS-	GRYASYRICM	
P20434	S. cerevisiae	5	PHD	DPVALYLGL-	----KRGEVV	KIIRKSETS-	GRYASYRICM	
P20434	S. cerevisiae	5	MOE	DPVALYLGL-	----KRGEVI	KIIRKSETS-	GRYASYRICM	
P20434	S. cerevisiae	5	BLAST	DPVALYLGL-	----KRGEVV	KIIRKSETS-	GRYASYRIC	
P20434	S. cerevisiae	5	BLAST2	DPVALYLGL-	----KRGEVV	KIIRKSETS-	GRYASYRIC	
P20434	S. cerevisiae	5	MPsrch	DPVALYLGL-	----KRGEVV	KIIRKSETS-	GRYASYRIC	
P20434	S. cerevisiae	5	Multa	DPVALYLGL-	----KRGEVV	KIIRKSETS-	GRYASYRICM	
Q9V116	P. abyssi	H	Multa	DPVAVKALGA-	----KPGDII	EIKRKSPTA-	GVYYYYRVVV	ED
O74019	P. horikoshii	H	Multa	DPVAVKALGA-	----KPGDII	EIKRKSPTA-	GVYYYYRVVV	ED
P31815	T. celer	H	Multa	DPVAVVALGA-	----KPGDVL	EIKRKSRTA-	GVYYYYRLVV	ED
Q9YAT3	A. pemix	H	Multa	DPIARLLKA-	----KPGDII	EITRRSYTA-	GEAKYYRFVV	AYQKGVK
P11521	S. acidocaldarius	H	Multa	DPVAKSIGA-	----KPGDII	KITRKSPTT-	GESVTYRYVI	TG
Q976A3	S. tokodaii	H	Multa	DPIAKLVGA-	----KPGDII	KITRKSLS-	GEVVVYRYVI	SG
Q980Q9	S. solfataricus	H	Multa	DPVARSINA-	----KPGDII	RIIRKSQLY-	GEVVSRYRYVI	SG
O28394	A. fulgidus	H	Multa	DPIAKEIGA-	----KPGDIV	KITRKSLS-	GESVFYRLVV	
O27122	M. thermoautotrophicum	H	Multa	DPVAKAIGA-	----KRGDIV	KIIRKSPTA-	EEFVTYRLVQ	D
Q58443	M. jannaschii	H	Multa	DPVIQEIGA-	----KEGDVV	RVIRKSPTA-	GVSIAYRLVI	KRII
P41559	M. vannielii	H	Multa	DPLVIEVGA-	----KVGDIV	KITRKSQTA-	GETVYFRLVV	SNSV
Q03588	T. acidophilum	H	Multa	DPAIKALEAV	HGKIKEGTII	KIVRKSPTM-	GHSVYYRVVA	SEVFK
Q979F2	T. volcanium	H	Multa	DPVIKALEAI	HGKIKDGTVI	KIIRNSPTM-	GHSVYYRVVA	SEVFK
Q9P4B9	K. marxianus	5	Multa	DPVALYLGL-	----KRGEVI	KIIRKSETS-	GRYASYRICL	
Q9ZVB9	A. thaliana	Multa	Multa	DPVARYLGL-	----KRGEVV	KIVRRSETS-	GRYNSYRICA	
Q09191	S. pombe	5	Multa	DPIARYFGL-	----KRQQVV	KIIRPSETA-	GRYVTYRYVV	
O81098	A. thaliana	Multa	Multa	DPVARYFGI-	----KRQQVV	KIIRPSETA-	GRYITYRLVQ	
P19388	H. sapiens	5	Multa	DPVARYFGL-	----KRQQVV	KIIRPSETA-	GRYISYRLVC	
Q9V5S0	D. melanogaster	Multa	Multa	DPVARYFGL-	----RRQQVV	KIIRPSETA-	GRYITYRLVV	
Q9N5K2	C. elegans	Multa	Multa	DPVAKYLGL-	----QRSEIV	KITRVSETS-	GKFITYRICI	
Q98RN7	G. theta	Multa	Multa	DPTAKYYGL-	----EKQQVV	EVTYKGESE	SDHVSRYCAW	
Q9LKX4	B. napus	Multa	Multa	DPLAQKAGA-	----KPGDVL	KIVRESPTA-	GEAVVYRLVV	PG

*Method:

BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD

PSIPRED and PHD = Structure Prediction

Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.87 Sequence alignment for *P. furiosus* RNAP subunit H (sheet 2)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	-	---,---10	-----20	-----30	-,,-----
Q8U0E8	P. furiosus	K	RMSD	M	FKY-TRFEKA	RIIGARALQI	SMGAPVLIDV	P--PGITPLE
Q8U0E8	P. furiosus	K	PSIPRED	M	FKY-TRFEKA	RIIGARALQI	SMGAPVLIDV	P--PGITPLE
Q8U0E8	P. furiosus	K	PHD	M	FKY-TRFEKA	RIIGARALQI	SMGAPVLIDV	P--PGITPLE
P20435	S. cerevisiae	6	RMSD	KAIPKQQRAT	TPYMTKYERA	RILGTRALQI	SMNAPVFVDL	E--GETDPLR
P20435	S. cerevisiae	6	PSIPRED	KAIPKQQRAT	TPYMTKYERA	RILGTRALQI	SMNAPVFVDL	E--GETDPLR
P20435	S. cerevisiae	6	PHD	KAIPKQQRAT	TPYMTKYERA	RILGTRALQI	SMNAPVFVDL	E--GETDPLR
P20435	S. cerevisiae	6	MOE	KAIPKQQRAT	TPYMTKYERA	RILGTRALQI	SMNAPVFVDL	E--GETDPLR
P20435	S. cerevisiae	6	BLAST		TKYERA	RILGTRALQI	SMNAPVFVDL	E--GETDPLR
P20435	S. cerevisiae	6	BLAST2		TKYERA	RILGTRALQI	SMNAPVFVDL	E--GETDPLR
P20435	S. cerevisiae	6	MPsrch		TKYERA	RILGTRALQI	SMNAPVFVDL	E--GETDPLR
P20435	S. cerevisiae	6	Multa		M-TKYERA	RILGTRALQI	SMNAPVFVDL	-E-GETDPLR
Q98RN6	G. theta	6	Multa		L-TKYEKA	RILGARALQI	SMGAPILTEV	-N-TLSDSLD
Q17684	C. elegans	6	Multa	HVTT	PFM-TKYERA	RVLGTRALQI	AMGAPVMVEL	-E-GETDPLE
Q24320	D. melanogaster	6	Multa	ITT	KYM-TKYERA	RVLGTRALQI	AMCAPIMVEL	-D-GETDPLQ
P36595	S. pombe	6	Multa		M-TKYERA	RILGTRALQI	SMNAPVLVDL	-E-GETDPLQ
O26148	M. thermoautotrophicum	K	Multa	HMAS	KKL-TRFERA	RLIGARALQI	SMGARPLVEI	-K-ESLDPVD
Q9HJ14	T. acidophilum	K	Multa		KY-TKFEKA	RIIGARALQI	AMGAPVIIDV	PK-NIIDPVD
P58232	T. volcanium	K	Multa		KY-TKFEKA	RIIGARALQI	AMGAPVIIDI	PK-NVIDPVD
Q9YDJ9	A. pernix	K	Multa	MKIGP	PYL-TKYERA	RIIGVRAFQL	SIGAPPLVDP	ERAGSRNPLD
Q8ZTF9	P. aerophilum	K	Multa		RL-TKYETA	RIIGARALQL	AMGAQPLVDI	QEVGSVDPLV
Q972B1	S. tokodaii	K	Multa		RL-TKYEIA	RIISARALQL	AMGASPLIDT	NNLPFNDVIS
Q57650	M. jannaschii	K	Multa		MKL-TKFEIA	RILGARSLQI	SSGAYATJET	K--CD-SSLK

*Method:
 BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
 PSIPRED and PHD = Structure Prediction
 Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.88 Sequence alignment for *P. furiosus* RNAP subunit K (sheet 1)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	-40-----	-50-----	57
Q8U0E8	P. furiosus	K	RMSD	AAILEFEKGV	IPITVIR--P	S
Q8U0E8	P. furiosus	K	PSIPRED	AAILEFEKGV	IPITVIR--P	S
Q8U0E8	P. furiosus	K	PHD	AAILEFEKGV	IPITVIR--P	S
P20435	S. cerevisiae	6	RMSD	IAMKELAEKK	IPLVIRR--Y	LPDGSFEDWS VEELIVDL
P20435	S. cerevisiae	6	PSIPRED	IAMKELAEKK	IPLVIRR--Y	LPDGSFEDWS VEELIVDL
P20435	S. cerevisiae	6	PHD	IAMKELAEKK	IPLVIRR--Y	LPDGSFEDWS VEELIVDL
P20435	S. cerevisiae	6	MOE	IAMKELAEKK	IPLVIRRYLP	DGSFEDWSVE ELIVDL
P20435	S. cerevisiae	6	BLAST	IAMKELAEKK	IPLVIRR	
P20435	S. cerevisiae	6	BLAST2	IAMKELAEKK	IPLVIRR	
P20435	S. cerevisiae	6	MPsrch	IAMKELAEKK	IPL-VIR	
P20435	S. cerevisiae	6	Multa	IAMKELAEKK	IPLVIRR	
Q98RN6	G. theta	6	Multa	IASKELYERK	IPITIRR	
Q17684	C. elegans	6	Multa	IARKELKQRR	IPIIVRR	
Q24320	D. melanogaster	6	Multa	IAMKELKQKK	IPIIIRR	
P36595	S. pombe	6	Multa	IAMKELAQKK	IPLLVRR	
O26148	M. thermoautotrophicum	K	Multa	IARKELEKKV	MPLDVRR	
Q9HJ14	T. acidophilum	K	Multa	IAMLEFENNV	IPITIKK	
P58232	T. volcanium	K	Multa	IATIEFENNV	IPITIKR	
Q9YDJ9	A. pernix	K	Multa	IARYEVDNGI	LPVSIYR	
Q8ZTF9	P. aerophilum	K	Multa	IAMEELRRGL	LDFIIVRE	
Q972B1	S. tokodaii	K	Multa	IAEELKRGV	LPITVRR	
Q57650	M. jannaschii	K	Multa	IAYEEIKQ GK	VPLKPIRPVK	A

*Method:

BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
 PSIPRED and PHD = Structure Prediction
 Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.89 Sequence alignment for *P. furiosus* RNAP subunit K (sheet 2)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	- - - - -10- - - - -20- - - - -30-
Q8U4N1	<i>P. furiosus</i>	L	RMSD	M KIEVIKKEEN LLEFYLEGED HTFANLLVET
Q8U4N1	<i>P. furiosus</i>	L	PSIPRED	M KIEVIKKEEN LLEFYLEGED HTFANLLVET
Q8U4N1	<i>P. furiosus</i>	L	PHD	M KIEVIKKEEN LLEFYLEGED HTFANLLVET
P38902	<i>S. cerevisiae</i>	11	RMSD	MNAPDRFEL FLLGEGESKL KIDPDTKAPN AVVITFEKED HTLGNLIRAE
P38902	<i>S. cerevisiae</i>	11	PSIPRED	MNAPDRFEL FLLGEGESKL KIDPDTKAPN AVVITFEKED HTLGNLIRAE
P38902	<i>S. cerevisiae</i>	11	PHD	MNAPDRFEL FLLGEGESKL KIDPDTKAPN AVVITFEKED HTLGNLIRAE
P38902	<i>S. cerevisiae</i>	11	MOE	MNAPDRFEL FLLGEGESKL KIDPDTKAPN AVVITFEKED HTLGNLIRAE
P38902	<i>S. cerevisiae</i>	11	BLAST	L KIDPDTKAPN AVVITFEKED HTLGNLIRAE
P38902	<i>S. cerevisiae</i>	11	BLAST2	L KIDPDTKAPN AVVITFEKED HTLGNLIRAE
P38902	<i>S. cerevisiae</i>	11	MPsrch	L KIDPDTKAPN AVVITFEKED HTLGNLIRAE
P38902	<i>S. cerevisiae</i>	11	Multa	MNAPDRFEL FLLGEGESKL KIDPDTKAPN AVVITFEKED HTLGNLIRAE
Q98SA6	<i>G. theta</i>	Y	Multa	MQNYTSSFD- -EVPSTSEKI KIINFNNKN TKEFLFENED ATMGEILVYK
P52435	<i>H. sapiens</i>	11	Multa	MNAPPAFES FLLFEGEKKI TINKDTKVPN ACLFTINKED HTLGNIIKSQ
O08740	<i>M. musculus</i>	11	Multa	MNAPPAFES FLLFEGEKKI TINKDTKVPN ACLFTINKED HTLGNIIKSQ
Q9VJE4	<i>D. melanogaster</i>	11	Multa	MNAPPTFES FLLYEGEKKI IKELDTKVIN AAIFTINKED HTLGNMIRNQ
Q9XVH6	<i>C. elegans</i>	11	Multa	MNAPPAFES FLLLD-DKKF YIEKDTKVPN AAIFTIMKED HTLGNMLKIQ
Q38859	<i>A. thaliana</i>	11	Multa	MNAPERYER FVVPEGTKKV SYDRDTKIIN AASTVERED HTIGNIVRMQ
P87123	<i>S. pombe</i>	11	Multa	MNQPERYEL IEL-MGLPKV TYELDSKSPN AAVVTLEKED HTLANMLANQ
Q57832	<i>M. jannaschii</i>	L	Multa	M EIKILERKDN LVEIELINED HSLPNLLKDI
Q96YA7	<i>S. tokodaii</i>	L	Multa	M EIKILRSGEN YLELQIDGEE HTVGNLLKGY
P46217	<i>S. acidocaldarius</i>	L	Multa	M EIKVIKEEQN YLELQIDGEE HTIGNLLKGM
Q980K0	<i>S. solfataricus</i>	L	Multa	M EIRILKSESN YLELEIEGED HTLGNLIAGT
Q9V2M4	<i>P. abyssi</i>	L	Multa	M KIEVIKKEQD LLEFYLEGED HTFANLLTET
P57671	<i>P. horikoshii</i>	L	Multa	M KIEVIKKEED MLEFYLEGED HTFANLLTEV
O30032	<i>A. fulgidus</i>	L	Multa	M EVKIIDIIGED YVRLVIKGEH HTYLNLLQHY
O27372	<i>M. thermoautotrophicum</i>	L	Multa	MEVILDKNR EMEIVFEGET HTLCNVLRSI

*Method:
 BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
 PSIPRED and PHD = Structure Prediction
 Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.90 Sequence alignment for *P. furiosus* RNAP subunit L (sheet 1)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	-----40- -----50 --,---,-- ---60----- ---70-----
Q8U4N1	P. furiosus	L	RMSD	LRNPHVKFT AYTIEHPITM AR-----KP RFRVVTDGEI TPEEALAEAA
Q8U4N1	P. furiosus	L	PSIPRED	LRNPHVKFT AYTIEHPITM AR-----KP RFRVVTDGEI TPEEALAEAA
Q8U4N1	P. furiosus	L	PHD	LRNPHVKFT AYTIEHPITM AR-----KP RFRVVTDGEI TPEEALAEAA
P38902	S. cerevisiae	11	RMSD	LLNDRKVLFA AYKVEHPF-F AR-----F- KLRIQTTEGY DPKDALKNAC
P38902	S. cerevisiae	11	PSIPRED	LLNDRKVLFA AYKVEHPF-F AR-----F- KLRIQTTEGY DPKDALKNAC
P38902	S. cerevisiae	11	PHD	LLNDRKVLFA AYKVEHPF-F AR-----F- KLRIQTTEGY DPKDALKNAC
P38902	S. cerevisiae	11	MOE	LLNDRKVLFA AYKVEHPF-F AR-----F- KLRIQTTEGY DPKDALKNAC
P38902	S. cerevisiae	11	BLAST	LLNDRKVLFA AYKVEHPFFA -----RF KLRIQTTEGY DPKDALKNAC
P38902	S. cerevisiae	11	BLAST2	LLNDRKVLFA AYKVEHPFFA RF-----KL RIQT-TEG-Y DPKDALKNAC
P38902	S. cerevisiae	11	MPsrch	LLNDRKVLFA AYKVEHPF-F AR-----F KLRIQTTEGY DPKDALKNAC
P38902	S. cerevisiae	11	Multa	LLNDRKVLFA AYKVEHPFFA -----RF KLRIQTTEGY DPKDALKNAC
Q98SA6	G. theta	Y	Multa	LENHKNVLCF GYKKLHPLDN -----NI FLKIKTNSFT DPVLELEWAI
P52435	H. sapiens	11	Multa	LLKDPQVLFA GYKVPHPLEH -----KI IIRVQTTPDY SPQEAFTNAI
O08740	M. musculus	11	Multa	LLKDPQVLFA GYKVPHPLEH -----KI IIRVQTTPDY SPQEAFTNAI
Q9VJE4	D. melanogaster	11	Multa	LLKDPNVLF A GYKVPHPLEH -----KF VIRIQTADY SPQEAFTNAI
Q9XVH6	C. elegans	11	Multa	LLKDPNVLF A GYKVPHPLEH -----KI LLRIQTNNNT TPADALTTAI
Q38859	A. thaliana	11	Multa	LHRDENVLF A GYQLPHPLKY -----KI IVRIHTTSQS SPMQAYNQAI
P87123	S. pombe	11	Multa	LLSDERVLFA GYKVPHPINH -----NF ILRVQTVEDC SPKQIVVDAA
Q57832	M. jannaschii	L	Multa	LLTKEGVKMA SYSIDHPLLH PETGRYISNP KITIIITEEGT DPLEVLEKEGL
Q96YA7	S. tokodaii	L	Multa	LLKVPGVKFA SYSKPHPLI- -D-----SI ILKIMTDGSI SPKEALVKAI
P46217	S. acidocaldarius	L	Multa	LLKVPGVKFA AYSLPHPLI- -T-----SI TIKILTDGSI SAREALIKAI
Q980K0	S. solfataricus	L	Multa	LRRISGVSAF SYQPHPLS- -D-----KI IVKILTDGSI TPKDALLKAI
Q9V2M4	P. abyssi	L	Multa	LRNPHVKFV AYTIEHPITM AR-----KP RFRVVTDGKI TPEKALEEAA
P57671	P. horikoshii	L	Multa	LREDPHVKFV AYTIEHPITM AR-----KP RFRVVTDGKI TPEKALEEAA
O30032	A. fulgidus	L	Multa	LVEDEDVIVA RYNIPHPLI- -----EP EIYIKT-SGV DPLEAVKRAN
O27372	M. thermoautotrophicum	L	Multa	LMEDKVKAA AYSIDHPIVG -----EP QLYIRA-GS- -PKKSLKAAA

*Method:
 BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
 PSIPRED and PHD = Structure Prediction
 Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.91 Sequence alignment for *P. furiosus* RNAP subunit L (sheet 2)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method*	---80-----	--90-----95
Q8U4N1	P. furiosus	L	RMSD	KKIFERAKEV	LEAWEKAVKS
Q8U4N1	P. furiosus	L	PSIPRED	KKIFERAKEV	LEAWEKAVKS
Q8U4N1	P. furiosus	L	PHD	KKIFERAKEV	LEAWEKAVKS
P38902	S. cerevisiae	11	RMSD	NSIINKLGAL	KTNFETEWNL QTL
P38902	S. cerevisiae	11	PSIPRED	NSIINKLGAL	KTNFETEWNL QTL
P38902	S. cerevisiae	11	PHD	NSIINKLGAL	KTNFETEWNL QTL
P38902	S. cerevisiae	11	MOE	NSIINKLGAL	KTNFETEWNL QTL
P38902	S. cerevisiae	11	BLAST	NSIINK	
P38902	S. cerevisiae	11	BLAST2	NSIINK	
P38902	S. cerevisiae	11	MPsrch	NSIINK	
P38902	S. cerevisiae	11	Multa	NSIINKLGAL	KTNFETEWNL QTLAADDFAF
Q98SA6	G. theta	Y	Multa	KEIYVDLSYI	QTNLEFSS
P52435	H. sapiens	11	Multa	TDLISELSLL	EERFRVAIKD KQEGIE
O08740	M. musculus	11	Multa	TDLISELSLL	EERFRGAIKD KQEGIE
Q9VJE4	D. melanogaster	11	Multa	TDLLAELSFL	EERFKDAIKE KKEGGD
Q9XVH6	C. elegans	11	Multa	TDLVGELSLL	EHRIDAAIKK CTQSGDQERG YN
Q38859	A. thaliana	11	Multa	NDLDKELDYL	KNQFEAEVAK FSNQF
P87123	S. pombe	11	Multa	KSLITHLEEI	KVNFMEWEL KMISVEGVEM EFS
Q57832	M. jannaschii	L	Multa	RDIKMCDDL	LDELKEKK
Q96YA7	S. tokodaii	L	Multa	ELAEEDTNKF	IEEVKSIEKR
P46217	S. acidocaldarius	L	Multa	ELAENYANLF	IDEVKKI
Q980K0	S. solfataricus	L	Multa	ENIRGMTSHY	IDEIKGLTK
Q9V2M4	P. abyssi	L	Multa	KKIFDRAKEV	LDAWEKAIKG
P57671	P. horikoshii	L	Multa	RKIFNRAKEV	LDAWEKVVKVKS
O30032	A. fulgidus	L	Multa	EKI IAACTNL	LEQL
O27372	M. thermoautotrophicum	L	Multa	ETLRDRCDEF	RRLIESL

*Method:

BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD
 PSIPRED and PHD = Structure Prediction
 Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.92 Sequence alignment for *P. furiosus* RNAP subunit L (sheet 3)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method	----- --10----- --20-----
P60292	P. furiosus	N	RMSD	MGASPV IVPVRC FTCG KVIGDK YYEF
P60292	P. furiosus	N	PSIPRED	MGASPV IVPVRC FTCG KVIGDK YYEF
P60292	P. furiosus	N	PHD	MGASPV IVPVRC FTCG KVIGDK YYEF
P22139	S. cerevisiae	10	RMSD	M IVPVRC FSCG KVVGD KWESY
P22139	S. cerevisiae	10	PSIPRED	M IVPVRC FSCG KVVGD KWESY
P22139	S. cerevisiae	10	PHD	M IVPVRC FSCG KVVGD KWESY
P22139	S. cerevisiae	10	MOE	M IVPVRC FSCG KVVGD KWESY
P22139	S. cerevisiae	10	BLAST	M IVPVRC FSCG KVVGD KWESY
P22139	S. cerevisiae	10	BLAST2	M IVPVRC FSCG KVVGD KWESY
P22139	S. cerevisiae	10	MPsrch	M IVPVRC FSCG KVVGD KWESY
P22139	S. cerevisiae	10	Multa	M IVPVRC FSCG KVVGD KWESY
Q98S56	G. theta	10	Multa	M IIPVRC FCTCG KIIGN KWEIY
Q9GR61	C. elegans		Multa	M IIPIRC FCTCG KVIGD KWET
Q9VC49	D. melanogaster		Multa	M IIPIRC FCTCG KVIGN KWESY
P52436	H. sapiens/M. musculus		Multa	M IIPVRC FCTCG KIVGN KWEAY
Q9SYA6	A. thaliana		Multa	M IVPVRC FCTCG KVIGN KWD
Q39290	B. napus		Multa	M IIPVRC FCTCG KVIGN KWDAY
O13877	S. pombe		Multa	M IIPIRC FSCG KVIGD KWD
Q9N9Q1	L. major		Multa	M IIPVRC FCTCG NVIAD KYLLY
Q9YB47	A. pernix	N	Multa	M LPPVRC FCTCG RPLGHL WPRF
O29135	A. fulgidus	N	Multa	MDVPRNENSS DFPIRC FSCG AVIGH LYDEY
Q979K0	T. volcanium	N	Multa	M IIPVRC FSCG RVIAS DYGRY
Q9HL09	T. acidophilum	N	Multa	M IIPVRC FSCG RVIAS DYGRY
P29199	H. marismortui	N	Multa	M MFPVRC FCTCG NVVGE HWEEF
Q57649	M. jannaschii	N	Multa	M MFPIRC FSCG NVIAE VFE
O26147	M. thermoautotrophicum	N	Multa	MIPVRC LSCG KPVS AYFNEY
Q8ZYP9	P. aerophilum	N	Multa	M IIPIRC FCTCG KPLGHL YAVF
Q96YW4	S. tokodaii	N	Multa	M IIPIRC FCTCG SLIGD KWEPF
Q980Z8	S. solfataricus	N	Multa	M LIPIRC FCTCG SLIAD KWQSF
P39472	S. acidocaldarius	N	Multa	M IIPIRC FCTCG AVVAD KWEPF
P60291	P. abyssi	N	Multa	M IVPVRC FCTCG KVIGD KYYEF
O59298	P. horikoshii	N	Multa	M IIPVRC FCTCG KVIGD KYYEF

*Method:

BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD

PSIPRED and PHD = Structure Prediction

Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.93 Sequence alignment for *P. furiosus* RNAP subunit N (sheet 1)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method	--30----	-----	-----	--40-----
P60292	P. furiosus	N	RMSD	KRRVEAGE-D	-----	P	EKVLDLGLG
P60292	P. furiosus	N	PSIPRED	KRRVEAGE-D	-----	P	EKVLDLGLG
P60292	P. furiosus	N	PHD	KRRVEAGE-D	-----	P	EKVLDLGLG
P22139	S. cerevisiae	10	RMSD	LNLLQEDEL	-----	E	GTALSRLGLK
P22139	S. cerevisiae	10	PSIPRED	LNLLQEDEL	-----	E	GTALSRLGLK
P22139	S. cerevisiae	10	PHD	LNLLQEDEL	-----	E	GTALSRLGLK
P22139	S. cerevisiae	10	MOE	LNLLQEDEL	-----	E	GTALSRLGLK
P22139	S. cerevisiae	10	BLAST	LNLLQEDEL	-----	E	GTALSRLGLK
P22139	S. cerevisiae	10	BLAST2	LNLLQEDEL	-----	E	GTALSRLGLK
P22139	S. cerevisiae	10	MPsrch	LNLLQEDEL	-----	E	GTALSRLGLK
P22139	S. cerevisiae	10	Multa	LNLLQEDEL	D-----	E	GTALSRLGLK
Q98S56	G. theta	10	Multa	LRLNENY-S	-----	E	SSALDTLGLQ
Q9GR61	C. elegans		Multa	LGFLQSEY-S	-----	E	GDALDALGLR
Q9VC49	D. melanogaster		Multa	LGLLQAEY-T	-----	E	GDALDALGLK
P52436	H. sapiens/M. musculus		Multa	LGLLQAEY-T	-----	E	GDALDALGLK
Q9SYA6	A. thaliana		Multa	LELLQADY-A	-----	E	GDALDALGLV
Q39290	B. napus		Multa	LDLLQLDY-T	-----	E	GDALDALNLV
O13877	S. pombe		Multa	LTLQEDN-T	-----	E	GEALDKLGLQ
Q9N9Q1	L. major		Multa	LDLVSQGV-S	-----	E	EDAMNAFHLE
Q9YB47	A. pernix	N	Multa	RELVASGK-T	-----	P	GEALDELGV
O29135	A. fulgidus	N	Multa	KQKLSEGG-S	-----	P	KEALDELGIE
Q979K0	T. volcanium	N	Multa	LRRINEIR-S	EGREPTAEEI		EKIFDDLGE
Q9HL09	T. acidophilum	N	Multa	IKRVNEIK-A	EGRDPSPEEI		EKIFDDLGE
P29199	H. marismortui	N	Multa	KARTREAE-E	-----	PEDP	EKVLDELGE
Q57649	M. jannaschii	N	Multa	KERILKGE-N	-----	P	KDVLDDLGIK
O26147	M. thermoautotrophicum	N	Multa	QRRVADGE-D	-----	P	KDVLDDLGLK
Q8ZYP9	P. aerophilum	N	Multa	KRRVLAGE-H	-----	P	GRVLDDLGV
Q96YW4	S. tokodaii	N	Multa	ITRVSNGE-D	-----	P	GKVLDDLGVK
Q980Z8	S. solfataricus	N	Multa	ITRVNAGE-N	-----	P	GKVLDDLGVK
P39472	S. acidocaldarius	N	Multa	SNRVMGGE-D	-----	P	EKVLTELGVN
P60291	P. abyssi	N	Multa	KRRVEAGE-D	-----	P	EKVLDLGLG
O59298	P. horikoshii	N	Multa	KRRVEAGE-D	-----	P	EKVLDLGLG

*Method:

BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE = BLAST/MOE RMSD = MOE Minimized RMSD

PSIPRED and PHD = Structure Prediction

Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.94 Sequence alignment for *P. furiosus* RNAP subunit N (sheet 2)

RNAP Subunit Alignments for *Pyrococcus furiosus*

Accession #	Source	SU	Method	--50-----	--60-----	--70
P60292	P. furiosus	N	RMSD	RYCCRRMLLS	HVELIDDDIMH	YRVY
P60292	P. furiosus	N	PSIPRED	RYCCRRMLLS	HVELIDDDIMH	YRVY
P60292	P. furiosus	N	PHD	RYCCRRMLLS	HVELIDDDIMH	YRVY
P22139	S. cerevisiae	10	RMSD	RYCCRRMILT	HVDLIEKFLLR	YNP
P22139	S. cerevisiae	10	PSIPRED	RYCCRRMILT	HVDLIEKFLLR	YNP
P22139	S. cerevisiae	10	PHD	RYCCRRMILT	HVDLIEKFLLR	YNP
P22139	S. cerevisiae	10	MOE	RYCCRRMILT	HVDLIEKFLLR	YNP
P22139	S. cerevisiae	10	BLAST	RYCCRRMILT	HVDLIEKFLLR	YNP
P22139	S. cerevisiae	10	BLAST2	RYCCRRMILT	HVDLIEKFLLR	YNP
P22139	S. cerevisiae	10	MPsrch	RYCCRRMILT	HVDLIEKFLLR	YNP
P22139	S. cerevisiae	10	Multa	RYCCRRMILT	HVDLIEKFLLR	YNFLEKRD
Q98S56	G. theta	10	Multa	RYCCRRMILT	HVDLIEKLLQ	YNGYINFKNS
Q9GR61	C. elegans		Multa	RYCCRRMLLA	HVDLIEKLLN	YHPLEK
Q9VC49	D. melanogaster		Multa	RYCCRRMLLG	HVDLIEKLLN	YAPLEK
P52436	H. sapiens/M. musculus		Multa	RYCCRRMLLA	HVDLIEKLLN	YAPLEK
Q9SYA6	A. thaliana		Multa	RYCCRRMLMT	HVDLIEKLLN	YNTMEKSDPN
Q39290	B. napus		Multa	RYCCRRMLMT	HVDLIEKLLN	YNTLEKSDNS
O13877	S. pombe		Multa	RYCCRRMILT	HVDLIEKLLC	YNPLSKQKNL
Q9N9Q1	L. major		Multa	RFCCRRMLLT	HVDMTDLLLK	FNPDATNVLG
Q9YB47	A. pernix	N	Multa	RYCCRRTLFT	SVTYIEHAAK	YRVGIHVPPE
O29135	A. fulgidus	N	Multa	RYCCRRMFIT	HKSIVIKELSR	PHGAVG
Q979K0	T. volcanium	N	Multa	RYCCRRMIIS	HVDLINEIMP	FS
Q9HL09	T. acidophilum	N	Multa	RYCCRRMIVS	HVDLISEIMP	FS
P29199	H. marismortui	N	Multa	RHCCRRMLVS	HKDLVDIVSP	YQ
Q57649	M. jannaschii	N	Multa	KYCCRRMFIS	YRIGEDGREI	IDETIAHDER
O26147	M. thermoautotrophicum	N	Multa	RYCCRRMLIS	HVETW	
Q8ZYP9	P. aerophilum	N	Multa	RYCCRRTLMA	HVEWIDDVLL	YERRS
Q96YW4	S. tokodaii	N	Multa	RYCCRRMLLS	HVDIIEKVIH	YTRPI
Q980Z8	S. solfataricus	N	Multa	RYCCRRMLLS	HVDIINEVIH	YTRPI
P39472	S. acidocaldarius	N	Multa	RYCCRRMLLS	HVNIIREIIH	YTRPI
P60291	P. abyssi	N	Multa	RYCCRRMLLS	HVELIDDDIMH	YRVY
O59298	P. horikoshii	N	Multa	RYCCRRMLLS	HVELIDDDIMH	YKVY

*Method:

BLAST = NCBI BLAST BLAST2 = EBI WU BLAST2 MPsrch = EBI MPsrch Multa = Multalin MOE =

BLAST/MOE RMSD = MOE Minimized RMSD

PSIPRED and PHD = Structure Prediction

Red = Helix Green = Strand Blue = Turn Gray = No coord

Figure C.95 Sequence alignment for *P. furiosus* RNAP subunit N (sheet 3)