Structural Bioinformatics Analysis of the HSP40 and HSP70 Molecular Chaperones from Humans

A mini-thesis submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE OF

RHODES UNIVERSITY

Grahamstown • 6140 • South Africa

by

Coursework / Thesis

in

Bioinformatics and Computational Molecular Biology

In the Department of Biochemistry, Microbiology & Biotechnology

Faculty of Science

By

Samson Adebowale Adeyemi

March, 2013

ABSTRACT

HSP70 is one of the most important families of molecular chaperone that regulate the folding and transport of client proteins in an ATP dependent manner. The ATPase activity of HSP70 is stimulated through an interaction with its family of HSP40 co-chaperones. There is evidence to suggest that specific partnerships occur between the different HSP40 and HSP70 isoforms. While some of the residues involved in the interaction are known, many of the residues governing the specificity of HSP40-HSP70 partnerships are not precisely defined. It is not currently possible to predict which HSP40 and HSP70 isoforms will interact. We attempted to use bioinformatics to identify residues involved in the specificity of the interaction between the J domain from HSP40 and the ATPase domain from the HSP70 isoforms from humans. A total of 49 HSP40 and 13 HSP70 sequences from humans were retrieved and used for subsequent analyses. The HSP40 J domains and HSP70 ATPase domains were extracted using python scripts and classified according to the subcellular localization of the proteins using localization prediction programs. Motif analysis was carried out using the full length HSP40 proteins and Multiple Sequence Alignment (MSA) was performed to identify conserved residues that may contribute to the J domain – ATPase domain interactions. Phylogenetic inference of the proteins was also performed in order to study their evolutionary relationship. Homology models of the J domains and ATPase domains were generated. The corresponding models were docked using HADDOCK server in order to analyze possible putative interactions between the partner proteins using the Protein Interactions Calculator (PIC). The level of residue conservation was found to be higher in Type I and II HSP40 than in Type III J proteins. While highly conserved residues on helixes II and III could play critical roles in J domain interactions with corresponding HSP70s, conserved residues on helixes I and IV seemed to be significant in keeping the J domain in its right orientation for functional interactions with HSP70s. Our results also showed that helixes II and III formed the interaction interface for binding to HSP70 ATPase domain as well as the linker residues. Finally, data based docking procedures, such as applied in this study, could be an effective method to investigate protein-protein interactions complex of biomolecules.

DECLARATION

I, the undersigned, declare that this dissertation and the work contained herein being submitted to Rhodes University for the degree of Master of Science in Bioinformatics and Computational Biology in the Faculty of Science is my original work with the exception of the citations. I also declare that this work has not been submitted to any other university in part or entirety for the award of any degree

ADEYEMI SAMSON ADEBOWALE

SIGNATURE

08/03/2013

DATE

DEDICATION

With gratitude to God, this research work is dedicated to my grandmother: Mrs Morenike Christianah Adebisi who crafted in me the quest for academic excellence.

ACKNOWLEDGEMENT

“Seest thou a man diligent in his business? He shall stand before kings; he shall not stand before mean men”

- Proverb 22:29

“Indolence is a delightful but distressing state; we must be doing something to be happy. Action is no less necessary than thought to be instinctive tendencies of the human frame”

- *Mahatma Gandhi*

The above thoughts aptly describe the cascade mechanism of events that made this project work a reality. Although it has been a Herculean task, this research study is a product of meticulous research, painstaking planning, brainstorming sessions and punctilious editing through the synergistic effort of some of the most brilliant minds I ever come across.

My sincere and profound gratitude goes to my supervisor, Dr Adrienne. L. Edkins, and co-supervisor; Dr Taştan Bishop Özlem for their relentless guidance, corrections and support which have led to the successful completion of this study.

I cannot but say thank you to the Ntintilis, the Klaas’, the Fayemis, and brethren in discipleship across South Africa. You are highly acknowledged for all your cares, advices and mentorship during the course of my stay over here in Grahamstown. Your company, mutual relationship and lovely gestures cannot and will ever remain fresh in my memory.

Special thanks to Rhodes University, for the Henderson Bioinformatics Rhodes University Prestigious Scholarship 2012 award during the period of this study and Rhode University Bioinformatics (RUBi) unit for providing the platform for this research.

To my colleagues, the members of Rhodes University Bioinformatics Group (RUBi) and Biomedical Biotechnology Research Unit (BioBRU), for the harmonious working relationship we had together, for all our constructive arguments and intellectual disputes which have positively impacted the quality of this dissertation, my gratitude knows no bounds.

This research work is the product of a dream. To those who nurtured the dream from conception to birth: my mum, my siblings, and my discipler (Mr. Klaas Lulamile), you are highly appreciated for always being there for me. Special thanks to Harris Onywera and Aquillah Kanzi, you duo are truly brothers from another mother.

To my ever loving wife, Ooreofe Adeyemi, thank you for your tender and loving care, patience, trust, determination and encouragement without which the completion of this research would not have been possible. You ever remain dear to my heart.

Above all, to God be the Glory for the strength and grace to pull through this study

Table of Contents

ABSTRACT…………….. i

DECLARATION.. ii

DEDICATION... iii

ACKNOWLEDGEMENT.. iv

Table of Contents.. vi

List of Figures... ix

List of Tables.... xi

List of Electronic Data.. xiii

List of Abbreviations... xiv

CHAPTER ONE: Literature review.. 1

1.1Heat Shock Proteins 1

1.2 Molecular Chaperones... 1

1.3 HSP70 as a Molecular Chaperone... 2

1.4 Structure and Organization of J domains in the DNAJ/HSP40 family 3

1.4.1 Type III HSP40... 5

1.5 Mechanism of Action of HSP70-HSP40 Chaperone complex................................ 6

1.6 Specific Interaction of HSP40 and HSP70 Partnership. 8

1.7 Structure of HSP70-HSP40 complex 11

1 .8 Knowledge Gap... 12

1.9 Hypotheses 13

1.10 Aim and objectives... 13

CHAPTER TWO: Sequence Analysis of the Human HSP40s and HSP70s.......................... 15

2.1 Introduction.. 15

2.2 Methods.. 15

2.2.1 Sequence Retrieval 15

2.2.2 Sub-cellular Localization Predictions 16

2.2.3 Motif Analysis using the Full Length Proteins 16

2.2.4 Multiple Sequence Alignment of HSP40s and HSP70s 16

2.2.5 Phylogenetic Inference of HSP40 Proteins and HSP70 Isoforms 17

2.3 Results and Discussion 17

2.3.1 Overview of the HSP40 and HSP70 genes 17

2.3.2 Subcellular Localization Predictions of HSP40s and HSP70s 21

2.3.3 Motif Analysis of Full Length HSP40 Sequences 26

2.3.4 Sequence conservation of the J-domain in HSP40s 32

2.3.5 Consensus Sequence Analysis of Human HSP40 J-domain 37

2.3.6 Sequence conservation of the ATPase domain in human HSP70s 45

2.3.7 Analysis of the Phylogenetic inference of HSP40 Genes and HSP70 Complements .. 48

CHAPTER THREE: Homology Modelling of HSP40 J domain and HSP70 ATPase domain 55

3.1 Introduction 55

3.2 Methodology 56

3.2.1 Target Sequence and Template Structure Selection 56

3.2.2 Template Validation 57

3.2.3 Template-Target Sequence Alignment 59

3.2.4 Homology Model Building and Refinement 59

3.2.5 Homology Model Validation 59

3.3 Results and Discussion 60

3.3.1 Template Search, Selection and Validation 60

3.3.2 Homology Model Validation 67

3.3.3: Structural Analysis of Calculated Models of HSP40 J domains 70

CHAPTER FOUR: Protein-Protein Interactions of human HSP40-HSP70 complex 77

4.1 Introduction 77

4.2 Methodology 77

4.2.1 Generation of HSP70-HSP40 complexes 77

4.2.2 Protein Interaction Calculator (PIC) 78

4.3 Results and Discussion 79

4.3.1 Predicted J domain-ATPase domain linker complexes 79

4.3.2 Identification of Intermolecular Interface in the Predicted Complex Structure 81

4.3.3 Hydrophobic Interactions within 5Å 86

4.3.4 Hydrogen bonds 87

4.3.5 Ionic Interactions within 6Å 89

4.3.6 Cation–π Interactions within 6Å in Protein–protein Interface 91

4.3.7 Prediction of Residues Critical for J-domain:ATPase domain_linker region Interactions (HSPA8-DNAJC19) 92

CHAPTER FIVE: Conclusions and Future Prospects 98

REFERENCES 100

List of Figures

Figure 1 Structure of HSP70 (2KHO) ……………………………………….......................2

Figure 2 Domain architecture and classification of motifs found in HSP40…………..…...5

Figure 3 Canonical models showing the mechanism of action of HSP70 in protein folding involving interaction with partner HSP40………………………………………………7

Figure 4 Cartoon representation of HSP70-HSP40 complex .…………………………........11

Figure 5 Block diagram of motifs present within full lengths HSP40 sequences using MEME………………………………………………………………………….....28

Figure 6 Multiple sequence alignment of Type I HSP40 J domain……………………........32

Figure 7 Multiple sequence alignment of Type II HSP40 J domain……………………....33

Figure 8 Multiple sequence alignment of Type III HSP40 J domain……………………...34

Figure 9 Multiple sequence alignment of combined HSP40 J domain………………........35

Figure 10 Multiple sequence alignment of consensus sequences…………………….….38

Figure 11 Multiple sequence alignment of Type III HSP40 based on subcellular localization. Alignment of proteins predicted to be localized in the cytosol...39

Figure 12 Multiple sequence alignment of Type III HSP40 based on subcellular localization. Alignment of proteins predicted to be localized in the endoplasmic reticulum…………………………………………………………………………40

Figure 13 Multiple sequence alignment of Type III HSP40 based on subcellular localization. Alignment of proteins predicted to be localized in the mitochondrial…………………………………………..............................41

Figure 14 Multiple sequence alignment of Type III HSP40 based on subcellular localization. Alignment of proteins predicted to be localized in the nucleus............................42

Figure 15 Multiple sequence alignment of Type III HSP40 based on subcellular localization consensus sequences…..………………………………………………………43

Figure 16 Multiple sequence alignment of HSP70 ATPase domain _linker……………….…46

Figure 17 Molecular phylogenetic analysis by maximum likelihood method for J domain of DNAJA proteins……………………………………….................................49

Figure 18 Molecular phylogenetic analysis by maximum likelihood method for J domain of DNAJB proteins…………………………………………………………………….50
Figure 19 Molecular phylogenetic analysis by maximum likelihood method for J domain of DNAJC proteins…………………………………………..................................51
Figure 20 Molecular phylogenetic analysis by maximum likelihood method for HSP40 J domains .……………………………………………………………........................52
Figure 21 Molecular phylogenetic analysis by maximum likelihood method for HSP70 ATPase domain_Linker……………………………………………….…………….53
Figure 22 Template selection, alignment and secondary selection of targeted HSP40 J domains……………………………………………………..………………65
Figure 23 Predicted orientation of conserved polar residues in human HSP40 J domains…74
Figure 24 Predicted orientation of conserved hydrophobic residues in human HSP40 J domains…..75
Figure 25 Cation-π interactions found among exposed residues at the interface of the complex between HSPA14 and DNAJC2……………………………..…..…..91
Figure 26 HSPA8 ATPase domain_Linker: DNAJC19 J domain complex ………………..95
Figure 27 Protein-protein interactions of HSPA8-DNAJC19…………………………….96

.

List of Tables

Table 1 Defined residues that are important for J domain function aside the HPD motif…………………………………………………………………………......9

Table 2 Overview of Type I&II HSP40 genes in human...………………………….........18

Table 3 Overview of Type III HSP40 genes in human……………………………...…19

Table 4 Overview of HSP70 genes in human…………………………………….........20

Table 5 Predictions of human HSP40 (DNAJA&B) subcellular localization…...............23

Table 6 Predictions of human HSP40 (DNAJC) subcellular localization………..........24

Table 7 Predictions of human HSP70 subcellular localization…………………………...25

Table 8 Motifs analysis of full length protein sequence of DNAJA & DNAJB using MEME………………………………………………………………………29

Table 9 Motifs analysis of full length protein sequence of DNAJC using MEME…........30

Table 10 Protein targets for homology modelling based on known HSP40-HSP70 interactions………………………………………………………………........56

Table 11 Template selection for homology modelling of HSP40s using HHpred server……………………………………………………………………….60

Table 12 Template selection for homology modelling of HSP70s ATPase-linker region using HHpred server……………….………………………………………..61

Table 13 Template validation using different structure assessment programs for homology modelling of HSP40 J domains…………………………….……………………66

Table 14 Template validation using different structure assessment programs for homology modelling of HSP70 ATPase domains……………………….………………………..66

Table 15 Model validation of predicted HSP40 J domain and HSP70 ATPase domain_Linker using various quality assessment programs……………………..68

Table 16 Ramachandran plot statistical result of the predicted models from homology modelling….………………………………………………………...……69

Table 17 Conserved, charged and polar residues in human HSP40 J domains……….......73

Table 18 Conserved, hydrophobic non-polar residues in human HSP40 J domains………73

Table 19 DOPE Z scores for predicted complexes………….……………………………..79

Table 20 Statistical analysis of HSP40 predicted complex structures using HADDOCK server…………………………………………………….................................82

Table 21 Intermolecular interface residues of J domain: ATPase_Linker region complexes…………….…………………………………………………………..83

List of Electronic Data

SCRIPTS……………………………………………………………………………………1

ATPase_Linker_sword.py……………………………………………………………....i

homology.py………………………………………………………………………………ii

J_slicer.py…………………………………………………………………………….....iii

DOPE_Z_score.py………....…………………………………………………………...iv

Renumbering_all_files.py…………………………………………………………………v

APPENDICES…………………………………………………………………………………2

CHAPTER TWO

Appendix I: Sequence logo of motifs found in full length HSP40s using MEME.

CHAPTER THREE

Appendix I: Possible templates search for homology modelling of HSP40 proteins using HHpred server

Appendix II: HSP70 ATPase domains Templates search and alignments using HHpred server

Appendix III: Predicted model structures of selected human HSP40 J domains

Appendix IV: Predicted structural models of selected human HSP70 ATPase_linker regions

Appendix V: Model quality assessment using Anolea and Qmean evaluations for selected human HSP40 J domains

Appendix VI: Model quality assessment using Anolea and Qmean evaluations for selected human HSP70 ATPase-linker regions

CHAPTER FOUR

Appendix I: Predicted complex model structures of ATPase domain_linker region and J domain of HSP70 and HSP40 respectively

Appendix II: Clustering and energy scores evaluations for predicted model complexes in HADDOCK

Appendix III: Experimental structures of 2WO, 2QWP, 2QWQ and 2QWR

List of Abbreviations

acr	= Acrosome
ADP	= Adenosine Diphosphate
AIC	= Alkaike Information Criteria
ANOLEA	= Atomic Non-Local Environment Assessment
ATP	= Adenosine Triphosphate
BIC	= Bayesian Information Criteria
CPORT	= Consensus Prediction Of interface Residues in Transient complexes
C-terminal	= Carboxyl terminal
cyt	= Cytosol
DFIRE2	= Distant-scaled, Finite, Ideal-gas Reference state 2
EBI	= European Bioinformatic Institute
E.coli	= *Escherichia coli*
ED	= Endosome
ER	= Endoplasmic Reticulum
ext	= Extracellular
FCCs	= Fraction of Common Contacts
FFT	= Fast Fourier Transform
GDT-TS	= Global Distance Test-Total score
GF	= Glycine/ Phenylalanine
gol	= Golgi
HGNC	= Human Gene Numenclature Committe
HMM	= Hidden Markovs Model
HPRD	= Human Protein Reference Database
HSC	= Heat Shock Cognigate
HSP	= Heat Shock Protein
KDa	= Kilo Dalton
MAFFT	= Multiple Alignment with Fast Fourier Transform
MAST	= Motif Alignment and Search Tool
MEGA5	= Molecular Evolutionary Genetics Analyses

MEME	= Multiple EM for Motif Elicitation
Mit	= Mitochondrial
MQAP	= Model Quality Assessment Program
NBD	= Nucleotide Binding Domain
NCBI	= National Center for Biotechnology Information
NEF	= Nucleotide Exchange Factor
NMR	= Nuclear Magnetic Resonance
nuc	= Nucleus
PDB	= Protein Data Bank
per	= Perixosome
PIC	= Protein Interactions Calculator
QMEAN	= Qualitative Model Energy ANalyses
RSMD	= Root Square Mean Deviation
SBD	= Substrate Binding Domain
Sec	= Secretory
SV40	= Semian Virus 40
TPR	= Tetracopeptide Repeat
WAG	= Whelan And Goldman

Amino acid abbreviations

3-letter word	1-letter word	Meaning
ALA	A	Alanine
ARG	R	Arginine
ASP	D	Aspartic Acid
ASN	N	Asparagine
CYS	C	Cysteine
GLN	Q	Glutamine
GLU	E	Glutamic Acid
GLY	G	Glycine
HIS	H	Histidine
ILE	I	Isoleucine
LEU	L	Leucine
LYS	K	Lysine
MET	M	Methionine
PHE	F	Phenylalanine
PRO	P	Proline
SER	S	Serine
THR	T	Threonine
TRP	W	Tryptophan
TYR	Y	Tyrosine
VAL	V	Valine

CHAPTER ONE: Literature review

1.1Heat Shock Proteins

Heat Shock Protein (HSP) is the collective name given to a group of ubiquitous and highly conserved proteins having essential roles in physiological and stressful cellular environments (Feder and Hofmann, 1999). While some HSPs are induced by stress, several others are constitutively expressed. The classification and nomenclature of HSPs into various groups is based on sequence homology and typical molecular weights. For instance, HSP70 isoforms have approximately a molecular weight of 70kDa, while HSP40 isoforms are assumed to be approximately 40kDa in size (Sterrenberg *et al*., 2011).

The ability of HSPs to act as molecular chaperones is integral to their protein folding and protective roles in the cell. HSPs rarely function alone; instead the interactions between different HSP classes occur to modulate distinct chaperone functions (Ohtsuka and Suzuki, 2000; Feder and Hofmann, 1999). However, not all molecular chaperones are heat shock proteins.

1.2 Molecular Chaperones

Molecular chaperones are proteins that coordinate protein homeostasis throughout their life span. They help in regulating the conformation of nascent proteins either in their native cellular environment or under inducible conditions (Sterrenberg *et al*., 2011; Kampinga and Craig, 2010). While the information needed for the native conformation of a polypeptide is contained in its primary amino acid sequence and enables for its protein folding *in vitro*, the situation is different *in vivo* (Lee and Tsai, 2005). Molecular chaperones are involved in diverse key cellular functions under both physiological and stressful conditions, including the prevention of protein aggregation, facilitating the folding of nascent and damaged proteins, aiding the transport of previously synthesized proteins across membranes, identification of targeted proteins for degradation and enhancing protein-protein interactions by guiding their conformational changes (Kampinga and Craig, 2010). Because the rate at which proteins aggregate increases when they are denatured under stressful condition, some molecular chaperones are classified as heat shock proteins due to their ability to prevent the aggregation of newly synthesized polypeptides and assemble subunits into nonfunctional structures under stressful condition.

Molecular chaperones usually undergo a continuous repeat of client binding and release cycles until the client has acquired its final active conformation or has found its way into the proteolytic system (Sterrenberg *et al*., 2011; Kampinga and Craig, 2010).

Most often, molecular chaperones do not perform their function as individual proteins. They function in partnership with other chaperones and co-chaperones in a multi-protein complex as well as direct interactions with client protein substrates (Freeman and Morimoto, 1996).

1.3 HSP70 as a Molecular Chaperone

One of the most important families of molecular chaperones is the HSP70 family. HSP70s are made up of a highly conserved 44 kDa ATPase domain and a 15 kDa peptide-binding domain. it should be noted that the ATPase domain is also referered to as the nucleotide binding domain while the peptide binding domain is also known as the substrate binding domain and are used interchangeably in this thesis. They are also characterized by the presence of a 10 kDa C-terminal region (Suh *et al*., 1998). HSP70s as molecular chaperones, have been implicated in a wide range of folding processes spanning from the folding and assembly of newly synthesized proteins, protein translocation, prevention of protein denaturation and misfolding during cellular stress, proteins degradation as well as the control of the activity of regulatory proteins (Mayer and Bukau, 2005; Suh, Lu, and Gross, 1999). Figure 1 shows the structure of the ATPase domain linked to the substrate (peptide) binding domain by a hydrophobic linker region proposed to be necessary for communication between these two domains of HSP70 (Jiang *et al*., 2007; 2005; 2003).

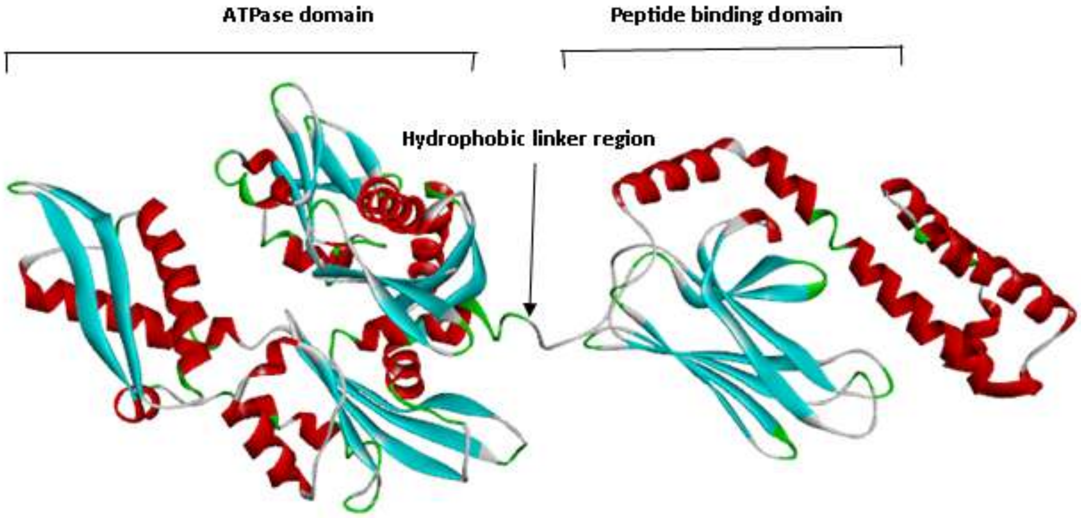


Figure 1: Structure of HSP70 (PDB ID: 2KHO). The ATPase binding domain and the peptide-binding domain of HSP70 protein (2KHO) linked together by short hydrophobic residues; thought to contribute to the ATPase activity of HSP70s and the secondary structures are displayed as residue hydrophobisity. The solid red cylinders represent the alpha helixes while those coloured in blue stand for the beta strands. Figure was generated using Discovery studio 3.5 visualizer.

1.4 Structure and Organization of J domains in the DNAJ/HSP40 family

HSP40 family of proteins has been described as the largest and most diverse family of co-chaperones. HSP40s stimulate the ATPase activity of HSP70 and also serve as substrate scanners for HSP70 (Jen-Sing *et al*., 1998). In humans, there are currently 49 genes coding for members of HSP40 family (Ohtsuka and Hata, 2000). These are grouped based on presence or absence of conserved domains with similarity to the canonical *Escherichia coli* HSP40, DnaJ (Cheetham, 1998)(Figure 2A). These categories include Type I (DNAJA, 4 members), Type II (DNAJB, 13 members) and Type III (DNAJC, 32 members) (Sterrenberg *et al*., 2011). Type I HSP40s contain four primary domains: an N-terminal J-domain, a glycine/phenylalanine (GF)-rich region, Cysteine repeat region (a zinc finger domain) and a C-terminal domain. Type II HSP40s are made up of an N-terminal J-domain, a GF-rich region and a C-terminal domain. Type III HSP40s possess only the J-domain which can be located at any of the position within the protein. Many of the Type III HSP40s differ in molecular size, sequence and structural architecture and possess specialized domains whose functions are usually different from that of Type I and II DNAJ. The G/F rich domain has been proposed to be in contact with HSP70 and contribute to the stability of the HSP70-client complex during HSP70-HSP40 partnership (Kampinga and Craig, 2010; Cheetham, 1998; Pellecchia *et al*., 1996). Type IV HSP40 are a distinct subtype that do not possess the HPD motif in the J domain. Type IV HSP40 have predominantly been described in plasmodium species and are rare in humans (Botha *el al.*, 2007). Human DNAJB13 could be considered a Type IV member in human due to the replacement of the aspartic acid residue in the HPD motif with a leucine residue (Jikui Guan and Li Yuan, 2008).

The J-domain is a specific feature that defines a protein as a member of the HSP40 family (DNAJ) (Suh *et al*., 1998; Cyr *et al*., 1994). The J-domain is believed to play important role for the interaction and stimulation of HSP70. However, the exact mechanism by which J-domain stimulates the ATPase activity of HSP70 for conformational changes resulting in the stabilization of HSP70-client protein interaction remains poorly characterised. Nuclear Magnetic Resonance (NMR) structures of J domain reveal the presence of four α-helices (I – IV) and a loop region between helices II and III which contain the highly conserved tripeptide histidine-proline-aspartic acid (HPD) motif that is essential for the stimulation of ATPase activity of HSP70 (Figure 2B). Mutations in this motif terminates the stimulation of HSP70 ATPase activity (Suh *et al.*, 1998). The majority of the substitutions performed on the J-domain have involved mutations in the HPD motif (Caplan *et al*., 1998). However, substitution in the other sections of the J-domain have been investigated (Hennessy *et al*., 2005). Other residues have also been identified to be important for J-domain function. These are grouped into two categories; charged residues/motifs and hydrophobic residues. Hennessy *et al*. (2000) reported that the highly conserved charged residues/motifs could be responsible for J domain function, while the conserved hydrophobic residues are likely to be mainly critical for maintaining the structural integrity of the J domain

Helices II and III are structurally conserved in all known J-domain and helix II in particularly is thought to contain positively charged residues that interact with the negatively charged residues at the undercleft of the ATPase domain of HSP70 thereby enhancing ATP hydrolysis (Sterrenberg *et al*., 2011; Hennessy *et al*., 2005a; Suh *et al*., 1999). A previous study proposed that the residues of helix IV are not essential to the co-chaperone function of DnaJ (Genevaux *et al*., 2002). However, other studies of (Garimella *et al*., 2006; Hennessy *et al*., 2005) suggested that helix IV may contribute to the specificity of J-domains as a secondary site of contact for their partnership with HSP70s.

1.4.1 Type III HSP40

Type III HSP40 proteins only have the J-domain in common with *E. coli* DnaJ which may be located at any position within the sequence of the proteins. Type III HSP40 members are the most diversified sub-type of the HSP40 and contain proteins with additional distinct motifs or domains, such as trans-membrane helices (*E. coli* DjlA, yeast Sec63, human DjC9/hSec63, yeast Mdj2), tetratricopeptide repeat domains (TPRs; mouse DjC2/Zrf1/Mida1, human DjC3/hp58 and DjC7/hTpr2), and cysteine-rich regions which are polypalmitoylated (cysteine string proteins). Some HSP40s have been reported to have a wider substrate specificity, such as *E. coli* DnaJ and yeast Ydj-I, while others have more restricted substrate binding spectrum especially among the Type III proteins (Kampinga and Craig, 2010). Previous studies have shown that the Type III class of the DnaJ proteins bind to a restricted number of substrates or may sometimes not bind to substrates directly but are positioned very closely to substrates and recruits substrates to partner HSP70 protein. In human, there are 32 Type III HSP40 genes localized at various positions within the cell (Hageman *et al*., 2011; Kampinga and Craig, 2010; Mayer and Bukau, 2005).

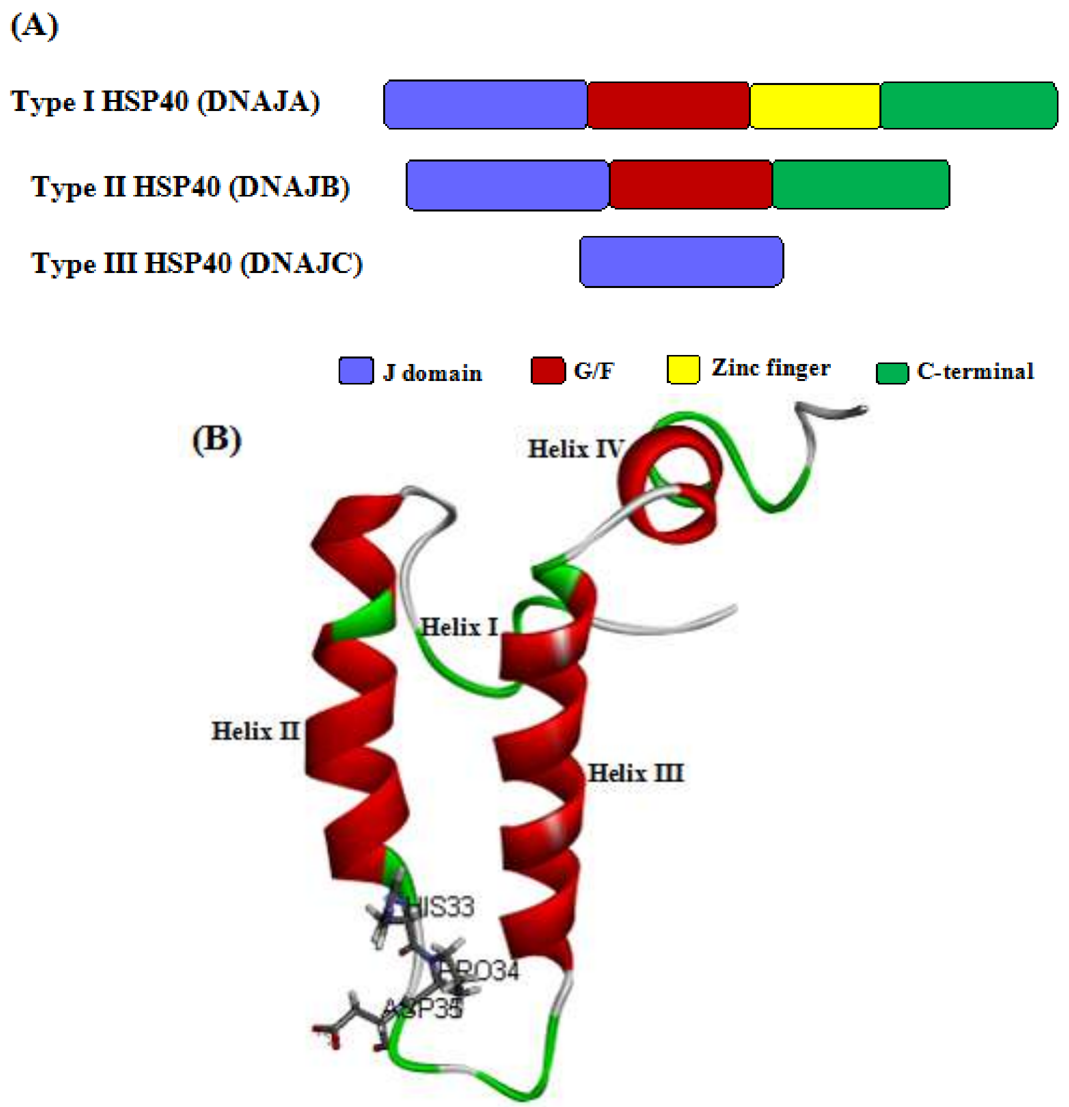


Figure 2: Domain architecture and classification of motifs found in HSP40 (A) Functional domains present in HSP40. Classification is based on the presence or absence of the four domains: J-domain, glycine/phenylalanine rich region (G/F), the cysteine repeats (Zinc finger motif) and the largely uncharacterized C-terminal region (Cheetham, 1998) (B) The three dimensional J-domain structure (*E. coli* J-domain; PDB ID: 1XBL) that is currently used to define the HSP40 family. The conserved HPD motif is shown in stick format and labeled. The four helices are labeled accordingly. The figures were generated using Discovery studio and Microsoft publisher 2010. Adapted from (Sterrenberg *et al*., 2011; Hennessy *et al*., 2005a).

1.5 Mechanism of Action of HSP70-HSP40 Chaperone complex

HSP70 interacts with hydrophobic peptide regions of a client protein in an ATP-dependent process. The molecular chaperone activity of HSP70 requires partnership with HSP40 co-chaperones and the nucleotide exchange factors (NEF) (Kampinga and Craig, 2010; Mayer and Bukau, 2005). The affinity of HSP70 for client substrate is modulated by ATP binding and hydrolysis. The mechanism of action of the polypeptide binding and release of HSP70 is coupled to the ATPase cycle which consists of a switch between the ATP bound state and the ADP bound state (Figure 3). The ATP bound state has low affinity for substrates with a high substrate exchange rate while the ADP bound state has high affinity for substrates binding with stability than the ATP bound state (Suh *et al.*, 1999). The hydrolysis of ATP to ADP enhances the binding of the HSP70 to client protein, thereby facilitating the formation of a stable HSP70-client complex. NEFs catalyse the dissociation and release of the folded polypeptide, as well as increase the rate at which nucleotide is exchanged (Hennessy *et al.,* 2005*)*. NEFs have higher affinity for HSP70-ADP than HSP70-ATP, then bind to the HSP70-client complex and reverses the conformational shift, thus allowing for the dissociation of ADP and the release of the client polypeptide (Mapa *et al*., 2010). If the client polypeptide has not attained its native folding state on release, the J protein rebinds to its exposed hydrophobic regions and the cycle continues. Evidence from mutagenesis experiments has shown that the lower cleft of the N-terminal ATPase domain is a binding pocket for the J-domain-NBD interactions (Jiang *et al*., 2007; Hennessy *et al*., 2005; Suh *et al*., 1998, 1999).

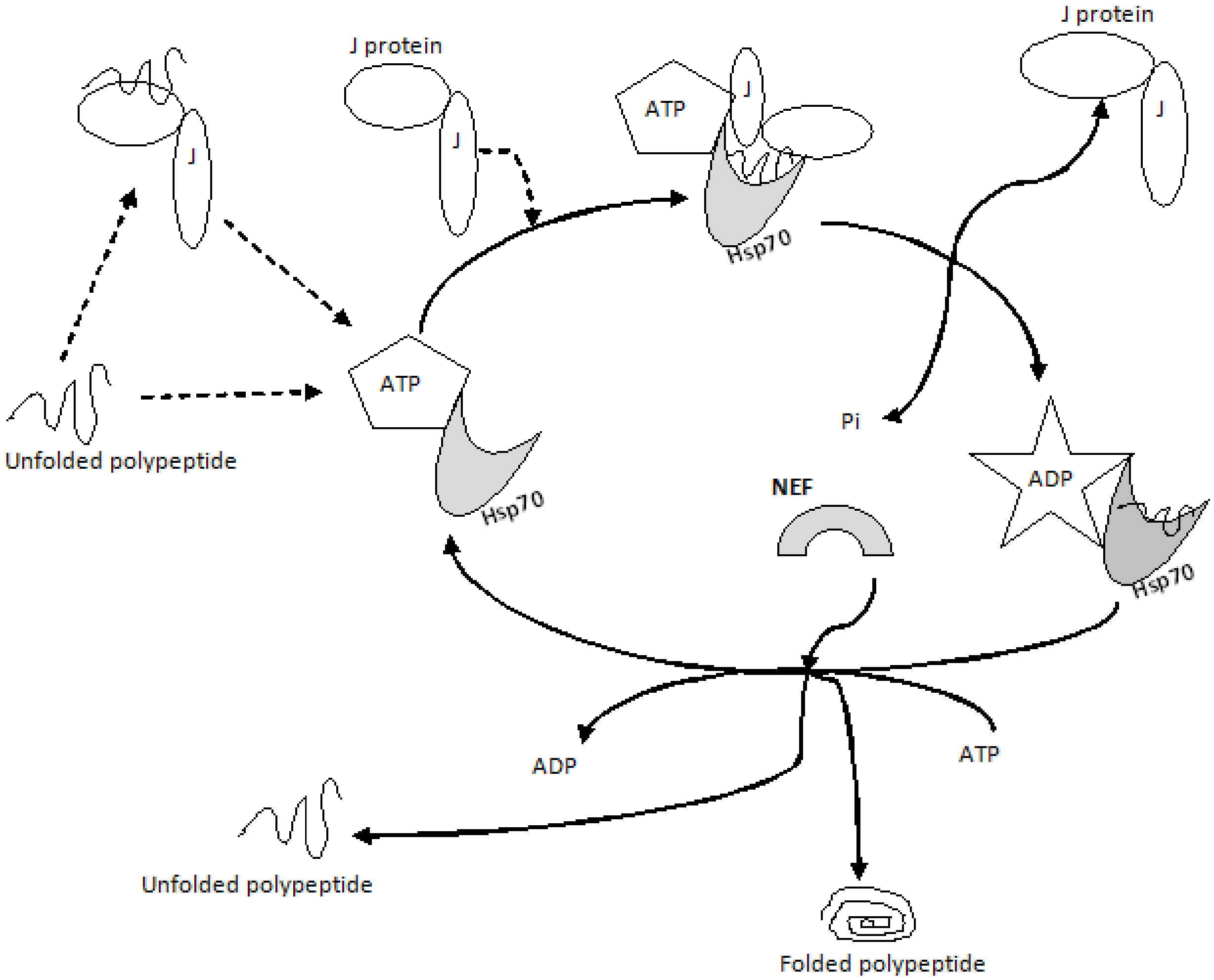


Figure 3: Canonical model showing the mechanism of action of HSP70 in protein folding involving the interaction with partner HSP40. The dotted lines show the different ways by which a client polypeptide and HSP40 protein can enter the cycle. A client protein can either be directly recognized by HSP70 protein, followed by the coming in of an HSP40 protein into the cycle or the client protein binds to the HSP40 protein and is subsequently presented to HSP70. ATP hydrolysis as stimulated by the J protein causes a conformational change in the peptide-binding domain of HSP70 protein, locking the client polypeptide within its cleft with a subsequent release of the J protein and an inorganic phosphate (Pi) from the complex. Nucleotide exchange factor (NEF), which has a high affinity for HSP70-ADP than HSP70-ATP, binds to the HSP70-client complex and reverses the conformational shift, thus allowing for the dissociation of ADP and the release of the client polypeptide. If the client polypeptide has not attained its native folding state on release, the J protein rebinds to its exposed hydrophobic regions and the cycle continues. Adapted from (Hageman *et al.*, 2011; Hennessy *et al.*, 2005; Suh *et al.*, 1999).

1.6 Specific Interaction of HSP40 and HSP70 Partnership.

There is evidence to suggest multiple binding sites among HSP70 and HSP40 proteins (Suh *et al.*, 1999; 1998). Evidence from J-domain swapping experiments has shown that specific partnership between HSP40-HSP70 interactions exist between co-localized HSP40 and HSP70 members as opposed to those localized in different subcellular locations within the cell. For instance, J-domains from endoplasmic reticulum (ER) localized HSP40 were able to bind and stimulate ER-localised HSP70 from different species (Sterrenberg *et al*., 2011; Nicoll *et al*., 2007; Hennessy *et al*., 2005; Schlenstedt *et al*., 1995;). *E coli* DnaJ was able to stimulate the ATPase activity of mammalian HSC70. However, Hdj1 in mammal cannot stimulate the ATPase activity of DnaK (Minami *el al*., 1996). In another experiment, the J-domain from yeast mitochondrial HSP40 protein Mdj1 (Type 1) could be interchanged with the J-domain of *E. coli* DnaJ (Hennessy *et al*., 2005). Genevaux *et al*., (2001) showed that the J domain from the Type I HSP40 protein Dj1A could effectively be substituted for the Type I *E. coli* DnaJ J domain and both can interact with the same HSP70 (DnaK). However, the J domain from another isoform of membrane-bound Type III J *E. coli* HSP40 protein (Dj1C) could not be interchanged with the cytosolic Type I *E. coli* J domain (DnaJ) *in vivo*. This suggested that Dj1C could not interact with DnaK but rather with HSC70. In turn, *E. coli* DnaJ could not interact with HSC70, but could interact with DnaK (Kluck *et al*., 2002; Minami *et al*., 1996).

The result of an *in vivo* complimentary assay by Nicoll *et al*. (2007) showed that ERj1, a membrane-bound Type III HSP40, was unable to substitute for the J-domain of Agt (a prokaryotic Type I HSP40). The degree to which a J-domain can be interchanged between different subcellular organelles from HSP40 proteins may depend at least in part on the kind of cellular processes in which the HSP40 proteins are involved and therefore the types of HSP70 isoforms involved (Nicoll *et al*., 2007; Genevaux *et al*., 2001). Certain HSP40 members will only bind to specific client substrates and present them to HSP70. However, some Type III HSP40s are thought not to interact with chaperone client proteins but rather use the J-domain to recruit HSP70 to a specific subcellular location for a discrete function. These HSP40s have been proposed to often consist of the J-domain in conjunction with other multiple non-classical HSP40 functional domains such as the trans-membrane domains. The diverse arrangement of the J-domain in Type III HSP40s and the presence of these non-classical domains indicate that the majority of Type III HSP40s may have defined functions in addition to HSP70 stimulation (Hageman *et al*., 2011; Sterrenberg *et al*., 2011).

While some HSP70s interact with specific HSP40, there are others that interact with more than one HSP40 protein (Jiang *et al*., 2007). However, the basis of such specification still remain unclear though Hennessey *et al* (2005b) showed that possible sequence variations within the J domain of HSP40 could be responsible for this partnership. The transient ATP dependent cycle of reaction processes between HSP70-HSP40 partnerships has made it difficult for the structural basis of these interactions to be studied experimentally. Studies have shown that the J-domain alone is not sufficient to stimulate the ATPase activity of DnaK but that J-domain stimulation is restored by the addition of a DnaK substrate peptide (Suh *et al.*, 1999; 1998). Substitution experiments carried out on the J-domain through mutations in the HPD motifs as well as investigation in the other parts of the J-domain have revealed some of the residues/motifs that could be responsible for specific HSP40-HSP70 interactions (Table 1) (Nicoll *et al.*, 2007; Hennessy *et al.*, 2000*;* Suh *et al.*, 1999). Thus, there may be features present within the J domain, especially with the Type III proteins, that mediates the specificity of binding between HSP70s and partner HSP40s. Meanwhile, there tend to be more HSP40s than HSP70 in cells and HSP40s seem to confer functional specilisation to HSP70s.

Table 1: Residues thought to be important for J domain function aside from the HPD motif

Amino acid residues	**Organism**	**References**
TYR 25, LYS 26, ARG 36, ASN 37, PHE 47	*Escherichia coli* DnaJ	(Genevaux *et al*., 2002)
TYR 7, LEU 10, ARG 26, LEU 57, ASP 59, ARG 63	*Agrobacterium tumefaciens*	(Hennessy *et al*.,2005b)
TYR 7, LEU 10, TYR 25, LYS 26, PHE 47, LYS 48, LEU 57, ASP 59, ARG 63	*Plasmodium falciparum*, trypanosomal, *homo sapiens* and murine	(Nicoll *et al*., 2007)
ASP 876, ASP 896, PHE 891, HIS 874, PRO 875, ARG 876 (CYS 876), MET 829, MET 889, THR 879, GLU 884	*Homo sapiens* DNAJC6 (auxilin)	(Jiang *et al*., 2003, 2007)
LYS 62 and ARG 63 in the QKRAA motif on helix IV	*Escherichia coli* DnaJ	(Suh *et al*., 1999; Auger and Roudier, 1997)

Although the regions of HSP70 that interact with the J domain are not yet fully elucidated, recent evidence from genetic and biochemical experiments have suggested that there is more than one site involved in HSP70-HSP40 interactions; namely, the lower cleft of the ATPase domain and at a point very close to the substrate binding site (Jiang *et al*., 2007; Suh *et al*., 1999). DnaJ interacts with the ARG 167 amino acid residue at the underside cleft of the ATPase domain and D206 residue which is part of the ATP binding site in *E. coli* DnaK. The conserved EEVD motif at the C-terminal region of the human HSP70 homolog has been shown to inhibit the ability of HSP40 (Hdj1) to catalyze ATP hydrolysis (Suh *et al*., 1999; Cheetham, 1998). Residues found in bovine HSC70 (HSPA8) involved in interactions with auxilin HSP40 J domain (DNAJC6) are; **LEU 170, LEU 380, LEU 393, ILE 179, ILE 181, ILE 216, VAL 388, ARG 171, ASP 152, GLU 175, SER 385, ASN 174, THR 177, TYR 371** (Jiang *et al*.,2007).

1.7 Structure of HSP70-HSP40 complex

Structures of the J domain from HSP40 and HSP40-like proteins have been experimentally determined. These includes *E.coli* DnaJ (Pellecchia *et al*.,1996), human Hdj1, *E.coli* HSC 20 (Cupp-vickery and Vickery, 2000), the large T antigen form murine polymavirus (Berjanskii *et al*., 2000), the large T antigen from SV40 in conjunction with the retinoblastoma tumor suppressor (Kim *et al*., 2001), and bovine auxilin (Jiang *et al*., 2003). Until now, a single crystal structure complex, that of the J domain of auxilin (DNAJC6) and Nucleotide Binding Domain (NBD) of bovine HSC70 has been described (Figure 4) (Jiang *et al*., 2007). While the J domain of HSP40 is the first primary contact that stimulates the ATPase activity of partner HSP70s, Jiang *et al*. (2007) argued that the NBD and Substrate Binding Domain (SBD) as well as the hydrophobic linker region between the former and the latter are responsible for its ATPase activity. The author observed that interaction of the J domain with the NBD alone could not stimulate ATPase activity of HSC70 whereas interaction of J domain with the NBD-linker region did (Figure 4). Previous report of Suh *et al.* (1999) also confirmed that the J domain alone neither stimulated ATP hydrolysis nor bound to DnaK in an *in vitro* study. Thus, this observation suggested that the linker region connecting the NBD and the SBD plays an important role in the stimulation of the ATPase activity of HSP70 proteins by the partner HSP40 J domain. The NBD serves as the primary recipient of the J domain signal which results in a transient conformational change in the linker region, this then causes a shift in the SBD to allow for the capture of the polypeptide substrate (Figure 4) (Jiang *et al*., 2007). Contrary to the report of Swain *et al.* (2007) that the SBD only interacts with the NBD in the ATP-bound state, Jiang *et al.* (2007) reported that the NBD of HSC70 and its SBD interact in the ADP-bound state of the chaperone. The latter author further argued that both the J domain and the nucleotide exchange factors are responsible for modulating the NBD-SBD and NBD-linker interactions of HSC70 protein to regulate its ATPase activity which may vary in different HSP70s.

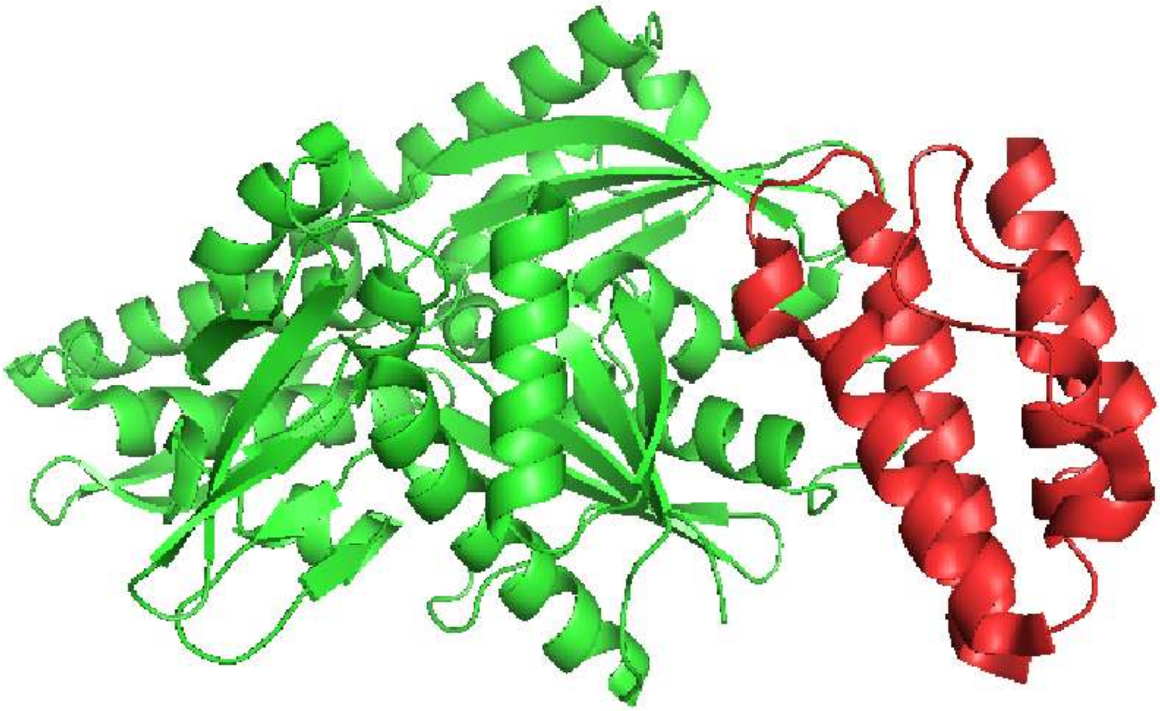


Figure 4: Cartoon representation of HSP70-HSP40 complex (PDB ID: 2QWO). NBD_linker domain is colored green while the J domain is colored red. Interaction interface residues between the complex structure of NBD_Linker region of bovine HSC70 protein in the ADP-bound state and the J domain of auxilin HSP40 protein are listed in Table 1 and section 1.6 above. (Kampinga and Craig, 2010; Jiang *et al*., 2007).

1 .8 Knowledge Gap

The mechanism by which the J-domain stimulates the ATPase activity of HSP70 for conformational changes resulting in the stabilization of HSP70-client protein interaction remains unclear. Despite the key role of HSP40 in the regulation of HSP70 functions, little is known about the molecular determinants that mediate binding of the HSP40 J domain to the HSP70 NBD (Suh *et al*., 1998). The interaction surface amino acid residues between these domains in HSP40-HSP70 partnerships are not precisely defined. The type and number of motifs contained in an HSP40 protein will govern its function. However, while there is information available on certain critical motifs required for HSP70-HSP40 function (e.g HPD motif), there is little known about the motifs that dictate the specificity of the different partnerships between HSP40 and HSP70. It is not currently possible to predict which HSP40 and HSP70 isoforms will interact, without testing the individual interactions experimentally (Cyr *et al*., 1994; Kampinga and Craig, 2010; Sterrenberg *et al*., 2011). In humans, there are 13 different HSP70 and 49 different HSP40 genes of which 32 are Type III HSP40. Thus, it is important to discriminate between general binding determinants that are important in the majority of HSP40-HSP70 partnerships, and specificity determinants, which are important in specific HSP40-HSP70 relationships (Hennessy *et al*., 2005). Protein functions and interactions are best studied when their structures are determined. However, there is only one available structure of the complex between HSP40-HSP70 proteins (Jiang, *et al*., 2007). HSP70 has been described as an emerging chaperone drug target in cancer and the HSP40 function has been linked to a number of human diseases including cancer, malaria, neurodegenerative diseases and viral infection. The HSP70-HSP40 interaction is regarded as a potential target for anti-cancer drug development (Sterrenberg *et al*., 2011; Kampinga and Craig, 2010).

1.9 Hypotheses

There is evidence to suggest specific partnerships between the different HSP40 and HSP70 isoforms; and that not all J domains will be able to interact with every HSP70. We therefore hypothesized that:

- Specific isoforms of human HSP40 will interact in partnership with specific isoforms of HSP70 for its ATPase activity.
- The specificity of interactions between Type III HSP40 and HSP70 will be determined by the J-domain of HSP40 isoforms using the biochemical data from literature.

1.10 Aim and objectives

This project aimed to identify interaction surfaces that may govern the specificity of the partnership between the J domains from Type III HSP40 and the N-terminal nucleotide binding (ATPase) domain from the HSP70 isoforms from humans. The following objectives were defined:

- Sequence analysis and generation of structural models for J domains and ATPase domains from the different HSP40 types and HSP70 isoforms.
- Define residues or motifs that could be used to predict unknown partnerships between HSP40 and HSP70 isoforms in humans.

CHAPTER TWO: Sequence Analysis of the Human HSP40s and HSP70s

2.1 Introduction

HSP70s are molecular chaperones with special functions in protein folding and aggregation of non-native proteins among other cellular processes in which they are involved. HSP40 proteins help in stimulating ATP hydrolysis of HSP70 proteins for its ATPase activity, thereby increasing its affinity for binding to client polypeptides (Kampinga and Craig, 2010). The J domain houses the tripeptide HPD motif located in the loop region between helices II and III of the four helices present in the NMR structure of *E. coli* J domain (Pellecchia *et al*.,1996). Mutation of the amino acid residues of this motif truncated the stimulation of HSP70 ATPase domain by partner HSP40s. This chapter sought to gain more insight to determining possible interacting partners and residues that might be responsible for HSP40-HSP70 specific interactions through multiple sequence alignment of the two proteins, identification of motifs within the HSP40s, as well as analyses of the level of conservation and relatedness between the different HSP40 J domains.

2.2 Methods

2.2.1 Sequence Retrieval

The 49 human HSP40 and 13 HSP70 genes were retrieved from the National Centre for Biotechnological Information (NCBI) databank and recent publications (Hageman and Kampinga, 2009). The standard accession number (gene symbol) as established by the Human Gene Nomenclature Committee (HGNC) of the proteins was presented in Table 2 – 4. These proteins were accessed for their family signatures from the Human Protein Reference Database (HPRD) (Keshava *et al*., 2009) and HGNC database. A BLAST search was performed in HPRD in order to identify isoforms of the proteins. Gene symbol, the old or alternative names, protein molecular weight, gene map locus, sequence length, positions of the J domain as well as the experimental localization predictions for each of the protein families were determined from HPRD. Python scripts: *J_slicer.py and ATPase_Linker_sword.py* (see electronic data/SCRIPTS) were written, and they were used to extract the J domain and the ATPase domain from the HSP40s and HSP70s respectively. Up to 83 amino acid residues were obtained from each HSP40 sequence to represent its J domain using the HPD motif as the anchor residues. The first 395 amino acid residues, representing the ATPase domain linker region of each HSP70s were extracted. This allowed for an efficient alignment of the HSP40 proteins since the sequence length varied from each other.

2.2.2 Sub-cellular Localization Predictions

Predictions of the sub-cellular localization of the HSP40 and HSP70 genes were performed using online localization prediction methods. PSORT II (Horton *et al*., 2007), pTarget (Guda, 2006), CELLO (Yu *et al.*, 2006), Multiloc1 (Höglund *et al*., 2006) and Multiloc2 (Blumer *et al*., 2009) were employed. A consensus localization result was selected for each HSP40 and HSP70 proteins following a previously published method (Hageman and Kampinga, 2009). Localization prediction that was consistent among the prediction programs was selected as the consensus result. In cases where the experimental localization of the proteins have been determined, the experimental results were selected.

2.2.3 Motif Analysis using the Full Length Proteins

Multiple EM for Motif Elicitation (MEME) suite was employed for scanning and identification of possible motifs present within the full length protein sequences of HSP40s (Bailey *et al*., 2006; 1998). The distribution of motif occurrences was set as any number of repetitions, the number of different motifs to find was set at 20, minimum motif width was set at 6 and the maximum motif width was set to 50.

2.2.4 Multiple Sequence Alignment of HSP40s and HSP70s

Clustal Omega (Sievers *et al*., 2011), PROMALS3D (Pei *et al*., 2008) and Multiple Alignment with Fast Fourier Transform (MAFFT) (Katoh and Frith, 2012) were employed for the alignment analysis. Clustal Omega and MAFFT alignment tools from the European Bioinformatics Institute (EBI) server were used. Clustal Omega alignment is based on the Hidden Markov's Model (HMM) while MAFFT employs the Fast Fourier Transform (FFT) method for multiple sequence alignment. Promals3D identifies homology with known 3-D structures for the input sequences for multiple sequence alignment. Alignment results from the three selected methods were compared and the most suitable outcome was used in this study. Two iterations were performed for both Clustal Omega and MAFFT alignment and other parameters were set as default; scoring matrix (BLOSSUM 62), gap penalty (1.53), gap extension penalty (0.123). In each of the alignment experiments, the results were obtained as Clustal output. The aligned sequences were visualized in JALVIEW (Waterhouse *et al*., 2009).

2.2.5 Phylogenetic Inference of HSP40 Proteins and HSP70 Isoforms

Phylogenetic analysis of the HSP40 and HSP70 proteins were performed using Molecular Evolution Genetic Analysis (MEGA5) tool following the method described by (Tamura *et al*., 2011). Substitution model of evolution for the multiple sequence alignment dataset was calculated and the best statistically fit model (usually with the lowest Bayesian information criterion (BIC) and correct (lowest) Akaike information criterion (AIC) values) was chosen for the phylogeny inference for each analysis. Maximum Likelihood method was employed to infer the evolutionary relationship of the proteins. 1000 bootstrap replicates were set to assess the statistical support of the inferred tree to the dataset.

2.3 Results and Discussion

2.3.1 Overview of the HSP40 and HSP70 genes

An overview of features of the HSP40 and HSP70 genes are presented in Table 2 - 4. Human chaperones are generally classified into various families based on their molecular weight. While there are other domains present within both HSP40s and HSP70s, the J domain and the ATPase domain remain the main signature that defines their identity and classifications. Variations occur in molecular weights especially within the J proteins. For instance, while some of the HSP40 genes have molecular weight around 40000 Dalton (40 KDa), the weight of the proteins ranges between 504.6 KD in DNAJC29 to 12.5 KD in DNAJC19 with noticeable variations in their sequence lengths. In all, 4 Type I HSP40s were retrieved, 16 Type II members, 30 Type III members and 13 HSP70s. The J domain position in most of the Type I and II HSP40 was located at the N-terminal region (Table 2). The position of the J domain varied in the Type III members. DNAJC3, DNAJC6, DNAJC22, DNAJC27 and DNAJC29 (Table 3) have the J domain located at C-terminal region while both DNAJC13 and DNAJC14 have the J domain located in the middle of the protein. While majority of the J proteins are located on different positions on the q locus on the chromosome, some of the proteins are on different positions on the p locus on the chromosome including; DNAJA1, DNAJA3, DNAJB1, DNAJB4, DNAJB, DNAJC1, DNAJC6, DNAJC8, DNAJC16, DNAJC21, DNAJC23, DNAJC26 and DNAJC27. It remained to be investigated wheather those proteins on the same locus on the chromosome share similar functional characteristics and cellular localizations.

Some of the HSP70-HSP40 interacting partners are shown in Table 4, including HSPA1A with DNAJA3, DNAJB11 and DNAJC3, HSPA1B with DNAJA1, HSPA5 with DNAJC1 & DNAJC10, HSPA8 with DNAJA3, DNAJC2 and DNAJC6, HSPA14 with DNAJC2. HSPA2, HSPA5, HSPA6, HSPA7, HSPA9 and HSPA12A were located on the q locus on the chromosome while HSPA1A, HSPA1B and HSPA1L, HSPA12B and HSPA14 genes were found on the p locus of the chromosome (Table 4). Interestingly, HSPA12A which was an isoform of HSPA12B was found located on the q locus as opposed to HSPA12B. Previous studies have shown that both HSPA6 and HSPA7 genes were only present in humans (Hageman and Kampinga, 2009) and HSPA7 gene contains a frameshift and therefore might be a pseudogene. A full length gene without the frameshift has been shown to be an homolog of HSPA1A (Brocchieri *et al*., 2008).

Table 2: Overview of Type I and Type II HSP40 genes in human

Genes symbol	Alternative name(s)	Position of J-domain	HPRD ID	Gene map location	MW (Da)	Length
DNAJA1	HDJ2, HSJ2	5-60	04159	9p13-p112	44868	397
DNAJA2	CPR3,HIRIP 4,Dnj3	8-70	07105	16q12.1	45746	412
DNAJA3	hTid1, TiD1	92-150	09758	16p13.3	52489	480
DNAJA4	PRO1472	5-60	09920	15q25.1	47963	397
DNAJB1	HDJ1	3-60	05198	19p13.2	38044	340
DNAJB2	HSPF3	2-61	07249	2q32-q34	30568	277
DNAJB3	HCG3	2-61	13638	2q37	16559	145
DNAJB4	HLJ1	3-60	07486	1p31.1	37807	337
DNAJB5	HSC40	3-60	07106	9p13.3	39133	348
DNAJB6	MRJ, HSJ-2, MRJ-1	3-60	07107	7q36.3	36087	326
DNAJB7	HSC3	2-61	07010	22q13.2	35435	309
DNAJB8	MGC33884	2-61	09921	3q21.3	24686	232
DNAJB9	ERdj4, UNQ743/PRO1471	25-82		7q31;14q24.2 – q24.3	25518	223
DNAJB11	ErJ3, ERdj3	24-82	07485	3q27.3	40574	358
DNAJB12	DJ10, FLT20027	109-166	07086	10q22.1	45490	409
DNAJB13	TSARG5, TSARG6, FLJ46748	3-60	15573	11q13.4	36118	316
DNAJB14a	FLJ14281, PRO34683	107-164	07013	4q23	42516	379

Table 3: Overview of Type III HSP40 genes in human

Genes	Alternative name(s)	Position of J-domain	HPRD ID	Gene map location	MW(Da)	length
DNAJC1	HTJ1, ERdj1	63-138	09922	10p12.31	63883	554
DNAJC2	ZUO1, MPP11	87-153	19072	7q22	71996	621
DNAJC3	PRKR1, P58	393-454	03114	13q32.1	57580	504
DNAJC4	HSPf2, MCG18	19-50	09170	11q13	27593	241
DNAJC5	CSP, FLJ00118	14-72	08539	20q13.33	22149	198
DNAJC6	Auxilin, K1AA0473	848-909	16326	1pterq31.3	99996	913
DNAJC7	TPR2	380-443	07046	17q11.2	56441	484
DNAJC8	HSPC 315, SPF31	56-115	13236	1p35.3	29842	264
DNAJC9	JDD1	14-74	13237	10q22.2	29910	260
DNAJC10	ERDJ5	34-92	09722	2q32.1	91079	793
DNAJC11	FLJ10737	13-82	07112	1q36031	63278	559
DNAJC12	JDP1	13-71	06930	10q22.1	12456	198/107
DNAJC13	FLJ25863, K1AA0678	1300-1358	10915	3q22.1	254414	2243
DNAJC14	DR1P78, HDJ3	442-499	12082	12q13.2	78569	702
DNAJC15	MCJ	96-149	13238	13q14.1	16383	150
DNAJC16	RP4-680D5.1	28-85	17202	1p36.1	90591	782
DNAJC17	FLJ10634	11-76	07111	15q15.1	34687	304
DNAJC18	MGC29463	81-138	-	5q31.2	41551	358
DNAJC19	TIM14	61-115	12349	3q26.33	12499	116
DNAJC20	HSCB, JAC1	71-136	16289	22q12.1	27422	235
DNAJC21	DNAJA5	2-61	14056	5p13.2	67141	576/531
DNAJC22	FLJ13236	276-335	08580	12q13.12	38086	341
DNAJC23	SEC63	103-157	09783	6p21	87997	760
DNAJC24	Zinc finger CSL domain	10-74	15705	11p13	17139	149
DNAJC25	DNAJ-Like protein, Ba16L21.21	48-116	18685	9q31.3	42404	360
DNAJC26	GAK	1252-1329	143200	4p16		1311
DNAJC27	RBJ protein	216-273	15221	2p23.3	30855	273

DNAJC28	C21 or F55 protein, FLJ20461	50-108	10752	21q22.11	45806	454
DNAJC29	ARSACS, Sacsin	4357-4440	05135	13q11	504600	4432
DNAJC30	WBSCR18	48-106	10303	7q11.23	25961	226

Table 4: Overview of HSP70 genes in human

Genes Symbol	Alternative name(s)	HPRD ID	Gene map location	MW(Da)	Length	Published Interaction with HSP40	References
HSPA1A	HSP70-1, HSP72, HSPA1	00774	6p21.3	70052	641	DNAJC3,DNAJB11, DNAJA3	(Diefenbach *et al.*, 2000), (Sarkar *et al.*, 2001)
HSPA1B	HSP70-2	06784	6p21.3	70052	641	DNAJA1	(Imai *et al.*, 2002)
HSPA1L	HSP70-HOM	00776	6p21.3	70375	641	-	
HSPA2	Heat shock related 70 KDA protein 2	07174	14q24.1	70021	639	-	
HSPA5	GRP78, BIP	00682	9q33-q34.1	72333	654	ERDJ5(DNAJC10), DNAJC1	(Hellman *el al.*, 1999), (Chevalier *et al.*, 2000)
HSPA6	HSP70B	00775	1q23	71028	643	-	
HSPA7	-	-	1q23.3	?	?	-	
HSPA8	HSC70, HSC71, HSC73, HSPA10	07205	11q24.1	53517	646/493	DNAJA2, DNAJA3, DNAJC6	(Scheele *et al.*, 2001), (Sarkar *et al.*, 2001), (Jiang *et al.*, 2007)
HSPA9	GRP75, Mortalin 2	02770	5q31.1	73681	679	-	
HSPA12A	FLJ13874, KIAA0417	-	10q26.12	141000	1296	-	
HSPA12B	C20orf60	13683	20p13	75687	686	-	
HSPA13	Microsomal Stress 70 protein ATPase core	03061	21q11.1; 21q11	51927	471	-	
HSPA14	HSP70-4, HSP70L1	07021	10p13	54794	509	ZU01(DNAJC2)	(Otto *et al.*, 2005)

2.3.2 Subcellular Localization Predictions of HSP40s and HSP70s

Knowledge of the sub-cellular localization of proteins will enhance proper understanding of their biochemical functioning as co-localized genes ought to share similar biochemical functions (Hageman and Kampinga, 2009). Sub-cellular localization signals share the same characteristics. This has allowed for the use of various computational methods in predicting the localization of proteins within the cell. Experimentally determined localization sites (see Table 5 – 7) for HSP40s and HSP70s within the Human Protein Reference Database (HPRD) were retrieved for comparison with those predicted by the various prediction programs in other to ascertained their efficiency. Experimental localization sites have not yet been investigated for the remaining J proteins including DNAJB3, DNAJB4, DNAJB7, DNAJB8, DNAJB12, DNAJB13, DNAJB14, DNAJC2, DNAJC4, DNAJC9, DNAJC11, DNAJC12, DNAJC15, DNAJC16, DNAJC18, DNAJC21, DNAJC22, DNAJC24, DNAJC27, DNAJC28 and DNAJC29. Overall, most of the prediction programs were able to make predictions in line with those that have been previously established by experimental methods (see Table 5, 6 and 7). For instance, DNAJA3 and DNAJB9 were predicted to be localized in the mitochondria and endoplasmic reticulum respectively, which correlated with the experimental localization result (Table 5). However, there were some discrepancies between the experimentally determined localizations and the predictions by the computational methods. For example, DNAJA1 was thought to be localized in the acrosome, nucleus or golgi apparatus by experimental methods, whereas most of the prediction methods predicted it to be localized in the cytosol. Of note also was the divergence in the localization of DNAJC14 predicted to be localized in the nucleus as opposed to experimental prediction of being localized in endoplasmic reticulum (Table 6). Thus, prediction programs should be used carefully as some of these programs change overtime. Each prediction program was designed for specific purposes and target specific localization signal. It could also be that while some of the HSPs are resident at some positions within the cell, they are transported to another location under specific conditions (Qiu *et al*., 2006). In this study, in cases where the experimental subcellular localization result differed from that predicted by the various prediction programs, the experimental localization result was chosen. Based on consensus localization result from both the computational methods and the experimental procedure, the majority of the Type I and Type II HSP40s (DNAJA and DNAJB) members were predicted to be localized in the cytosol (Table 5). DNAJA3 was localized in the mitochondrial while both DNAJB9 and DNAJB11 were shown to be experimentally localized in the endoplasmic reticulum. In contrast to the sub-cellular localization of the Type I and II HSP40s, most of the Type III members were predicted to be localized in the nucleus as seen in Table 6. DNAJC4, DNAJC19, DNAJC20 and DNAJC28 were predicted as mitochondrial localized, while DNAJC10, DNAJC16, DNAJC23, DNAJC25, and DNAJC30 were predicted to be localized in the endoplasmic reticulum. This was consistent with previous review of HSP40 sub-cellular localization (Kampinga and Craig, 2010).

A large number of the HSP70 members were localized in the cytosol/nucleus (Table 7). This could suggest that the majority of the HSP70 genes were not products of gene duplication as a result of cellular compartmentalization (Hageman and Kampinga, 2009). This could also explain conversely, the reason for the high level of divergence in the Type III HSP40 genes. There might not have been much pressure on the Type III genes to retain the sequence identity as seen in both Type I and II (Hennessy *et al*.,2000). Only HSPA9 was localized in the mitochondria. HSPA5 and HSPA13 were predicted to be localized in the endoplasmic reticulum. HSPA2, HSPA6 and HSPA8 were experimentally predicted to be localized in the nucleus while HSPA1A, HSPA1B and HSPA1L were localized in the cytosol.

Within the human HSP70 family as presented in Table 7, majority of the proteins were localized in the cytosol including; HSPA1A, HSPA1B, HSPA1L, HSPA2, and HSPA6. HSPA5 and HSPA13 were endoplasmic reticulum localized and only HSP9 was localized in the mitochondria. The experimental localization of HSPA12A, HSPA12B and HSPA14 have not been determined. It was interesting to note that while HSPA1A, HSPA1B and HSPA1L were located on the same position on the chromosome and share similar cellular localizations, they do not all interact with the same HSP40s even though they are isoforms (Table 7).

The number of HSP40 genes out-numbers that of HSP70 genes. For example, in all, while there are nine HSP40s localized wihin the endoplasmic reticulum, only two HSP70s share the same localization. Similarly, there were three HSP40s localized within the mitochondrial whereas only one HSP70 share similar subcellular ocalization.

Table 5: Predictions of human HSP40 (DNAJA&B) subcellular localization

Gene symbol	Psort II	Ptarget	Multiloc2	Cello	Consensus	Experimental prediction	References
DNAJA1	cyt	cyt	cyt	nuc/cyt	cyt	acr/nuc/cyt	(Røsok *et al*., 1999), (Davis *et al*., 1998)
DNAJA2	cyt	cyt	cyt	nuc/cyt	cyt	cyt	(Terada *et al*., 2000)
DNAJA3	mit	mit	mit	mit	mit	mit	(Syken *et al*.,1999)
DNAJA4	cyt	cyt	cyt	nuc/cyt	cyt	cyt/pla	(Terada *et al*., 2002)
DNAJB1	cyt	nuc	cyt	cyt	cyt	cyt	(Freeman *et al*., 1996)
DNAJB2	nuc	nuc	cyt	nuc	nuc	cyt/nuc	(Chapple and Cheetham, 2003)
DNAJB3	cyt	cyt	cyt	cyt	cyt	-	
DNAJB4	nuc	nuc	cyt	cyt	nuc/cyt	-	
DNAJB5	cyt	nuc	cyt	cyt	cyt	cyt	(Ohtsuka *et al*., 2000)
DNAJB6	nuc	nuc	nuc	nuc/cyt	nuc	nuc/cyt	(Izawa *et al*., 2000)
DNAJB7	nuc	cyt	nuc	nuc	nuc	-	
DNAJB8	cyt	cyt	cyt	nuc/cyt	cyt	-	
DNAJB9	nuc	ER	sec	nuc/ext	nuc/ER	ER/nuc	(Haslam *et al*.,2000)
DNAJB11	ext	ER	sec	cyt	-	ER/mit	(Yu *et al*., 2000), (Mayya *et al*., 2007)
DNAJB12	nuc	nuc	mit	nuc	nuc	-	
DNAJB13	cyt	cyt	cyt	cyt	cyt	-	
DNAJB14	nuc	cyt	nuc	nuc	nuc	-	

Legend: **cyt** = cytoplasmic, **ER** = endoplassmic reticulum, **acr** = acrosome, **ext** = extracellular, **mit** = mitochondrial, **nuc** = nuclear, **pla** = plasma membrane, **sec** = secretory pathway.

Table 6: Predictions of human HSP40 (DNAJC) subcellular localization

Genes	Psort II	Ptarget	Multiloc1	Multiloc2	Cello	Consensus	Experimental prediction	References
DNAJC1	pla	nuc	ER	nuc	nuc	nuc/ER	ER/nuc/cyt	(Kroczynska *et al*., 2004) (Olsen *et al*., 2006)
DNAJC2	nuc	per	nuc	nuc	nuc	nuc	-	
DNAJC3	cyt	ER	ER	sec	cyt	cyt/ER	ER/cyt	(Korth *et al*., 1996)
DNAJC4	mit	gol	mit	mit	nuc	mit	-	
DNAJC5	nuc	cyt	ext	cyt	ext	cyt	cyt	(Zhang *et al*., 2002)
DNAJC6	nuc	nuc	nuc	cyt	nuc	nuc	nuc/cyt	(Ohtsuka *et al*.,2000)
DNAJC7	nuc	cyt	cyt	cyt	nuc	cyt	cyt	(Xiang *et al*., 2001)
DNAJC8	nuc	nuc	nuc	nuc	nuc	nuc	nuc	(Andersen *et al*., 2002)
DNAJC9	cyt	per	cyt	cyt	cyt	cyt	-	
DNAJC10	ER	gol	gol	cyt	cyt	cyt/gol/ER	ER	(Cunnea *et al*., 2003)
DNAJC11	cyt	cyt	nuc	cyt	nuc	cyt/nuc	-	
DNAJC12	nuc	nuc	nuc	cyt	nuc	nuc	-	
DNAJC13	pla	cyt	cyt	cyt	cyt	cyt	ED	
DNAJC14	nuc	nuc	nuc	nuc	nuc	nuc	ER/nuc/cyt	(Chen *et al*., 2003; Olsen *et al*., 2006)
DNAJC15	cyt	per	nuc	cyt	nuc	cyt/nuc	-	
DNAJC16	ER	ER	gol	sec	pla	ER	-	
DNAJC17	nuc	cyt	cyt	cyt	nuc	cyt/nuc	nuc	(Olsen *et al*., 2006)
DNAJC18	nuc	cyt	nuc	cyt	nuc	nuc	-	
DNAJC19	cyt	mit	mit	mit	mit	mit	mit	
DNAJC20	nuc	ER	mit	mit	nuc	mit	mit	(Cupp-vickery *et al* 2000)
DNAJC21	nuc	nuc	nuc	nuc	cyt	nuc	-	
DNAJC22	ER	lys	lys	mit	pla	lys	-	
DNAJC23	vac	cyt	ER	cyt	nuc	ER	ER	(Kurihara & Silver, 1993)
DNAJC24	cyt	cyt	cyt	cyt	nuc	cyt	-	
DNAJC25	ER	gol	ER	mit	pla	ER/pla	ER/pla	(Zhang *et al*., 2002)
DNAJC26	pla	lys	nuc	cyt	nuc	cyt	cyt	(Greener *et al*., 2000)
DNAJC27	cyt	cyt	cyt	cyt	cyt	cyt	-	
DNAJC28	mit	mit	mit	sec	nuc	mit	-	
DNAJC29	nuc	cyt	nuc	cyt	nuc	nuc/cyt	-	
DNAJC30	nuc	lys	ext	mit	mit	mit	ER/gol/nuc	(Simpson *et al*.,2009)

Legend: **cyt** = cytoplasmic, **ER** = endoplasmic reticulum, **ED** = Endosome, **ext** = extracellular, **gol** = golgi, **lys** = lysosome, **mit** = mitochondrial, **nuc** = nuclear, **per** = peroxisome, **pla** = plasma membrane, **sec** = secretory pathway.

Table 7: Prediction of Human HSP70 subcellular localization

Genes	Psort II	Ptarget	Multiloc1	Multiloc2	Cello	Consensus	Experimental prediction	References
HSPA1A	cyt	per	per	cyt	cyt	cyt	cyt	(Nogami *et al*., 2000)
HSPA1B	cyt	cyt	per	cyt	cyt	cyt	cyt	(Feng *et al*., 2001)
HSPA1L	cyt	cyt	per	cyt	cyt	cyt	cyt	(Fourie *et al*., 2001)
HSPA2	cyt	cyt	per	cyt	cyt	cyt	nuc/cyt	(Allen *et al*., 1996)
HSPA5	ER	per	ER	cyt	ER	ER	ER	(Morris *et al*., 1997)
HSPA6	cyt	cyt	nuc	cyt	cyt	cyt	nuc/cyt	(Mercier *et al*., 1999)
HSPA7	cyt	cyt	per	cyt	cyt	cyt	-	
HSPA8	cyt	cyt	per	cyt	cyt	cyt	nuc/cyt	(Rosorius *et al*., 2000), (Andersen *et al*., 2005)
HSPA9	mit	mit	mit	mit	mit	mit	mit	(Bhattacharyya *et al*., 1995)
HSPA12A	cyt	per	per	cyt	cyt	cyt/per	-	
HSPA12B	cyt	nuc	per	cyt	mit	cyt	-	
HSPA13	cyt	ER	ER	sec	cyt	cyt/ER	ER	(Otterson *et al*., 1994)
HSPA14	cyt	per	cyt	cyt	cyt	cyt	-	

Legend: **cyt** = cytoplasmic, **ER** = endoplasmic reticulum, **ED** = Endosome, **ext** = extracellular, **gol** = golgi, **lys** = lysosome, **mit** = mitochondrial, **nuc** = nuclear, **per** = peroxisome, **pla** = plasma membrane, **sec** = secretory pathway.

2.3.3 Motif Analysis of Full Length HSP40 Sequences

Homologous protein sequences that share the same ancestry ought to share similar functional characteristics since homology correlates strongly with the structure and function of a macromolecule (Bailey and Gribskov, 1998). A motif is a key functional part of a protein molecule which can be used in defining the characteristics of a protein family. A total of 20 motifs were searched for in the protein sequences. Those motifs with significant p-value better than 0.0001, did not overlap with others with significant occurrences as shown in the combined block diagram in Figure 5, were presented in Table 8 and 9. The sequence logos for the motifs found by MEME were presented in **Chapter 2**, **appendix I**. The picture observed from the motif analysis presented a clear classification of the HSP40 family. Motif 1 and 2 were part of the J domain previously reported as a main signature domain that defined all HSP40s. These motifs were found in all the types demonstrating that the J domain is conserved across all HSP40s. Motif 3, 4 and 5 were very similar to the Glycine/Phenylalanine (G/F) rich region in HSP40s and were present within the Type I and II proteins and absent in Type III (Table 8). Motif 7 which is the cysteine repeat region was also found only in the Type I proteins. Motif 8 is very similar to motif 7 as both are characterized with a high content of cysteine repeats with a long stretch of varied residues in between the motif. It is called the cysteine-rich region and previously proposed to be present in Type III HSP40s (Zhang *et al*., 2002). This motif was only present within DNAJC10 and DNAJC29. These observations were consistent with literature that the Zinc finger motif and the G/F domain were not present in the Type III proteins while the G/F domain is absent in Type II proteins (Hennessy *et al*., 2005). Surprisingly, the Zinc finger repeat was not found in the member 3 homolog (DNAJA3) of Type I whereas all other motifs found in other members of the subgroup were present in the protein. However, this region was found in the multiple sequence alignment analysis of the full length protein of the Type I HSP40s (data not shown). DNAJA1, DNAJA2 and DNAJA4 were all predicted to be localized in the cytosol while DNAJA3 is mitochondrial localized. Motif 9 was only found among the Type II proteins including DNAJB2, DNAJB6, DNAJB7 and DNAJB8. Also, motif 10 was found in DNAJB12, DNAJB14 and DNAJC18. Motif 11 and 16 were only present in DNAJC29. Motifs 12 and 20 were only found in DNAJC6 and DNAJC26. Motif 13 was only found in DNAJC6, DNAJC20 and DNAJC26. Interestingly, these three proteins were localized at different positions in the cell. For instance, while DNAJC6 is localized in the nucleus/cytosol, DNAJC20 is localized in mitochondrial and DNAJC26 is localized in the cytosol. This might indicate that while these proteins were localized differently within the cell, they may share similar functions. For example, both DNAJC6 and DNAJC26 were both involved in uncoathing of clathrin (Greener *et al.*, 2000). Highly conserved motif 14 was present in DNAJC6, DNAJC8, DNAJC13 and DNAJC26. Motif 15 was found in DNAJC1, DNAJC2, DNAJC9 and DNAJC29. Motif 17 was the tetratricopeptide repeat (TPR) domain which was only present repeatedly in DNAJC3 and DNAJC7 which were predicted to be localized in the endoplasmic reticulum and cytosol respectively. Highly conserved motif 18 was found only in DNAJC6, DNAJC13 and DNAJC26. Motif 19 was only found in DNAJB14, DNAJB12, DNAJC13 and DNAJC18. Apart from motifs 1 – 7, which constitute the domains frequently used in classifying HSP40 family namely the J domain, Glycine/Phenylalanine domain as well as the Cysteine repeat region, all other motifs found were first characterized in this study. Further detailed analysis of these motifs may provide more insight into the functional properties of HSP40s. Based on the combination of motifs found within the J proteins, DNAJA1, DNAJA2 and DNAJA4 were most similar (Table 8). DNAJB4 and DNAJB5 contained similar motifs while DNAJB6, DNAJB7 and DNAJB8 were more closely related having similar motif combinations. Both DNAJB13 and DNAJB14 contained the same set of motifs. There was a high level of variations among the Type III proteins (Table 9). However, motifs 1 and 2, which constitute the J domain, were present among the proteins. DNAJC1 and DNAJC2 contained the same number of similar motifs. Interestingly, while DNAJC1 is predicted to be endoplasmic reticulum localized, DNAJC2 is predicted to be localized in the nucleus. Similarly, DNAJC6 and DNAJC26 were the most closely related as they possess similar motifs and both proteins are known to be involved in similar cell functions in the cytoplasm (see Table 6).

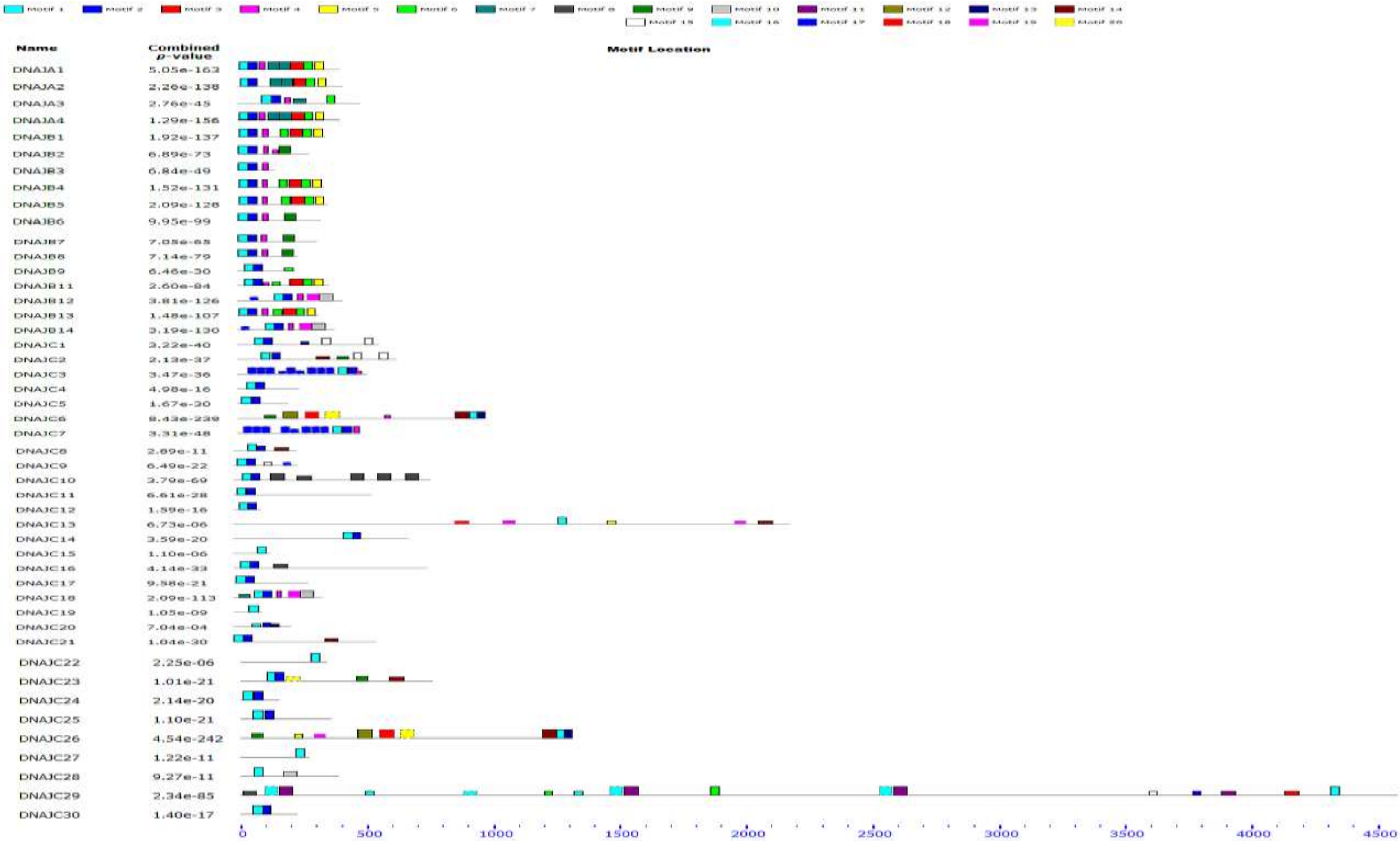
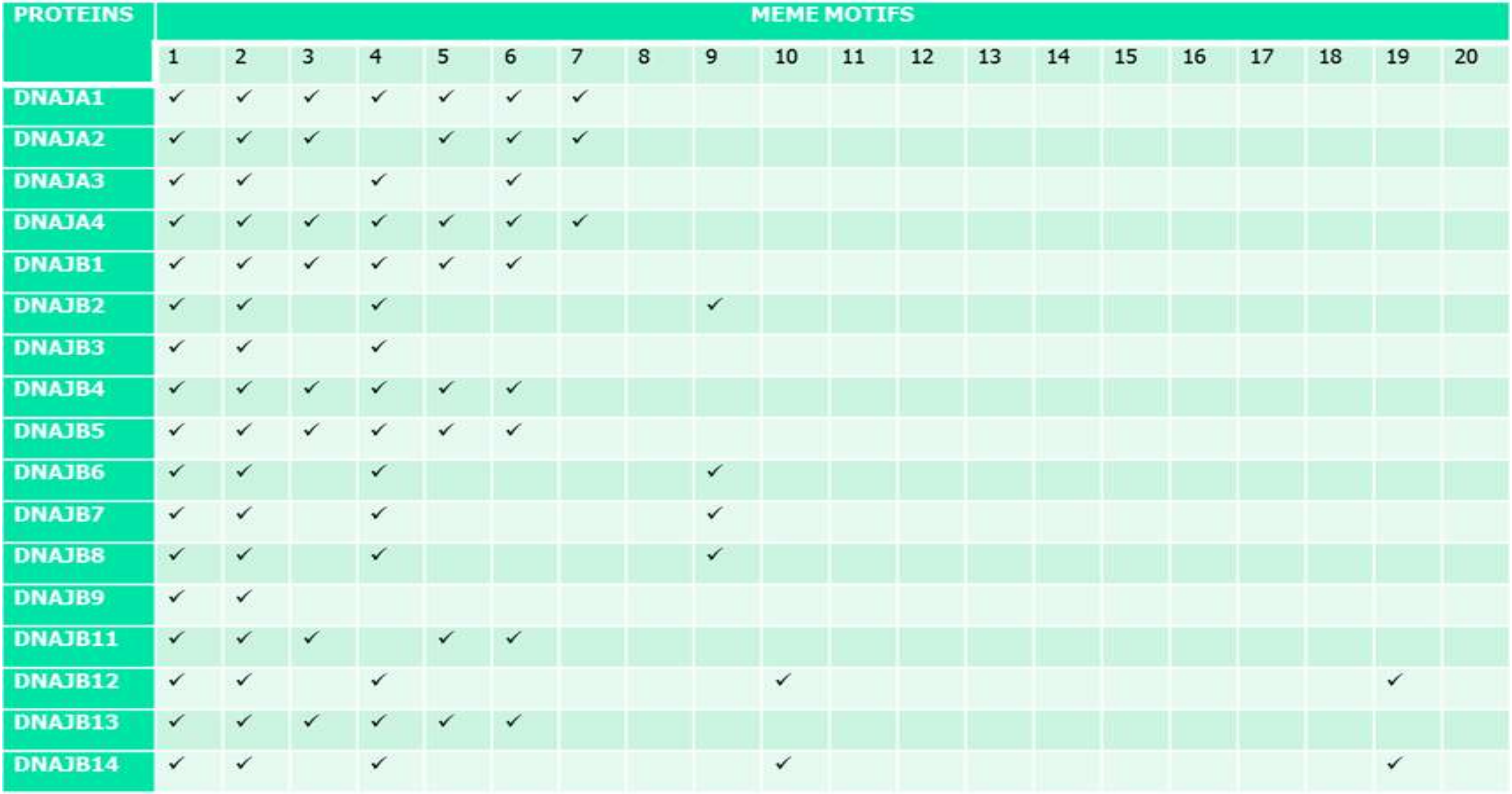


Figure 5: Block diagrams of motifs present within full length HSP40 sequences using MEME. 20 motifs were searched for within the full length proteins of the different types of the human HSP40s. while some of the motifs found were distincts, majority of them were parts of previously characterized domains that defined HSP40s. In all, motif 1 & 2 which constitute the J domain were present across the proteins.

Table 8: Motif analysis of full length protein sequence of DNAJA & DNAJB using MEME



PROTEINS	MEME MOTIFS																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
DNAJA1	✓	✓	✓	✓	✓	✓	✓													
DNAJA2	✓	✓	✓		✓	✓	✓													
DNAJA3	✓	✓		✓		✓														
DNAJA4	✓	✓	✓	✓	✓	✓	✓													
DNAJB1	✓	✓	✓	✓	✓	✓														
DNAJB2	✓	✓		✓					✓											
DNAJB3	✓	✓		✓																
DNAJB4	✓	✓	✓	✓	✓	✓														
DNAJB5	✓	✓	✓	✓	✓	✓														
DNAJB6	✓	✓		✓					✓											
DNAJB7	✓	✓		✓					✓											
DNAJB8	✓	✓		✓					✓											
DNAJB9	✓	✓																		
DNAJB11	✓	✓	✓		✓	✓														
DNAJB12	✓	✓		✓						✓									✓	
DNAJB13	✓	✓	✓	✓	✓	✓														
DNAJB14	✓	✓		✓						✓									✓	

Table 9: Motif analysis of full length protein sequence of DNAJC using MEME

PROTEINS	MEME MOTIFS																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
DNAJC1	✓	✓													✓					
DNAJC2	✓	✓													✓					
DNAJC3	✓	✓															✓			
DNAJC 4	✓	✓																		
DNAJC 5	✓	✓																		
DNAJC 6	✓											✓	✓	✓				✓		✓
DNAJC 7	✓	✓		✓													✓			
DNAJC 8	✓	✓												✓						
DNAJC 9	✓	✓													✓					
DNAJC 10	✓	✓						✓												
DNAJC 11	✓	✓																		
DNAJC 12	✓	✓																		
DNAJC 13	✓				✓									✓				✓	✓	
DNAJC 14	✓	✓																		
DNAJC 15	✓																			
DNAJC 16	✓	✓																		
DNAJC 17	✓	✓																		
DNAJC 18	✓	✓		✓						✓									✓	
DNAJC 19	✓																			
DNAJC 20	✓	✓											✓							
DNAJC 21	✓	✓																		
DNAJC 22	✓																			
DNAJC 23	✓	✓																		
DNAJC 24	✓	✓																		
DNAJC 25	✓	✓																		
DNAJC 26	✓											✓	✓	✓				✓		✓
DNAJC 27	✓																			
DNAJC 28	✓									✓										
DNAJC 29	✓	✓	✓			✓		✓			✓				✓	✓				
DNAJC 30	✓	✓																		

2.3.4 Sequence conservation of the J-domain in HSP40s

The J domain structure (see Figure 2B) is conserved in all known HSP40 proteins and contains the highly conserved HPD motif that is required for the stimulation of HSP70 ATPase activity (Genevaux *et al.*, 2002). The structure of the J domain consists of four helixes and the loop region located between helixes II and III. The level of J domain conservation is higher in the Type I and II HSP40 sub-families (Figure 6 and 7) than in the Type III family (Figure 8). The length and residue composition in the loop region also varied across the different types (Figure 9). Type I and II seem to have a high level of ASN, PRO and GLU residues in the loop region. The level of GLY residue was higher at the beginning of helix III region of Type I, while an ALA residue is found to have high level of conservation among the Type II and Type III proteins though with a lower level in Type III. Interestingly, a high level of residue variation was observed at the start of helix III of Type III proteins with no significant bias to any residue. The ALA residue at the start of helix III was replaced with a SER residue in DNAJB13, DNAJC15, DNAJC19, DNAJC20 and DNAJC27 respectively (Figure 7 and 8). The reason for these variations has not been fully clarified, although many of the Type III J proteins may be products of gene duplication events since most of the proteins are localized in different positions within the cell (Hageman *et al*., 2011; Hageman and Kampinga, 2009). Aside the highly conserved HPD motif in the loop region, other conserved residues were found with high level of conservation especially in Type I and II HSP40 including the LEU-GLY-VAL residues on helix I, LYS-LYS-ALA-TYR quartet and LEU-ALA residues on helix II. The LYS-PHE-LYS (KFK) motif and the ALA-TYR-GLU-VAL-LEU-SER signature residues on helix III are also highly conserved. Of note also were the LYS and ARG residues as well as TYR-ASP residues located on helix IV (Figure 6). The KFK motif was less conserved across the Type II proteins (Figure 7) and almost absent in Type III except in DNAJC5 and DNAJC7 (Figure 8). The PHE on the KFK motif was highly conserved across all the sub-families (Figure 9). Of interest however was the replacement of the PHE residue on helix III with a SER residue in DNAJC20. Interestingly, all these highly conserved residues across the four J domain helixes have been proposed to be involved in J domain interactions with partner HSP70s (Jiang *et al*., 2007; Nicoll *et al*., 2007; Hennessy *et al*., 2005; 2000).

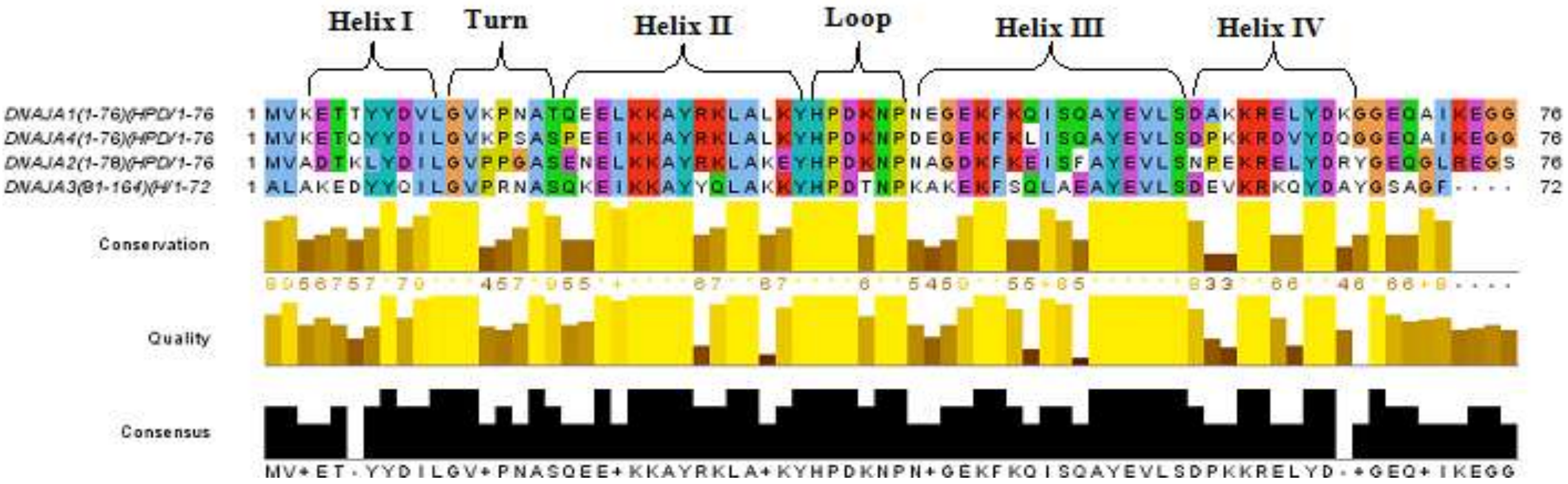

Helix I
Turn
Helix II
Loop
Helix III
Helix IV
DNAJA1(1-76)(HPD/1-76 1 MVKETTYYDVLGVKPNATQEELKKAYRKLALKYHPDKNPNEGEKFKQISQAYEVLSDAKKRELYDKGGEQAIKEGG 76
DNAJA4(1-76)(HPD/1-76 1 MVKETQYYDILGVKPSASPEEIKKAYRKLALKYHPDKNPDEGEKFKLISQAYEVLSDPKKRDVYDQGGEQAIKEGG 76
DNAJA2(1-78)(HPD/1-76 1 MVADTKLYDILGVPPGASENELKKAYRKLAKEYHPDKNPNAGDKFKEISFAYEVLSNPEKRELYDRYGEQGLREGS 76
DNAJA3(81-164)(H/1-72 1 ALAKEDYYQILGVPRNASQKEIKKAYYQLAKKYHPDTNPKAKEKFSQLAEAYEVLSDEVKRKQYDAYGSAGF---- 72
Conservation
Quality
Consensus
MV+ET-YYDILGV+PNASQEE+KKAYRKLA+KYHPDKNPN+GEKFKQISQAYEVLSDPKKRELYD-+GEQ+IKEGG

Figure 6: Multiple sequence alignment of Type I HSP40 J domain. The positions of the four helixes present within the J domain as well as the turn and loop regions between helixes I & II and helixes II & III respectively were highlighted as shown in the figure. Aside the HPD motif, highly conserved residues were also present across the helixes. The presence of an higly conserved glycine residue immediately after helix IV showed the beginning of the GLY/PHE rich region; a typical domain present in Type I & II HSP40s. The standard nomenclature for the proteins, the positions of the J domain and the number of aligned residues in each of the proteins were also highlighted on the left hand side of the figure. The level of conservation of each amino acid residues, the quality of the conservation as well as the consensus residues across the protein sequences were depicted in the figure as shown above. Figure was generated using JALVIEW (Waterhouse *et al*., 2009).

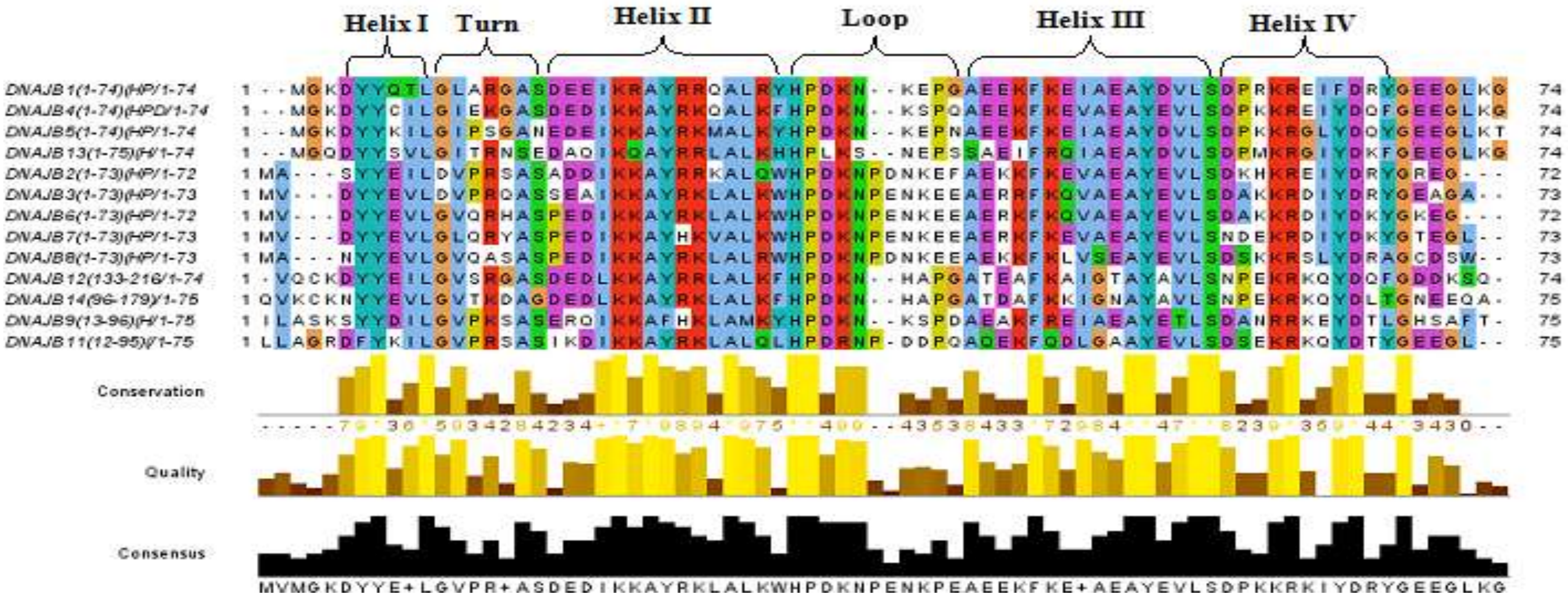


Figure 7: Multiple sequence alignment of Type II HSP40 J domain. The positions of the four helixes present within the J domain as well as the turn and loop regions between helixes I & II and helixes II & III respectively were highlighted as shown in the figure. Aside the HPD motif, highly conserved residues were also present across the helixes. However, the length and residue composition in the loop region varied as observed in the Type I J domains. The presence of an higly conserved glycine residue immediately after helix IV showed the beginning of the GLY/PHE rich region; a typical domain present in Type I & II HSP40s. The standard nomenclature for the proteins, the positions of the J domain and the number of aligned residues in each of the proteins were also highlighted on the left hand side of the figure. The level of conservation of each amino acid residues, the quality of the conservation as well as the consensus residues across the protein sequences were depicted in the figure as shown above. Figure was generated using JALVIEW (Waterhouse *et al*., 2009).

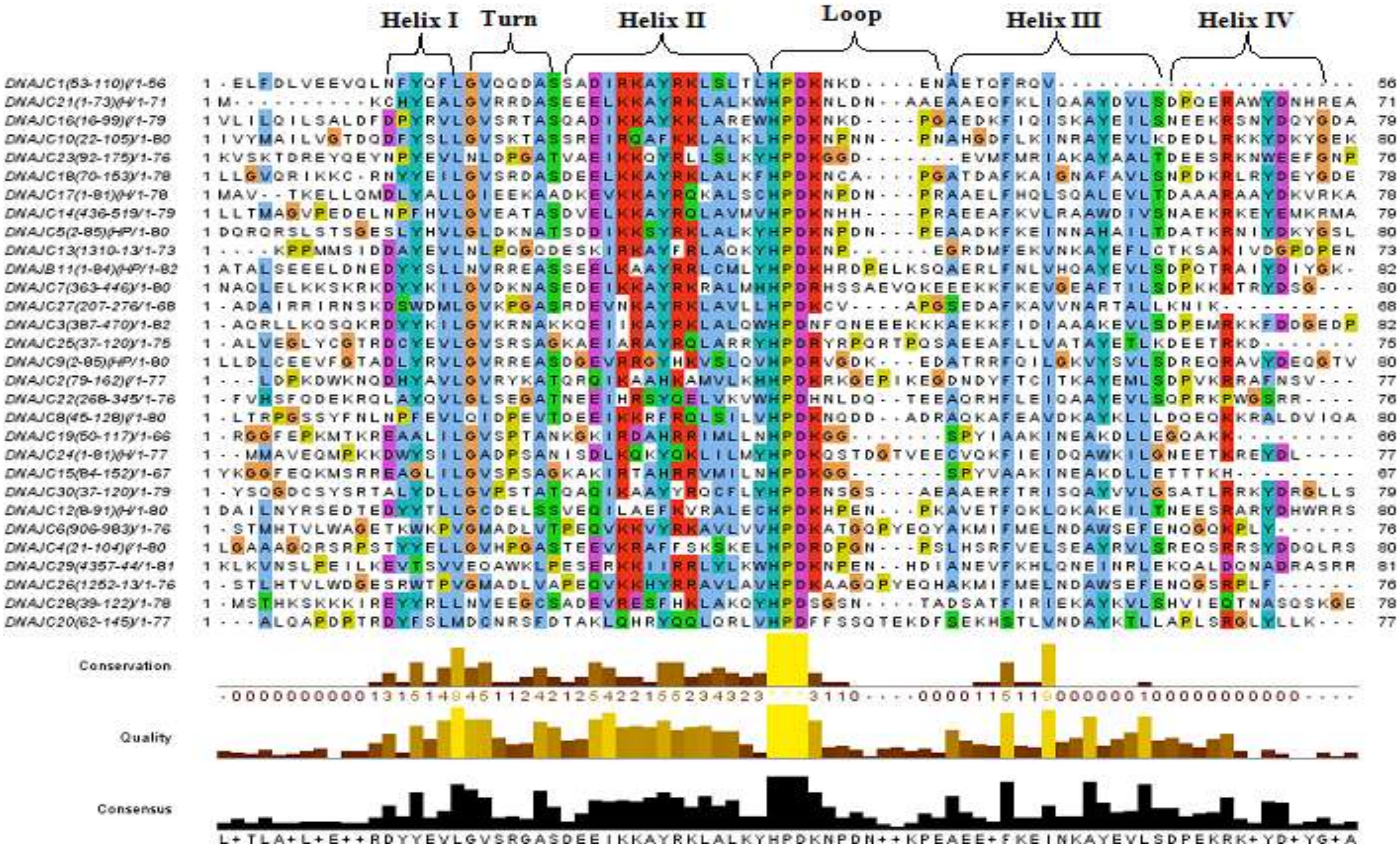


Figure 8: Multiple sequence alignment of Type III HSP40 J domain. The positions of the four helixes present within the J domain as well as the turn and loop regions between helixes I & II and helixes II & III respectively were highlighted as shown in the figure. Aside the HPD motif, the level of conservation was higher in both helixes I & II than observed in helixes III & IV as well as in the loop region. However, the residues conservation observed were considerably lower than found among the Type I & II J domians. These could probably explained why Type I & II J proteins may not be interchanged with Type III proteins for functioning in domain swapping experiments. The standard nomenclature for the proteins, the positions of the J domain and the number of aligned residues in each of the proteins were also highlighted on the left hand side of the figure. The level of conservation of each amino acid residues, the quality of the conservation as well as the consensus residues across the protein sequences were depicted in the figure as shown above. Figure was generated using JALVIEW (Waterhouse *et al*., 2009).

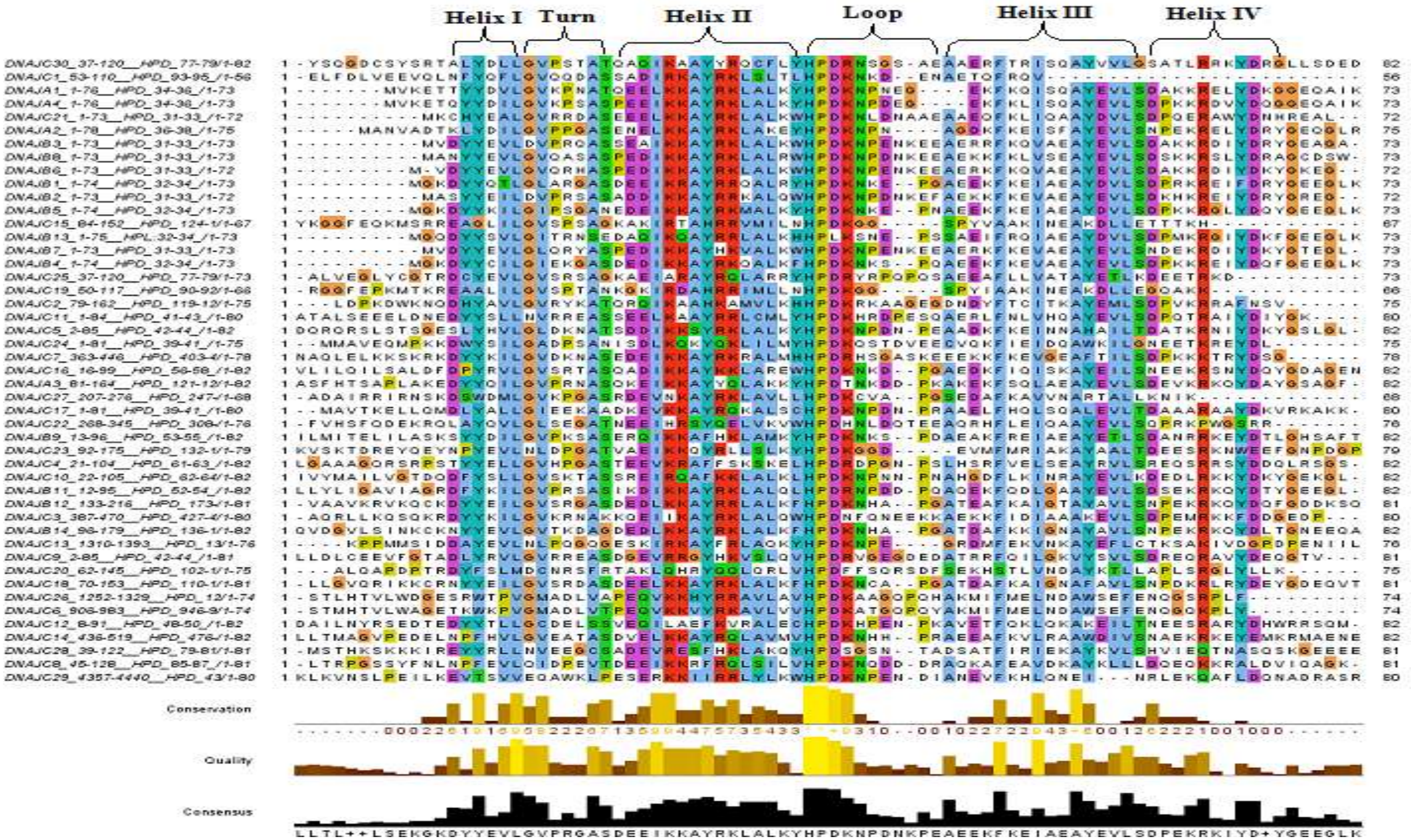


Figure 9: Multiple sequence alignment of combined HSP40 J domain. Alignment of the combined J domain from Type I, II and I. The positions of the four helixes present within the J domain as well as the turn and loop regions between helixes I & II and helixes II & III respectively were highlighted as shown in the figure. Aside the HPD motif, the level of conservation was higher in both helixes I & II than observed in helixes III & IV. The standard nomenclature for the proteins, the positions of the J domain and the number of aligned residues in each of the proteins were also highlighted on the left hand side of the figure. The level of conservation of each amino acid residues, the quality of the conservation as well as the consensus residues across the protein sequences were depicted in the figure as shown above. Figure was generated using JALVIEW (Waterhouse *et al*., 2009).

2.3.5 Consensus Sequence Analysis of Human HSP40 J-domain

Consensus sequences from each of the HSP40 sub-families were aligned and the positions of residues that could be important for HSP40-HSP70 interactions were identified on some of the available 3-dimensional structures of Human HSP40s using PROMAL3D (Figure 10). 2QWO is the only available crystal structure of an HSP40-HSP70 complex, showing the complex between the J domain of auxilin (DNAJC6) and Bovine HSC70 (HSPA8) (Jiang *et al*., 2007). Other structures of J domains alone include 1HDJ, which is a Type II HSP40 (DNAJB1), 2CTQ, which is a Type III member (DNAJC12) and 2CTW, which is a homolog subfamily C member 5 HSP40 from mouse. There are highly conserved residues that could be involved in maintaining the structural integrity and stability of the J domain for interactions with partner HSP70. These are: the PHE at position 49 on the combined consensus sequence (position 45 in 1HDJ) (F891 in 2QWO) which is part of the tripeptide LYS-PHE-LYS (KFK) motif in the middle of helix III; the ALA residue at consensus position 55 (position 51 in 1HDJ) and the TYR residue at consensus position 8 (position 6 in 1HDJ). Others conserved residues that could mediate general binding and interactions with partner HSP70 include the highly conserved tripeptide HPD motif in the loop region and the GLU at positions 18, 43 and 47 respectively in IHDJ. Also notable are the ASP at positions 57 and 65 respectively. Highly conserved LYS residues at consensus positions 23, 24, 28, 32 and 64 as well as the ARG residue at position 65 are positively charged residues that have been previously reported to interact with the negatively charged residues at the under cleft region of HSP70 ATPase domain (Hennessy *et al*., 2005b; Nicoll *et al*., 2007; Suh *et al*., 1999). The last category includes those residues with low level of conservation which could probably define specific HSP40-HSP70 interactions. Significant among those residues is the ALA residue at position 53 in the combined consensus sequence as well as in the Type II (DNAJB) and 1HDJ structures. This ALA residue was replaced by a SER residue in Type I (DNAJA) but an ASN residue mainly across all the Type III sequences as shown in Figure 10. Both LYS 64 and ARG 65 in the combined consensus sequence were parts of the residues (EKRKI), corresponding to the QKRAA motif in *E. coli* J domain (Auger and Roudier, 1997; Genevaux *et al*., 2002). Interestingly, these two residues seemed to be highly conserved and thereby could function in the general binding of HSP40s to HSP70s while the GLU and ILE residues are less conserved. Surprisingly, while the GLU was only retained in the combined consensus sequence and in DNAJC (Type III) consensus sequence, it was replaced with a LYS residue in both DNAJA (Type I) and DNAB (Type II) consensus sequences. Of note also was the PRO residue at position 14 in the combined consensus sequence as well as in DNAJB (Type II) consensus sequence. This residue was replaced with a LYS residue in DNAJA (Type I) consensus sequence and a SER residue in DNAJC (Type III) consensus sequence as shown in Figure 10.

An overview of the sequence alignment of the Type III HSP40 based on their sub-cellular localizations (Figure 11 - 15) showed high level of variation in the KFK motif in most of the localization groups except for those localised in the cytosol particularly DNAJC5 and DNAJC7 (Figure 11). The two LYS residues on this motif were completely absent in most of the Type III proteins (DNAJC10, DNAJC23 & DNAJC25) that are localized in the endoplasmic reticulum (Figure 12) and this tripeptide motif (KFK) was also completely absent in those proteins localized in the endoplasmic reticulum as seen in the overall consensus' sequence alignment in (Figure 15). This might explain why the J domains of endoplasmic reticulum proteins could not be interchanged with those localized in the cytosol in yeast HSP40s in J domain swapping experiment (Schlenstedt *et al*., 1995). However, the level of residue conservation was higher in those proteins localized in the endoplasmic reticulum (ER) than those in the cytosol (Figure 11 and Figure 12). This suggested that the endoplasmic reticulum localized proteins contain other highly conserved residues that may not be present in those proteins localized in the cytosol *(*Hennessy *et al*., 2000), thus proteins in the cytosol may fulfil more diverse functions while the functions in the endoplasmic reticulum are likely to be more restricted. For example, the two GLU on helix IV at positions 52 and 53 in the ER localized consensus sequence (Figure 15) were highly conserved in those proteins localized in the endoplasmic reticulum as opposed to those localized in the cytosol.

There was a high degree of variation in the residue compositions in the loop regions even among those proteins in the same sub-cellular localization. There was high LYS and ALA residues composition in loop region of those proteins localized in the cytosol and nucleus though with higher level of ASN residue in the latter. The composition outside the HPD motif was biased toward a high ASN and GLU residues in those localized in the endoplasmic reticulum while the residue composition was biased toward SER and GLY residues in those localized in the mitochondria.

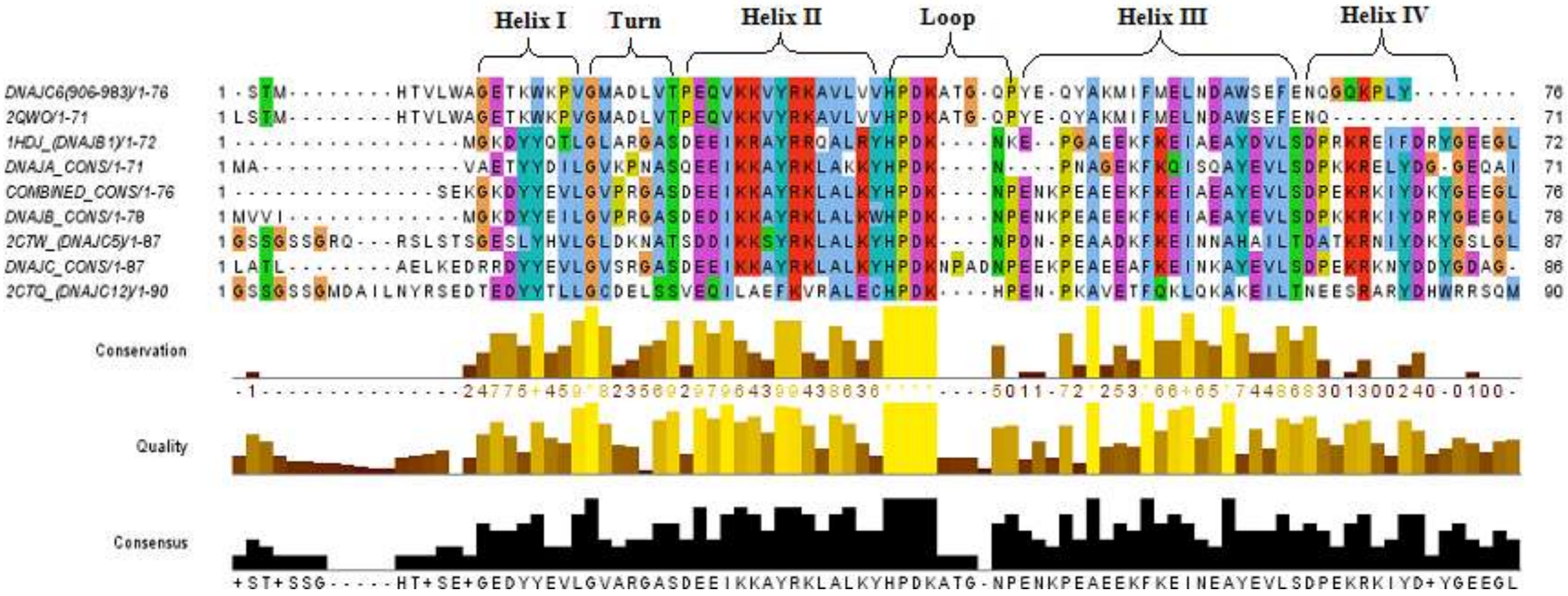

Helix I
Turn
Helix II
Loop
Helix III
Helix IV
DNAJC6(906-983)/1-76 1 -STM---------HTVLWAGETKWKPVGMADLVTPEQVKKVYRKAVLVVHPDKATG-QPYE-QYAKMIFMELNDAWSEFENQGQKPLY--------- 76
2QWO/1-71 1 LSTM---------HTVLWAGETKWKPVGMADLVTPEQVKKVYRKAVLVVHPDKATG-QPYE-QYAKMIFMELNDAWSEFENQ-------------- 71
1HDJ_(DNAJB1)/1-72 1 -----------------MGKDYYQTLGLARGASDEEIKRAYRRQALRYHPDK----NKE--PGAEEKFKEIAEAYDVLSDPRKREIFDRYGEEGL 72
DNAJA_CONS/1-71 1 MA---------------VAETYYDILGVKPNASQEEIKKAYRKLAKKYHPDK----N----PNAGEKFKQISQAYEVLSDPKKRELYDG-GEQAI 71
COMBINED_CONS/1-76 1 ---------------SEKGKDYYEVLGVPRGASDEEIKKAYRKLALKYHPDK----NPENKPEAEEKFKEIAEAYEVLSDPEKRKIYDKYGEEGL 76
DNAJB_CONS/1-78 1 MVVI-------------MGKDYYEILGVPRGASDEDIKKAYRKLALKWHPDK----NPENKPEAEEKFKEIAEAYEVLSDPKKRKIYDRYGEEGL 78
2C7W_(DNAJC5)/1-87 1 GSSGSSGRQ---RSLSTSGESLYHVLGLDKNATSDDIKKSYRKLALKYHPDK----NPDN-PEAADKFKEINNAHAILTDATKRNIYDKYGSLGL 87
DNAJC_CONS/1-87 1 LATL---------AELKEDRRDYYEVLGVSRGASDEEIKKAYRKLALKYHPDKNPADNPEEKPEAEEAFKEINKAYEVLSDPEKRKNYDDYGDAG- 86
2CTQ_(DNAJC12)/1-90 1 GSSGSSGMDAILNYRSEDTEDYYTLLGCDELSSVEQILAEFKVRALECHPDK----HPEN-PKAVETFQKLQKAKEILTNEESRARYDHWRRSQM 90
Conservation
-1---------------24775+459*823569297964399438636****----5011-72*253*66+65*744868301300240-0100-
Quality
Consensus
+ST+SSG-----HT+SE+GEDYYEVLGVARGASDEEIKKAYRKLALKYHPDKATG-NPENKPEAEEKFKEINEAYEVLSDPEKRKIYD+YGEEGL

Figure 10: Multiple sequence alignment of consensus sequences from the different HSP40 types. Consensus sequences were derived from the alignment of each of the HSP40 sub-family (Type I, II & III) and from the alignment of the combined J domain from all HSP40s in human. Sequences from the structures of DNAJC6 (2QWO), DNAJB1 (1HDJ), DNAJC5 (2CTW) from Mouse ortholog, and DNAJC12 (2CTQ) were also used in the alignment to locate the positions of the highly and less conserved residues critical for J domain interactions with HSP70s. The four helixes were highlighted as well as the turn region and loop region as shown in the figure above. The standard nomenclature for the proteins and the number of aligned residues in each of the proteins were also highlighted on the left hand side of the figure. The level of conservation of each amino acid residues, the quality of the conservation as well as the consensus residues across the protein sequences were depicted in the figure as shown above. Figure was generated using JALVIEW (Waterhouse *et al*., 2009).

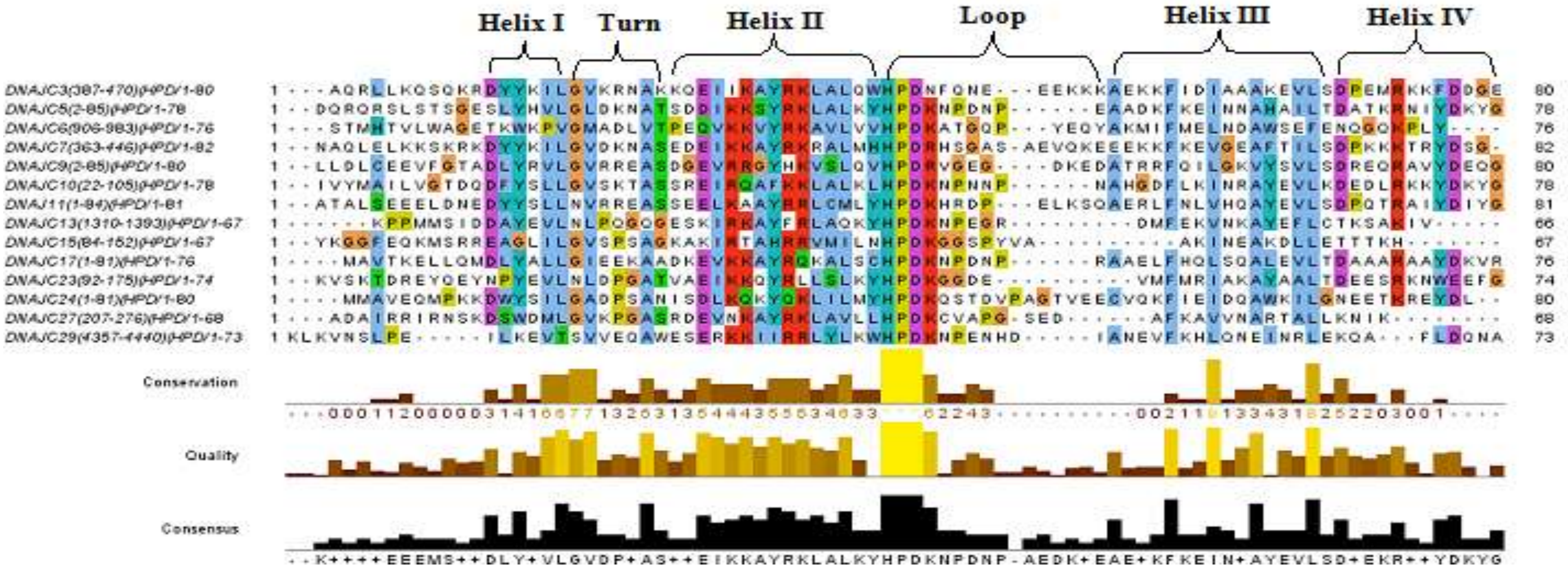


Figure 11: **Multiple sequence alignment of Type III HSP40 based on sub-cellular localizations of proteins predicted to be localized in the cytosol.** The positions of the four helixes present within the J domain as well as the turn and loop regions between helixes I & II and helixes II & III respectively were highlighted as shown in the figure. Aside the HPD motif, the level of conservation was higher in both helixes I & II than observed in helixes III & IV. The standard nomenclature for the proteins, the positions of the J domain and the number of aligned residues in each of the proteins were also highlighted on the left hand side of the figure. The level of conservation of each amino acid residues, the quality of the conservation as well as the consensus residues across the protein sequences were depicted in the figure as shown above. Figure was generated using JALVIEW (Waterhouse *et al*., 2009).

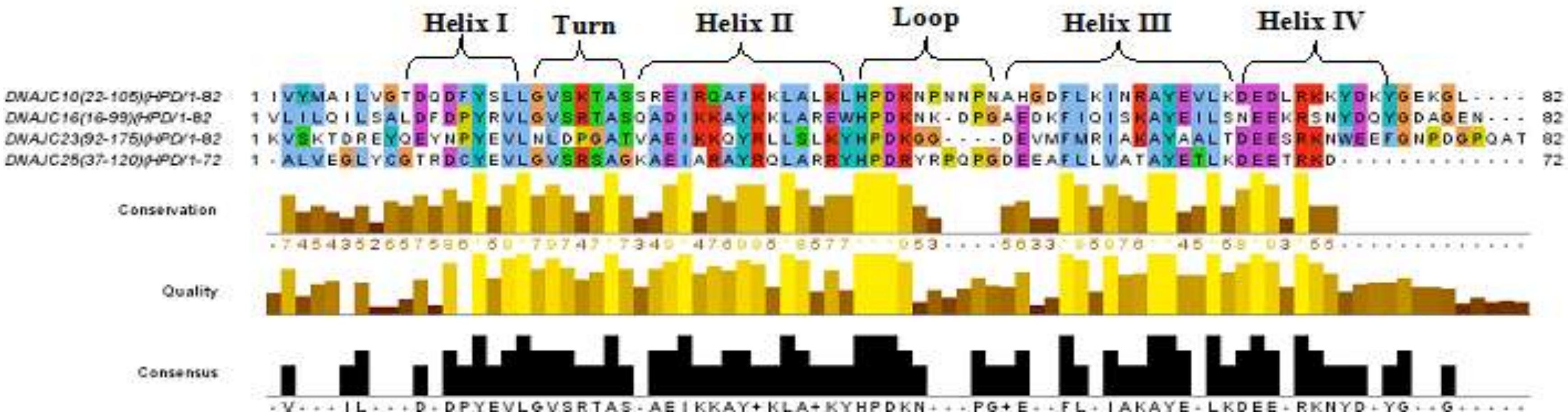

Helix I
Turn
Helix II
Loop
Helix III
Helix IV
DNAJC10(22-105)/HPD/1-82 1 IVYMAILVGTDQDFYSLLGVSKTASSREIRQAFKKLALKLHPDKNPNNPNAHGDFLKINRAYEVLKDEDLRKKYDKYGEKGL---- 82
DNAJC16(16-99)/HPD/1-82 1 VLILQILSALDFDPYRVLGVSRTASQADIKKAYKKLAREWHPDKNK-DPGAEDKFIQISKAYEILSNEEKRSNYDQYGDAGEN--- 82
DNAJC23(92-175)/HPD/1-82 1 KVSKTDREYQEYNPYEVLNLDPGATVAEIKKQYRLLSLKYHPDKGG----DEVMFMRIAKAYAALTDEESRKNWEEFGNPDGPQAT 82
DNAJC25(37-120)/HPD/1-72 1 -ALVEGLYCGTRDCYEVLGVSRSAGKAEIARAYRQLARRYHPDRYRPQPGDEEAFLLVATAYETLKDEETRKD------------- 72
Conservation
Quality
Consensus
-V---IL---D-DPYEVLGVSRTAS-AEIKKAY+KLA+KYHPDKN---PG+E--FL-IAKAYE-LKDEE-RKNYD-YG--G-----

Figure 12: Multiple sequence alignment of Type III HSP40 based on sub-cellular localizations of proteins predicted to be localized in the endoplasmic reticulum. The positions of the four helixes present within the J domain as well as the turn and loop regions between helixes I & II and helixes II & III respectively were highlighted as shown in the figure. Aside the HPD motif, the level of conservation was higher in both helixes I & II than observed in helixes III & IV. However, there appeared to be more conserved residues present within endoplasmic reticulum than other proteins localized at the other positions within the cell especially on helix III. The standard nomenclature for the proteins, the positions of the J domain and the number of aligned residues in each of the proteins were also highlighted on the left hand side of the figure. The level of conservation of each amino acid residues, the quality of the conservation as well as the consensus residues across the protein sequences were depicted in the figure as shown above. Figure was generated using JALVIEW (Waterhouse *et al*., 2009).

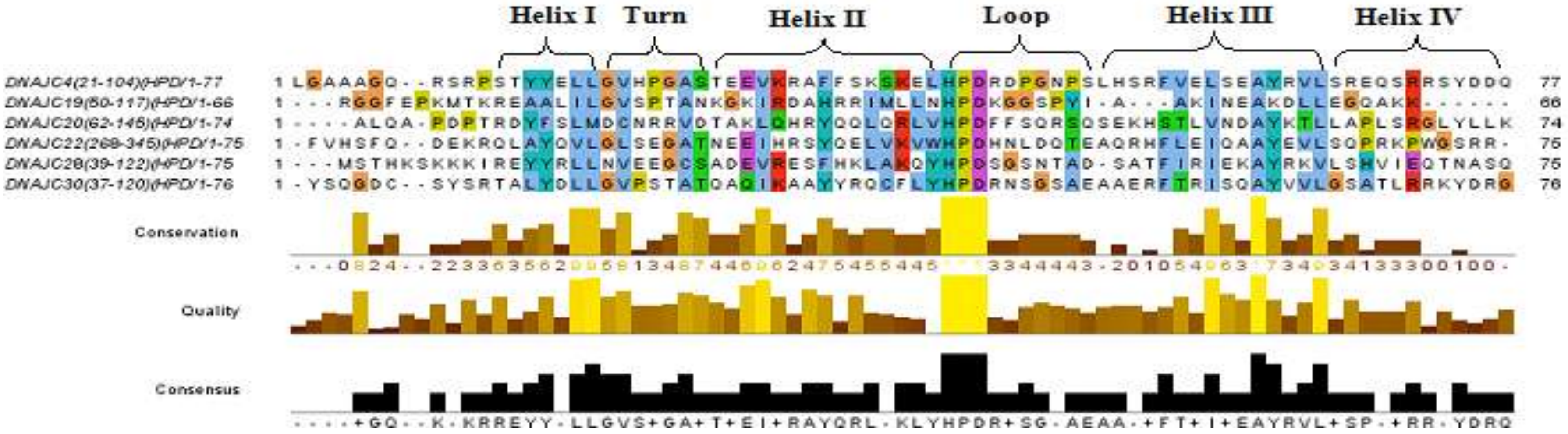

Helix I
Turn
Helix II
Loop
Helix III
Helix IV
DNAJC4(21-104)/HPD/1-77 1 LGAAAGQ--RSRPSTYYELLGVHPGASTEEVKRAFFSKSKELHPDRDPGNPSLHSRFVELSEAYRVLSREQSRRSYDDQ 77
DNAJC19(50-117)/HPD/1-66 1 ---RGGFEPKMTKREAALILGVSPTANKGKIRDAHRRIMLLNHPDKGGSPYI-A---AKINEAKDLLEGQAKK------ 66
DNAJC20(62-145)/HPD/1-74 1 ----ALQA-PDPTRDYFSLMDCNRRVDTAKLQHRYQQLQRLVHPDFFSQRSQSEKHSTLVNDAYKTLLAPLSRGLYLLK 74
DNAJC22(268-345)/HPD/1-75 1 -FVHSFQ--DEKRQLAYQVLGLSEGATNEEIHRSYQELVKVWHPDHNLDQTEAQRHFLEIQAAYEVLSQPRKPWGSRR- 75
DNAJC28(39-122)/HPD/1-75 1 ---MSTHKSKKKIREYYRLLNVEEGCSADEVRESFHKLAKQYHPDSGSNTAD-SATFIRIEKAYRKVLSHVIEQTNASQ 75
DNAJC30(37-120)/HPD/1-76 1 -YSQGDC--SYSRTALYDLLGVPSTATQAQIKAAYYRQCFLYHPDRNSGSAEAAERFTRISQAYVVLGSATLRRKYDRG 76
Conservation
---0824--223363562996581348744696247545544511334444 3-201054963173493413330 0100-
Quality
Consensus
----+GQ--K-KRREYY-LLGVS+GA+T+EI+RAYQRL-KLYHPDR+SG-AEAA-+FT+I+EAYRVL+SP-+RR-YDRQ

Figure 13: Multiple sequence alignment of Type III HSP40 based on sub-cellular localizations of proteins predicted to be localized in the mitochondrial. The positions of the four helixes present within the J domain as well as the turn and loop regions between helixes I & II and helixes II & III respectively were highlighted as shown in the figure. Aside the HPD motif, the level of conservation was higher in both helixes I & II than observed in helixes III & IV. The standard nomenclature for the proteins, the positions of the J domain and the number of aligned residues in each of the proteins were also highlighted on the left hand side of the figure. The level of conservation of each amino acid, the quality of the conservation as well as the consensus residues across the protein sequences were depicted in the figure as shown above. Figure was generated using JALVIEW (Waterhouse *et al*., 2009).

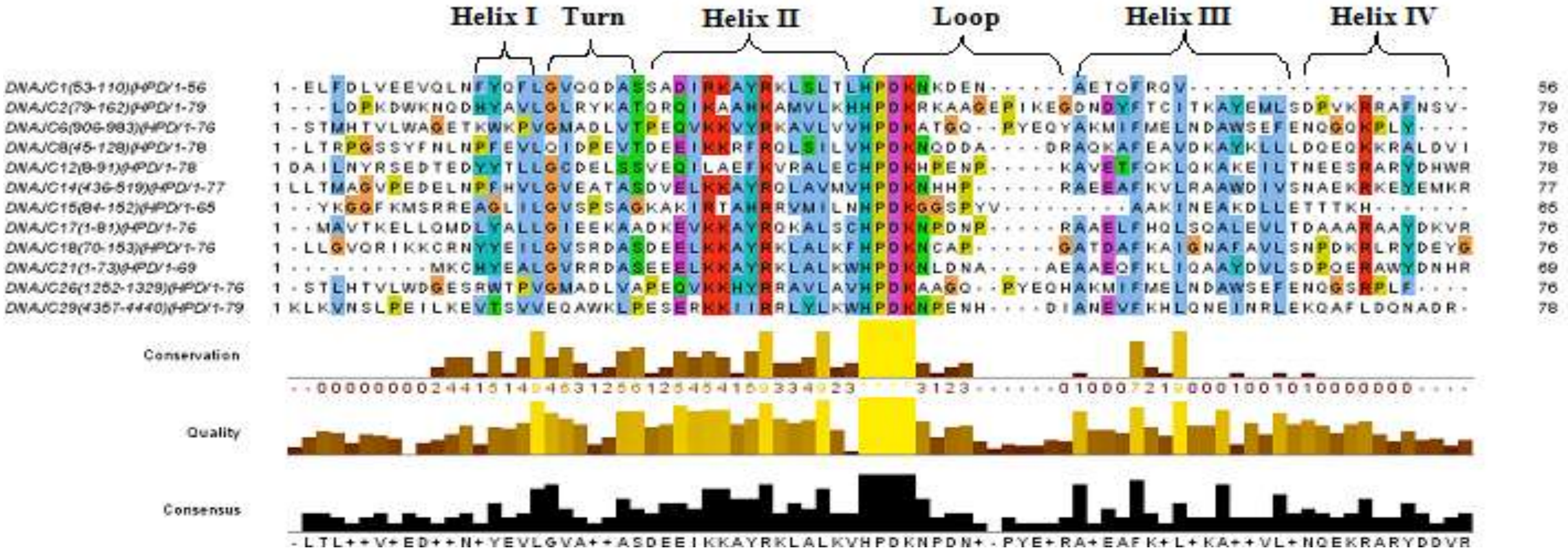


Figure 14: Multiple sequence alignment of Type III HSP40 based on sub-cellular localizations of proteins predicted to be localized in the nucleus. The positions of the four helixes present within the J domain as well as the turn and loop regions between helixes I & II and helixes II & III respectively were highlighted as shown in the figure. Aside the HPD motif, the level of conservation was higher in both helixes I & II than observed in helixes III & IV. The standard nomenclature for the proteins, the positions of the J domain and the number of aligned residues in each of the proteins were also highlighted on the left hand side of the figure. The level of conservation of each amino acid residues, the quality of the conservation as well as the consensus residues across the protein sequences were depicted in the figure as shown above. Figure was generated using JALVIEW (Waterhouse *et al*., 2009).

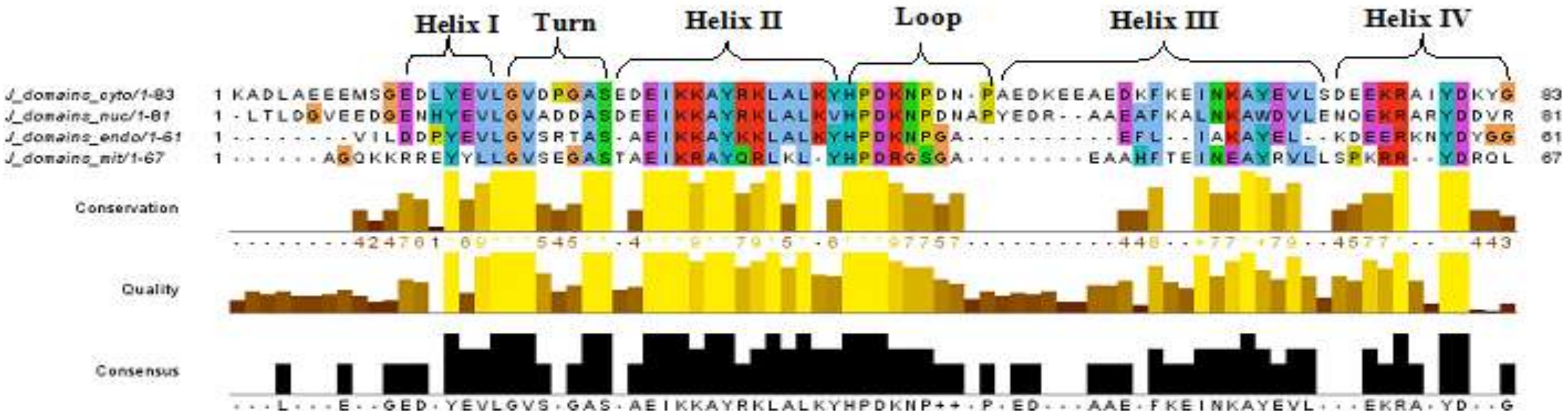

Helix I
Turn
Helix II
Loop
Helix III
Helix IV
J_domains_cyto/1-83
J_domains_nuc/1-81
J_domains_endo/1-61
J_domains_mit/1-67
1 KADLAEEEMSGEDLYEVLGVDPGASEDEIKKAYRKLALKYHPDKNPDN-PAEDKEEAEDKFKEINKAYEVLSDEEKRAIYDKYG 83
1 -LTLDGVEEDGENHYEVLGVADDASDEEIKKAYRKLALKVHPDKNPDNAPYEDR--AAEAFKALNKAWDVLENQEKRARYDDVR 81
1 ---------VILDDPYEVLGVSRTAS-AEIKKAYKKLALKYHPDKNPGA-----------EFL--IAKAYEL--KDEERKNYDYGG 61
1 ------AGQKKRREYYLLGVSEGASTAEIKRAYQRLKL-YHPDRGSGA---------EAAHFTEINEAYRVLLSPKRR--YDRQL 67
Conservation
Quality
Consensus
---L---E--GED-YEVLGVS-GAS-AEIKKAYRKLALKYHPDKNP++-P-ED---AAE-FKEINKAYEVL---EKRA-YD--G

Figure 15: Multiple sequence alignment of Type III HSP40 based on subcellular localizations of consensus sequences derived from each of the sub-cellular localization groups. The positions of the four helixes present within the J domain as well as the turn and loop regions between helixes I & II and helixes II & III respectively were highlighted as shown in the figure. Aside the HPD motif, the level of conservation was higher in both helixes I & II than observed in helixes III & IV. The consensus sequence from the proteins localized in different regions of the cell and the number of aligned residues in each of the censesus sequences were also highlighted on the left hand side of the figure. The level of conservation of each amino acid residues, the quality of the conservation as well as the consensus residues across the protein sequences were depicted in the figure as shown above. Figure was generated using JALVIEW (Waterhouse *et al*., 2009).

2.3.6 Sequence conservation of the ATPase domain in human HSP70s

The levels of sequence conservation in human HSP70s are high. The stretch of residues between ILE 172 to THR 177 (indicated within the first black squared box in Figure 16) were residues previously found to be involved in HSP70-HSP40 interactions (Jiang *et al*., 2007; Suh *et al*., 1999). These residues formed part of the under cleft pocket in HSP70 ATPase domain. Of note among these residues was the highly conserved GLU 175 across all the human HSP70 proteins, which has been previously proposed to be of catalytic importance in HSP40-HSP70 interaction (Jiang *et al*., 2007). Mutation of this residue abolished HSP40-HSP70 partnership (Jiang *et al*., 2007). Also of interest was the ARG residue at position 171 which corresponds to ARG 167 in *E. coli* proposed to be critical for DnaK:DnaJ interaction (Suh *et al*., 1998). However, this residue has been substituted with a PRO residue in HSPA7, ILE in HSPA12A and a LEU residue in HSPA12B. ILE 216, LEU 170, LEU 380, ILE 181, VAL 388 and LEU 393 have also been previously reported to be involved in HSP40-HSP70 interactions (Jiang *et al*., 2007; Suh *et al*., 1998). VAL 388 and LEU 393 were part of the hydrophobic residues at the linker region connecting the ATPase domain and the substrate binding domain proposed to be important for HSP70 interaction with the J domain of partner HSP40. Mutations of these linker residues reduced or abrogated J domain stimulation of the ATPase activity (Jiang *et al*., 2007). The highly conserved VAL 388 was replaced with THR and ASP residues respectively in HSPA5 and HSPA12A. Both of these residues were polar as opposed to VAL which is hydrophobic. HSPA5 is endoplasmic reticulum localized while HSPA12A is localized in the cytosol. Whereas the ATPase domains of HSP70 homologs were very similar, there exist minor differences that allowed for classification of the entire HSP70 family (Mayer and Bukau, 2005). Variations in the residues proposed to be at the exposed loop region in the subdomain IIB in the HSP70 ATPase domain structure near the nucleotide binding cleft are important in highlighting the subtle differences in HSP70 family (Mayer and Bukau, 2005). These stretch of residues includes SER 276 – ARG 302 in Bovine HSC70 which corresponds to ALA 276 – ARG 302 in *E.coli* DnaK with subfamily-specific sequence. Notable among these variations were the substitutions of the ILE 290 in HSPA8 which was replaced with GLU HSPA5, SER in HSPA9 and GLN in both HSPA13 and HSPA14. Previous study has shown that this residue is part of the residues in the loop region that constituted a device that allowed rapid association of ATP and slow dissociation of ATP to ADP and Pi (Mayer and Bukau, 2005). Also, the GLU 288 was replaced by MET 288 in HSPA9 (mitochondrial localized). This and many other variations of residues in this loop region were proposed to be involved in HSP70 nucleotide exchange rates (Mayer and Bukau, 2005).

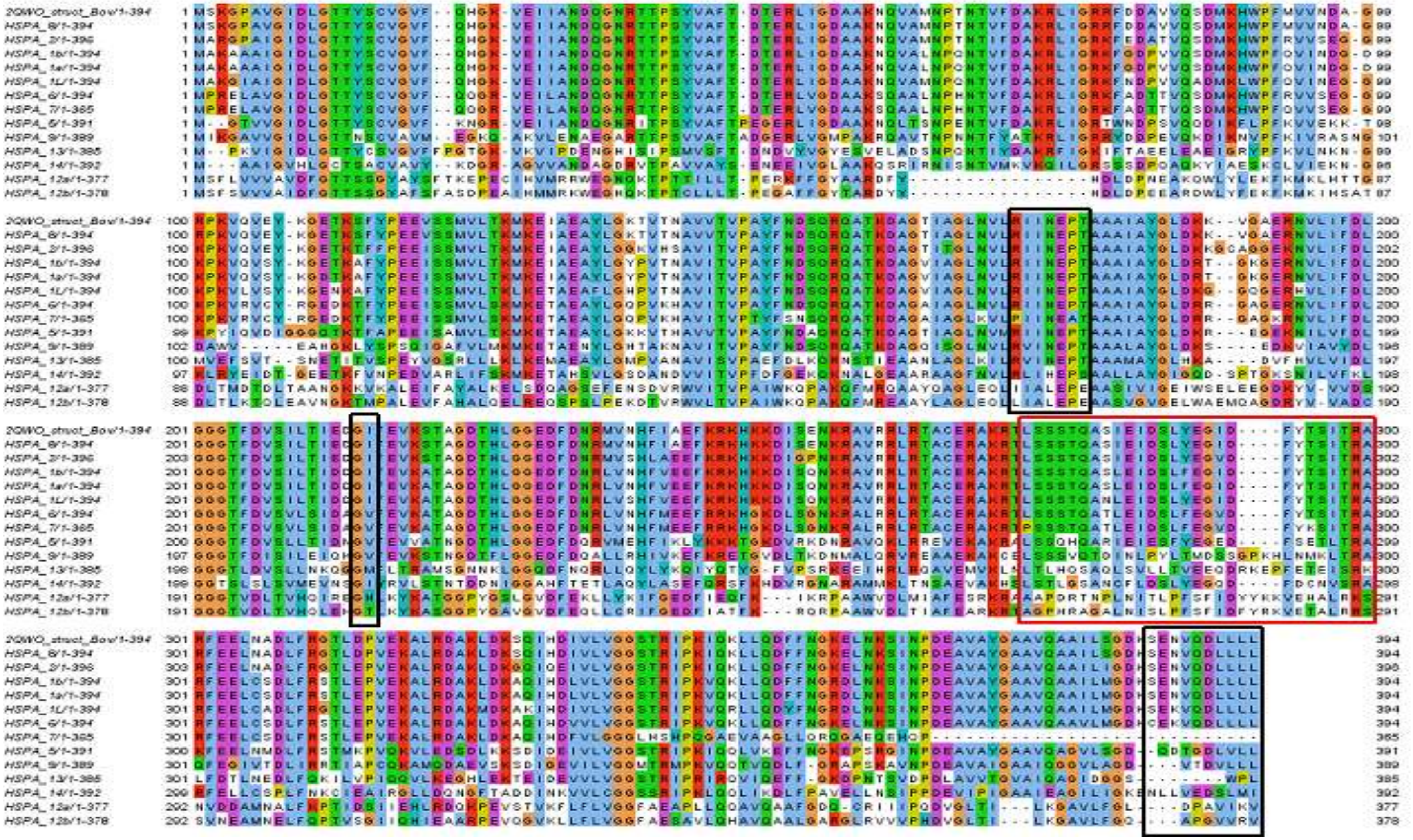


Figure 16: Multiple sequence alignment of HSP70 ATPase domain_linker. The structure sequence of HSPA8 (2QWO) was included in order to monitor the positions of the highly conserved residues on the protein. Residues highlighted within the black-coloured square brackets are part of the residues at the under cleft region of the ATPase domain and the linker region proposed to form a binding interface for HSP40 J domain interactions. Residues within the red-coloured square bracket showed regions of major variations within the HSP70s.

2.3.7 Analysis of the Phylogenetic inference of HSP40 Genes and HSP70 Complements

Homologous HSPs ought to cluster together and co-localised HSPs should share similar biochemical functions. Figure 17 – 19 represent the clustering patterns of the J domains from the phylogeny of Type I, II and III HSP40s respectively and Figure 21 showed the phylogeny of the HSP70 ATPase domains. Multiple sequence alignments of the HSP40 J domains and the HSP70 ATPase domains were performed using PROMALS3D and used for the phylogenetic analysis. 1000 bootstrap replicates of the J domain sequences for each types were computed using the best substitution model calculated prior to building the tree using MEGA5 as earliar discussed in the methodology. Overall, the statistical value obtained for the bootstrap consensus trees were low especially among the Type II and Type III J proteins (Figure 18 and 19) than observed in Type I. DNAJA3, a mitochondrial localized Type I member is the most divergent among the Type I proteins while the other members (DNAJA1, DNAJA2 and DNAJA4) (Figure 17), which were localized in the cytosol, were very similar and clustered together. Both DNAJA1 and DNAJA4 shared 98% identity. The same trend was observed among the Type II subfamily. Three major clusters were found within the phylogeny (Figure 18): (i) DNAJB6, DNAJB7 DNAJB3, DNAJB8, DNAJB2, (ii) DNAJB11, DNAJB12, DNAJB14 and (iii) DNAJB13, DNAJB1, DNAJB4, DNAJB5. Most of the proteins within the first major cluster were localized in either the cytosol or nucleus and shared 87% identity as seen in Figure 18. Whereas both DNAJB12 and DNAJB14 within the second cluster were predicted to be nucleus localized and shared 100% similarity, DNAJB11 has been experimentally shown to be localized in the endoplasmic reticulum (see Table 5) and share 49% identity with both DNAJB12 and DNAJB14. DNAJB9, predicted to be localized in endoplasmic reticulum, did not cluster with any other protein though was close to DNAJB11 with similar localization.

Although most of the Type III proteins also clustered according to their cellular localizations, some discrepancies were identified (Figure 19). High residue variations have been reported among the Type III proteins (Hennessy *et al*., 2000) and these could probably be responsible for the discrepancies observed using phylogeny. For instance, both DNAJC5 and DNAJC24 were experimentally shown and predicted respectively to be cytosolic (see Table 6) and clustered together with very low similarity value of 7%. DNAJC15 and DNAJC19 predicted to be localized in the mitochondrial clustered together with 97% and both DNAJC18 and DNAJC21 shared 40% identity and were predicted to be localized in the nucleus (see Table 6). However, while both DNAJC4 and DNAJC28 were predicted to be mitochondrial localized and shared a very low similarity value of 12%, they did not clustered with both DNAJC15 and DNAJC19 with similar localization. Also, while DNAJC3 and DNAJC7 clustered together with 14% identity. DNAJC3 has been experimentally shown to be localized in either the endoplasmic reticulum or the cytosol while DNAJC7 was predicted to be in the cytosol. This might probably account for their low similary value though the two proteins contained the tetratricopeptide (TRP) motif (see Table 9). Similarly, both DNAJC6 and DNAJC26 clustered together with 93% similarity (Figure 19). While DNAJC6 was both nucleus and cytosol localized, DNAJC26 was localized in the cytosol and have both been shown to perform similar functions in clathrin uncoating (Greener *et al*., 2000).

From the clustering pattern of the J domain alone across all human HSP40 family presented in Figure 20, a similar trend was observed in which most of the J domains clustered together based on their localizations and sub-family types. However, some of the J domains clustered together based on their cellular sub-localization regardless of their sub-family types. Of note are DNAJB6, DNAJB3, DNAJB7, DNAJB8, DNAJB2, DNAJC7 and DNAJC3 at the top of the phylogram in Figure 20 predicted to be localized in the cytosol/nucleus/ER but clustered together at the same evolutionary distance. Conversely, DNAJA3 (mitochondrial localized); DNAJB9 (endoplasmic reticulum localized) and *E. coli* DnaJ (cytoplasm localized) were all clustered in the same clade.

The clustering pattern of the HSP70 genes revealed that most of the HSP70 members that clustered together were localised in the cytosol/nucleus, represented in the square bracket in Figure 21, (HSPA1A, HSPA1B, HSPA1L, HSPA6, HSPA7, HSPA2 and HSPA8) as expected sharing 89% similarity (Figure 21) (see also Table 7). HSPA1A and HSPA1B shared 99% similarity while HSPA1L shared 45% identity with the two proteins. Both HSPA6 and HSPA7 clustered at 98% similarity and both HSPA2 and HSPA8 shared 82% level of similarity. Distantly related HSPA12A and HSPA12B clustered differently from the remaining proteins but the two isoforms shared 100% identity. HSPA5 (Mitochondrial localized) and HSPA9 (endoplasmic reticulum localized) clustered together with 56% similarity value though the two proteins were localized at different positions within the cell. HSPA13 (endoplasmic reticulum localized) clustered relatively close together with both HSPA5 and HSPA9 with 61% identity while HSPA14 did not cluster with any of the proteins though predicted to be localized in the cytosol.

The clustering pattern observed for the J domain of the different HSP40 types (Figure 17 - 20) and the HSP70 ATPase domain (Figure 21) were found to be very similar to the pattern obtained with the previous study of Hageman and Kampinga (2009) using the full length protein sequences. Most of the proteins that clustered together in the full length protein were also found clustered together in the phylogenetic tree using the J domain. Despite the sequence variations observed within the Type III proteins, some of them clustered together. For instance, DNAJC15 and DNAJC19 were clustered together with 97% similarity value (Figure 20) in both the J domain phylogeny as well as the full length protein sequence (data not shown). Both DNAJC15 and DNAC19 were predicted to be mitochondrial localized. Similarly, DNAJC4 and DNAJC28, predicted to be localized in mitochondria, clustered together with 25% similarity (Figure 20). Also, DNAJC6 and DNAJC26 clustered together significantly with 96% similarity (Figure 20) both in the phylogenetic tree based on their J domains and the full length proteins. The same trend was also seen within the Type I and II proteins. The fact that some of the proteins predicted to be localized in the same sub-cellular position were not clustered together in the phylogeny could mean that while some of the proteins were resident in some regions within the cell, they were being expressed at another location within the cell or probably catalysed similar functions while at different locations (Qiu *et al.*, 2006).

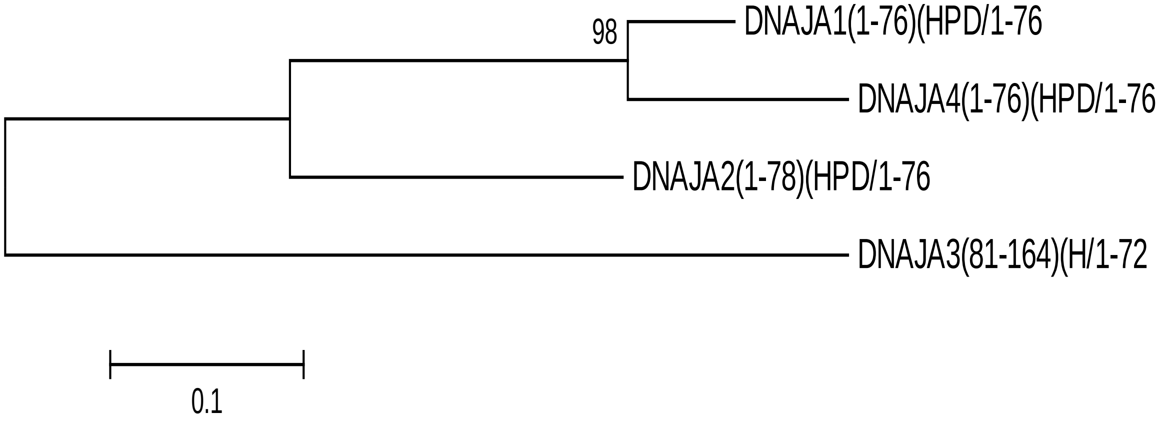


Figure 17: Molecular Phylogenetic analysis by Maximum Likelihood method for J domain of DNAJA proteins showing the bootstrap consensus tree. Whelan And Goldman (WAG) model (Whelan Liò and Goldman, 2001) of substitution was employed to calculates the evolutionary relationship among the proteins. 1000 bootsrap replicates (Felsenstein, 1985) were calculated and the level of relationships among the proteins were shown as percentage next to the branches. 4 amino acid sequences were analyzed and evolutionary analyses were investigated using MEGA5 (Tamura *et al.*, 2011).

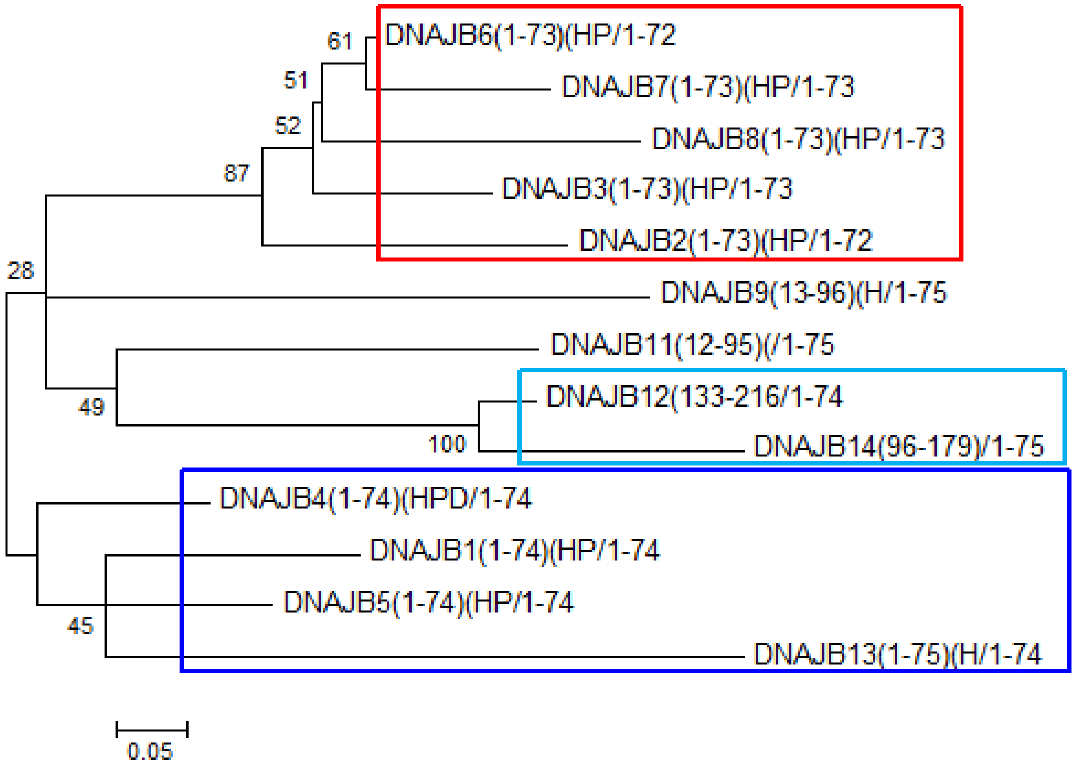


Figure 18: Molecular Phylogenetic analysis by Maximum Likelihood method for J domain of DNAJB proteins showing the bootstrap consensus tree. Proteins within the coloured square brackets clustered together in both trees based on full length protein sequences and J domain and also share similar subcellular localizations prediction. Whelan And Goldman (WAG) model (Whelan Liò and Goldman 2001) of substitution was employed to calculates the evolutionary relationship among the proteins. 1000 bootsrap replicates (Felsenstein, 1985) were calculated and the level of relationships among the proteins were shown as percentage next to the branches. 13 amino acid sequences were analyzed and evolutionary analyses were investigated using MEGA5(Tamura *et al*., 2011).

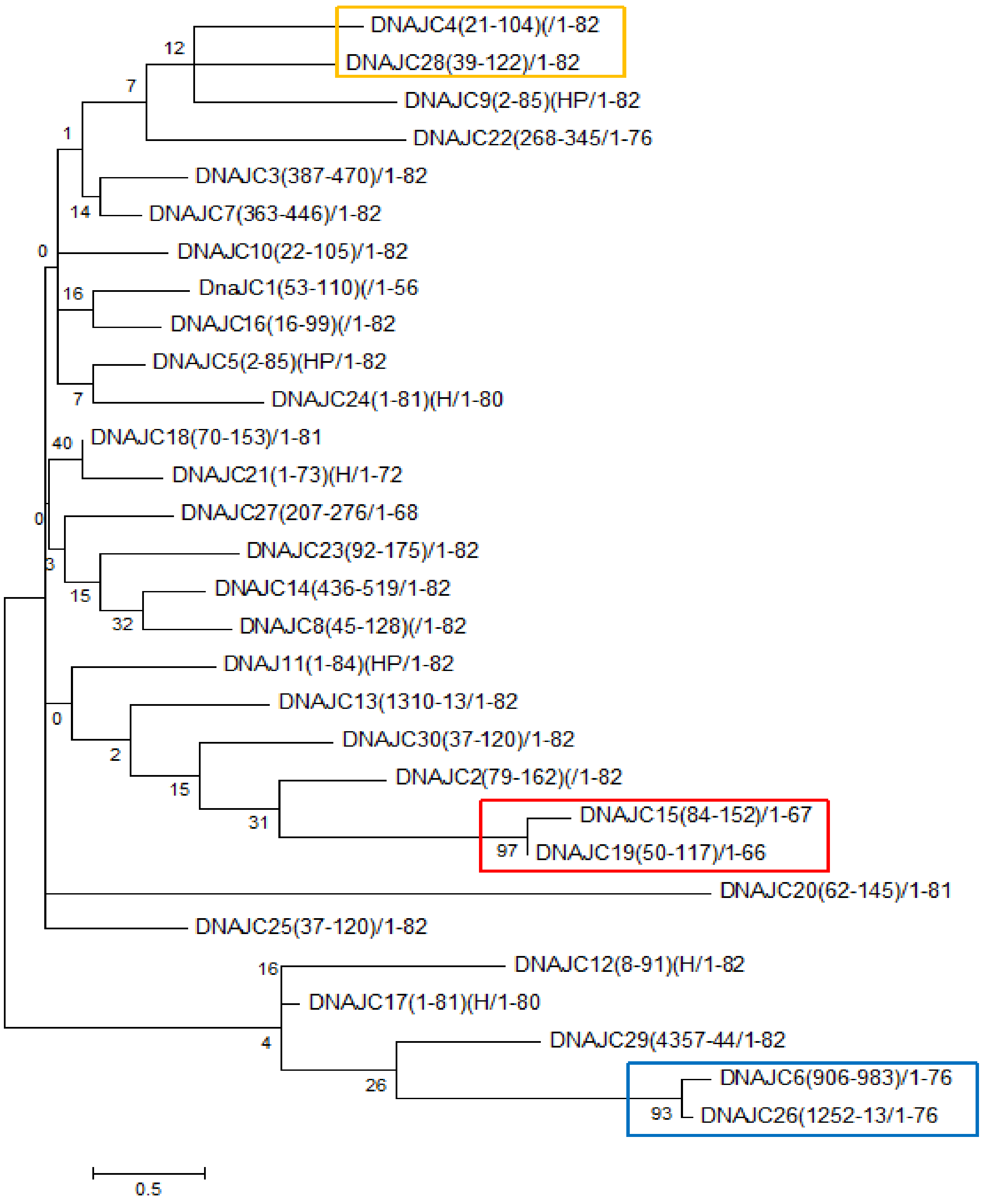


Figure 19: Molecular Phylogenetic analysis by Maximum Likelihood method for the J domain of DNAJC proteins showing the boostrap consensus tree. Proteins within the coloured square brackets clustered together in both trees based on full length protein sequences and J domains and also share similar subcellular localizations prediction except for DNAJC6 and DNAJC26 which were localized in the nucleus/cytosol and cytosol respectively. General Reverse Transcriptase (rtREV) model (Dimmic *et al*., 2002)of substitution was employed to calculates the evolutionary relationship among the proteins. 1000 bootsrap replicates (Felsenstein, 1985) were calculated and the level of relationships among the proteins were shown as percentage next to the branches. 30 amino acid sequences were analyzed and evolutionary analyses were investigated using MEGA5 (Tamura *et al*., 2011).

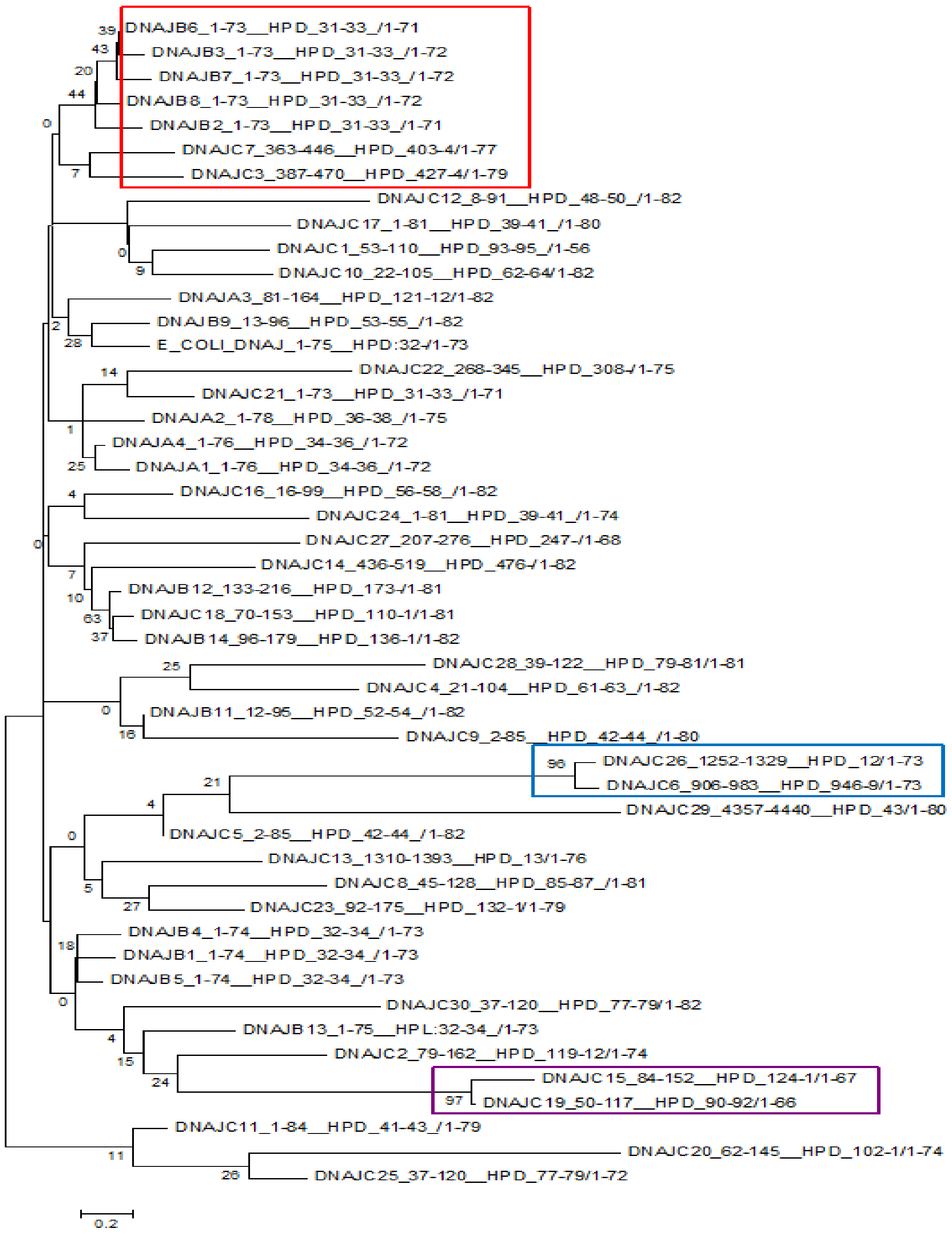

39
43
20
44
0
7
DNAJB6_1-73__HPD_31-33_/1-71
DNAJB3_1-73__HPD_31-33_/1-72
DNAJB7_1-73__HPD_31-33_/1-72
DNAJB8_1-73__HPD_31-33_/1-72
DNAJB2_1-73__HPD_31-33_/1-71
DNAJC7_363-446__HPD_403-4/1-77
DNAJC3_387-470__HPD_427-4/1-79
DNAJC12_8-91__HPD_48-50_/1-82
DNAJC17_1-81__HPD_39-41_/1-80
0
DNAJC1_53-110__HPD_93-95_/1-56
9
DNAJC10_22-105__HPD_62-64/1-82
DNAJA3_81-164__HPD_121-12/1-82
2
DNAJB9_13-96__HPD_53-55_/1-82
28
E_COLI_DNAJ_1-75__HPD:32-/1-73
14
DNAJC22_268-345__HPD_308-/1-75
DNAJC21_1-73__HPD_31-33_/1-71
DNAJA2_1-78__HPD_36-38_/1-75
1
DNAJA4_1-76__HPD_34-36_/1-72
25
DNAJA1_1-76__HPD_34-36_/1-72
4
DNAJC16_16-99__HPD_56-58_/1-82
DNAJC24_1-81__HPD_39-41_/1-74
0
DNAJC27_207-276__HPD_247-/1-68
7
DNAJC14_436-519__HPD_476-/1-82
10
DNAJB12_133-216__HPD_173-/1-81
63
DNAJC18_70-153__HPD_110-1/1-81
37
DNAJB14_96-179__HPD_136-1/1-82
25
DNAJC28_39-122__HPD_79-81/1-81
DNAJC4_21-104__HPD_61-63_/1-82
0
DNAJB11_12-95__HPD_52-54_/1-82
16
DNAJC9_2-85__HPD_42-44_/1-80
96
DNAJC26_1252-1329__HPD_12/1-73
21
DNAJC6_906-983__HPD_946-9/1-73
4
DNAJC29_4357-4440__HPD_43/1-80
0
DNAJC5_2-85__HPD_42-44_/1-82
DNAJC13_1310-1393__HPD_13/1-76
5
DNAJC8_45-128__HPD_85-87_/1-81
27
DNAJC23_92-175__HPD_132-1/1-79
DNAJB4_1-74__HPD_32-34_/1-73
18
DNAJB1_1-74__HPD_32-34_/1-73
DNAJB5_1-74__HPD_32-34_/1-73
0
DNAJC30_37-120__HPD_77-79/1-82
DNAJB13_1-75__HPL:32-34_/1-73
4
15
DNAJC2_79-162__HPD_119-12/1-74
24
DNAJC15_84-152__HPD_124-1/1-67
97
DNAJC19_50-117__HPD_90-92/1-66
DNAJC11_1-84__HPD_41-43_/1-79
11
DNAJC20_62-145__HPD_102-1/1-74
26
DNAJC25_37-120__HPD_77-79/1-72
0.2

Figure 20: Molecular Phylogenetic analysis by Maximum Likelihood method for HSP40 J-domains. Proteins within the colored square backets were clustered together regardless of their types, and also share similar sub-cellular localizations. General Reverse Transcriptase (rtREV) model (Dimmic *et al*., 2002) of substitution was employed to calculates the evolutionary relationship among the proteins. 1000 bootsrap replicates (Felsenstein, 1985) were calculated and the level of relationships among the proteins were shown as percentage next to the branches. 47 amino acid sequences were analyzed and evolutionary analyses were investigated using MEGA5 (Tamura *et al*., 2011).

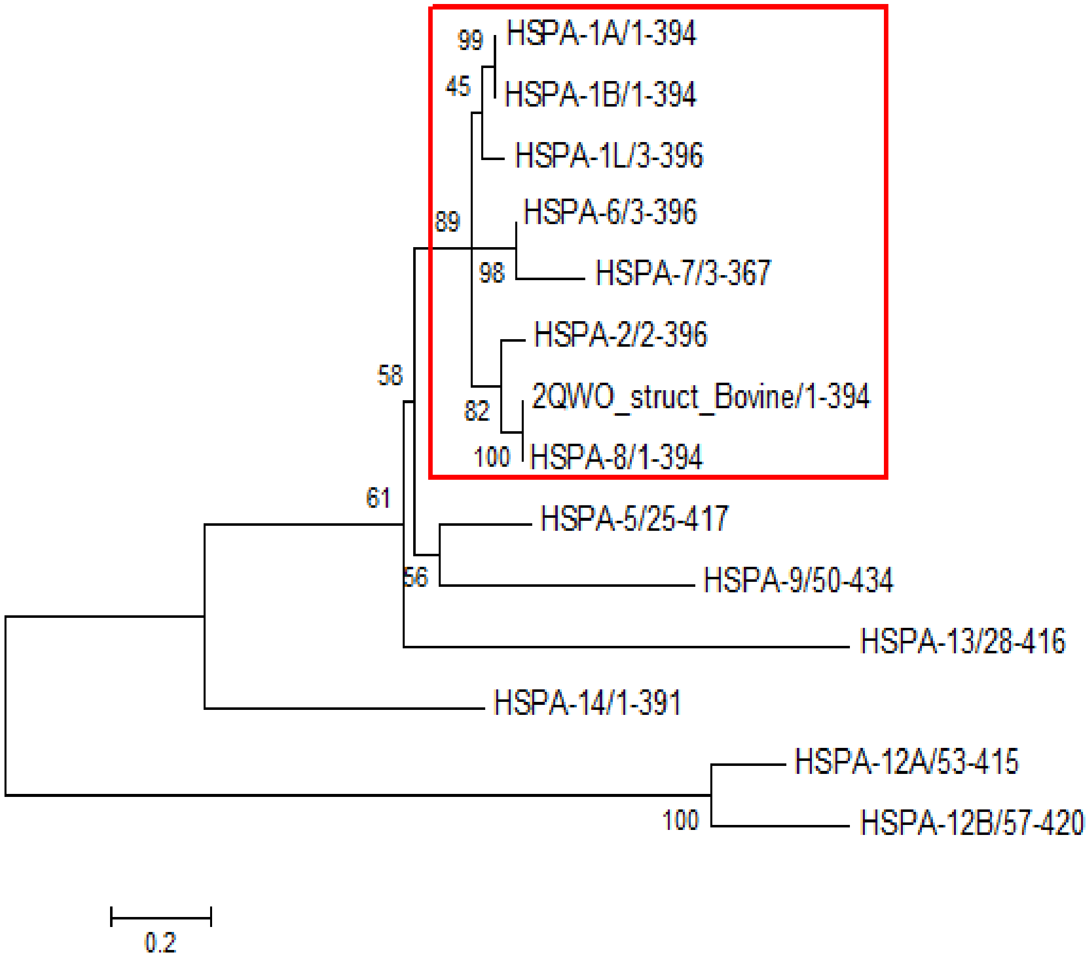


Figure 21: Molecular Phylogenetic analysis by Maximum Likelihood method for HSP70 ATPase domain_Linker. Proteins within the square bracket share similar localization predictions. Whelan And Goldman (WAG) model (Whelan Liò and Goldman 2001) of substitution was employed to calculates the evolutionary relationship among the proteins. 1000 bootsrap replicates (Felsenstein, 1985) were calculated and the level of relationships among the proteins were shown as percentage next to the branches. 14 amino acid sequences were analyzed and evolutionary analyses were investigated using MEGA5 (Tamura *et al*., 2011).

CHAPTER THREE: Homology Modelling of HSP40 J domain and HSP70 ATPase domain

3.1 Introduction

The concept of homology modelling is based on the observation that protein tertiary structure is better conserved than amino acid sequence. Therefore, proteins with appreciable diverse sequence identity but having a measure of sequence similarity that falls within the *safe zone* will also share common structural properties most especially in their folding (di Luccio and Koehl, 2011; Elmar Krieger and Sander Nabuurs, 2003). Experimental procedures, such as NMR spectroscopy and X-ray crystallography have been widely employed in determining protein structures. However, these procedures are time-consuming and are not completely free from various experimental errors and limitations for every protein of interest. Advances in genome sequencing technology have led to an exponential increase in the number of protein sequences available in various databases such as NCBI. Notwithstanding, the number of protein structures that have been characterized through experimental procedures and deposited in Protein Data Bank (PDB) are minimal compared to the available gene sequences (di Luccio and Koehl, 2011; Tastan Bishop *et al.*, 2008). Thus, there is a need for an *in silico* method to generate 3-D structures of protein to complement experimental techniques in order to bridge this gap (di Luccio and Koehl, 2011; Melo, 2007). *In silico* protein structure prediction can be sub-divided into three approaches, the *ab-initio* folding method, threading technique and homology modelling. Homology modelling involves the prediction of the 3-D structure of a given sequence (target) based on sequence similarity to one or more known protein structures (templates). If the percentage similarity between the target sequence to be modeled and the template sequence is detected, structural similarity can be assumed. On a general note, 30% sequence identity is required to generate a useful model (di Luccio and Koehl, 2011; Tastan Bishop *et al*., 2008). Homology modelling is of importance to applications including structure-guide design of mutagenesis experiments, design of *in vitro* test assays, structured-based prediction of drug metabolism and toxicity, functional information about protein-ligand complexes such as the location of the ligand, receptor active site residues and interactions with ligand if the protein is an enzyme (Tastan Bishop *et al.*, 2008). Steps in homology modelling can be divided into four major steps including (i) template identification, (ii) alignment (iii) model building and refinement, and (iv) model validation. Different types of computational software are available for the calculation of a homology model at the various stages including stand-alone programs such as; MODELLER, WHATIF as well as web-based servers like HHpred, SWISS MODEL to mention but a few (Tastan Bishop *et al*., 2008).

A good understanding of the 3-D structure of a protein will facilitate knowledge of its functional specificity and interactions (Faure *et al*., 2008). With the increase in the number of available crystal structures from experimental procedures, *in silico* approach through homology modelling is becoming a powerful tool to predict and study the 3-D structure of proteins. As opposed to the time-consuming and expensive experimental techniques such as X- ray crystallography and Nuclear Magnetic Resonance (NMR) which are fairly accurate and without which homology modelling could not be performed, homology modelling is cost effective with minimum time requirement in predicting the structure of a putative protein sequence from previously determined 3-D structure (Sahay and Shakya, 2010). It is based on the assumption that protein sequences that share minimum homology (sequence identity), usually greater than 30%, will possess similar structural properties (Wiltgen and Tilz, 2009; Tastan Bishop *et al.*, 2008). Since there is evidence to suggest sufficient similarity between protein sequences in the same family, accurate structural molecular models of proteins can be generated using homology modelling (Wiltgen and Tilz, 2009). A great challenge in homology modelling is the prediction of models with sub-optimum bond angle and length as opposed to having global minimum energy in all possible conformations (Tastan Bishop *et al*., 2008). Optimizing the predicted structure will allow it to possess a lower energy conformation which is more similar to its nascent protein geometry.

This chapter aimed to employ homology modelling to predict the 3-D structures of selected HSP40 J domains and HSP70 ATPase-linker regions with a view to gain insights for their possible interactions (see Chapter 4).

3.2 Methodology

3.2.1 Target Sequence and Template Structure Selection

Target sequences for homology modelling were selected based on previous report of interacting partners of HSP40 and HSP70 as presented in Table 10. Template search was performed using HHpred server for homology detection (http://toolkit.tuebingen.mpg.de/hhpred) (**Chapter 3,**

appendix I). HHpred is based on two search engines; HHsearch and HHblits. It employs hidden Markov's model (HMM) to searche for homologous proteins with known structures to the protein of interest from different protein databases (Hildebrand *et al*., 2009). The best top four templates for each target protein were selected from the HHpred search (**Chapter 3, appendix I**) and the best template was chosen for the homology modelling of each selected HSP40 J domains and HSP70 ATPase-linker domains respectively.

Table 10: Protein targets for homology modelling based on known HSP40-HSP70 interactions

HSP40	**HSP70**	**References**
DNAJA1	HSPA1B, HSPA8	(Imai *et al*., 2002), (Takayama *et al*., 1997)
DNAJA2, DNAJA3, DNAJC6	HSPA8	(Scheele *et al*., 2001), (Sarkar *et al*., 2001) (Jiang *et al*., 2007)
DNAJC2	HSPA14	(Otto *et al*., 2005)
DNAJA2, DNAJA3, DNAJC3, DNAJB11	HSPA1A	(Diefenbach & Kindl, 2000),(Sarkar *et al*., 2001) (Melville *et al*., 1999), (Lau *et al*., 2001),
DNAJC1, DNAJC10	HSPA5	(Chevalier *et al*., 2000), (Hellman *et al*., 1999)

3.2.2 Template Validation

An initial validation of the template structure prior to its use for model building was performed in order to ascertain its structural accuracy and quality. Various validation programs such PROCHECK (Laskowski *et al*., 1993), ANOLEA (Melo and Feytmans, 1998), PROSA II (Wiederstein and Sippl, 2007), QMEAN6 (Benkert *et al*., 2011; Arnold *et al*., 2006), METAMQAP (Pawlowski *et al*., 2008) and DFIRE2 (Yang and Zhou, 2008) were employed to access the quality of the template structures (Table 13 and 14). PROCHECK evaluates the stereochemical parameters of the 3-D structure of the template protein or model. These parameters includes; Ramachandran plot and a list of residue-residue values. These are generated from high resolution experimentally determined structures with which comparisons are made with the template structure. ANOLEA (Atomic Non-Local Environment Assessment) is based on the assessment of the energy of non-local interactions of heavy atoms within a protein structure and employs a very accurate and sensitive Atomic Mean Force Potential (AMFP) to calculate the non-local energy profile of a protein structure in evaluating its quality. PROSA II compares the Z score between a target and the structure of a template protein. The Z score of a protein represents the overall quality of the model and measures the deviation of the overall energy of the model with respect to random conformations of experimentally determined structures. Z score that is not within the range of characteristics for native proteins symbolises a bad structural model (Wiederstein and Sippl, 2007). QMEAN6 estimates the absolute model reliability of a model structure between 0 and 1. Protein structures with QMEAN value within this range are said to be error free. Its model quality assessment is derived from six different structural features descriptors including C-beta interaction energy, all-atom pairwise energy, solvation energy, torsion angle energy, secondary structure agreement and solvent accessibility agreement. The quality scores of each descriptor are expressed as Z-scores and compared to scores derived from the evaluation of high-resolution experimental structures from PDB. MetaMQAP as a meta-predictor is based on a multivariate regression model which employs scores from eight different model quality assessment programs with the regulation of some important parameters for the assessment of the local quality of models. It also calculates the absolute deviation (Ǻ) of individual C-α atom between the target model and the unknown true structure as well as the global deviation (expressed as a root mean square deviation and GDT_TS scores). GDT_TS scores above 40% and less than 90% symbolize a very good model while a GDT_TS value above 90% slightly decreases in model quality (greater than 10Å) (Pawlowski *et al*., 2008). Individual residue prediction accuracy is visualized as a colour in a spectrum between blue (predicted high accuracy) and red (predicted low accuracy) as presented in the B-factor column of the coordinate file. It is used to assign different confidence to regions that are of particular interest for the prediction of biological function of the modelled protein. DFIRE2 refers to distance-scaled, finite, ideal-gas reference state (DFIRE) which is an all-atom statistical energy function. It performs a global energy minimization of short unfolded segments having secondary structure as a direct test of the energy function of a protein. It employs an *ab initio* refolding method in assessing the energy function of unfolded segments of a protein structure while the other folded segments maintain their native conformations. It evaluates the accuracy of the refolded segments in terms of a local root-mean-squared distance (lrmsd), which is calculated by superposing the unfolded segment to that of native protein structures.

3.2.3 Template-Target Sequence Alignment

In this study, Multiple Sequence Alignment tool using Fast Fourier Transform (MAFFT) (Katoh and Frith, 2012) was employed for aligning the selected target HSP40 J domain and HSP70 ATPase domain sequences with the selected template sequences respectively. The fasta sequences of the selected template structures were retrieved from the Protein Data Bank (PDB). These sequences were aligned with the target protein sequences using MAFFT and the alignment results were viewed using JALVIEW (Waterhouse *et al.*, 2009) and saved as Protein Identification Resource (PIR) format for homology modelling. The results were compared with the alignment results using HHpred server (Figure 22) (see also **Chapter 3, appendix II**) in order to ascertain their accuracy.

3.2.4 Homology Model Building and Refinement

Once template structure has been identified and its sequenced properly aligned with the protein sequence of interest (target protein), the model building phase is the next crucial step. The PDB coordinate files of the best selected templates were retrieved from PDB and their coordinates were visualized using Discovery Studio as well as manually investigated in an editor (gedit) with respect to the alignment prior to modelling as some of the template residues are sometimes wrongly numbered. In such cases, the templates residues were renumbered using a python script; R*enumbering_all_files.py* (see electronic data/SCRIPTS). A stand-alone MODELLER version 9.7 (Sali, 2010) was employed in this study for building the homology models of the selected HSP40 J domains and HSP70 ATPase-linker regions. Python scripts, *homology.py and DOPE_Z_score.py*, (see electronic data/ SCRIPTS) were used to generate 100 models each of the target proteins as well as calculate and select the best model with the least DOPE Z score (i.e the model with the most negative DOPE Z score tends to be very similar to the native structure).

3.2.5 Homology Model Validation

Various quality assessment programs are available for checking the quality of a model ranging from the estimation of different stereochemical parameters such as bond angles, bond lengths, dihedral angles and residue planarity, to analysing the energy function of the protein such as the DOPE Z-score. In this study, the DOPE Z-score from MODELLER (Sánchez and Sali, 1997) using python scripts, the GTS_TS value and RSMD from MetaMQAP server (Pawlowski *et al.*, 2008), as well as the DFIRE2 total energy score (Yang and Zhou, 2008) were employed in assessing the quality of the homology built models. ANOLEA mean force potential (Francisco Melo, 2007) for each residue and Ramachandran plot using PROCHECK (Laskowski *et al.*, 1993) from the SWISS-MODEL (Arnold *et al.*, 2006) workspace server were also used in validating how reliable and realistic the models are (**Chapter 3, appendices V and VI**).

3.3 Results and Discussion

3.3.1 Template Search, Selection and Validation

The result of the template search and selection for homology modelling of selected HSP40 J domains and HSP70 ATPase domains are presented in Table 11 and 12 respectively. Five of the templates were crystal structures, while the remaining two (PDB ID: 2DN9 and IHDJ) were from NMR experiment. The resolution of the crystal structures was within the range of high-resolution (< 3Å). The BLAST E-value shows the likelihood that the sequence alignment result between the template sequence and the target sequence occurred by chance and randomly. Thus, a lower E-value is significant and suggests a high probability that the two proteins are similar (Wiltgen and Tilz, 2009). The E-value, sequence identity and alignment coverage between the templates and target sequences in this study (Table 11 and 12) were significant for building a good homology model. The template-target alignments result of selected HSP40 J domains and HSP70 ATPase domains for homology modelling are presented in Figure 22 and **Chapter 3, appendix II** respectively. Accurate and reliable models are often determined by the level of the sequence identity between the template sequence and the target sequence as well as the quality of the crystal structure. Template protein structures with high sequence identity to the target sequences, usually above 30%, will produce protein models with high structural and functional quality in comparison to high resolution experimentally determined protein structures (Wiltgen and Tilz, 2009). Other important parameters to consider in choosing a good template for homology modelling includes: availability of crystal structure with high resolution (usually lower than 3Å), as well as a maximum alignment coverage length between the target protein sequence and the template protein structure. The template structure should also be checked if in complex with any ligand or not from the PDB (Tastan Bishop *et al.*, 2008). An appropriate template sequence-target sequence alignment will enhance the building of an accurate model. Inability to choose the most suitable template structure as well as an incorrect alignment between the target and the template, are mainly the most common source of errors encountered during homology modelling

(Arnold *et al*., 2006). Thus, manual inspection of the template-target alignment has been recommended for building models with good structural accuracy and meaningful biological functions (Tastan Bishop *et al*., 2008).

The results of the quality assessment analysis for the various templates used for homology modelling are shown in Table 13 for HSP40s as well as Table 14 for HSP70s. The GDS_TS score, RMSD (according to METAMQ) and DOPE Z score value fall within the range of reliable experimental native crystal structures. Preliminary quality assessment and validation of the template structure is necessary since experimental techniques are not completely error free (Wiederstein and Sippl, 2007). This will not only enhance the quality of the predicted models but serves as a quick check in identifying the source of any problem encountered during homology modelling.

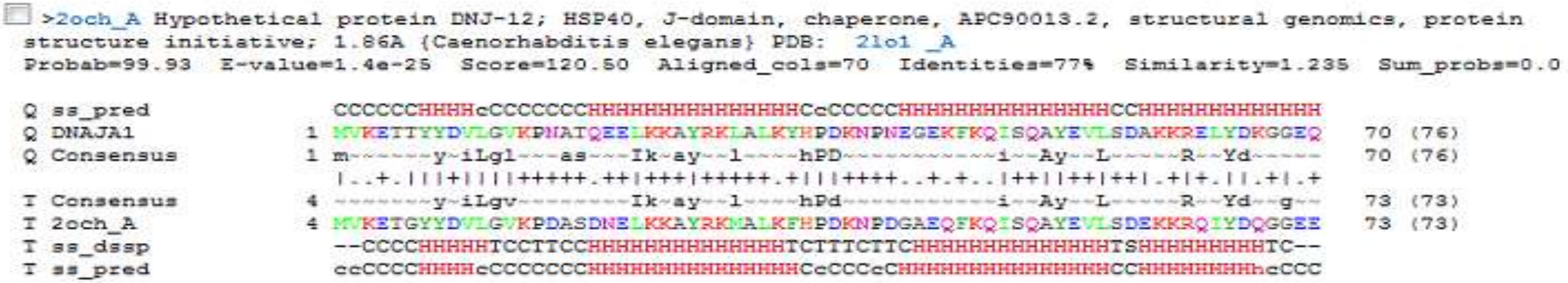
Table 11: Template selection for homology modelling of HSP40s using HHpred server

Target		**Template**					
Protein	**Alignment coverage**	**PDB ID**	**Organism**	**Resolution (A)**	**E-value**	**Sequence Identity**	**Alignment coverage**
DNAJA1	1 – 70 (76)	**2OCH**	*C.elegans*	1.86	$1.4e^{-25}$	77%	4 – 73 (73)
DNAJA2	4 – 72 (78)	**2OCH**	*C.elegans*	1.86	$5.9e^{-24}$	61%	5 – 73 (73)
DNAB11	6 – 82 (82)	**2DN9**	*H.sapiens*	-	$1.6e^{-24}$	60%	1 – 77 (79)
DNAJC2	3 – 82 (82)	**1HDJ**	*H.sapiens*	-	$1.5e^{-22}$	41%	1 – 71 (88)
DNAJC3	1 – 79 (82)	**2Y4T**	*H.sapiens*	3.00	$1.2e^{-13}$	100%	372 – 450 (450)
DNAJC6	1 – 69 (76)	**2QWO**	*B.taurus*	1.70	$3.3e^{-23}$	99%	23 – 91 (92)
DNAJC10	12 – 82 (82)	**3APQ**	*M.musculus*	1.84	$2.7e^{-21}$	99%	2 – 72 (210)
DNAJC19	1 – 66 (66)	**2GUZ**	*S.cerevisiae*	2.00	$1.1e^{-21}$	56%	4 – 70 (71)

Table 12: Template selection for homology modelling of HSP70 ATPase-linker region using HHpred server

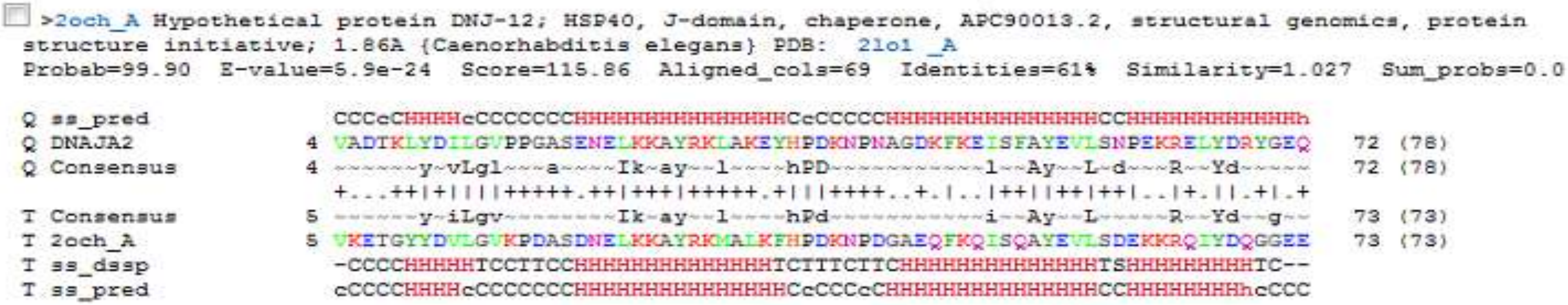
Target		Template					
Protein	**Alignment coverage**	**PDB ID**	**Organism**	**Resolution (A)**	**E-value**	**Sequence Identity**	**Alignment coverage**
HSPA-IA	1 – 453 (453)	1YUW	*B.taurus*	2.60	7.2e^{-79}	89%	1–453 (554)
HSPA-IB	1 – 453 (453)	1YUW	*B.taurus*	2.60	1.6e^{-78}	89%	1–453 (554)
HSPA5	26– 406 (453)	3QFU	*S.cerevisiae*	1.80	8.7e^{-58}	72%	16–394 (394)
HSPA8	1 – 453 (453)	1YUW	*B.taurus*	2.60	1.0e^{-75}	100%	1–453 (554)
HSPA14	1 – 381 (453)	3I33	*H.sapiens*	1.30	9.6e^{-62}	40%	23–404 (404)

1



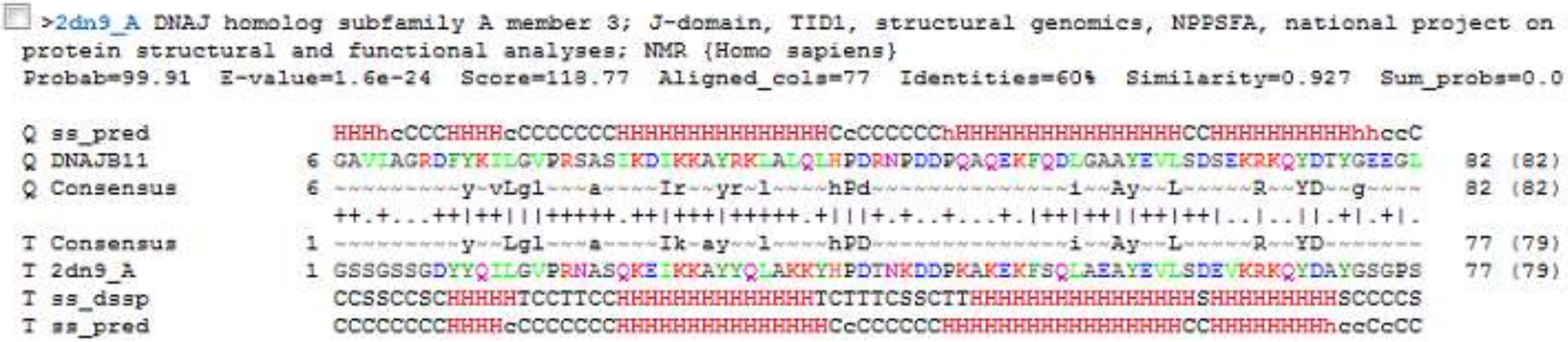
```
☐ >2och_A Hypothetical protein DNJ-12; HSP40, J-domain, chaperone, APC90013.2, structural genomics, protein
 structure initiative; 1.86A {Caenorhabditis elegans} PDB:  2lo1 _A
 Probab=99.93  E-value=1.4e-25  Score=120.50  Aligned_cols=70  Identities=77%  Similarity=1.235  Sum_probs=0.0

 Q ss_pred                CCCCCCHHHHcCCCCCCCHHHHHHHHHHHHHHHCcCCCCCHHHHHHHHHHHHHHHCCHHHHHHHHHHHHH
 Q DNAJA1               1 MVKETTYYDVLGVKPNATQEELKKAYRKLALKYHPDKNPNEGEKFKQISQAYEVLSDAKKRELYDKGGEQ   70 (76)
 Q Consensus            1 m~~~~~~y~iLgl~~~as~~~Ik~ay~~l~~~~hPD~~~~~~~~~~~~i~~Ay~~L~~~~~R~~Yd~~~~~   70 (76)
                          |..+.|||+||||+++++.++|+++|+++++.+|||++++..+.+..|++||++|++|.+|+.||.+|.+
 T Consensus            4 ~~~~~~~y~iLgv~~~~~~~~Ik~ay~~l~~~~hPd~~~~~~~~~~~~i~~Ay~~L~~~~~R~~Yd~~g~~   73 (73)
 T 2och_A               4 MVKETGYYDVLGVKPDASDNELKKAYRKMALKFHPDKNPDGAEQFKQISQAYEVLSDEKKRQIYDQGGEE   73 (73)
 T ss_dssp                --CCCCHHHHHTCCTTCCHHHHHHHHHHHHHHTCTTTCTTCHHHHHHHHHHHHHHTSHHHHHHHHHTC--
 T ss_pred                ccCCCCHHHHcCCCCCCCHHHHHHHHHHHHHHHCcCCCcCHHHHHHHHHHHHHHHCCHHHHHHHHhcCCC
```

2



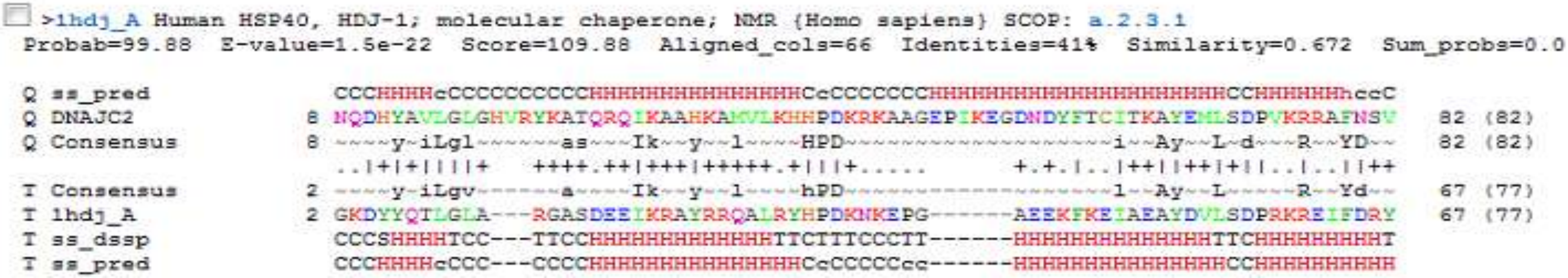
```
☐ >2och_A Hypothetical protein DNJ-12; HSP40, J-domain, chaperone, APC90013.2, structural genomics, protein
 structure initiative; 1.86A {Caenorhabditis elegans} PDB:  2lo1 _A
 Probab=99.90  E-value=5.9e-24  Score=115.86  Aligned_cols=69  Identities=61%  Similarity=1.027  Sum_probs=0.0

 Q ss_pred                CCCcCHHHHcCCCCCCCHHHHHHHHHHHHHHHCcCCCCCHHHHHHHHHHHHHHHCCHHHHHHHHHHHHh
 Q DNAJA2               4 VADTKLYDILGVPPGASENELKKAYRKLAKEYHPDKNPNAGDKFKEISFAYEVLSNPEKRELYDRYGEQ   72 (78)
 Q Consensus            4 ~~~~~~y~vLgl~~~a~~~~Ik~ay~~l~~~~hPD~~~~~~~~~~~~l~~Ay~~L~d~~~R~~Yd~~~~~   72 (78)
                          +...++|+||||+++++.++|+++|+++++.+|||++++..+.|..|++||++|++|..|+.||.+|.+
 T Consensus            5 ~~~~~~y~iLgv~~~~~~~~Ik~ay~~l~~~~hPd~~~~~~~~~~~~i~~Ay~~L~~~~~R~~Yd~~g~~   73 (73)
 T 2och_A               5 VKETGYYDVLGVKPDASDNELKKAYRKMALKFHPDKNPDGAEQFKQISQAYEVLSDEKKRQIYDQGGEE   73 (73)
 T ss_dssp                -CCCCHHHHHTCCTTCCHHHHHHHHHHHHHHTCTTTCTTCHHHHHHHHHHHHHHTSHHHHHHHHHTC--
 T ss_pred                cCCCCHHHHcCCCCCCCHHHHHHHHHHHHHHHCcCCCcCHHHHHHHHHHHHHHHCCHHHHHHHHhcCCC
```

3



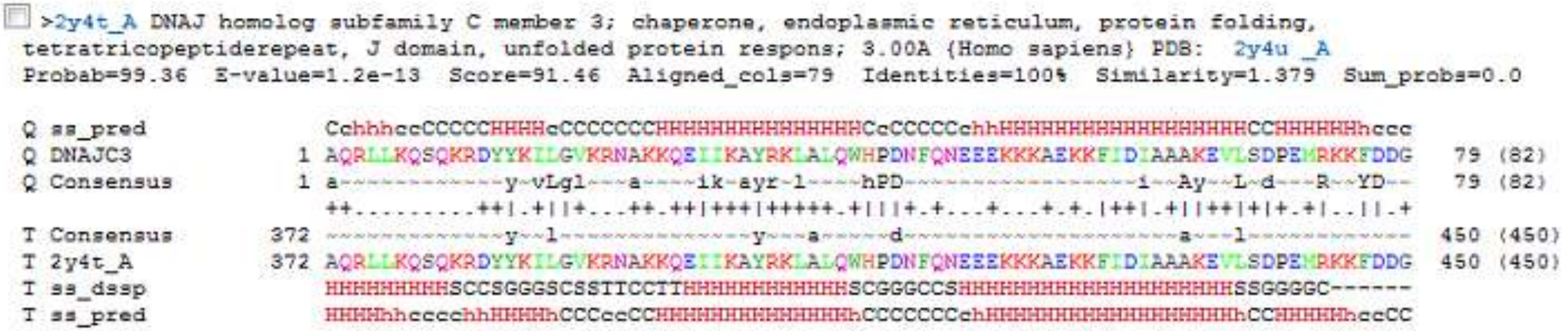
```
□ >2dn9_A DNAJ homolog subfamily A member 3; J-domain, TID1, structural genomics, NPPSFA, national project on
protein structural and functional analyses; NMR {Homo sapiens}
 Probab=99.91  E-value=1.6e-24  Score=118.77  Aligned_cols=77  Identities=60%  Similarity=0.927  Sum_probs=0.0

 Q ss_pred                HHHhcCCCHHHHcCCCCCCCHHHHHHHHHHHHHHHCcCCCCCCChHHHHHHHHHHHHHHHHCCHHHHHHHHHHhhccC
 Q DNAJB11             6 GAVIAGRDFYKILGVPRSASIKDIKKAYRKLALQLHPDRNPDDPQAQEKFQDLGAAYEVLSDSEKRKQYDTYGEEGL   82 (82)
 Q Consensus           6 ~~~~~~~~~y~vLgl~~~a~~~~Ir~~yr~l~~~~hPd~~~~~~~~~~~~~~~~i~~Ay~~L~~~~~R~~YD~~g~~~~   82 (82)
                         ++.+...++|++|||+++++.++|+++|+++++.+|||+.+..+...+.|++|++||++|++|..|..||.+|.+|.
 T Consensus           1 ~~~~~~~~~y~~Lgl~~~a~~~~Ik~ay~~l~~~~hPD~~~~~~~~~~~~~~~~i~~Ay~~L~~~~~R~~YD~~~~~~~~   77 (79)
 T 2dn9_A              1 GSSGSSGDYYQILGVPRNASQKEIKKAYYQLAKKYHPDTNKDDPKAKEKFSQLAEAYEVLSDEVKRKQYDAYGSGPS   77 (79)
 T ss_dssp               CCSSCCSCHHHHHTCCTTCCHHHHHHHHHHHHHHHTCTTTCSSCTTHHHHHHHHHHHHHHHHSHHHHHHHHSCCCCS
 T ss_pred               CCCCCCCCHHHHcCCCCCCCHHHHHHHHHHHHHHHCcCCCCCCHHHHHHHHHHHHHHHHHCCHHHHHHHHhccCcCC
```

4



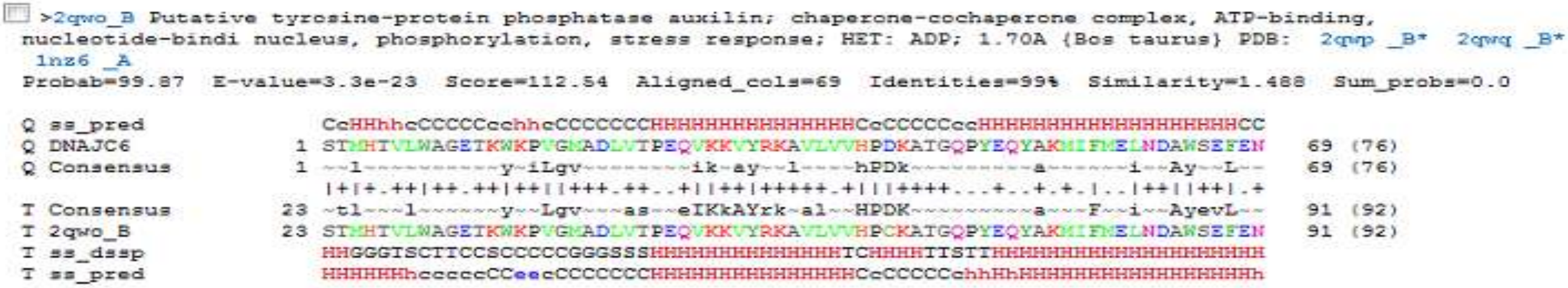
```
□ >1hdj_A Human HSP40, HDJ-1; molecular chaperone; NMR {Homo sapiens} SCOP: a.2.3.1
 Probab=99.88  E-value=1.5e-22  Score=109.88  Aligned_cols=66  Identities=41%  Similarity=0.672  Sum_probs=0.0

 Q ss_pred                CCCHHHHcCCCCCCCCCCHHHHHHHHHHHHHHHCcCCCCCCCHHHHHHHHHHHHHHHHHHHHCCHHHHHHHhccC
 Q DNAJC2              8 NQDHYAVLGLGHVRYKATQRQIKAAHKAMVLKHHPDKRKAAGEPIKEGDNDYFTCITKAYEMLSDPVKRRAFNSV   82 (82)
 Q Consensus           8 ~~~~y~iLgl~~~~~~~as~~~Ik~~y~~l~~~~HPD~~~~~~~~~~~~~~~~~~~~~~i~~Ay~~L~d~~~R~~YD~~   82 (82)
                         ..|+|+||||+   ++++.++|+++|+++++.+|||+.....      +.+.|..|++||++|+||..|..||++
 T Consensus           2 ~~~~y~iLgv~~~~~~~a~~~~Ik~~y~~l~~~~hPD~~~~~~~~~~~~~~~~~~~~~~l~~Ay~~L~~~~~R~~Yd~~   67 (77)
 T 1hdj_A              2 GKDYYQTLGLA---RGASDEEIKRAYRRQALRYHPDKNKEPG------AEEKFKEIAEAYDVLSDPRKREIFDRY   67 (77)
 T ss_dssp               CCCSHHHHTCC---TTCCHHHHHHHHHHHHHTTCTTTCCCTT------HHHHHHHHHHHHHTTCHHHHHHHHHHT
 T ss_pred               CCCHHHHcCCC---CCCCHHHHHHHHHHHHHHHCcCCCCCcc------HHHHHHHHHHHHHHCCHHHHHHHHHHH
```

5

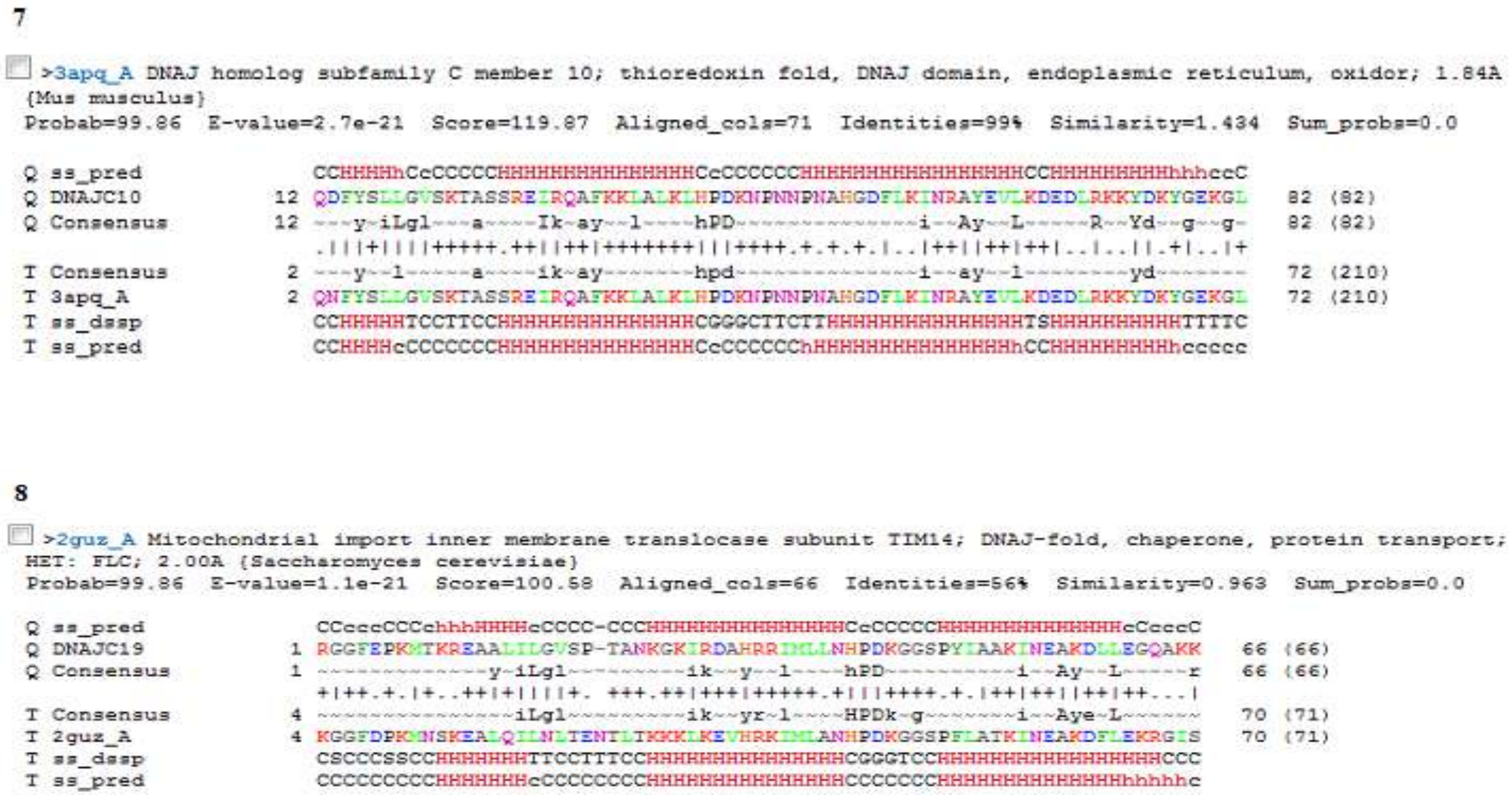


```
☐ >2y4t_A DNAJ homolog subfamily C member 3; chaperone, endoplasmic reticulum, protein folding,
 tetratricopeptiderepeat, J domain, unfolded protein respons; 3.00A {Homo sapiens} PDB:  2y4u _A
 Probab=99.36  E-value=1.2e-13  Score=91.46  Aligned_cols=79  Identities=100%  Similarity=1.379  Sum_probs=0.0

 Q ss_pred            CchhhccCCCCCHHHHcCCCCCCCHHHHHHHHHHHHHHHCcCCCCCchhHHHHHHHHHHHHHHHHHHCCHHHHHHhccc
 Q DNAJC3          1 AQRLLKQSQKRDYYKILGVKRNAKKQEIIKAYRKLALQWHPDNFQNEEEKKKAEKKFIDIAAAKEVLSDPEMRKKFDDG   79 (82)
 Q Consensus       1 a~~~~~~~~~~~~y~vLgl~~~a~~~~ik~ayr~l~~~~hPD~~~~~~~~~~~~~~~~~~~~~i~~Ay~~L~d~~~R~~YD~~   79 (82)
                     ++.........++|.+||+...++.++|+++|+++++.+|||+.+...+...+.+.|++|.+||++|+|+.+|..||.+
 T Consensus     372 ~~~~~~~~~~~~~~y~~l~~~~~~~~~~~~~~y~~~a~~~~~d~~~~~~~~~~~~~~~~~~~~~a~~~l~~~~~~~~~~~~~~~  450 (450)
 T 2y4t_A        372 AQRLLKQSQKRDYYKILGVKRNAKKQEIIKAYRKLALQWHPDNFQNEEEKKKAEKKFIDIAAAKEVLSDPEMRKKFDDG  450 (450)
 T ss_dssp            HHHHHHHHHSCCSGGGSCSSTTCCTTHHHHHHHHHHHHSCGGGCCSHHHHHHHHHHHHHHHHHHHHSSGGGGC------
 T ss_pred            HHHHhhcccchhHHHHhCCCccCCHHHHHHHHHHHHHHhCCCCCCCchHHHHHHHHHHHHHHHHHHhCCHHHHHhccCC
```

6



```
☐ >2qwo_B Putative tyrosine-protein phosphatase auxilin; chaperone-cochaperone complex, ATP-binding,
 nucleotide-bindi nucleus, phosphorylation, stress response; HET: ADP; 1.70A {Bos taurus} PDB:  2qwp _B*  2qwq _B*
  1nz6 _A
 Probab=99.87  E-value=3.3e-23  Score=112.54  Aligned_cols=69  Identities=99%  Similarity=1.488  Sum_probs=0.0

 Q ss_pred           CcHHhhcCCCCCcchhcCCCCCCCHHHHHHHHHHHHHHHCcCCCCCccHHHHHHHHHHHHHHHHHHHCC
 Q DNAJC6         1 STMHTVLWAGETKWKPVGMADLVTPEQVKKVYRKAVLVVHPDKATGQPYEQYAKMIFMELNDAWSEFEN   69 (76)
 Q Consensus      1 ~~l~~~~~~~~~~~~~y~iLgv~~~~~~~~~~ik~ay~~l~~~~hPDk~~~~~~~~~~~~a~~~~~~~i~~Ay~~L~~   69 (76)
                    |+|+.++|++.++|++||+++.++..+||++|+++++.+|||++++...+..+.+.|..|++||++|.+
 T Consensus     23 ~tl~~~l~~~~~~~~y~~Lgv~~~as~~eIKkAYrk~al~~HPDK~~~~~~~~~~~~a~~~F~~i~~AyevL~~   91 (92)
 T 2qwo_B       23 STMHTVLWAGETKWKPVGMADLVTPEQVKKVYRKAVLVVHPCKATGQPYEQYAKMIFMELNDAWSEFEN   91 (92)
 T ss_dssp          HHGGGTSCTTCCSCCCCCGGGSSSHHHHHHHHHHHHHHTCHHHHTTSTTHHHHHHHHHHHHHHHHHHHH
 T ss_pred          HHHHHHhcccccCCeecCCCCCCCHHHHHHHHHHHHHHHCcCCCCCchhHhHHHHHHHHHHHHHHHHHh
```



7

```
☐ >3apq_A DNAJ homolog subfamily C member 10; thioredoxin fold, DNAJ domain, endoplasmic reticulum, oxidor; 1.84A
 {Mus musculus}
 Probab=99.86  E-value=2.7e-21  Score=119.87  Aligned_cols=71  Identities=99%  Similarity=1.434  Sum_probs=0.0

 Q ss_pred               CCHHHHhCcCCCCCHHHHHHHHHHHHHHHCcCCCCCCHHHHHHHHHHHHHHHHHHCCHHHHHHHHHhhhccC
 Q DNAJC10         12 QDFYSLLGVSKTASSREIRQAFKKLALKLHPDKNPNNPNAHGDFLKINRAYEVLKDEDLRKKYDKYGEKGL   82 (82)
 Q Consensus       12 ~~~y~iLgl~~~a~~~~~Ik~ay~~l~~~~hPD~~~~~~~~~~~~~~~~i~~Ay~~L~~~~~~R~~Yd~~g~~g~   82 (82)
                       .|||+||||+++++.++||++|+++++++|||++++.+.+.+.|..|++||++|++|..|..||.+|..|+
 T Consensus        2 ~~~y~~l~~~~~a~~~~~ik~ay~~~~~~~~hpd~~~~~~~~~~~~~~~i~~ay~~l~~~~~~~~~yd~~~~~~~   72 (210)
 T 3apq_A           2 QNFYSLLGVSKTASSREIRQAFKKLALKLHPDKNPNNPNAHGDFLKINRAYEVLKDEDLRKKYDKYGEKGL   72 (210)
 T ss_dssp             CCHHHHHTCCTTCCHHHHHHHHHHHHHHHCGGGCTTCTTHHHHHHHHHHHHHHHTSHHHHHHHHHHTTTTC
 T ss_pred              CCHHHHcCCCCCCCHHHHHHHHHHHHHHHCcCCCCCChHHHHHHHHHHHHHHHhCCHHHHHHHHHhccccc
```

8

```
☐ >2guz_A Mitochondrial import inner membrane translocase subunit TIM14; DNAJ-fold, chaperone, protein transport;
 HET: FLC; 2.00A {Saccharomyces cerevisiae}
 Probab=99.86  E-value=1.1e-21  Score=100.58  Aligned_cols=66  Identities=56%  Similarity=0.963  Sum_probs=0.0

 Q ss_pred               CCcccCCCchhhHHHHcCCCC-CCCHHHHHHHHHHHHHHCcCCCCCHHHHHHHHHHHHHHcCcccC
 Q DNAJC19          1 RGGFEPKMTKREAALILGVSP-TANKGKIRDAHRRIMLLNHPDKGGSPYIAAKINEAKDLLEGQAKK   66 (66)
 Q Consensus        1 ~~~~~~~~~~~~~~~y~iLgl~~~~~~~~~~~ik~~y~~l~~~~hPD~~~~~~~~~~~~i~~Ay~~L~~~~~~r   66 (66)
                       +|++.+.|+..++|+||||+. +++.++|+++|+++++++.+|||++++.+.|++|++||++|++...|
 T Consensus        4 ~~~~~~~~~~~~~~~~~iLgl~~~~~~~~~~~ik~~yr~l~~~~HPDk~g~~~~~~~~i~~Aye~L~~~~~~~~   70 (71)
 T 2guz_A           4 KGGFDPKMNSKEALQILNLTENTLIKKKLKEVHRKIMLANHPDKGGSPFLATKINEAKDFLEKRGIS   70 (71)
 T ss_dssp             CSCCCSSCCHHHHHHHTTCCTTTCCHHHHHHHHHHHHHHHCGGGTCCHHHHHHHHHHHHHHHHHCCC
 T ss_pred              CCCCCCCCCHHHHHHHcCCCCCCCCHHHHHHHHHHHHHHHCCCCCCCHHHHHHHHHHHHHHhhhhhc
```

Figure 22: Template selection, alignment and secondary structure prediction of targeted HSP40 J domains using HHpred. (1) – (2) 2OCH (3) 2DN9 (4) IHDJ (5) 2Y4T (6) 2QWO (7) 3APQ (8) 2GUZ respectively. Sellected templates were used for predicting the model structures of the target proteins using homology modeling.

Table 13: Template validation using different structure assessment programs for homology modelling of HSP40 J domains

Template	ProSA Z-score	Q-mean6 Score	GDT_TS	RMSD	DFIRE2 Energy
2QWO_B	-6.85	0.758	81.250	1.437	-158.113
2GUZ_A	-5.73	0.944	71.127	3.136	-94.634
2DN9	-6.50	1.018	58.861	3.345	-103.301
3APQ_B	-5.87	0.765	70.548	1.935	-106.777
2OCH	-6.30	0.951	-	-	-92.979
1HDJ	-5.85	0.835	55.519	2.788	-104.351
2Y4T_A	-4.20	0.781	33.929	5.257	-106.589

Table 14: Template validation using different structure assessment programs for homology modelling of HSP70 ATPase domains

Template	ProSA Z-score	Q-mean6 Score	GDT_TS	RMSD	DFIRE2 Energy
1YUW	-11.05	0.755	46.255	4.290	-888.463
3QFU	-11.36	0.761	49.472	3.628	-657.85
3I33	-11.53	0.719	-	-	-607.532

3.3.2 Homology Model Validation

Assessing the overall quality and how realistic both experimental and theoretical models of protein structures are, remains an important procedure in order to check and ascertain the structural accuracy of such models, whether they are of any biological significance as well as check for potential errors (Wiederstein and Sippl, 2007). The quality of homology model is evaluated by comparing its geometry with that of well-defined native high-resolution crystal structures in protein structure databases such as PDB (Arnold *et al*., 2006). A total of 100 models were built for each selected HSP40 J domains and HSP70s. The most reliable model was chosen for each protein based of their DOPE Z score i.e, model with the lowest DOPE Z score shares the highest similarity to the native structure (Table 15) (**Chapter 3,** a**ppendices III & IV**). Z score shows the overall model quality of a protein structure and measures how close or distant apart, the total energy of a predicted model structure is with respect to an energy distribution of native proteins derived from random conformations. The DOPE Z score is calculated from the statistics of the raw DOPE scores computed using a python script in MODELLER (Sali, 2010). Negative scores of -1 or below are usually a measure of accurate and reliable models similar to native structures. As can be seen in the result presented in Table 15, the predicted models were accurate and reliable when compared to the structures of the native proteins. The RMSD value between the template and the predicted model is below 1 showing a higher similarity between the two structures. The GDS_TS value above 40% indicated that the predicted model was realistic and the RSMD value below 3.5Å correlated with native crystal structures having high resolution (Pawlowski *et al*., 2008). Overall, the GDS_TS value of the predicted models was above 40% except for DNAJC2 having an average of 25%. This was expected since the sequence identity between the template structure (1HDJ) as shown in the Figure 22 (4) is the lowest (41%) and a gap was also present in the template structure around the loop region. Structural information of this segment was omitted by modeller in building the model since residues of this region are missing in the template PDB file. The high level of sequence variation in the Type III HSP40 especially in the loop region and helix III (Hennessy *et al*., 2000) has made it difficult to identify a perfect template with high sequence identity for the proteins. However, sequence identity of 41% is good enough to build a theoretically accurate model. 1HDJ is a crystal structure for DNAJB1, the first member protein of the Type II HSP40 family and shares the highest sequence identity with DNAJC2 based on HHpred search as shown in Figure 22. Interestingly, the percentage of residues in the most favoured regions in the Ramachandran plots (Table 16) using PROCHECK is 91.4 which is consistent with the cut-off value (90%) for good quality and reliable models (Laskowski *et al*., 1993). Loop refinement could probably increase the quality of the model (Sánchez and Sali, 1997).

The model assessment result using ANOLEA and QMEAN for the models of selected J and ATPase domains are presented in **Chapter 3, appendices V & VI**. The results showed the quality of each of the residues in the model. While there were some dissimilarities between the evaluation score by ANOLEA and QMEAN, the majority of the residues fall within the reliable regions in the model. It should also be noted that different quality assessment programs employ different and unique energy evaluation parameters. Overall, we concluded that the models showed a high level of accuracy and could be used for further analysis.

Table 15: Model validation of predicted HSP40 J domains and HSP70 ATPase domain_Linker using various quality assessment programs

Proteins	Best model number	Normalized Z-score	GTS_TS	RMSD (Å)	Template-model (RMSD)	DFIRE2 total energy score
DNAJA1	71	-2.808	69.014	2.433	0.201	-100.684
DNAJA2	17	-3.112	70.070	2.579	0.208	-102.304
DNAJB11	24	-1.309	47.840	4.723	0.527	-105.030
DNAJC2	2	-0.961	25.316	6.687	0.288	-94.964
DNAJC3	1	-1.044	42.123	4.247	0.370	-95.454
DNAJC6	87	-1.781	49.342	3.724	0.215	-105.02
DNAJC10	36	-1.731	56.173	3.684	0.362	-111.092
DNAJC19	81	-1.685	65.909	2.672	0.197	-80.743
HSPA1A	70	-1.132	85.316	1.648	0.172	-659.538
HSPA1B	54	-1.143	85.696	1.632	0.176	-660.346
HSPA5	95	-1.669	85.549	1.238	0.192	-651.539
HSPA8	51	-1.147	85.506	1.660	0.187	-651.030
HSPA14	36	-1.204	84.960	1.402	0.158	-644.621

Table 16: Ramachandran plot statistical result showing the most favored, additional allowed, generously allowed and disallowed regions respectively of the predicted models from homology modelling

Predicted model ID	Residues in the most favored regions [A, B, L] (%)	Residues in additional allowed regions [a, b, l, p] (%)	Residues in generously allowed regions [~a, ~b, ~l, ~p] (%)	Residues in disallowed regions (%)
DNAJA1	(59) 95.2	(3) 4.8	(0) 0.0	(0) 0.0
DNAJA2	(57) 96.6	(2) 3.4	(0) 0.0	(0) 0.0
DNAJB11	(65) 94.2	(3) 4.3	(1) 1.4	(0) 0.0
DNAJC2	(64) 91.4	(5) 7.1	(1) 1.4	(0) 0.0
DNAJC3	(65) 95.6	(2) 2.9	(1) 1.5	(0) 0.0
DNAJC6	(64) 98.5	(1) 1.5	(0) 0.0	(0) 0.0
DNAJC10	(67) 94.4	(3) 4.2	(1) 1.4	(0) 0.0
DNAJC19	(50) 94.3	(2) 3.8	(1) 1.9	(0) 0.0
HSPA1A	(329) 93.5	(21) 6.0	(1) 0.3	(1) 0.3
HSPA1B	(327) 92.9	(22) 6.2	(2) 0.6	(1) 0.3
HSPA5	(322) 96.4	(12) 3.6	(0) 0.0	(0) 0.0
HSPA8	(329) 93.2	(23) 6.5	(0) 0.0	(1) 0.3
HSPA14	(317) 94.6	(18) 5.4	(0) 0.0	(0) 0.0

***values above 90% in the most favored regions correlates to a accurate and reliable model**

3.3.3: Structural Analysis of Calculated Models of HSP40 J domains

The conservation of protein structure is much greater than sequence (Krieger *et al.*, 2003). Functional specificity and interactions of a macromolecule are strongly correlated to its 3-D structure. This is because protein residues that are responsible for its function are best arranged in space according to their geometry which in turn allows for interactions with other proteins at the structural level. Thus, in understanding the function of a protein, its structure is far more informative than the sequence (Wiltgen and Tilz, 2009). The homology models for selected human HSP40s and HSP70s are presented in **Chapter 3, appendices III & IV** respectively. The positions of highly conserved residues from the multiple sequence alignment and motif analysis (Table 17 and 18) were mapped on the homology models of DNAJA1, DNAJB11 and DNAJC10 as a representative member of Type I, II and III HSP40s respectively as shown in Figure 23 and 24. At a first glance, the number of conserved charged residues on helixes II and III varies in the different sub-family types with more residues found in helix II. No conserved charged residues were found on the helix III of DNAJC10 (Figure 23). This was in line with previous report of high residue variations in Type III HSP40s more especially on helix III (Hennessy *et al.*, 2000;

2005). However, whereas the positions of the conserved residues vary among the different J domain types, most of the conserved amino acid residues were the same or share similar physicochemical characteristics. Only in few cases were there additional residues present in one type than found in the other types.

The orientation of the highly conserved TYR residue in helix I at position 6 on DNAJA1 and position 14 on both DNAJB11 and DNAJC10 was found to project towards the residues on helix IV in between helixes II and III. Thus, this residue could be critical in maintaining the structural integrity of the J domain together with other hydrophobic residues on helix IV.

Interestingly, all of the conserved charged residues in helix II are projected outward from the solvent exposed surface of the helix. Of note is the ARG residue at position 25, 33 and 26 in DNAJA1, DNAJB11, and DNAJC10 (Figure 23) respectively, the orientation of which is directed towards the solvent exposed surface area. Studies have shown that this residue as well as others of the same structural equivalents is critical for the correct functioning of the J domain (Genevaux *et al*., 2002; Hennessy *et al*., 2005b). Other charged conserved residues in helix II are the GLU and LYS at position 19 and 21 respectively on DNAJA1 (Figure 23). The GLU is negatively charge while the LYS residue is positively charged. However, substitutions of these charged residues in scanning mutagenesis experiments at these positions had no detectable effect on the function of the J domain even though they are projected outward from the solvent surface of the helix (Hennessy *et al*., 2005b). The GLU residue was not conserved in DNAJB11 whereas the LYS residue was conserved at position 32 and 33 respectively in DNAJB11 and DNAJC10. The TYR residues on helix II of DNAJA1 and DNAJB11 were all seen to not be solvent exposed. Their orientations were projected inwardly between helixes II and III. Thus, they might play a role in the structural stability of the J domain in order for it to be in the proper orientation. Positively charged residues in helix II have been reported to interact with the negatively charged residues at the underside cleft pocket of the ATPase domain of partner HSP70s (Suh *et al*., 1999). Thus, those highly conserved residues that were exposed to the solvent were likely to be involved in functional interactions with partner HSP70s rather than maintaining the structural fitness of the protein.

The orientation of the HIS and ASP residues (position 32 and 34 respectively in DNAJA1 and position 40 and 42 respectively both in DNAJB11 and DNAJC10) in the loop region seem to protrude towards the solvent accessible area. This orientation is necessary for the interactions with partner HSP70, as mutations of these residues showed complete alterations in the region and abolished interaction of *Agrobacterium tumefaciens* (Agt) DnaJ with DnaK (Hennessy *et al*., 2005b). It is not yet clear what the role of other conserved charged residues found in the loop region are but as can be seen in the structure of DNAJA1 and DNAJB11 (Figure 23), the orientation of ASN 36 and 44 in the two J domains respectively is pointed inwards in the loop region and in network with the PRO residues in the region. Therefore, they could probably play important role in maintaining the structure of the J domain.

The picture presented by the conserved residues on helix III showed that most of the conserved residues in Type II and III HSP40s are hydrophobic in nature. Very few of the conserved residues were charged as opposed to Type I proteins (Table 17). The orientation of LYS 44 and GLU 51 in the helix III of DNAJA1 as shown in Figure 23 projects outwards from the solvent surface area of the helix and thus is orientated such that it might be able to interact with partner HSP70. Substitution of LYS 44 (LYS 48 in *Agrobacterium tumefaciens*) compromised the function of Agt DnaJ (Hennessy *et al*., 2005b). LYS 42, TYR 50 and SER 54 are all solvent exposed but project in between helixes II and III. Both TYR 61 and SER 65 in the helix III of DNAJB11 (Figure 23) were observed to have their orientation projected toward the ARG residue at position 70 in helix IV. We therefore proposed that these residues together with other hydrophobic residues in the helix as well as helixes I and IV form the network of residues that are responsible for the structural stability of the J domain.

The network of conserved residues on helix IV across all the J domains considered in this study were relatively polar and solvent exposed as shown in Table 17 and Figure 23. Their orientations were protruded toward residues on helix I and could probably share some interactions with the conserved residues on helix I, which help in stabilizing the J domain in its proper conformation for interactions with partner HSP70. While there are yet no functional roles attributed to these residues, the fact they are polar and higly conserved may suggest that they could play a role in HSP40-HSP70 interactions which have not yet been characterized (Genevaux *et al*., 2002). However, previous studies have shown ASP 55 and ARG 59 in DNAJA1 (positions 65 and 70 in DNAJB11 and DNAJC10 respectively) to be implicated as important residues in J domain functioning (Hennessy *et al*., 2005b). ARG 59 is part of the QKRAA motif in *E.coli* DnaJ (Genevaux *et al*., 2002). More importantly was the aspartic acid at position 57 (DNAJA1) which is located at the beginning of helix IV. Its position in between helixes III and IV makes it of both functional and structural significance. Any mutation or substitution that results in the loss of its side chain could result in the loss of the structural integrity of the J domain.

Highly conserved residues that are hydrophobic could play critical roles in maintaining the structural integrity of the J domain. As presented in Table 18 and Figure 24, majority of conserved residues that could be important for the structural fitness of the J domain are found on helixes II and III and others in the turn between helixes I and II most especially in Type I HSP40 J domains. Key among those that have been previously reported as being important for maintaining the structural integrity of the J domain are LEU 9, PHE 43, ALA 49 and LEU 53 (DNAJA1) (Table 18, Figure 24) (Hennessy *et al*., 2005b; Hennessy *et al*., 2000). LEU 9 in helix I as well as VAL 11 and ALA 15 in the turn between helixes I and II (DNAJA1) project outwards from the J domain. Their orientation lies in-between helixes II and III and they seemed to make contact with other residues within helixes I, II and III. These interactions could be significant for keeping the J domain in shape and in the correct conformation. Of note also were the orientations of LEU 27 in helix II and PHE 43 in helix III protruding towards the centre of the two helixes. PHE 43 has been previously predicted to have several potential interactions with HIS 32 in the HPD motif. It is highly conserved across all J domains and located within the highly conserved tripeptide, KFK motif majorly in Type I and II HSP40s (Hennessy *et al*., 2000). The projection of LEU 53 (position 64 in DNAJB11 and DNAJC10) into the interior of the J domain could likely be crucial for holding helixes II and III together. Also, ALA 49 (position 50 in DNAJB11 and DNAJC10) has been implicated to be important in J domain structure and function since the substitution of the corresponding residue in *S. cerevisiae* Sec63p with THR resulted in a translocation defect (Lyman and Schekman, 1995).

Conclusively, while the majority of the conserved residues have been characterized, the role of GLY 10 at the turn region in DNAJA1 (position 18 at the turn region in both DNAJB11 and DNAJC10) (Table 18) remained to be documented. Interestingly, this residues was highly conserved across all HSP40 J domain types as seen in the multiple sequence alignment in Figure 9.

Table 17: Conserved, charged and polar residues in human HSP40 J domains. The exact positions of the residues in the protein sequences prior to alignment are included in brackets.

Protein name	Sub-family type	Helix I	Turn	Helix II	Loop	Helix III	Helix IV
DNAJA1	I	TYR6(8)	-	GLU 19(21), LYS 21(23), TYR 24(26), TYR 31(33), ARG 25(27)	HIS 32(34), ASP 34(36), ASN 36(38)	LYS 42(44), LYS 44(46), TYR 50(52), GLU 51(53), SER 54(56)	ASP 57(59), LYS 58(60), ARG 59(61), TYR 62,(64) ASP 63(65)
DNAJB11	II	TYR 14(15)	-	LYS 29(30), TYR 32(33), ARG 33(34)	HIS 40(41), ASP 42(43), AGR 43(44), ASN 44(45)	TYR 61(62), SER 65(66)	ASP 66(67), LYS 71(72), ARG 70(71), TYR 73(74), ASP 74(75)
DNAJC10	III	TYR 14(15)	-	ARG 26(27), GLU 27(28), LYS 33(35)	HIS 40(41), ASP 42(43)	-	ASP 66(67), ARG 70(71), TYR 73(74), ASP 74(75)

Table 18: Conserved, hydrophobic non-polar residues in human HSP40 J domains. The exact positions of the residues in the protein sequences prior to alignment are included in brackets.

Protein name	Sub-family type	Helix I	Turn	Helix II	Loop	Helix III	Helix IV
DNAJA1	I	LEU 9(11)	GLY 10(12), VAL 11(13), ALA 15(17)	ALA 23(25), LEU 27(29), ALA 28(30)	PRO 33(35), PRO 37(39)	PHE 43(45), ALA 49(51), VAL 52(54), LEU 53(55)	GLY 65(68)
DNAJB11	II	LUE 17(18)	GLY 18 (19), ALA 23 (24)	ILE 28(29), ALA 31(32), ALA 36(37), LEU 37(38)	PRO 41(42)	ALA 50(51), PHE 54(55), ALA 60(61), LEU 64(65)	-
DNAJC10	III	LEU 17(18)	GLY 18 (19), ALA 23(24)	ALA 31(32), PHE 32(33)	PRO 41(42)	ALA 50(51), PHE 54(55), ILE 57(58), ALA 60(61), LEU 64(65)	-

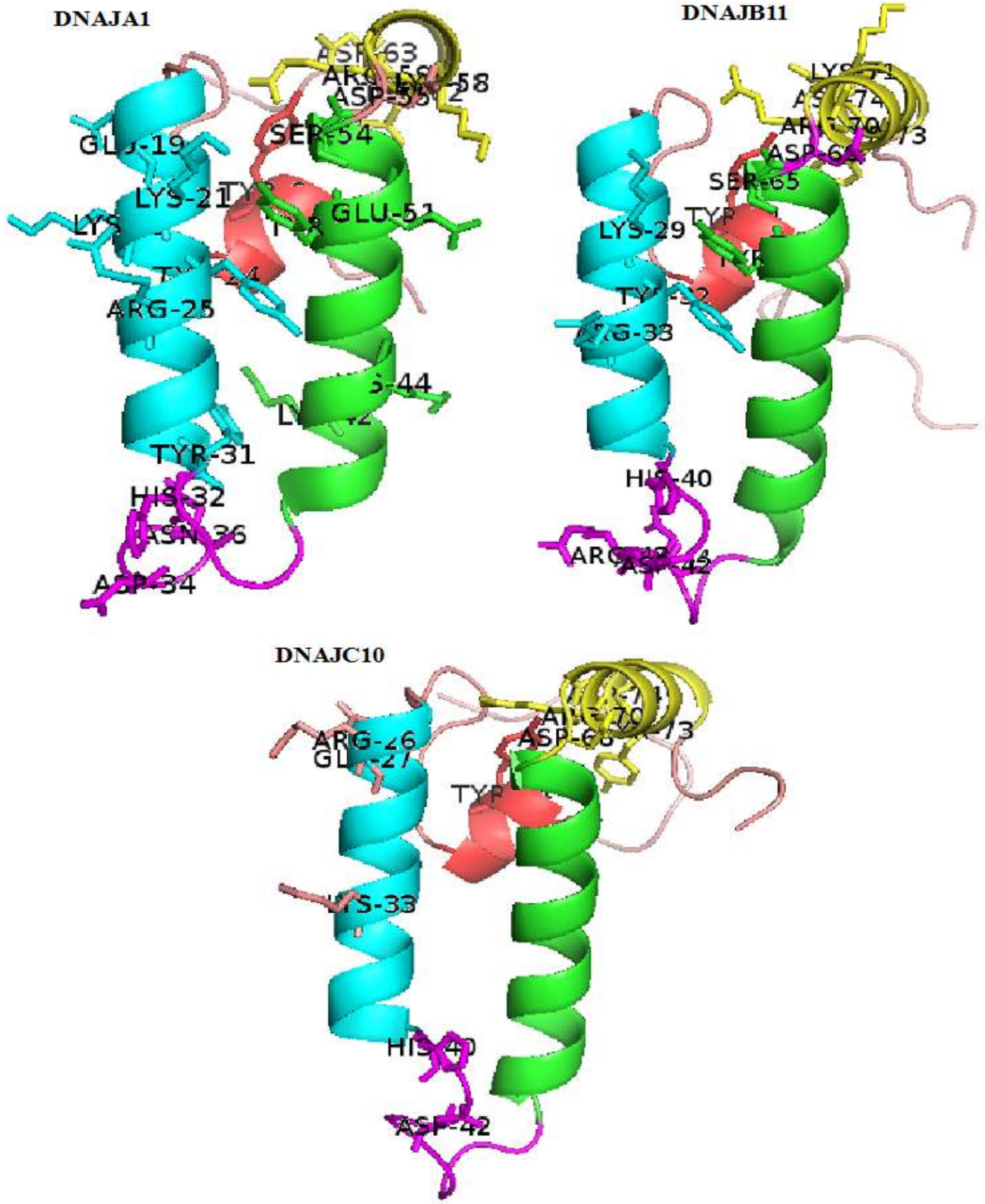


Figure 23: Predicted orientation of conserved polar residues in human HSP40 J domains. Structures of DNAJA1, DNAJB11 and DNAJC10 respectively are represented in cartoon format and conserved polar residues found in the multiple sequence alignment analysis are mapped on the structures as sticks as shown in the figure. Figures were generated in PyMol (Delano and Bromberg, 2004) .

2004).

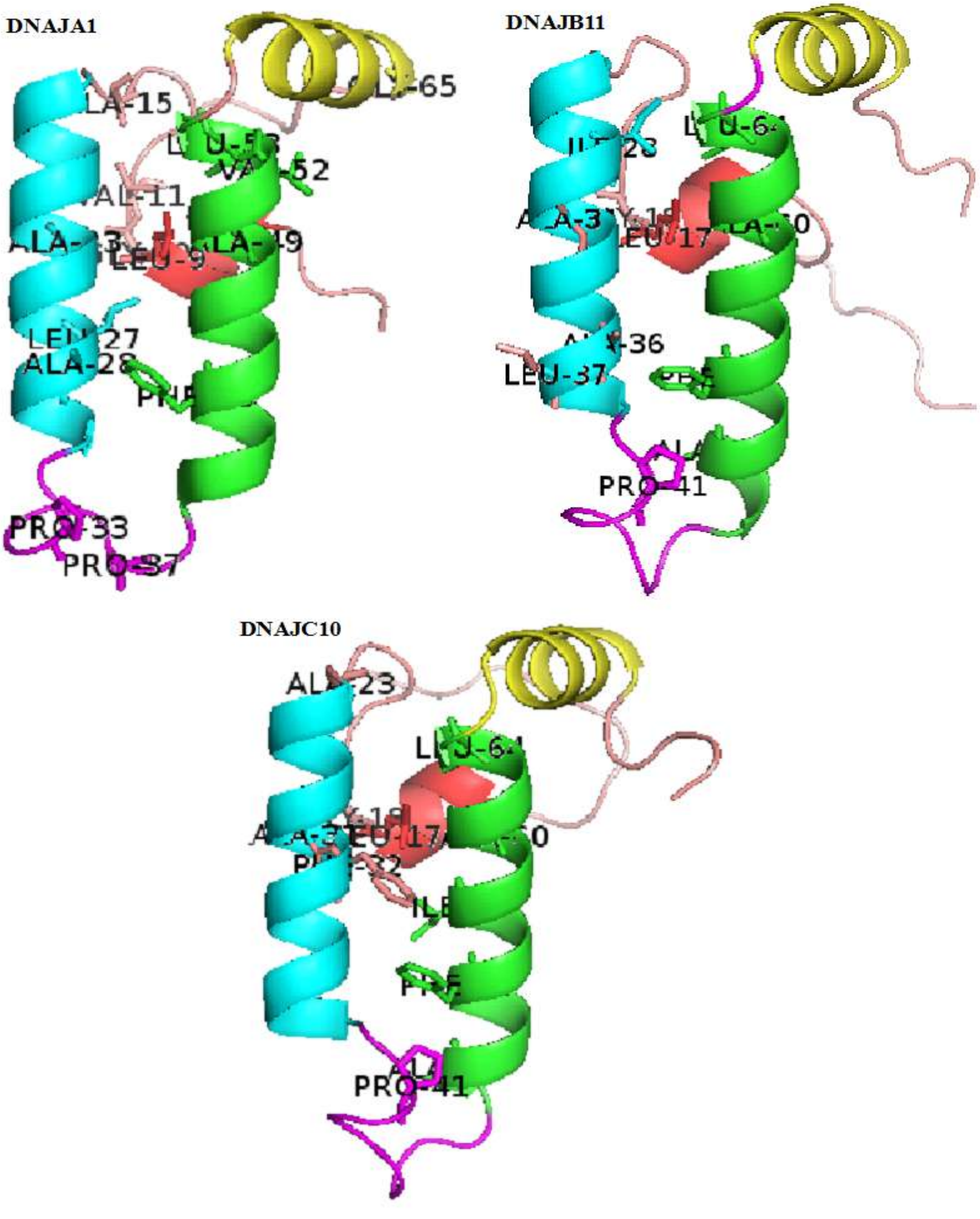


Figure 24: Predicted orientation of conserved hydrophobic residues in human HSP40 J domains. Structures of DNAJA1, DNAJB11 and DNAJC10 respectively are represented in cartoon format and conserved hydrophobic residues found in the multiple sequence alignment analysis are mapped on the structures as sticks as shown in the figure. Figures were generated in PyMol (Delano and Bromberg, 2004).

CHAPTER FOUR: Protein-Protein Interactions of human HSP40-HSP70 complex

4.1 Introduction

Interactions between two proteins play an important role in various biochemical activities (e.g signal transduction). This is because protein complex formation has functional consequences. HSP70-HSP40 partnerships have been widely reported, as the ATPase activity of HSP70 is stimulated by the J domain of HSP40 (Jiang *et al.*, 2007; Nicoll *et al*., 2007). Various techniques are available for predicting the structure of a protein-protein complex at the atomic level. Most of these methods make use of the atomic coordinates of unbound proteins previously determined by experimental methods including X-ray crystallography or NMR. A major challenge in solving the 3-D structure of a complex by X-ray crystallography is the difficulties in crystallising the complex (de Vries *et al*., 2010). This is because the nature of the intermolecular interface of many protein complexes is transient. Many of the proteins structures in the PDB which are able to generate protein-protein complexes are non-obligates complexes (i.e complexes with non-permanent interaction between the monomers; it is possible for the component proteins to exist independently) (Smith and Sternberg, 2002). Docking methods are getting more accurate with new algorithms. Protein-protein docking can provide substantial structural knowledge about complexes, as well as a detailed description of the interactions between the proteins which could give functional information or guidance for further experimental design. This chapter aimed to use on-line molecular docking method to generate possible HSP40-HSP70 complexes with a view to elucidate the interaction interface of the complexes, as well as predict residues and intermolecular interactions that could be critical for such partnership.

4.2 Methodology

4.2.1 Generation of HSP70-HSP40 complexes

Homology models of HSP40 J domains and HSP70 ATPase domains were built (see Chapter three). CPORT (Consensus Prediction Of interface Residues in Transient complexes) server was employed in predicting the interface residues that could be critical for the docking of the two unbound proteins during the complex development process (http://haddock.science.uu.nl/services/CPORT/). CPORT is a Meta server based on consensus method that combines the interface residue prediction scores from six different prediction servers namely; WHISCY, PIER, ProMate, cons-PPISP, SPPIDER, and PINUP (de Vries and Bonvin, 2011). CPORT predictions were used to dock the HSP40 J domain and the ATPase-linker region of HSP70 in this study. Known HSP70-HSP40 J domain interacting residues through previous *in vitro* studies, including ARG 171 in HSP70 and ASP 34 in HSP40, predicted to be involved in direct J domain-ATPase domain interactions were specifically set as active residues during the docking experiment in order to aid the accuracy of the possible orientation of the predicted complex in HADDOCK (de Vries, *et al*., 2010) (http://haddock.science.uu.nl/services/HADDOCK/haddock.php). Experimental data in form of active and passive residues using the prediction interface option, predicted by CPORT, were automatically converted into Ambiguous Interaction Restraints (AIRs) and employed in driving the docking experiment in HADDOCK. The topology of the proteins to be docked is automatically generated in HADDOCK. Three major automated stages are systematically followed during the docking experiment namely: a rigid body energy minimization, a semi-flexible refinement in torsion angle space and refinement in explicit solvent. Interface-ligand RMSD (iL-RMSD) is used in HADDOCK for clustering purposes and employs the Fraction of Common Contacts (FCCs) algorithm written in python language.

4.2.2 Protein Interaction Calculator (PIC)

Protein Interactions Calculator (PIC) server (Tina *et al*., 2007) (http://pic.mbu.iisc.ernet.in/) was used to predict the possible intermolecular interactions between the J domain and the ATPase-linker domain of the predicted docked complex structures. PIC sever is designed to recognise various kinds of interaction including hydrophobic interactions (5Å), disulphide bridges, main-chain-main-chain hydrogen bonds, main chain-side chain hydrogen bonds, side chain-side chain hydrogen bonds, ionic interactions (6Å), aromatic-aromatic interactions (4.5Å to 7Å), aromatic-sulphur interactions (5.3Å) and cation-π interactions (6Å). The coordinate files of the predicted docked complexes were submitted to the server and the aforementioned intermolecular interactions with the default parameters were set for the docked complex structures. Similar calculations were also performed for interactions of the exposed residues at the complex interface.

4.3 Results and Discussion

4.3.1 Predicted J domain-ATPase domain linker complexes

The results of the complexes generated using HADDOCK server are presented in **Chapter 4, appendix I.** The program generates 10,000 structures for each of the complexes during the rigid body stage, out of which the best 400 were refined. The best structures were scored and arranged according to their HADDOCK scores after each stage and prior to the next stage. The weighted sum of the van der Waals energy, electrostatic energy, desolvation energy, the energy from restraint violations and the buried surface area was computed as the HADDOCK score. All these calculations were done automatically by the webserver. The statistics of the best clusters for each complex are shown in Table 20. Usually, the cluster with the lowest HADDOCK score (the lowest negative score) among the clusters for each complex contains the most reliable predicted complex structures (de Vries *et al*., 2010) (**Chapter 4, Appendix II**). Within each cluster, the program provides the best four predicted structures, which were ranked according to their prediction accuracy based on the evaluation of the Van der Waals, electrostatic and the desolvation energies with the topmost model being the best predicted model complex structure. To allow for a better comparison between the predicted docked complexes and the experimental complex crystal structures, the DOPE Z score energy using a python script; (see electronic data/SCRIPTS/*DOPE_Z_score.py*) in MODELLER were calculated for all the predicted complexes in each cluster as well as all the experimental crystal complex structures of (Jiang *et al*., 2007) as presented in Table 19. The lower the Z-score, the better the predicted complex structure generated by HADDOCK and the models in each cluster were ranked as such. As can be seen, the experimental crystal structures have the lowest energy values compared to the predicted docked complexes though with a minimal difference. Whereas all the complexes from each cluster predicted by HADDOCK were aligned in the same orientation (**see Chapter 4, appendix I**), there were minor differences in their energies (Table 19). A comparison of the predicted interactions of the various models in each cluster was performed and the complex with the highest accuracy especially in line with known experimental prediction data was selected in each cluster for subsequent analyses (Table 19). Interestingly, the predicted complex model in each cluster with the lowest HADDOCK score for each HSP70-HSP40 docked complexes showed the highest protein-protein interactions.

Table 19: DOPE Z scores for the predicted complexes within each cluster ranked best by HADDOCK. Models having the lowest HADDOCK scores are highlighted in red colour.

Ranking	Protein complexes	DOPE Z score
1	2QWO	-1.7793
2	2QWP	-1.7450
3	HSPA5_DNAJC10_cluster1.2	-1.7260
4	HSPA5_DNAJC10_cluster1.3	-1.6847
5	HSPA5_DNAJC10_cluster1.4	-1.6691
6	2QWQ	-1.6522
7	2QWR	-1.6435
8	HSPA5_DNAJC10_cluster1.1	-1.6246
9	HSPA8_DNAJA2_cluster1.1	-1.4537
10	HSPA8_DNAJA2_cluster1.2	-1.4360
11	HSPA1B_DNAJA1_ cluster1.2	-1.4195
12	HSPA8_DNAJA2_ cluster1.3	-1.4187
13	HSPA1A_DNAJC3_ cluster1.4	-1.4155
14	HSPA8_DNAJA2_cluster1.4	-1.4045
15	HSPA1A_DNAJC3_cluster1.3	-1.3947
16	HSPA1B_DNAJA1_cluster1_3	-1.3871
17	HSPA1A_DNAJC3_cluster1.2	-1.3690
18	HSPA1A_DNAJB11_cluster1_2	-1.3627
19	HSPA1B_DNAJA1_cluster1_4	-1.3617
20	HSPA1B_DNAJA1_cluster1_1	-1.3596
21	HSPA1A_DNAJC3_cluster1.1	-1.3563
22	HSPA1A_DNAJB11_cluster1_3	-1.3499
23	HSPA1A_DNAJB11_cluster1_4	-1.3392
24	HSPA1A_DNAJB11_cluster1_1	-1.3339
25	HSPA8_DNAJC6_cluster1_3	-1.2935
26	HSPA8_DNAJC6_cluster1_1	-1.2757
27	HSPA8_DNAJC6_cluster1_4	-1.2662
28	HSPA8_DNAJC19_cluster1.3	-1.2638
29	HSPA8_DNAJC6_cluster1_2	-1.2594
30	HSPA8_DNAJC19_cluster1.1	-1.2520
31	HSPA8_DNAJC19_cluster1.2	-1.2489
32	HSPA8_DNAJC19_cluster1.4	-1.2478
33	HSPA14_DNAJC2_cluster1.3	-1.1741
34	HSPA14_DNAJC2_cluster1.2	-1.1703
35	HSPA14_DNAJC2_cluster1.1	-1.1684
36	HSPA14_DNAJC2_cluster1.4	-1.1345

4.3.2 Identification of Intermolecular Interface in the Predicted Complex Structure

Interface prediction is crucial in order to identify residues on the protein structure that interact with another protein. It is mainly based on the extraction and combination of distinct features from protein sequences and structures, which in turn provides biological information for running docking experiments (Vries and Bonvin, 2008). The intermolecular interactions in the docked complexes buried surface area ranging between 1150 Å in HSPA1B-DNAJA1 and 1600 Å in HSPA8-DNAJC6 (Table 20). In order to assess if these complexes represented functional HSP70-HSP40 interactions, the interface exposed residues of the complex models were determined using the Protein Interactions Calculation (PIC) server and all the possible intermolecular interactions were calculated. These should be in agreement with previously documented interactions of HSP70-HSP40 and predict previously unidentified interactions should the predicted docked models captured the functional orientations of both the ATPase domain of HSP70 as well as HSP40 J domain in the complex. As shown in Table 20, analyses of the interactions of the exposed interface residues of the predicted complex models were in line with previously identified J domain-ATPase domain interactions including ARG 171, GLY 215, ILE 216, GLU 386, VAL 388, GLN 389, ASP 395 in HSP70s (Jiang *et al.*, 2007, 2005; Suh *et al.*, 1999) and LYS 25, ARG 29, ASP 30, ARG 34, LEU 37, HIS 40, PRO 41, ASP 42, LYS 57 in HSP40s (Hennessy *et al.*, 2005b). Interestingly, more intermolecular interactions were found in the predicted docked complexes than in the available experimental crystal complex structure (2QWO) previously reported (Jiang *et al.*, 2007). The intermolecular interactions in 2QWO (Jiang *et al.*, 2007) buried 1028Å of protein surface whereas the least intermolecular interface in the docked complex, HSP1B-DNAJAI, buried a protein surface of about 1150Å (Table 20). Surprisingly, HSPA8-DNAJC6 predicted docked complex in this study, comprised of the same set of proteins as in the experimental crystal structure (2QWO), that of bovine HSC70 and auxilin J domain. To compare the orientations predicted by HADDOCK in the complexes with the crystal structure, the unbound protein coordinates from the PDB for bovine HSC70 (2QWL) and human auxilin (1NZ6), named as HSPA8-DNAJC6-exp in this study, were docked using HAADOCK. Interestingly, the docked complex of the proteins aligned in the same orientation with almost all the predicted docked complexes considered in this study and buried 1571.6±66.9Å of the protein surface area (Table 20). However, the orientation of the J domain in the predicted docked complex models was different from that observed in the experimental crystal structures as shown in **Chapter 4, appendix III**. This and the fact that the intermolecular interfaces captured in the predicted docked complexes were greater than that captured in the experimental crystals could suggest that the orientation captured by the predicted docked model for the ATPase and J domains in this study could probably present a J domain-ATPase interface that could define an alternative binding interface for their interactions. The crystal structure (2QWO) could have captured a non-functional orientation of the J domain as the result of the rotation caused by the disulphide linkage introduced between the HSC70:auxilin complex during crystallization due to the transient nature and ATP dependent requirement of J domain:HSP70 interactions (Jiang *et al*., 2007). This suggested that although the interface identified in the docking experiments was different to the crystal structure, it may represent an alternative binding interface. Interestingly, majority of the residues at the binding interface on the J domain were conserved residues on helix II and the tripeptide HPD motif in the loop region. Only LYS 57 among the residues was found on helix III of the J domain. This suggested that Helix II together with the HPD motif in the loop region forms the primary binding interface with partner HSP70 ATPase domain_linker domain. Also, inter-domain linker residues on HSP70 (VAL 388, GLU 386, GLN 389 and ASP 395) could be involved in binding and interactions with corresponding HSP40 J domain.

Table 20: Statistical analysis of HSP70-HSP40 predicted complex structures using HADDOCK server

Predicted complex	Best cluster number /size	HADDOCK score	RMSD	Van der Waals energy	Electrostatic energy	Desolvation energy	Restraints violation energy	Buried Surface Area	Z-score
HSPA1B-DNAJA1	1 (98)	-75.1±8.3	1.6±1.0	-12.7±4.3	-601.3±54.2	56.6±8.2	13.2±22.80	1148.4±223.7	-1.6
HSPA8-DNAJA2	1 (69)	-101.9±3.1	1.3±0.9	-15.4±1.2	-711.8±48.5	49.9±8.9	58.6±11.10	1199.9±57.5	-1.7
HSPA14-DNAJC2	1(142)	-91.0±2.6	0.8±0.6	-24.9±5.7	-518.1±34.9	35.6±5.5	19.8±19.65	1563.1±138.2	-1.4
HSPA1A-DNAJC3	1(96)	-71.0±3.2	5.7±0.6	-16.4±7.9	-586.1±73.1	61.2±10.7	13.0±15.20	1195.5±154.5	-1.2
HSPA1A-DNAJB11	1(62)	-72.2±9.1	2.2±1.6	-22.6±8.7	-581.10±46.8	66.4±11.1	1.9±2.21	1305.5±166.9	-1.3
HSPA8-DNAJC6	1(95)	-110.1±6.4	1.3±0.8	-22.1±14.2	-646.6±134.3	39.7±11.7	16.7±15.93	1525.5±138.3	-1.3
HSPA5-DNAJC10	1(46)	-70.3±1.1	16.4±0.1	-41.8±2.6	-214.9±22.4	10.7±2.0	38.2±27.86	1190.6±14.0	-0.1
HSPA8-DNAJC19	1(106)	-81.0±5.4	1.3±0.8	-19.7±6.9	-509.0±62.6	37.0±10.0	34.6±30.82	1282.4±135.0	-1.8
HSPA8-DNAC6-exp.	1(28)	-90.7±12.7	0.7±0.5	-28.9±5.4	-584.4±84.3	55.0±4.8	1.4±0.64	1571.6±66.9	-2.4

Table 21: Intermolecular interface residues of complexes predicted using Protein Interactions Calculator (PIC) server

Protein complexes	Hydrophobic Interactions (5Å)		Main Chain-Side Chain Hydrogen Bonds			Side Chain-Side Chain Hydrogen Bonds			Ionic Interactions (6Å)		Cation-*π* interactions (6Å)	
	ATPase domain	**J Domain**	**ATPase domain**	**J domain**	**Bond**	**ATPase domain**	**J domain**	**Bond**	**ATPase domain**	**J domain**	**ATPase domain**	**J domain**
HSPA1B-DNAJA1	ILE 216	PRO 37	ARG 171	TYR 31	NH1 to O	ASP 160	LYS 22	OD1 to NZ OD2 to NZ	LYS 3	ASP 55	-	-
	PHE 217	PRO 33	VAL 219	ASP 34	N to OD2	ASP 152	LYS 30, LYS 35	OD1 to NZ OD2 to NZ	ASP 152	LYS 30, LYS 35,	-	-
	-	-	-	-		-	-		ASP 160 GLU 218	LYS 22 ASP 34		
HSPA8-DNAJA2	-	-	ASP 395	LYS 22	OXT to NZ	ASP 214, ASP 395	LYS 22	OD1 to NZ OD2 to NZ	GLU 192	ARG 25	-	-
	-	-	-	-		ASP 214, ASP 395	LYS 26	OD1 to NZ OD2 to NZ	GLU 213	LYS 22, ARG 25	-	-
	-	-	-	-		GLU 192	ARG 25	OE2 to NH2	ASP 214	LYS 22, LYS 26, ARG 25	-	-
	-	-	-	-		GLU 213	ARG 25	OE1 to NH1, NH2	ASP 395	LYS 22, LYS 26	-	-
HSPA1A-DNAJB11	ILE 216	TYR 32, TYR 61	GLN 156	GLU 68	N to OE1	LYS 159	GLU 68	NZ to OE1 NZ to OE2	ASP 152	LYS 71	-	-
	VAL 388	PRO 41	LEU 393	ASP 42, ARG 43	N to OD2 O to NH2	ARG 171	GLU 62	NE to OE1	LYS 159	GLU 68	-	-
	LEU 393	LEU 37	LEU 170	LYS 69	O to NZ	GLU 218	LYS 26	OE1 to NZ	ARG 171	GLU 62, ASP 66	-	-
	-	-	-	-		ASP 213, GLU 218	LYS 29	OD1 to NZ OE1 to NZ	GLU 192	LYS 29	-	-
	-	-	-	-		ASP 213	ARG 33	OD2 to NH1	ASP 213	LYS 29 ARG 33	-	-
	-	-	-	-		ASP 214	ARG 33	OD2 to NH2	ASP 214	ARG 33, HIS 40	-	-
	-	-	-	-		ASP 395	ARG 43	OD2 to NH1	GLU 218	LYS 26, LYS 29	-	-
	-	-	-	-		ASP 152	LYS 71	OD2 to NZ	ASP 395	ARG 43		
HSPA14-DNAJC2	ILE 214	PHE 57	VAL 166	LYS 31	O to NZ	ARG 168	TYR 64	NE to OH	ASP 138	ARG 24	ARG168	TYR 64
	ILE 379	PRO 70	PHE 145, ASP 146, PHE 147	LYS 43	O to NZ O to NZ O to NZ	HIS 171	ASP 40,	NH2 to OD2	GLU 149	LYS 41	PHE 145	LYS 43
	-	-	GLU 149	ALA 45	OE1 to N	ASN137, ASP 138, ASN 165	ARG 24	OD1 to NH2 OD2 to NH1 ND2 to NH2	HIS 171	ASP 40	-	-
	-	-	-	-		GLU 149	LYS 41	OE1 to NZ	-	-	-	-

HSPA1A-DNAJC3	ILE 216	LEU 29	ARG 171	PRO 33	NH1 to O NH2 to O	LYS 384	ASN 36	NZ to OD1	LYS 384	GLU 39	PHE 217	LYS 35
	-	-	ASP 395	LYS 22, LYS 26	O to NZ OXT to NZ	ASP 395	LYS 22, LYS 26	OD2 to NZ OD1 to NZ OD2 to NZ	ASP 395	LYS 22, LYS 26	-	-
	-	-	GLY 215	LYS 35	O to NZ	-	-		-	-	-	-
HSPA8-DNAJC6	ILE 216	TYR 32, PHE 57	ASP 395	LYS 30	OXT to NZ	LYS 159	GLN 51	NZ to OE1	GLU 192	LYS 73	ARG171	PHE 57
	VAL 388, LEU 393	LEU 37	LEU 394	ARG 33	O to NHI O to NH2	LYS 220	GLU 68	NZ to OE1	GLU 213, ARG 214	ARG 33	-	-
	-	-	LEU 170	LYS 54	O to NZ	ASP 395	LYS 30, LYS 34	OD1 to NZ OD2 to NZ	LYS 220	GLU 68	-	-
	-	-	GLU 192	LEU 75	OE1 to N	GLU 213	ARG 33	OE1 to NH1	GLU 386	LYS 43	-	-
	-	-	-	-		ASP 214	ARG 33	OD1 to NH1	ASP 395	LYS 30, LYS 34	-	-
	-	-	-	-		GLU 386	LYS 43	OE1 to NZ	-	-	-	-
	-	-	-	-		GLU 192	LYS 73	OE1 to NZ OE2 to NZ	-	-	-	-
2QWO	-	-	LEU 170	CYS 876	O to SG	SER 385	ASP 896	OG to ODI OG to OD2	ASP 214	LYS 816	-	-
	-	-	-	-		-	-		GLU 386	ARG 828	-	-
HSPA5-DNAJC10	VAL 216	LEU 37	GLU 322, ASP 323	ARG 29	O to NH1 O to NH2 O to NE	ASP 153	HIS 51	OD1 to NE2	ASP 153	HIS 51	-	-
	-	-	GLY 215	LYS 43	O to NZ	ASP 323, ASP 325	ARG 29	OD2 to NH2 OD1 to NH1	GLU 192	LYS 33	-	-
	-	-				GLU 192	LYS 33	OE1 to NZ	ASP 323	ARG 29, LYS 65	-	-
	-	-				ASP 323	LYS 65	OD1 to NZ	ASP 325	ARG 26, ARG 29	-	-
HSPA8-DNAJC19	ILE 216	PRO 41	ARG 171	PRO 41	NH1 to O	ARG 171	ASP 42	NH1 to OD1 NH2 to OD2	ARG 171	ASP 42	-	-
	VAL 388	LEU 37	ASP 395	LYS 25, ARG 29, LYS 57	O to NZ O to NE O to NH2 OXT to NZ	ASP 395	LYS 25, LYS 57	OD2 to NZ OD1 to NZ OD2 to NZ	GLU 386	ARG 34		
	-	-	GLY 215	HIS 40	O to NE2	GLN 389	ARG 34	OE1 to NH2	ASP 395	LYS 25, ARG 29, ASP 30, LYS 57	-	-

4.3.3 Hydrophobic Interactions within 5Å

Hydrophobicity has been described as the major driving force in protein folding and stability (Gromiha and Selvaraj, 2004). As presented in Table 21, hydrophobic interactions among the exposed residues in the predicted complexes showed more interactions in HSPA1A-DNAJB11 and HSPA8-DNAJC6 respectively than found in the other complexes. Hydrophobic interactions within 7Å between residues have been reported as significant (Gromiha and Selvaraj, 2004). All the hydrophobic interactions found in this study were within 5Å as calculated using PIC. ILE 216 and PHE 217 in HSPA1B formed hydrophobic interactions with PRO 37 and PRO 33 respectively in DNAJA1 J domain. There were no hydrophobic interactions found among the exposed residues at the interface region of the predicted docked complex between HSPA8 and DNAJA2. Looking at the complex formed between HSPA1A and DNAJB11, there were two hydrophobic interactions between ILE 216 (HSPA1A) and TYR 32 and TYR 61 (DNAJB11) respectively. Also, both VAL 388 and LEU 393 (HSPA1A) interacted with PRO 41 and LEU 37 (DNAJB11) respectively**.** For the complex between HSPA14 and DNAJC2, both ILE 214 and ILE 379 (HSPA14) formed hydrophobic interactions with PHE 57 and PRO 70 (DNAJC2) respectively**.** One hydrophobic interaction was observed each in HSPA1A-DNAJC3 complex between ILE 216 and LEU 29 as well as between VAL 216 and LEU 37 in HSPA5-DNAJC10 complex respectively (Table 21).

Interestingly, more hydrophobic interactions were found in the docked complex of HSPA8 and DNAJC6 than in the experimental crystal complex structure of 2QWO (Jiang *et al*., 2007) (Table 21). While no hydrophobic interactions were found among the exposed interface residues in the crystal structure of 2QWO, ILE 216 (HSPA8) formed two interactions with TYR 32 and PHE 57 (DNAJC6) respectively. Also, LEU 37 (DNAJC6) formed two interactions with VAL 388 and LEU 393 respectively in HSPA8. Whereas both the ATPase domain in 2QWO and HSPA8-DNAJC6 docked complex model aligned in the same orientations, the two J domains in the complexes were in different orientations (**Chapter 4,** a**ppendix IV**). This could probably account for the differences observed in the number of interactions between the two complexes. Two hydrophobic interactions were found in HSPA8-DNAJC19 docked complex model between ILE 216, VAL 388 (HSPA8) and PRO 41, LEU 37 (DNAJC19) respectively (Table 21). Interestingly**,** ILE 216, VAL 388 and LEU 393 were conserved across all human HSP70s and have been previously reported to be implicated in HSP40-HSP70 interactions (Jiang *et al*., 2007).

TYR 32, PRO 33, LEU 29 and 37, PRO 41, PHE 57 and TYR 61 are part of the highly conserved hydrophobic residues in the loop region, helixes II, III and IV respectively. These were parts of previously predicted conserved residues implicated to be involved in maintaining the structural stability of J domain in order to be in its correct orientation for interactions with partner HSP70 (Nicoll *et al*., 2007; Hennessy *et al*., 2005b).

4.3.4 Hydrogen bonds

Hydrogen bonds between main chain-side chain interactions, as well as side chain-side chain interactions among the exposed residues at the interface of the predicted docked complex structures, are presented in Table 21. In the complex between HSPA1B and DNAJA1, both ARG 171 and VAL 219 (HSPA1B) formed hydrogen bonds with the side chains of TYR 31 and ASP 34 (DNAJA1) respectively. ASP160 (HSPA1B) interacted with LYS 22 while ASP 152 formed two hydrogen bonds with both LYS 30 and LYS 35 in a side chain-side chain hydrogen bonding interactions. For the HSPA8-DNAJA2 docked complex, ASP 395 formed an hydrogen bond with the side chain of LYS 22 in DNAJA2. Interestingly, the side chains of both LYS 22 and LYS 26 (DNAJA2) formed two hydrogen bonds with ASP 214 and ASP 395 (HSPA8) respectively. Also, the side chain of AGR 25 in DNAJA2 J domain formed two hydrogen bonds with the two GLU residues at positions 192 and 213 (HSPA8) respectively. Looking at the protein-protein interactions within the complex structure of HSPA1A-DNAJB11, the main chain of GLN 156 (HSPA1A) interacted with the side chain of GLU 68 in DNJAJB11. LEU 393 in HSPA1A formed two hydrogen bonds with the side chains of ASP 42 and ARG 43 (DNAJB11). Also, the main chain of LEU 170 interacted with the side chain of LYS 69 (DNAJB11) through an hydrogen bond. Scores of side chain-side chain hydrogen bond interactions were observed between LYS 159, ARG 171, GLU 218, ASP 395, ASP 152 in HSPA1A and GLU 68, GLU 62, LYS 26, ARG 43 and LYS 71 in DNAJB11 respectively. LYS 29 (DNAJB11) forms two side chain hydrogen bonds with the side chains of both ASP 213 and GLU 218 (HSPA1A). Furthermore, the side chain of ARG 33 (DNAJB11) forms two hydrogen bonds with the side chains of the two Aspartic acids at positions 213 and 214 of HSPA1A.

Main chain-side chain hydrogen bonding interactions in HSPA14-DNAJC2 complex included VAL 166 and LYS 31, GLU 149 and ALA 45 respectively. Also, main chain of LYS 43 (DNAJC2) formed three hydrogen bonds with PHE 145, ASP 146 and PHE 147 in HSPA14 respectively. Hydrogen bonding interactions were also found between ARG 168, HIS 171, GLU 149 (HSPA14) and TYR 64, ASP 40 and LYS 41 (DNAJC2) respectively. The ARG residue at position 24 in DNAJC2 formed three side chain hydrogen bonds with the side chains of ASN 137, ASP 138 and ASN 165 respectively in HSPA14.

For the complex structure between HSPA1A and DNAJC3, the main chain of ARG 171 and GLY 215 (HSPA1A) formed hydrogen bonds with PRO 33 and LYS 35 (DNAJC3) respectively. ASP 395 in HSPA1A interacted with side chains of both LYS 22 and 26 (DNAJC3) in two separate hydrogen bonds. The side chain of LYS 384 (HSPA1A) formed a hydrogen bond with that of ASN 36 in DNAJC3 whereas the side chain ASP 395 (HSPA1A) also formed two hydrogen bonds with both LYS 22 and LYS 26 (DNAJC3) as presented in Table 21.

The hydrogen bond interactions observed in HSPA8-DNAJC6 complex appeared to be the highest number of protein-protein interactions analysed in this study. Main chain-side chain hydrogen bonds were formed between ASP 395, LEU 394, LUE 170, and GLU 192 in HSPA8 with LYS 30, ARG 33, LYS 54 and LEU 75 in DNAJC6 respectively. Also, the side chains of LYS 159, LYS 220, GLU 386 and GLU 192 (HSPA8) formed hydrogen bonds with the side chains of GLN 51, GLU 68, LYS 43, and LYS 73 (DNAJC6) respectively. The side chains of both GLU 213 and ASP 214 interacted with the side chain ARG 33 (DNAHC6) while that of ASP 395 (HSPA8) formed a hydrogen bond with the side chains of both LYS 30 and LYS 34 in DNAJC6.

Endoplasmic reticulum localized docked complex of HSPA5 and DNAJC10 showed hydrogen bonds between the main chain of GLY 215 (HSPA5) and the side chain of LYS 43 (DNAJC10), whereas the ARG 29 (DNAJC10) formed two hydrogen bonds with the side chains of GLU 322 and ASP 323 respectively in HSPA5. Side chain-side chain hydrogen bonds were found between ASP 153, GLU 192, ASP 323 in HSPA5 and HIS 51, LYS 33 and LYS 65 in DNAJC10 respectively. The side chain of ARG 29 (DNAJC10) formed two hydrogen bonds with both ASP 323 and 325 in HSPA5.

In the complex model between HSPA8 and DNAJC19, both the main chains of ARG 171 and GLY 215 formed hydrogen bonds with the side chains of PRO 41 and HIS 40 respectively. The main chain of ASP 395 (HSPA5) formed three hydrogen bonds with LYS 25, ARG 29 and LYS 57 respectively in DNAJC10. Also, the side chains of both ARG 171 and GLN 389 in HSPA5 interacted with the side chains of ASP 42 and ARG 34 in DNAJC10 respectively. The side chains LYS 25 and 57 formed hydrogen bonds with that of ASP 395 (HSPA5). Interestingly, the majority of these residues were charged and highly conserved in both HSP70-HSP40. Of note is the interaction between ARG 171 (ARG 167 in *E coli*) in HSP70 and the Aspartic acid (ASP 43 in DNAJA1, ASP 42 in DNAJB11 and ASP 40 in DNAJC2) in the loop region within the tripeptide HPD signature located between helixes II and III in HSP40 J domain. This interaction has been widely reported to be critical for HSP40-HSP70 partnership (Nicoll *et al*., 2007; Hennessy *et al*., 2005b; Genevaux *et al* 2002; Schwager *et al*., 2002; Suh *et al*., 1998). Also, most of these residues found on the exposed surfaces of the J domain in the complexes were highly conserved positively charged residues on helix II. Of interest was the LYS residue on helix II (e.g LYS 29 in the J domain of HSPA1A-DNAJB11 complex structure), this lysine residue was highly conserved across the J proteins and solvent exposed regardless of its position in the different HSP40 J domains considered in this study. Highly conserved positively charged residues on helix II, particularly LYS 26 in *E. coli* DnaJ (Genevaux *et al*., 2002) have been previously implicated to be important for J domain function. This and other highly conserved residues like the ARG on the same helix II (position 25 in DNAJA2, position 33 in DNAJB11 & DNAJC6, position 23 in DNAJC2, and position 34 in DNAJC19) on helix II probably formed the recognition interface for binding with the negatively charge regions of HSP70 ATPase domain. This is in line with previous report of Hennessy *et al*., 2005b that the arginine at position 26 in *Agrobacterium tumefaciens* made a network of interactions with DnaK and its alteration could inhibit the correct functioning of the J domain of *A. tumefaciens* (Hennessy *et al*., 2005b).

4.3.5 Ionic Interactions within 6Å

Ionic interactions within 6Å were found among conserved and exposed residues at the interface of the predicted docked complexes as presented in Table 21. Within the complex structure of HSPA1B-DNAJA1, ionic interactions were observed between LYS 3, ASP 160, GLU 218 in HSPA1B and ASP 55, LYS 22, ASP 34 in DNAJA1 respectively. Two ionic interactions were also found between ASP 152 (HSPA1B) and the two LYS residues at positions 30 and 35 (DNAJA1)**.** More ionic interactions were found in the complex structure of HSPA8-DNAJA2 than observed in HSPA1B-DNAJA1. GLU 192 in HSPA8 interacted with ARG 25 in DNAJA2. Also, two ionic interactions were present with GLU 213 (HSPA8) interacting with both LYS 22 and ARG 25 (DNAJA2). ASP 214 formed three ionic bonds with LYS 22, ARG 25 and LYS 26.

The ASP at position 395 in HSPA8 formed two ionic bonds with both LYS 22 and LYS 26. The results presented by the complex between HSPA1A-DNAJB11 showed a number of ionic interactions including two ionic bonds each between ARG 171 (HSPA1A) and GLU 62 as well as ASP 66 (DNAJB11); ASP 213 (HSPA1A) and LYS 29 with ARG 33 (DNAJB11); ASP 214 (HSPA1A) and ARG 33 with HIS 40 (DNAJB11); GLU 218 (HSPA1A) and LYS 26 with LYS 29 (DNAJB11) within the complex. Also, ASP 152, LYS 159, GLU 192, and ASP 395 in HSPA1A formed ionic bonds with LYS 71, GLU 68, LYS 26 and ARG 43 in DNAJB11 respectively.

Three ionic interactions were found in the complex structure of HSPA14 and DNAJC2 shown in Table 21. ASP 138, GLU 149 and HIS 171 (ARG 171 in other HSP70s) formed ionic bonds with ARG 24, LYS 41 and ASP 40 in DNAJC2 respectively. LYS 384 in HSAP1A formed an ionic interaction with GLU 39 in DNAJC3. Also, ASP 395 in HSPA1A formed two ionic bonds with the two LYS residues at positions 22 and 26 in DNAJC3. From the docked complex model of HSPA8 and DNAJC6 as shown in Table 21, GLU 192, LYS 220 and GLU 386 in HSPA8 formed ionic bonds with LYS 73, GLU 68 and LYS 43 in DNAJC6 respectively. ARG 33 (DNAJC6) formed two ionic bonds with GLU 213 and ARG 214 in HSPA8. ASP 395 interacted with both LYS 30 and LYS 34 in DNAJC6.

The ionic interactions found in exposed interface residues of HSPA5-DNAJC10 complex showed that ASP 153 and GLU 192 interact with HIS 51 and LYS 33 respectively. ASP 323 (HSPA5) formed two ionic bonds with both ARG 29 and LYS 65 in DNAJC10. The ASP residue at position 325 (HSPA5) formed two ionic bonds with the two ARG residues at positions 26 and 29 in DNAJC10.

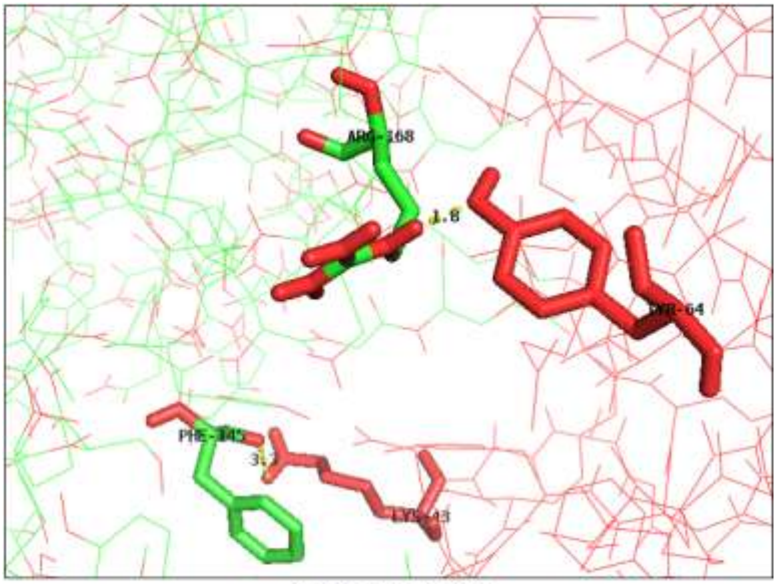
Finally, the complex structure of HSPA8 and DNAJC19 revealed four ionic bonds between ASP 395 (HSPA8) and LYS 25, ARG 29, ASP 30, LYS 57 in DNAJC19. An ionic bond was formed between ARG 171 (HSPA8) and ASP 42 (DNAJC19) as well as GLU 386 (HSPA8) and ARG 34 in DNAJC19.

It was interesting to note that almost all the residues predicted to be involved in ionic interactions at the complex interface were highly conserved charged residues as observed in the multiple sequence alignment of HSP40 J domains and the ATPase domain of HSP70s as shown in chapter two (see Figure 9 and Figure 16). It appeared as if more conserved charged residues were present at the interface of the complexes involving Type II and Type III HSP40 J domains than observed in the Type I. Surprisingly, most of these residues were positively charged residues, especially the LYS and ARG residues, located on the helix II of the J domain. These and other conserved residues have been previously reported to interact with the negatively charged pocket of the ATPase domain of HSP70s. Most of the interacting partners in the HSP70 counterparts were negatively charged GLU and ASP acid residues at the under cleft pocket of HSP70 ATPase domain with positively charged residues on helix II of HSP40 J domain. Of note were ASP 138, 152, 153, 160, 213, 214, 323, and 325 which were conserved across all the HSP70s considered in this study. Also GLU 149, 192, 213, and 218 were found conserved across the HSP70s in the predicted model complex. Highly conserved residues in the linker region between the ATPase and substrate binding domains have also been implicated to be critical for HSP40-HSP70 interactions. Mutagenesis experiments where these residues were absent abolished J domain-ATPase domain interactions (Jiang *et al*., 2007, 2005, 2003; Suh *et al*., 1998). In line with our findings, conserved residues at the linker region between the ATPase domain and Substrate Binding Domain **(**SBD) in HSP70s found important in the predicted complexes include: GLU 386, VAL 388, GLN 389, LEU 393, LEU 394 and ASP 395. These residues were all found conserved at the interface of the complexes majorly in hydrophobic and ionic interactions with the interface residues of partner HSP40 J domains. ASP 395 stands out among these linker residues as it was found to form either hydrogen bonds or ionic interactions with interface residues of J domain as previously highlighted most especially the conserved LYS and ARG residues on helix II.

4.3.6 Cation–π Interactions within 6Å in Protein–protein Interface

Cation- π interactions were found between HSPA14-DNAJC2, HSPA1A-DNAJC3 and HSPA8-DNAJC6 model complexes. Both the side chains of ARG 168 and PHE 145 in HSPA14 formed cation-π interactions with the side chains of TYR 64 and LYS 43 respectively in DNAJC2 (Figure 25). Also, the side chain of the phenylalanine at position 217 in HSPA1A interacted with the side chain of the lysine residue at position 35 in DNAJC3. Lastly, the side chain of ARG 171 forms a cation-π interaction with PHE 57 in DNAJC6 (Table 21).

Cation-π interactions with the side chains of aromatic residues have been reported as an important non-covalent interaction at the protein-protein interface (Crowley and Golovin, 2005). It involves interaction between the side chains of positively charged LYS, ARG or HIS residues with the side chains of any of the aromatic amino acids including PHE, TYR or TRP. ARG, being one of the most abundant residues at the interface of different types of protein-protein interactions, is usually favoured in most cation-π interactions. This is because its large side chain contributes to intermolecular interactions (Gallivan and Dougherty, 1999). Interestingly, these interactions were only found at the interface of complex structures between Type III HSP40 J domains and partner HSP70 ATPase domain. This could partly account for the reason why Type III HSP40s do not interact non-specifically in J domain swapping experiments. These interactions were not found among the Type I and II J domains analysed in this study. It therefore remained to be argued through various experimental studies such as site directed mutagenesis, if such interactions are important for J domain-ATPase interactions.



Cation-π interactions

Figure 25: Cation-π interactions found among exposed residues at the interface of the complex between HSPA14 and DNAJC2. Both HSP70 ATPase domain-linker and HSP40 J domain are displayed as lines and colored in green and red respectively. Exposed residues at the complex interface involved in the interactions were shown and labeled as sticks. The picture was rendered in PyMol (Delano and Bromberg, 2004).

4.3.7 Prediction of Residues Critical for J-domain:ATPase domain_linker region Interactions (HSPA8-DNAJC19)

In order to assess if the predicted complex model represented a functional model complex of J domain-ATPase domain interactions and could identify new interactions between the partner proteins, DNAJC19 and HSPA8 were docked in a complex structure since interactions between these two proteins have not been published. The complex model should predict known interactions of HSP40 J domain and HSP70 ATPase, should the model represent a functional HSP70-HSP40 partnership. It should also predict unknown interactions that could be important in defining the interactions between the two proteins in the docked complex model. The result of the protein-protein interactions of the exposed interface residues as presented in Table 21 were mapped on the predicted complex model as shown in Figure 26. The result showed conserved residues on both helixes II and III were mostly involved in HSP40-HSP70 interactions with more interactions with positively charged residues on helix II binding with the negatively charged residues at the under cleft pocket of the ATPase domain. Also, linker region residues particularly GLU 386, VAL 388, GLN 398 and ASP 395 interacted mainly with those conserved residues on helix II including LYS 25, ARG 29, ASP 30, ARG 34 and LEU 37. ARG 171, GLY 215 and ILE 216 at the HSP70 ATPase interface formed network of interactions with residues at the loop region particularly the residues in the HPD motif. Interestingly, these residues were highly conserved in the two proteins. While ARG 171 and ILE 216 from HSP70 have been widely reported in the literature to be involved in ATPase-J domain interactions, the role of GLY 215 in ATPase activities and interactions with HSP40 remains undocumented. However, this residue showed complete conservation across all human HSP70 as shown in the multiple sequence alignment result in chapter two (see Figure 16). It therefore remained to be investigated through site directed mutagenesis experiments if substitution of this residue could play a deleterious role in HSP70-J domain interactions.

As seen in Figure 27A, both ILE 216 and VAL 388 in HSPA8 formed hydrophobic interactions with PRO 41 and LEU 37 (DNAJC19) respectively. PRO 41 also interacted with the side chain of ARG 171. ARG 171 (equivalent to ARG 167 in *E.coli*) has been widely reported to be critical for HSP70 interactions with HSP40 J domain especially with the ASP 42 residue within the HPD signature (Suh *et al*., 1999). PRO 41 being located within the HPD motif which is highly conserved across all HSP40 J domains, could be critical in keeping the J domain in its proper orientation for interactions with partner HSP70. LEU 37 is located on helix II and shared 65% conservation across all human HSP40 J domain as seen in the multiple sequence alignment analysis in chapter two (see Figure 9). Both ILE 216 and VAL 388 (HSPA8) were highly conserved across all HSP70 (see Figure 16). The non-polar nature of these residues and the fact that they are not charged, confirmed their hydrophobic roles and importance in maintaining the structural integrity of both the ATPase and J domains for interactions. Any mutational changes that bring about conformational changes in the orientation at these residues may disrupt both the ATPase and J domain interface, thus affects J domain-ATPase_linker ineractions.

As with other studies, the highly conserved ARG 171 in HSPA8 interacts with the ASP 42 within the HPD motif of partner DNAJC19 J domain (Figure 27B). Interestingly, these and previous interactions found within this complex model corroborate the previous report that both ARG 171 and ASP 42 are fundamental for mediating the interactions between HSP70-HSP40 partnership. GLY 215 interacted with HIS 40 which is also part of the HPD network in the DNAJC19 J domain (Figure 27C). Both of these residues were highly conserved across all HSP70 and HSP40 respectively with 100% conservation as previously shown in the multiple sequence alignment analyses. Also, GLU 386 and GLN 389 in the ATPase domain were found to both interact with ARG 34 at the complex interface (Figure 27D).These residues were highly conserved both in HSP70 and HSP40 J domains respectively.

Of note was the network of interactions between ASP 395 and LYS 25, ARG 29, ASP 30, LYS 57. ASP 395 is conserved across HSP70s (see Figure 16). Both ARG 30 and ASP 29 (lysine residues across J domains as shown in the multiple sequence alignment (see Figure 9) were highly conserved and packed each other and formed ionic bonds with ASP 395 (Figure 27B). Surprisingly, LYS 25 located on helix II and LYS 57 on helix III were not conserved in Type III J domains as opposed to both in Type I and II (see Figure 6, 7 and 8). LYS 57 is part of the tripeptide KFK motif. Whereas interactions of the PHE in the KFK motif and the HIS residue within the HPD have been reported to be important for maintaining and stabilizing helixes II and III structure in addition to other anti-parallel bonding between them, the two LYS residues in the motif have been proposed to likely play a role in interactions with HSP70 (Genevaux *et al*., 2002; Hennessy *et al.*, 2000). Thus, LYS 57 could play significant roles in determining specific interactions of the J domain with corresponding HSP70s since it is not highly conserved across all HSP40s. Interestingly, these two LYS were replaced with ILE and ALA residues respectively in the sequence of DNAJC19 represented as IAA signature as opposed to the KFK motif highly conserved mainly in Type I and II J domains (see Figure 9). This could also probably explained why some Type III J proteins could not be swapped for functioning with other J proteins based on their sub-cellular localizations and vise visa. DNAJC19 is predicted to be localized in the mitochondrial.

Highly conserved residues have been investigated to be crucial for J domain-HSP70 functional interactions (Hennessy *et al*., 2000). We therefore proposed, in line with other studies, that highly conserved residues identified on helixes II and the tripeptide HDP motif in the loop region of the J domain could be critical for HSP40-HSP70 general interactions while less conserved residues on both helixes II and III could be involved in defining HSP70-J domain specific functions. Also, conserved residues at the linker region between the ATPase and substrate binding domain of HSP70s are critical for interactions with partner HSP40s especially ASP 395. The role of ASP 395 has not been reported in literature. We proposed that ASP 395 together with other highly conserved hydrophobic residues that have been previously reported in the HSP70 linker region formed a network of interactions with J domain and as such, could play important role in mediating interactions with partner HSP40s.

Finally, the predicted docked complex model confirmed functional interacting residues of known J domain-ATPase interactions as well as predicted helixes II and III of the J domain as the main binding interface with helix II as the main point of contact. It suggested that the lower cleft of the ATPase domain provided a binding pocket for J domain interactions and the linker residues could play crucial roles in J domain binding and interactions.

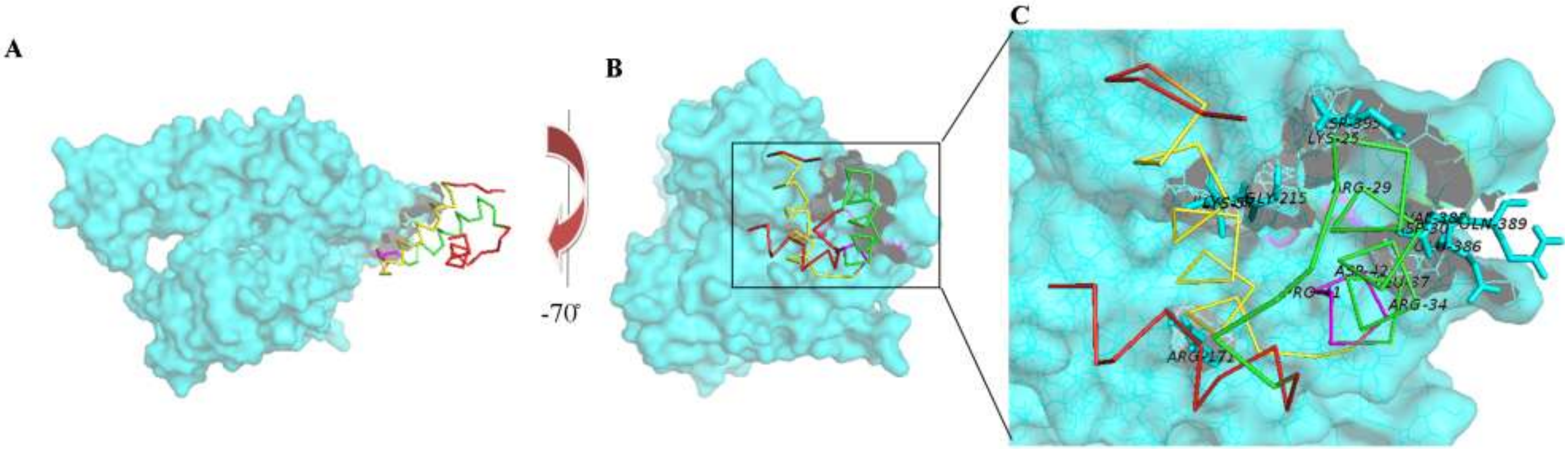
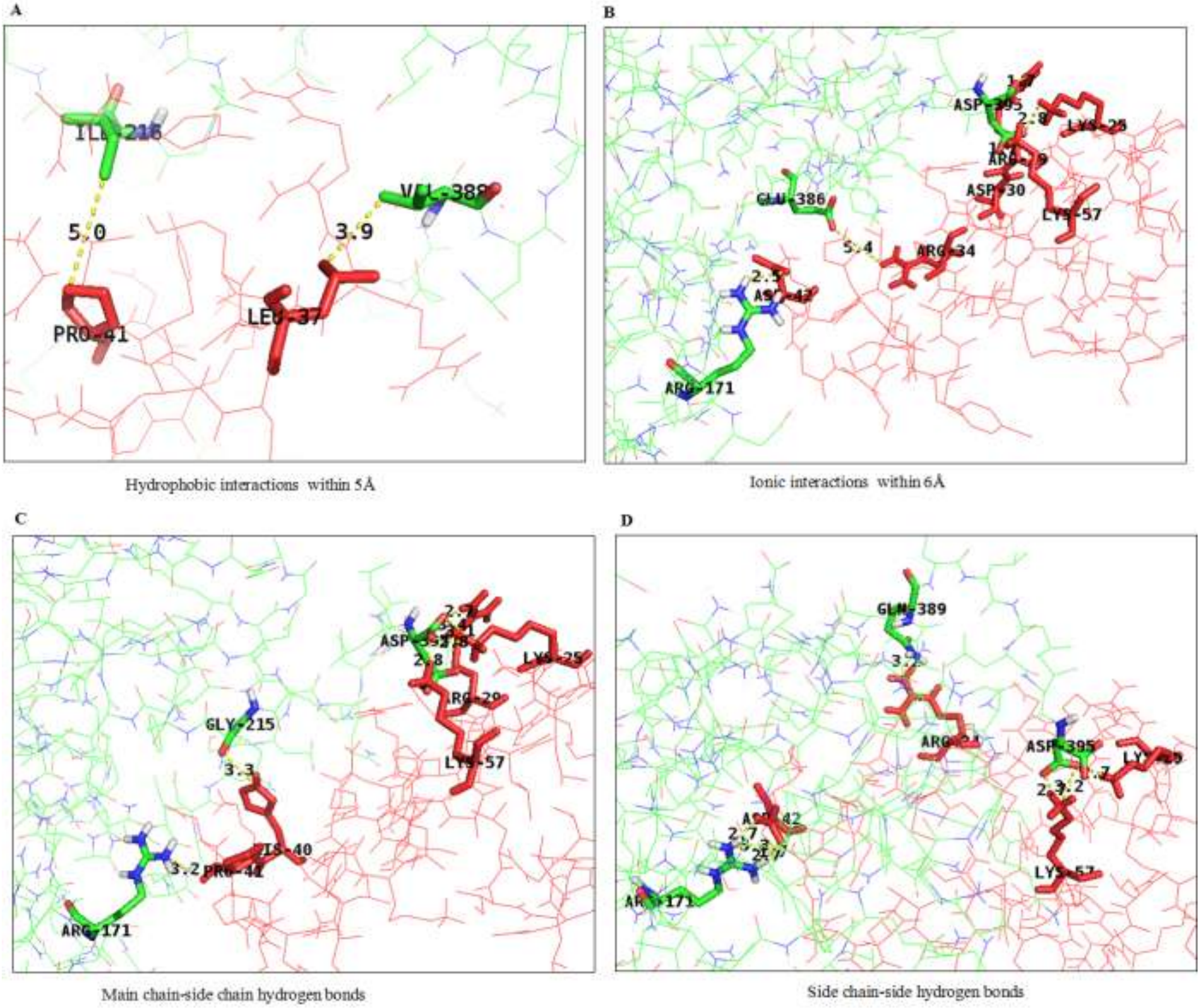


Figure 26: HSPA8 ATPase domain_Linker:DNAJC19 J domain Complex. (A) Complex structure of ATPase domain linker region represented as a transparent surface colored in cyan with J domain shown as lines. The HPD motif is colored magenta, helixes II and III as green and yellow respectively. Helixes I and IV are colored as red. (B) towards the Y-axis. (C) The regions demarcated within the box in (B) was zoomed out to show important residues at the exposed ATPase-J domain interface predicted to interact using the Protein Interaction Calculator server (Tina *et al*., 2007). HSP70 residues are displayed as sticks and various interacting residues were mapped and labeled in black accordingly on both domains. Figure was generated in PyMol (Delano and Bromberg, 2004).



Hydrophobic interactions within 5Å

Ionic interactions within 6Å

Main chain-side chain hydrogen bonds

Side chain-side hydrogen bonds

Figure 27: Protein-protein interactions of HSPA8-DNAJC19 complex. Both HSP70 ATPase domain-linker and HSP40 J domain are displayed as lines and colored in green and red respectively. Exposed residues at the complex interface predicted to be involved in the various intermolecular interactions using PIC were shown and labeled as sticks. Pictures were rendered in PyMol (Delano and Bromberg, 2004).

CHAPTER FIVE: Conclusions and Future Prospects

Highly conserved residues on HSP40 J domain have been identified. Of interest were those highly conserved residues outside the HPD motif. Variations in the tripeptide KFK motif in the sequence alignment across the Type III members and many others on helix III could be critical for defining specific HSP40-HSP70 partnership. Only in those proteins localized in the cytosol was this motif conserved and mainly absent in those localized in the endoplasmic reticulum. This may explain why endoplasmic reticulum localized proteins could not be swapped with those localized in the cytosol in domain swapping experiment (Schlenstedt *et al.*, 1995). While highly conserved residues on both helixes II and III could mediate the general interactions of J domain-ATPase activity, determinant residues for specific partnership could rely on those that are less conserved especially those on helix III (Hennessy *et al.*, 2000). This was in agreement with the binding interface found in the predicted docked complex models. Highly conserved residues on helix II bind to residues at the underside pocket of HSP70 ATPase domain as well as linker residues, especially ASP 395 whereas helix III residues also formed part of the interface architecture. High residue variation in Type III HSP40s J domain could be critical for such specificity since J domains structure is thought to be conserved (Hennessy *et al.*, 2000). Thus, while both Type I and II J proteins can bind HSP70 non-specifically, Type III J proteins may not. Also, highly conserved hydrophobic residues on both helixes I and IV were probably responsible for maintaining the structure of the J domain rather than mediate direct interactions with partner HSP70s. However, the highly conserved TYR 64 in DNAJC2 which formed a cation-π interactions with ARG 168 (HSPA14) on helix IV, found on the interface of the predicted complexes, could play a role in J domain-ATPase interaction. The clustering pattern observed from the phylogenetic analyses of the J domain was very similar to previous analysis of the full length protein sequences though some of the proteins did not cluster according to the predicted subcellular localizations (Hageman and Kampinga, 2009). This could suggest post-translational trafficking of proteins or possibly share common catalytic functions while localized at different positions within the cell (Qiu *et al.*, 2006). The high level of conservation in the ATPase domain in HSP70s allowed for the proteins to cluster based on their sub-cellular localization, suggesting that these proteins were not products of gene duplication as mostly found among J proteins and especially in Type III HSP40s. In all, the J domain and HSP70 ATPase domain could be the main factor for defining HSP40 and HSP70 families together with other domains present in the proteins.

Finally, it now remains to be investigated through *in vitro* experimental procedures such as site directed scanning mutagenesis analyses if the predictions from this study corroborate experimental results. Uncharacterized conserved residues including the GLY and SER residues in the turn between helixes II & III, the last SER residue on helix III just before the beginning of helix IV as well as the ASP residue at the beginning of helix IV should be investigated. Interestingly, all these residues were highly conserved across the different classes of HSP40 J proteins. The roles of both GLY 215 and ASP 395 in the HSP70s should be studied since their conservation and positions within the HSP70s lied at the interface of the protein and were involved in a network of interactions with the J domain residues as seen in this study.

REFERENCES

Abdul, K. M., Terada, K., Gotoh, T., Hafizur, R. M., & Mori, M. (2002). Characterization and functional analysis of a heart-enriched DnaJ/ Hsp40 homolog dj4/DjA4. *Cell stress chaperones*, *7*(2), 156–166.

Allen, J. W., Dix, D. J., Collins, B. W., Merrick, B. A., He, C., Selkirk, J. K., Poorman-Allen, P., et al. (1996). HSP70-2 is part of the synaptonemal complex in mouse and hamster spermatocytes. *Chromosoma*, *104*(6), 414–421.

Andersen, J. S., Lam, Y. W., Leung, A. K. L., Ong, S.-E., Lyon, C. E., Lamond, A. I., & Mann, M. (2005). Nucleolar proteome dynamics. *Nature*, *433*(7021), 77–83.

Andersen, J. S., Lyon, C. E., Fox, A. H., Leung, A. K. L., Lam, Y. W., Steen, H., Mann, M., et al. (2002). Directed proteomic analysis of the human nucleolus. *Current Biology*, *12*(1), 1–11.

Arnold, K., Bordoli, L., Kopp, J., & Schwede, T. (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics (Oxford, England)*, *22*(2), 195–201. doi:10.1093/bioinformatics/bti770

Auger, I., & Roudier, J. (1997). A Function for the QKRAA Amino Acid Motif: Mediating Binding of DnaJ to DnaK Implications for the Association of Rheumatoid Arthritis with HLA-DR4, *99*(8), 1818–1822.

Bailey, T L, & Gribskov, M. (1998). Combining evidence using p-values: application to sequence homology searches. *Bioinformatics (Oxford, England)*, *14*(1), 48–54.

Bailey, Timothy L, Williams, N., Misleh, C., & Li, W. W. (2006). MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic acids research*, *34*(Web Server issue), W369–73. doi:10.1093/nar/gkl198

Benkert, P., Biasini, M., & Schwede, T. (2011). Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics (Oxford, England)*, *27*(3), 343–50. doi:10.1093/bioinformatics/btq662

Berjanskii, M. V, Riley, M. I., Xie, A., Semenchenko, V., Folk, W. R., & Van Doren, S. R. (2000). NMR structure of the N-terminal J domain of murine polyomavirus T antigens. Implications for DnaJ-like domains and for mutations of T antigens. *The Journal of Biological Chemistry*, *275*(46), 36094–36103.

Bhattacharyya, T., Karnezis, A. N., Murphy, S. P., Hoang, T., Freeman, B. C., Phillips, B., & Morimoto, R. I. (1995). Cloning and subcellular localization of human mitochondrial hsp70. *The Journal of Biological Chemistry*, *270*(4), 1705–1710.

Blum, T., Briesemeister, S., & Kohlbacher, O. (2009). MultiLoc2: integrating phylogeny and Gene Ontology terms improves subcellular protein localization prediction. *BMC bioinformatics*, *10*, 274. doi:10.1186/1471-2105-10-274

Botha, M., Pesce, E. R., & Blatch, G. L. (2007). The Hsp40 proteins of Plasmodium falciparum and other apicomplexa: Regulating chaperone power in the parasite and the host. The International Journal of Biochemistry & Cell Biology 39, 1781–1803

Brocchieri, L., Conway de Macario, E., & Macario, A. J. L. (2008). hsp70 genes in the human genome: Conservation and differentiation patterns predict a wide array of overlapping and specialized functions. *BMC evolutionary biology*, *8*, 19. doi:10.1186/1471-2148-8-19

Chapple, J., & Cheetham, M. (2003). The chaperone environment at the cytoplasmic face of the endoplasmic reticulum can modulate rhodopsin processing and inclusion formation. *The Journal of Biological Chemistry*, *278*(21), 19087–19094.

Cheetham, M. E. and C. A. J. (1998). Structure, function and evolution of Dnaj: conservation and adaptation of chaperone function, 28–36.

Chen, J., Huang, Y., Wu, H., Ni, X., Cheng, H., Fan, J., Gu, S., et al. (2003). Molecular cloning and characterization of a novel human J-domain protein gene (HDJ3) from the fetal brain. *Journal of Human Genetics*, *48*(5), 217–221.

Chevalier, M., Rhee, H., Elguindi, E. C., & Blond, S. Y. (2000). Interaction of murine BiP/GRP78 with the DnaJ homologue MTJ1. *The Journal of Biological Chemistry*, *275*(26), 19620–19627.

Crowley, P. B., & Golovin, A. (2005). Cation-pi interactions in protein-protein interfaces. *Proteins*, *59*(2), 231–9. doi:10.1002/prot.20417

Cunnea, P. M., Miranda-Vizuete, A., Bertoli, G., Simmen, T., Damdimopoulos, A. E., Hermann, S., Leinonen, S., et al. (2003). ERdj5, an endoplasmic reticulum (ER)-resident protein containing DnaJ and thioredoxin domains, is expressed in secretory cells or following ER stress. *The Journal of Biological Chemistry*, *278*(2), 1059–1066.

Cupp-vickery, J. R., & Vickery, L. E. (2000). Crystal Structure of Hsc20 , a J-type Co-chaperone from Escherichia coli. *Structure*, *1*. doi:10.1006/jmbi.2000.4252

Cyr, D. M., Langer, T., & Douglas, M. G. (1994). DnaJ-like proteins: molecular chaperones and specific regulators of Hsp70. *Trends in biochemical sciences*, *19*(4), 176–81.

Davis, A. R., Alevy, Y. G., Chellaiah, A., Quinn, M. T., & Mohanakumar, T. (1998). Characterization of HDJ-2, a human 40 kD heat shock protein. *The international journal of biochemistry cell biology*, *30*(11), 1203–1221.

De Vries, S. J., & Bonvin, A. M. J. J. (2011). CPORT: a consensus interface predictor and its performance in prediction-driven docking with HADDOCK. *PloS one*, *6*(3), e17695. doi:10.1371/journal.pone.0017695

De Vries, S. J., Van Dijk, M., & Bonvin, A. M. J. J. (2010). The HADDOCK web server for data-driven biomolecular docking. *Nature protocols*, *5*(5), 883–97. doi:10.1038/nprot.2010.32

Delano, W. L., & Bromberg, S. (2004). PyMOL User's Guide. *DeLano Scientific LLC*, 1–66.

Di Luccio, E., & Koehl, P. (2011). A quality metric for homology modeling: the H-factor. *BMC bioinformatics*, *12*(1), 48. doi:10.1186/1471-2105-12-48

Diefenbach, J., & Kindl, H. (2000). The membrane-bound DnaJ protein located at the cytosolic site of glyoxysomes specifically binds the cytosolic isoform 1 of Hsp70 but not other Hsp70 species. *The Federation of European Biochemical Societies Journal*, *267*(3), 746–754.

Dimmic, M. W., Rest, J. S., Mindell, D. P., & Goldstein, R. A. (2002). rtREV□: An Amino Acid Substitution Matrix for Inference of Retrovirus and Reverse Transcriptase Phylogeny, 65–73. doi:10.1007/s00236-001-2304-y

Elmar Krieger, Sander B. Nabuurs, and G. V. (2003). *Homology modeling*. (P. E. B. and H. Weissig, Ed.)*Structural Bioinformatics* (pp. 507–521). Wiley-Liss, Inc.

Faure, G., Bornot, A., & De Brevern, A. G. (2008). Protein contacts, inter-residue interactions and side-chain modelling. *Biochimie*, *90*(4), 626–39. doi:10.1016/j.biochi.2007.11.007

Feder, M. E., & Hofmann, G. E. (1999). Heat-shock proteins, molecular chaperones, and the stress response: evolutionary and ecological physiology. *Annual review of physiology*, *61*(c), 243–82. doi:10.1146/annurev.physiol.61.1.243

Felsenstein, J. (1985). Confidence-Limits on Phylogenies - an Approach Using the Bootstrap. *Evolution*, *39*(4), 783–791. Retrieved from http://www.jstor.org/stable/2408678

Feng, H. L., Sandlow, J. I., & Sparks, A. E. (2001). Decreased expression of the heat shock protein hsp70-2 is associated with the pathogenesis of male infertility. *Fertility and Sterility*, *76*(6), 1136–1139.

Fourie, A. M., Peterson, P. A., & Yang, Y. (2001). Characterization and regulation of the major histocompatibility complex–encoded proteins Hsp70-Hom and Hsp70-1/2. *Cell stress chaperones*, *6*(3), 282–295.

Freeman, B. C., & Morimoto, R. I. (1996). The human cytosolic molecular chaperones hsp90, hsp70 (hsc70) and hdj-1 have distinct roles in recognition of a non-native protein and protein refolding. *The EMBO journal*, *15*(12), 2969–79.

Gallivan, J. P., & Dougherty, D. a. (1999). Cation-pi interactions in structural biology. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(17), 9459–64.

Garimella, R., Liu, X., Qiao, W., Liang, X., Zuiderweg, E. R. P., Riley, M. I., & Van Doren, S. R. (2006). Hsc70 contacts helix III of the J domain from polyomavirus T antigens: addressing a dilemma in the chaperone hypothesis of how they release E2F from pRb. *Biochemistry*, *45*(22), 6917–6929.

Genevaux, P, Wawrzynow, a, Zylicz, M., Georgopoulos, C., & Kelley, W. L. (2001). DjlA is a third DnaK co-chaperone of Escherichia coli, and DjlA-mediated induction of colanic acid capsule requires DjlA-DnaK interaction. *The Journal of biological chemistry*, *276*(11), 7906–12. doi:10.1074/jbc.M003855200

Genevaux, Pierre, Schwager, F., Georgopoulos, C., & Kelley, W. L. (2002). Scanning mutagenesis identifies amino acid residues essential for the in vivo activity of the Escherichia coli DnaJ (Hsp40) J-domain. *Genetics*, *162*(3), 1045–53.

Greener, T., Zhao, X., Nojima, H., Eisenberg, E., & Greene, L. E. (2000). Role of cyclin G-associated kinase in uncoating clathrin-coated vesicles from non-neuronal cells. *The Journal of Biological Chemistry*, *275*(2), 1365–1370.

Gromiha, M. M., & Selvaraj, S. (2004). Inter-residue interactions in protein folding and stability. *Progress in biophysics and molecular biology*, *86*(2), 235–77. doi:10.1016/j.pbiomolbio.2003.09.003

Guda, C. (2006). pTARGET: a web server for predicting protein subcellular localization. *Nucleic acids research*, *34*(Web Server issue), W210–3. doi:10.1093/nar/gkl093

Hageman, J., & Kampinga, H. H. (2009). Computational analysis of the human HSPH/HSPA/DNAJ family and cloning of a human HSPH/HSPA/DNAJ expression library. *Cell stress & chaperones*, *14*(1), 1–21. doi:10.1007/s12192-008-0060-2

Hageman, J., Van Waarde, M. a W. H., Zylicz, A., Walerych, D., & Kampinga, H. H. (2011). The diverse members of the mammalian HSP70 machine show distinct chaperone-like activities. *The Biochemical journal*, *435*(1), 127–42. doi:10.1042/BJ20101247

Harm H. Kampinga* and Elizabeth A. Craig. (2010). In format provided by Kampinga and Craig (August 2010), (August).

Hellman, R., Vanhove, M., Lejeune, a, Stevens, F. J., & Hendershot, L. M. (1999). The in vivo association of BiP with newly synthesized proteins is dependent on the rate and stability of folding and not simply on the presence of sequences that can bind to BiP. *The Journal of cell biology*, *144*(1), 21–30.

Hennessy, F, Cheetham, M. E., Dirr, H. W., & Blatch, G. L. (2000). Analysis of the levels of conservation of the J domain among the various types of DnaJ-like proteins. *Cell stress & chaperones*, *5*(4), 347–58.

Hennessy, Fritha, Boshoff, A., & Blatch, G. L. (2005). Rational mutagenesis of a 40 kDa heat shock protein from Agrobacterium tumefaciens identifies amino acid residues critical to its in vivo function. *The international journal of biochemistry & cell biology*, *37*(1), 177–91. doi:10.1016/j.biocel.2004.06.009

Hennessy, Fritha, Nicoll, W. S., Zimmermann, R., Cheetham, M. E., & Blatch, G. L. (2005). Not all J domains are created equal: Implications for the specificity of Hsp40 – Hsp70 interactions, 1697–1709. doi:10.1110/ps.051406805.Hsp70

Hildebrand, A., Remmert, M., Biegert, A., & Söding, J. (2009). Fast and accurate automatic structure prediction with HHpred. *Proteins*, *77 Suppl 9*, 128–32. doi:10.1002/prot.22499

Höglund, A., Dönnes, P., Blum, T., Adolph, H.-W., & Kohlbacher, O. (2006). MultiLoc: prediction of protein subcellular localization using N-terminal targeting sequences, sequence motifs and amino acid composition. *Bioinformatics (Oxford, England)*, *22*(10), 1158–65. doi:10.1093/bioinformatics/btl002

Horton, P., Park, K.-J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C. J., & Nakai, K. (2007). WoLF PSORT: protein localization predictor. *Nucleic acids research*, *35*(Web Server issue), W585–7. doi:10.1093/nar/gkm259

Imai, Y., Soda, M., Hatakeyama, S., Akagi, T., Hashikawa, T., Nakayama, K. I., & Takahashi, R. (2002). CHIP is associated with Parkin, a gene responsible for familial Parkinson's disease, and enhances its ubiquitin ligase activity. *Molecular Cell*, *10*(1), 55–67. doi:10.1016/S1097-2765(02)00583-X

Izawa, I., Nishizawa, M., Ohtakara, K., Ohtsuka, K., Inada, H., & Inagaki, M. (2000). Identification of Mrj, a DnaJ/Hsp40 family protein, as a keratin 8/18 filament regulatory protein. *The Journal of Biological Chemistry*, *275*(44), 34521–34527.

Jen-Sing L., Shu-Ru, K., Alexander, M. M., Douglas, M. C., Jack, D. G., Thomas, R. B., & Louise, T. C. (1998). Human Hsp70 and Hsp40 Chaperone Proteins Facilitate Human Papillomavirus-11 E1 Protein Binding to the Origin and Stimulate Cell-free DNA Replication. The Journal of Biological Chemistry, 273(46), 30704–30712.

Jiang, J., Maes, E. G., Taylor, A. B., Wang, L., & Hinck, A. P. (2007). Structural Basis of J Cochaperone Binding and Regulation of Hsp70. *molecular cell*, *28*(3), 422–433.

Jiang, J., Prasad, K., Lafer, E. M., & Sousa, R. (2005). Structural basis of interdomain communication in the Hsc70 chaperone. *Molecular cell*, *20*(4), 513–24. doi:10.1016/j.molcel.2005.09.028

Jiang, J., Taylor, A. B., Prasad, K., Ishikawa-Brush, Y., Hart, P. J., Lafer, E. M., & Sousa, R. (2003). Structure-function analysis of the auxilin J-domain reveals an extended Hsc70 interaction interface. *Biochemistry*, *42*(19), 5748–5753.

Jikul, G. & Li Y. (2008). A Heat-Shock Protein 40, DNAJB13, is an Axoneme-Associated Component in Mouse Spermatozoa. Molecular Reproduction and Development, 75, 1379–1386. doi 10.1002/mrd.20874

Kampinga, H. H., & Craig, E. A. (2010). The HSP70 chaperone machinery: J proteins as drivers of functional specificity. *Nature Reviews Molecular Cell Biology*, *11*(8), 579–592.

Katoh, K., & Frith, M. C. (2012). Adding unaligned sequences into an existing alignment using MAFFT and LAST. *Bioinformatics (Oxford, England)*, 1–3. doi:10.1093/bioinformatics/bts578

Keshava Prasad, T. S., Goel, R., Kandasamy, K., Keerthikumar, S., Kumar, S., Mathivanan, S., Telikicherla, D., et al. (2009). Human Protein Reference Database--2009 update. *Nucleic acids research*, *37*(Database issue), D767–72. doi:10.1093/nar/gkn892

Kim, H.-Y., Ahn, B.-Y., & Cho, Y. (2001). Structural basis for the inactivation of retinoblastoma tumor suppressor by SV40 large T antigen. *the The European Molecular Biology Organization Journal*, *20*(1&2), 295–304.

Kluck, C. J., Patzelt, H., Genevaux, P., Brehmer, D., Rist, W., Schneider-Mergener, J., Bukau, B., et al. (2002). Structure-function analysis of HscC, the Escherichia coli member of a novel subfamily of specialized Hsp70 chaperones. *The Journal of biological chemistry*, *277*(43), 41060–9. doi:10.1074/jbc.M206520200

Korth, M. J., Lyons, C. N., Wambach, M., & Katze, M. G. (1996). Cloning, expression, and cellular localization of the oncogenic 58-kDa inhibitor of the RNA-activated human and mouse protein kinase. *Gene*, *170*(2), 181–188.

Kroczynska, B., Evangelista, C. M., Samant, S. S., Elguindi, E. C., & Blond, S. Y. (2004). The SANT2 domain of the murine tumor cell DnaJ-like protein 1 human homologue interacts with alpha1-antichymotrypsin and kinetically interferes with its serpin inhibitory activity. *The Journal of Biological Chemistry*, *279*(12), 11432–11443.

Kurihara, T., & Silver, P. (1993). Suppression of a sec63 mutation identifies a novel component of the yeast endoplasmic reticulum translocation apparatus. *Molecular Biology of the Cell*, *4*(9), 919–930.

Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*, *26*(2), 283–291. doi:10.1107/S0021889892009944

Lau, P. P., Villanueva, H., Kobayashi, K., Nakamuta, M., Chang, B. H., & Chan, L. (2001). A DnaJ protein, apobec-1-binding protein-2, modulates apolipoprotein B mRNA editing. *The Journal of Biological Chemistry*, *276*(49), 46445–46452.

Lee, S., & Tsai, F. T. F. (2005). Molecular chaperones in protein quality control. *Journal of biochemistry and molecular biology*, *38*(3), 259–65.

Lyman, S. K., & Schekman, R. (1995). Interaction between BiP and Sec63p is required for the completion of protein translocation into the ER of Saccharomyces cerevisiae. *The Journal of cell biology*, *131*(5), 1163–71.

Mapa, K., Sikor, M., Kudryavtsev, V., Waegemann, K., Kalinin, S., Seidel, C. a M., Neupert, W., et al. (2010). The conformational dynamics of the mitochondrial Hsp70 chaperone. *Molecular cell*, *38*(1), 89–100. doi:10.1016/j.molcel.2010.03.010

May, A., & Zacharias, M. (2007). Energy minimization in low-frequency normal modes to efficiently allow for global flexibility during systematic protein–protein docking, (April), 794–809. doi:10.1002/prot

Mayer, M. P., & Bukau, B. (2005). Hsp70 chaperones: cellular functions and molecular mechanism. *Cellular and molecular life sciences□: CMLS*, *62*(6), 670–84. doi:10.1007/s00018-004-4464-6

Melo, F, & Feytmans, E. (1998). Assessing protein structures with a non-local atomic interaction energy. *Journal of molecular biology*, *277*(5), 1141–52. doi:10.1006/jmbi.1998.1665

Melo, Francisco. (2007). Fold assessment for comparative protein structure modeling, 2412–2426. doi:10.1110/ps.072895107.)

Melville, M. W., Tan, S. L., Wambach, M., Song, J., Morimoto, R. I., & Katze, M. G. (1999). The cellular inhibitor of the PKR protein kinase, P58(IPK), is an influenza virus-activated co-chaperone that modulates heat shock protein 70 activity. *The Journal of Biological Chemistry*, *274*(6), 3797–3803.

Mercier, P. A., Winegarden, N. A., & Westwood, J. T. (1999). Human heat shock factor 1 is predominantly a nuclear protein before and after heat stress. *Journal of Cell Science*, *112 (Pt 1*(Pt 16), 2765–2774.

Minami, Y., Höhfeld, J., Ohtsuka, K., & Hartl, F. U. (1996). Regulation of the heat-shock protein 70 reaction cycle by the mammalian DnaJ homolog, Hsp40. *The Journal of biological chemistry*, *271*(32), 19617–24.

Morris, J. A., Dorner, A. J., Edwards, C. A., Hendershot, L. M., & Kaufman, R. J. (1997). Immunoglobulin binding protein (BiP) function is required to protect cells from endoplasmic reticulum stress but is not required for the secretion of selective proteins. *The Journal of Biological Chemistry*, *272*(7), 4327–34.

Nicoll, W. S., Botha, M., McNamara, C., Schlange, M., Pesce, E.-R., Boshoff, a, Ludewig, M. H., et al. (2007). Cytosolic and ER J-domains of mammalian and parasitic origin can functionally interact with DnaK. *The international journal of biochemistry & cell biology*, *39*(4), 736–51. doi:10.1016/j.biocel.2006.11.006

Nogami, M., Takatsu, A., Endo, N., & Ishiyama, I. (2000). Immunohistochemical localization of heat shock protein 70 in the human medulla oblongata in forensic autopsies. *Legal medicine Tokyo Japan*, *2*(1), 198–203.

Ohtsuka, K, & Hata, M. (2000). Molecular chaperone function of mammalian Hsp70 and Hsp40--a review. *International journal of hyperthermia: the official journal of European Society for Hyperthermic Oncology, North American Hyperthermia Group*, *16*(3), 231–45.

Ohtsuka, K, & Suzuki, T. (2000). Roles of molecular chaperones in the nervous system. *Brain research bulletin*, *53*(2), 141–6.

Ohtsuka, Kenzo, & Hata, M. (2000). Mammalian HSP40/DNAJ homologs: cloning of novel cDNAs and a proposal for their classification and nomenclature. *Cell stress chaperones*, *5*(2), 98–112.

Olsen, J. V, Blagoev, B., Gnad, F., Macek, B., Kumar, C., Mortensen, P., & Mann, M. (2006). Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell*, *127*(3), 635–48. doi:10.1016/j.cell.2006.09.026

Otterson, G. A., Flynn, G. C., Kratzke, R. A., Coxon, A., Johnston, P. G., & Kaye, F. J. (1994). Stch encodes the “ATPase core” of a microsomal stress 70 protein. *the The European Molecular Biology Organization Journal*, *13*(5), 1216–1225.

Otto, H., Conz, C., Maier, P., Wölfle, T., Suzuki, C. K., Jenö, P., Rücknagel, P., et al. (2005). The chaperones MPP11 and Hsp70L1 form the mammalian ribosome-associated complex. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(29), 10064–10069.

Pawlowski, M., Gajda, M. J., Matlak, R., & Bujnicki, J. M. (2008). MetaMQAP: a meta-server for the quality assessment of protein models. *BMC bioinformatics*, *9*, 403. doi:10.1186/1471-2105-9-403

Pei, J., Kim, B.-H., & Grishin, N. V. (2008). PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic acids research*, *36*(7), 2295–300. doi:10.1093/nar/gkn072

Pellecchia, M., Szyperski, T., Wall, D., Georgopoulos, C., & Wüthrich, K. (1996). NMR structure of the J-domain and the Gly/Phe-rich region of the Escherichia coli DnaJ chaperone. *Journal of Molecular Biology*, *260*(2), 236–250.

Qiu, X.-B., Shao, Y.-M., Miao, S., & Wang, L. (2006a). The diversity of the DnaJ/Hsp40 family, the crucial partners for Hsp70 chaperones. *Cellular and molecular life sciences: CMLS*, *63*(22), 2560–70. doi:10.1007/s00018-006-6192-6

Qiu, X.-B., Shao, Y.-M., Miao, S., & Wang, L. (2006b). The diversity of the DnaJ/Hsp40 family, the crucial partners for Hsp70 chaperones. *Cellular and molecular life sciences: CMLS*, *63*(22), 2560–70. doi:10.1007/s00018-006-6192-6

Røsok, O., Pedeutour, F., Ree, A. H., & Aasheim, H. C. (1999). Identification and characterization of TESK2, a novel member of the LIMK/TESK family of protein kinases, predominantly expressed in testis. *Genomics*, *61*(1), 44–54.

Rosorius, O., Fries, B., Stauber, R. H., Hirschmann, N., Bevec, D., & Hauber, J. (2000). Identification of novel nuclear export and nuclear localization-related signals in human heat shock cognate protein 70. *The Journal of Biological Chemistry*, *275*(10), 12061–12068.

Sahay, A., & Shakya, M. (2010). In silico Analysis and Homology Modelling of Antioxidant Proteins of Spinach. *Journal of Proteomics & Bioinformatics*, *03*(05), 148–154. doi:10.4172/jpb.1000134

Sali, A. (2010). *MODELLER A Program for Protein Structure Modeling*.

Sánchez, R., & Sali, A. (1997). Evaluation of comparative protein structure modeling by MODELLER-3. *Proteins*, *Suppl 1*(s 1), 50–58.

Sarkar, S., Pollack, B. P., Lin, K. T., Kotenko, S. V, Cook, J. R., Lewis, A., & Pestka, S. (2001). hTid-1, a human DnaJ protein, modulates the interferon signaling pathway. *The Journal of Biological Chemistry*, *276*(52), 49034–49042.

Scheele, U., Kalthoff, C., & Ungewickell, E. (2001). Multiple interactions of auxilin 1 with clathrin and the AP-2 adaptor complex. *The Journal of Biological Chemistry*, *276*(39), 36131–36138.

Schlenstedt, G., Harris, S., Risse, B., Lill, R., & Silver, P. a. (1995). A yeast DnaJ homologue, Scj1p, can function in the endoplasmic reticulum with BiP/Kar2p via a conserved domain that specifies interactions with Hsp70s. *The Journal of cell biology*, *129*(4), 979–88.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology*, *7*(539), 539. doi:10.1038/msb.2011.75

Simpson, J. C., Wellenreuther, R., Poustka, A., Pepperkok, R., & Wiemann, S. (2009). Systematic subcellular localization of novel proteins identified by large-scale cDNA sequencing. *EMBO Reports*, *10*(12), 1363.

Smith, G. R., & Sternberg, M. J. E. (2002). Prediction of protein-protein interactions by docking methods. *Current opinion in structural biology*, *12*(1), 28–35.

Sterrenberg, J. N., Blatch, G. L., & Edkins, A. L. (2011). Human DNAJ in cancer and stem cells. *Cancer letters*, *312*(2), 129–42. doi:10.1016/j.canlet.2011.08.019

Suh, W. C., Burkholder, W. F., Lu, C. Z., Zhao, X., Gottesman, M. E., & Gross, C. a. (1998). Interaction of the Hsp70 molecular chaperone, DnaK, with its cochaperone DnaJ. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(26), 15223–8.

Suh, W. C., Lu, C. Z., & Gross, C. a. (1999). Structural features required for the interaction of the Hsp70 molecular chaperone DnaK with its cochaperone DnaJ. *The Journal of biological chemistry*, *274*(43), 30534–9.

Swain, J. F., Dinler, G., Sivendran, R., Montgomery, D. L., Stotz, M., & Gierasch, L. M. (2008). Hsp70 chaperone ligands control domain association via an allosteric mechanism mediated by the interdomain linker, *26*(1), 27–39.

Syken, J., De-Medina, T., & Münger, K. (1999). TID1, a human homolog of the Drosophila tumor suppressor l(2)tid, encodes two mitochondrial modulators of apoptosis with opposing functions. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(15), 8499–8504.

Takayama, S., Bimston, D. N., Matsuzawa, S., Freeman, B. C., Aime-Sempe, C., Xie, Z., Morimoto, R. I., et al. (1997). BAG-1 modulates the chaperone activity of Hsp70/Hsc70. *the The European Molecular Biology Organization Journal*, *16*(16), 4887–4896.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., & Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular biology and evolution*, *28*(10), 2731–9. doi:10.1093/molbev/msr121

Tastan Bishop. O, T. A. P. D. B. and J. F. (2008). Protein homology modelling and its use in South Africa 2. *South African Journal of Science*, *104*(February), 2–6.

Terada, K., & Mori, M. (2000). Human DnaJ homologs dj2 and dj3, and bag-1 are positive cochaperones of hsc70. *The Journal of Biological Chemistry*, *275*(32), 24728–24734.

Tina, K. G., Bhadra, R., & Srinivasan, N. (2007). PIC: Protein Interactions Calculator. *Nucleic acids research*, *35*(Web Server issue), W473–6. doi:10.1093/nar/gkm423

Vries, S. J. De, & Bonvin, A. M. J. J. (2008). How Proteins Get in Touch: Interface Prediction in the Study of Bio-molecular Complexes, 394–406.

Waterhouse, A. M., Procter, J. B., Martin, D. M. a, Clamp, M., & Barton, G. J. (2009). Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics (Oxford, England)*, *25*(9), 1189–91. doi:10.1093/bioinformatics/btp033

Whelan, S., Liò, P., & Goldman, N. (2001). Molecular phylogenetics: state-of-the-art methods for looking into the past. *Trends in genetics: TIG*, *17*(5), 262–72.

Wiederstein, M., & Sippl, M. J. (2007). ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic acids research*, *35*(Web Server issue), W407–10. doi:10.1093/nar/gkm290

Wiltgen, M., & Tilz, G. P. (2009). Homology modelling: a review about the method on hand of the diabetic antigen GAD 65 structure prediction. *Wiener medizinische Wochenschrift (1946)*, *159*(5-6), 112–25. doi:10.1007/s10354-009-0662-z

Xiang, S. L., Kumano, T., Iwasaki, S. I., Sun, X., Yoshioka, K., & Yamamoto, K. C. (2001). The J domain of Tpr2 regulates its interaction with the proapoptotic and cell-cycle checkpoint protein, Rad9. *Biochemical and Biophysical Research Communications*, *287*(4), 932–940.

Yang, Y., & Zhou, Y. (2008). Ab initio folding of terminal segments with secondary structures reveals the fine difference between two closely related all-atom statistical energy functions. *Protein science: a publication of the Protein Society*, *17*(7), 1212–9. doi:10.1110/ps.033480.107

Yu, C., Chen, Y., Lu, C., & Hwang, J. (2006). Prediction of Protein Subcellular Localization, *651*(December 2005), 643–651. doi:10.1002/prot

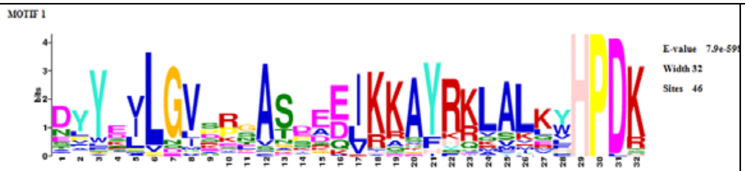
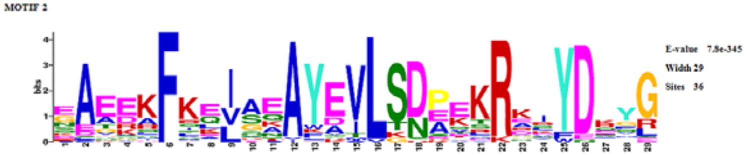
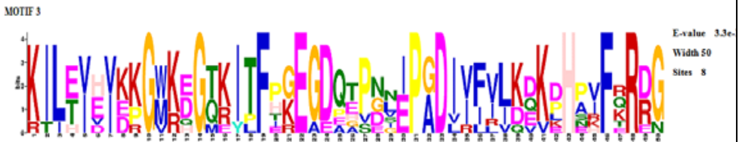
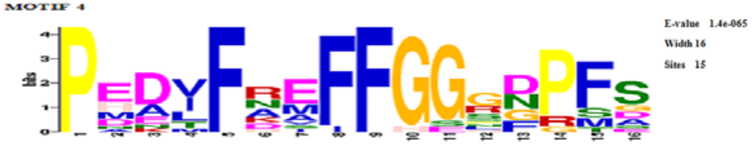
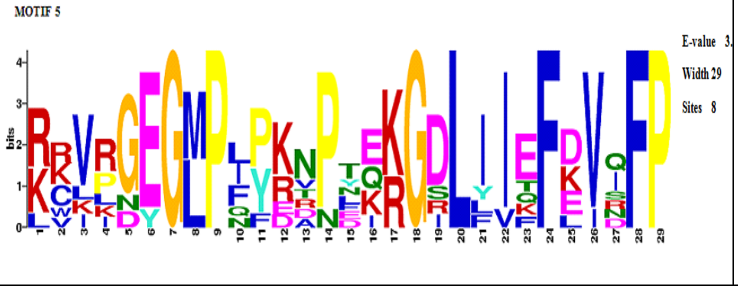
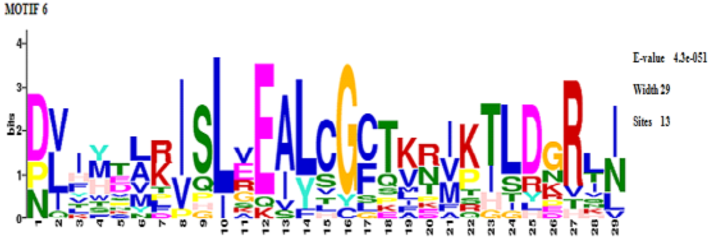
Yu, M., Haslam, R. H., & Haslam, D. B. (2000). HEDJ, an Hsp40 co-chaperone localized to the endoplasmic reticulum of human cells. *The Journal of Biological Chemistry*, *275*(32), 24984–24992.

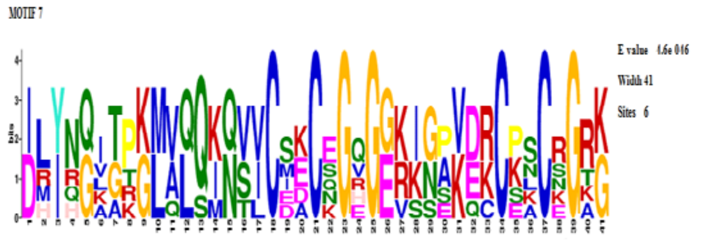
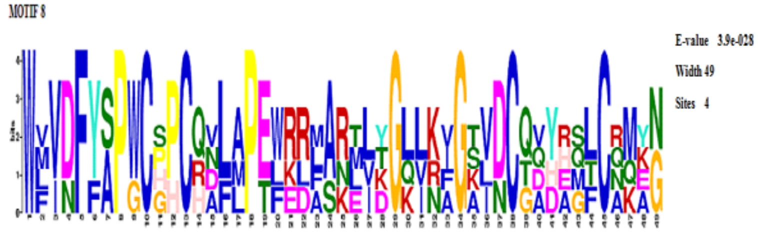
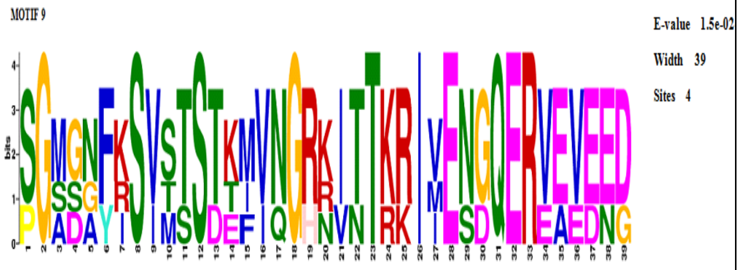
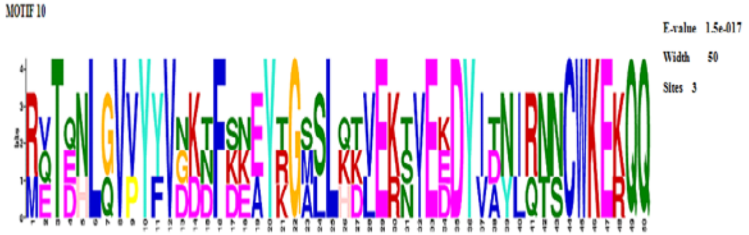
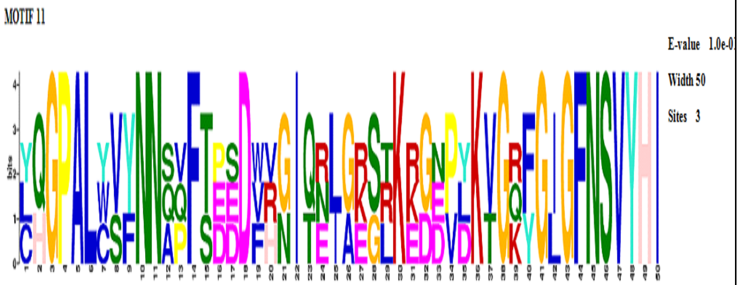
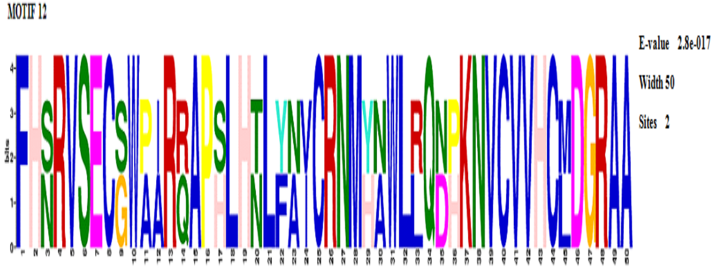
Zhang, H., Peters, K. W., Sun, F., Marino, C. R., Lang, J., Burgoyne, R. D., & Frizzell, R. A. (2002). Cysteine string protein interacts with and modulates the maturation of the cystic fibrosis transmembrane conductance regulator. *The Journal of Biological Chemistry*, *277*(32), 28948–58. doi:10.1074/jbc.M111706200

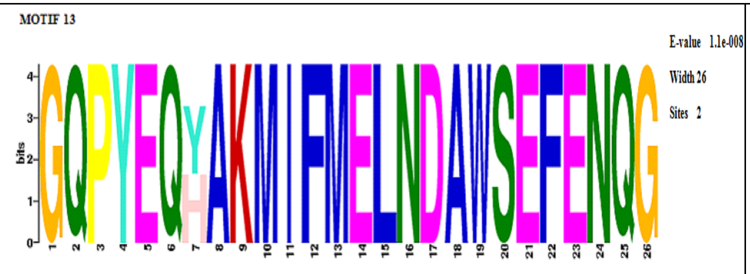
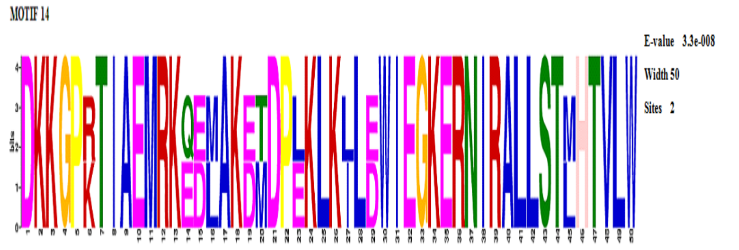
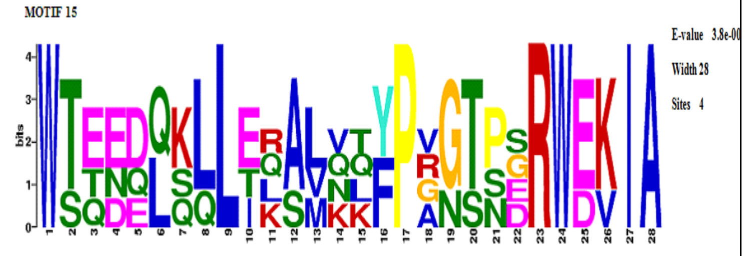
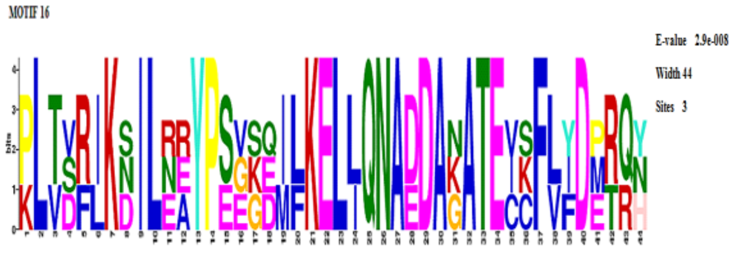
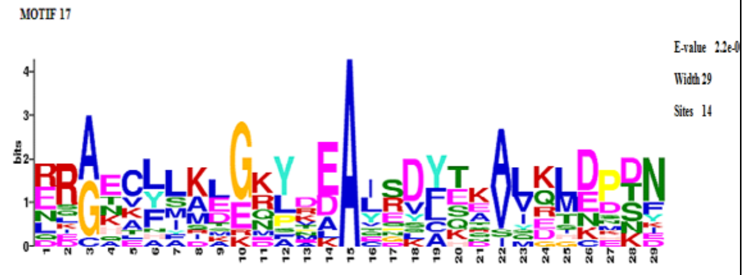
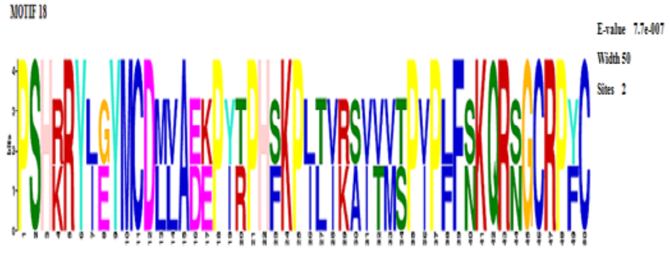
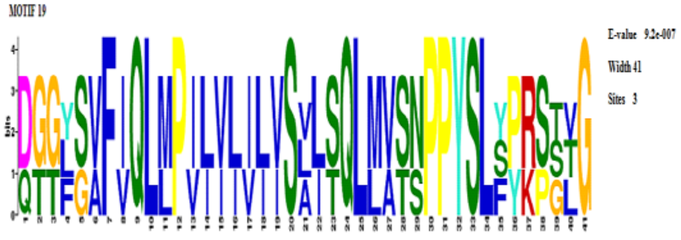
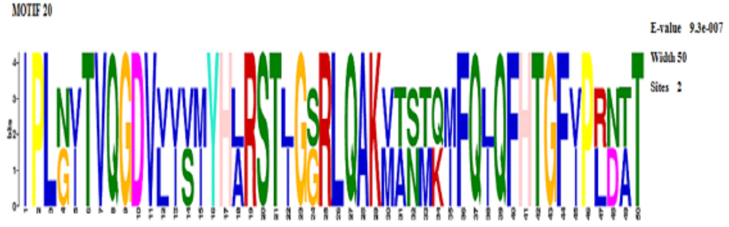
CHAPTER TWO

Appendix I: Sequence logo of motifs found in full length HSP40s using MEME.

Table 1: Sequence logo of motifs found in full length HSP40s using MEME

Motif 1		Motif 2	
Motif 3		Motif 4	
Motif 5		Motif 6	

Motif 7		Motif 8	
Motif 9		Motif 10	
Motif 11		Motif 12	

Motif 13		Motif 14	
Motif 15		Motif 16	
Motif 17		Motif 18	
Motif 19		Motif 20	

CHAPTER THREE

Appendix I: Possible templates search for homology modelling of HSP40 proteins using HHpred server

Table 2: Possible templates search for homology modelling of HSP40 proteins using HHpred server

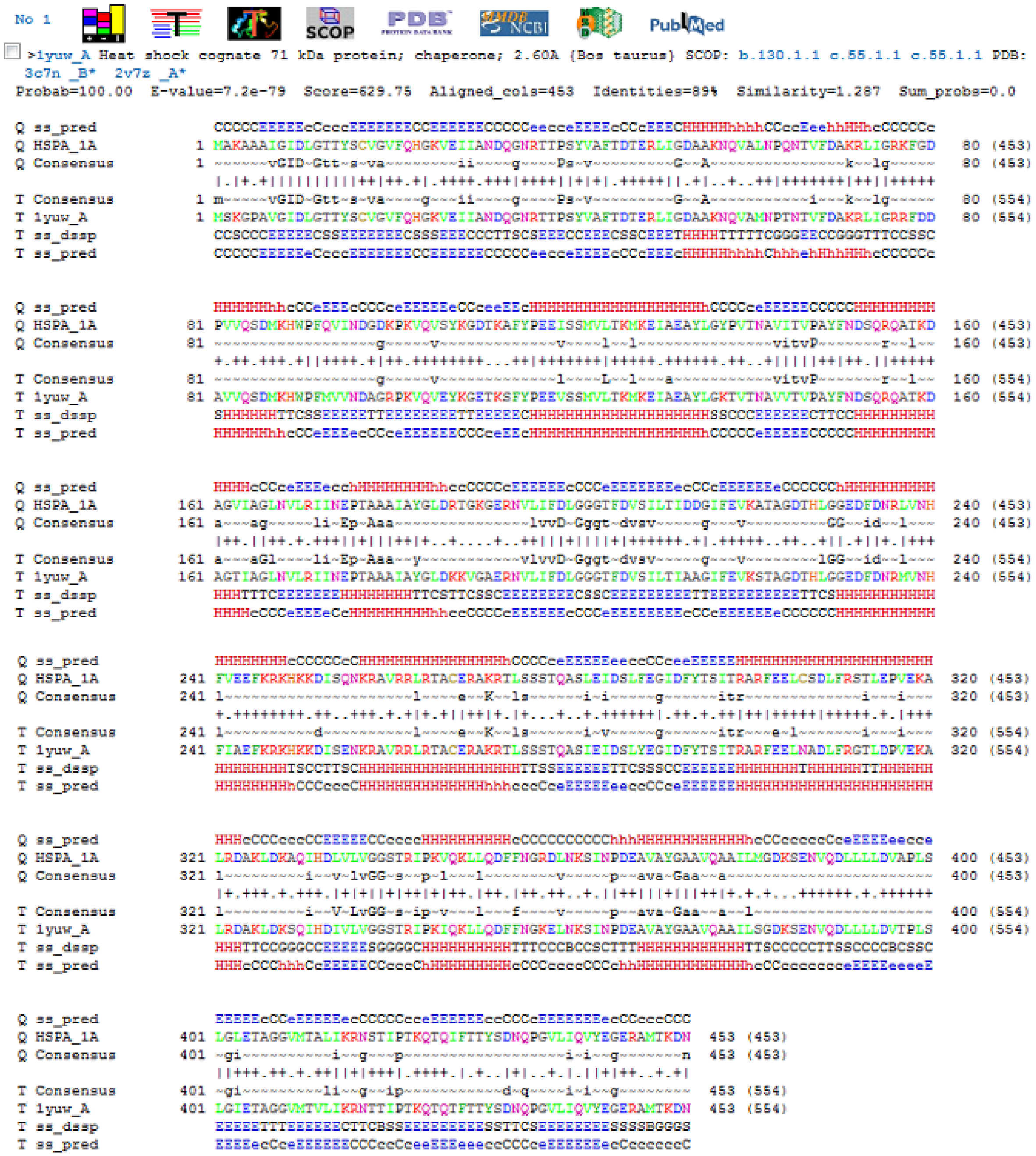
Proteins	**Family type**	**Subcellular localization**	**Possible templates**					
			PDB ID	**Organism**	**Sequence identity**	**E – value**	**Sequence coverage**	**Resolution**
DNAJA1	I	Cytosol	i) 1HDJ	*Homo sapien*	52%	8.8e-25	2 – 76 (77)	-
			ii)2OCH	*Caenorhabiditis elegan*	77%	1.4e-25	4 – 73 (73)	1.86A
			iii)2O37	*Homo sapien*	61%	4.0e-25	4 – 79 (92)	-
			iv)2CTR	*Homo sapien*	48%	1.6e-23	4 – 80 (88)	-
DNAJA2	I	Cytosol	i) 2OCH	*Caenorhabditis elegan*	61 %	$5.9e^{-24}$	5 – 73 (73)	1.86A
			ii) 2O37	*Saccharomycescerevisiae*	59 %	$5.6e^{-24}$	5 – 79 (92)	1.25A
			iii)1HDJ	*Homo sapien*	55 %	$2.4e^{-24}$	1 – 76 (77)	-
DNAJB11	II	Endoplasmic reticulum	i) 2IGW	*Homo sapien*	62%	$4.4e^{-23}$	2 – 73 (99)	-
			ii)2DN9	*Homo sapien*	60%	$1.6e^{-24}$	1 – 77 (79)	-
			iii)2CTP	*Homo sapien*	54%	$3.4e^{-23}$	1 – 76 (78)	-
			iv)1HDJ	*Homo sapien*	54%	$1.5e^{-23}$	1 – 72 (77)	-
DNAJC2	III	Nucleus	i)1HDJ	*Homo sapien*	41%	$1.5e^{-22}$	2- 67 (77)	-

			ii)2OCH	*Caenorhabiditis elegan*	38%	$1.1e^{-22}$	5 – 70 (73)	1.86A
			iii)2CTP	*Homo sapien*	34%	$7.9e^{-23}$	2 – 71 (78)	-
			iv)2DN9	*Homo sapien*	33%	$9.5e^{-24}$	1 – 72 (79)	-
DNAJC3	III	Endoplasmic reticulum	i)2Y4T	*Homo sapien*	100%	$1.2e^{-13}$	372 – 450 (450)	3.00A
			ii)2DN9	*Homo sapien*	47%	2.0^{e-24}	2 – 74 (79)	-
			iii)1HDJ	*Homo sapien*	46%	$2.8e^{-23}$	1 – 69 (77)	-
			iv)2OCH	*Caenorhabiditis elegan*	*42%*	$8.6e^{-25}$	4 – 72 (73)	1.86A
DNAJC6	III	Nucleus	i)1N4C	*Auxilin*	100%	$4.1e^{-23}$	107 – 182 (182)	-
			ii)2QW0	*Auxilin*	99%	$3.3e^{-23}$	23 – 91 (92)	1.70A
			iii)2AG7	*Arabidopsis thaliana*	30%	$6e^{-19}$	31 – 103 (106)	1.80A
DNAJC10	III	Endoplasmic reticulum	i)3APQ	*Mus musculus*	99%	$2.7e^{-21}$	2 – 72 (210)	1.84A
			ii)1BQO	*E. coli*	51%	$1.4e^{-24}$	1 – 73 (103)	-
			iii)2CTR	*Homo sapien*	48%	$1.4e^{-24}$	4 – 76 (88)	-
			iv)2EJ7	*Homo sapien*	45%	$1.1e^{-24}$	5 -80 (82)	-
DNAJC19	III	Mitochondrial	i)2GUZ	*Saccaromyces cerevisiae*	56%	$1.1e^{-21}$	4 – 70 (71)	2.00A

	ii)1HDJ	*Homo sapien*	32%	$4.6e^{-19}$	3 - 61 (77)	-
	iii)2037	*Saccaromyces cerevisiae*	31%	$2.4e^{-19}$	2 – 64 (92)	1.25A
	iv)2YS8	*Homo sapien*	31%	$3.7e^{-19}$	28 – 85 (90)	-

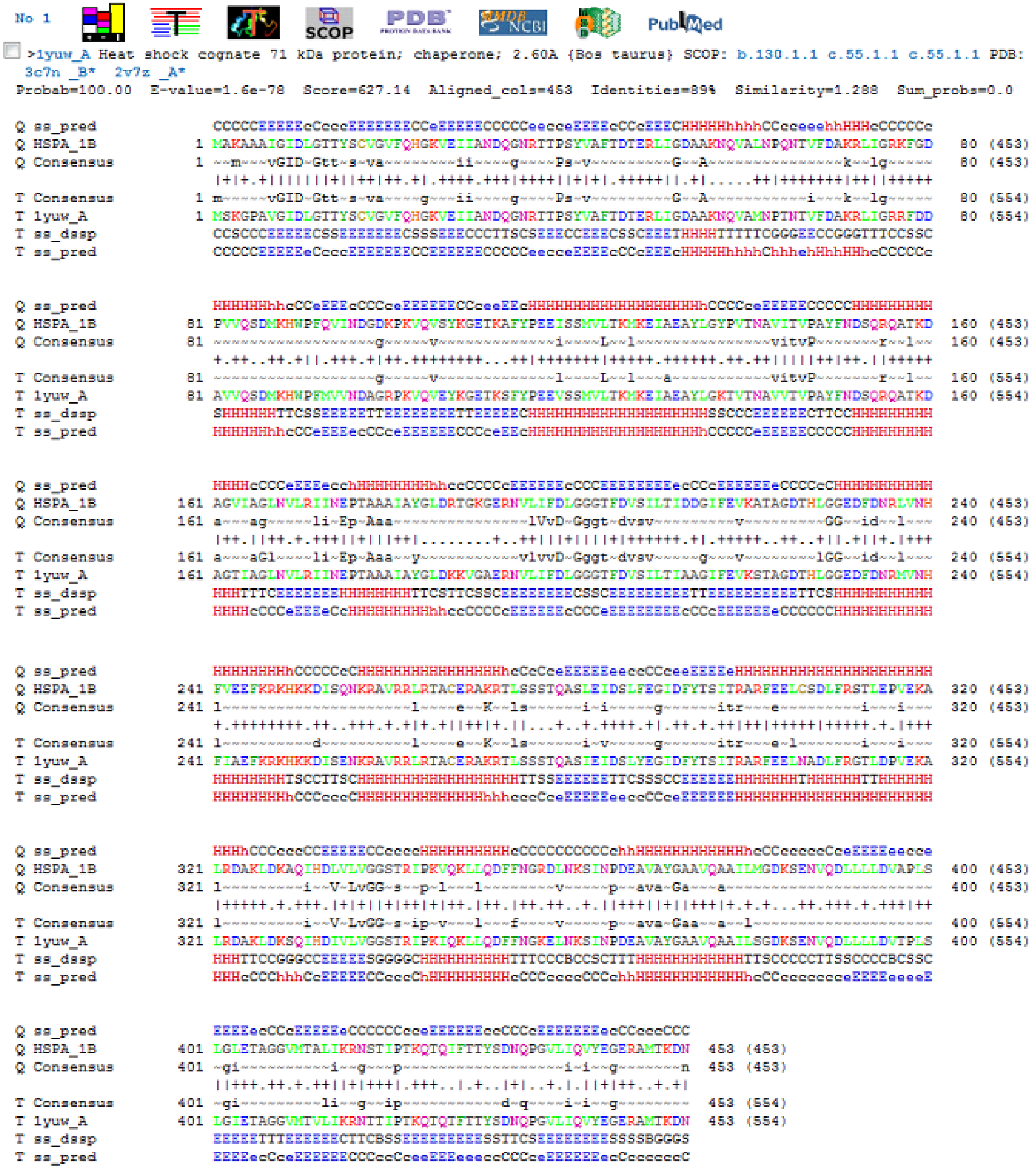
Appendix II: HSP70 ATPase domains Templates search and alignments using HHpred server

1



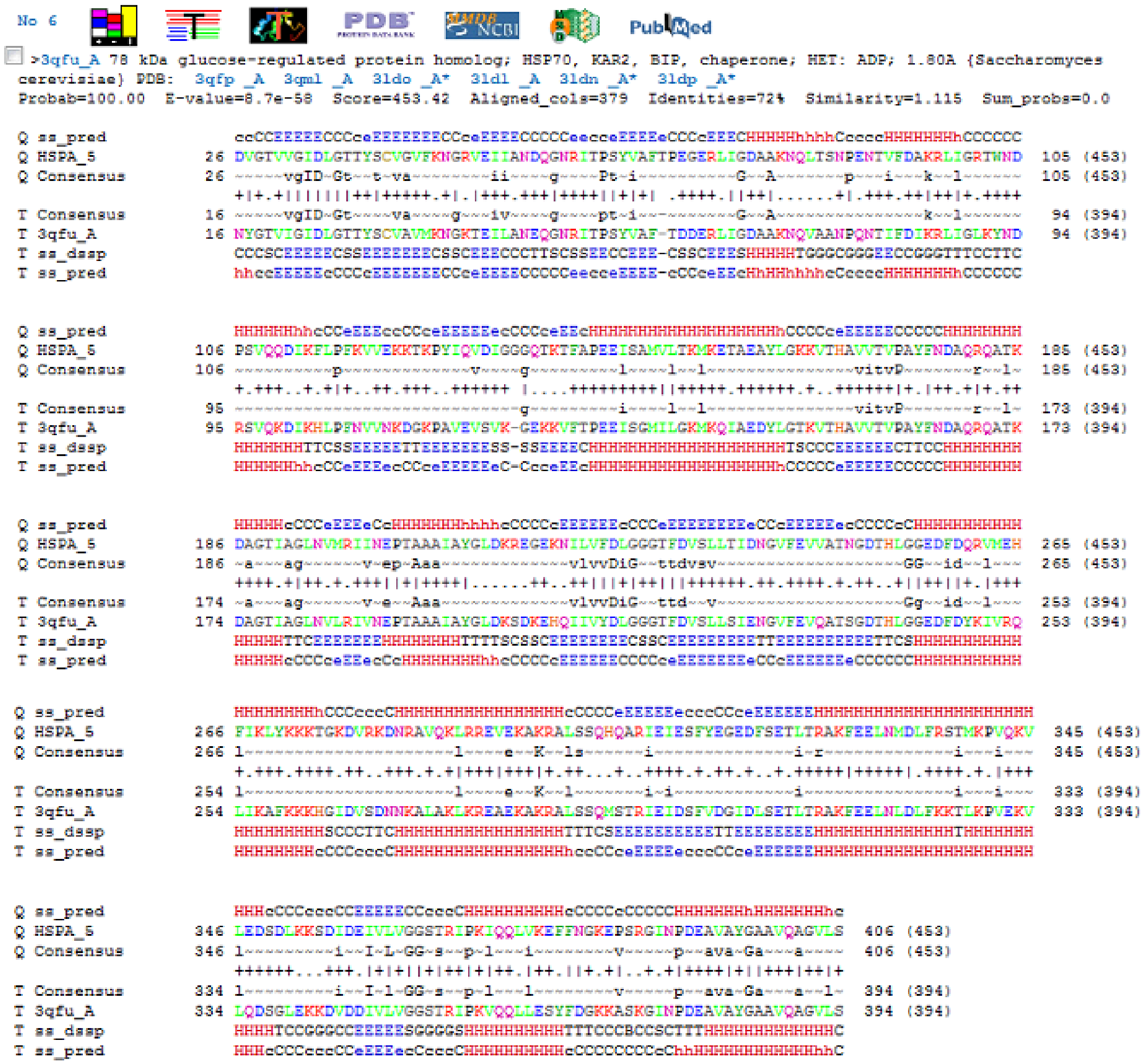
```
No 1

>1yuw_A Heat shock cognate 71 kDa protein; chaperone; 2.60A {Bos taurus} SCOP: b.130.1.1 c.55.1.1 c.55.1.1 PDB:
 3c7n _B*  2v7z _A*
Probab=100.00  E-value=7.2e-79  Score=629.75  Aligned_cols=453  Identities=89%  Similarity=1.287  Sum_probs=0.0

Q ss_pred               CCCCCEEEEEcCcccEEEEEEEECCEEEEEEECCCCCeecceEEEEcCCcEEECHHHHHhhhhCCccEeehhHHhcCCCCCc
Q HSPA_1A          1 MAKAAAIGIDLGTTYSCVGVFQHGKVEIIANDQGNRTTPSYVAFTDTERLIGDAAKNQVALNPQNTVFDAKRLIGRKFGD   80 (453)
Q Consensus        1 ~~~~~~vGID~Gtt~s~va~~~~~~~~~ii~~~~g~~~~Ps~v~~~~~~~~~~G~~A~~~~~~~~~~~~~~~k~~lg~~~~~   80 (453)
                       |.|+.+||||||||||++|++.+|.++++.+++|++++||+|+|.+++++||.+|..+..++|+++++++|++||+++++
T Consensus        1 m~~~~~vGID~Gtt~s~va~~~~g~~~ii~~~~g~~~~Ps~v~~~~~~~~~~G~~A~~~~~~~~~~~~~i~~~k~~lg~~~~~   80 (554)
T 1yuw_A         1 MSKGPAVGIDLGTTYSCVGVFQHGKVEIIANDQGNRTTPSYVAFTDTERLIGDAAKNQVAMNPTNTVFDAKRLIGRRFDD   80 (554)
T ss_dssp               CCSCCCEEEEECSSEEEEEEEECSSSEEECCCTTSCSEEECCEEECSSCEEETHHHHTTTTTCGGGEECCGGGTTTCCSSC
T ss_pred               CCCCCEEEEEeCcccEEEEEEEECCEEEEEEECCCCCeecceEEEEcCCcEEEcHHHHHhhhhChhhehHhhHHhcCCCCCc


Q ss_pred               HHHHHHhhcCCeEEEcCCCceEEEEEeCCceeEEcHHHHHHHHHHHHHHHHHHHHHhCCCCceEEEEECCCCCHHHHHHHHH
Q HSPA_1A         81 PVVQSDMKHWPFQVINDGDKPKVQVSYKGDTKAFYPEEISSMVLTKMKEIAEAYLGYPVTNAVITVPAYFNDSQRQATKD  160 (453)
Q Consensus       81 ~~~~~~~~~~~~~~~~~~~~~g~~~~~~v~~~~~~~~~~~~~~~~v~~~~l~~l~~~~~~~~~~~~~~~~~vitvP~~~~~~~~r~~l~~  160 (453)
                       +.++.+++.+||++++.+|++.++++++++...++|+++++++|+++++.++++++.++..+||||++|++.||+++++
T Consensus       81 ~~~~~~~~~~~~~~~~~~~~~g~~~~~~v~~~~~~~~~~~~~~~l~~~~L~~l~~~~a~~~~~~~~~~~~vitvP~~~~~~~~r~~l~~  160 (554)
T 1yuw_A        81 AVVQSDMKHWPFMVVNDAGRPKVQVEYKGETKSFYPEEVSSMVLTKMKEIAEAYLGKTVTNAVVTVPAYFNDSQRQATKD  160 (554)
T ss_dssp               SHHHHHHTTCSSEEEEEETTEEEEEEEEETTEEEEECHHHHHHHHHHHHHHHHHHHHHHSSCCCEEEEEECTTCCHHHHHHHHH
T ss_pred               HHHHHHhhcCCeEEEecCCceEEEEEECCCceEEcHHHHHHHHHHHHHHHHHHHHHhCCCCCeEEEEECCCCCHHHHHHHHH


Q ss_pred               HHHHcCCceEEEecchHHHHHHHHHhhccCCCCcEEEEEEcCCCeEEEEEEEEecCCcEEEEEEeCCCCCChHHHHHHHHHH
Q HSPA_1A        161 AGVIAGLNVLRIINEPTAAAIAYGLDRTGKGERNVLIFDLGGGTFDVSILTIDDGIFEVKATAGDTHLGGEDFDNRLVNH  240 (453)
Q Consensus      161 a~~~ag~~~~~li~Ep~Aaa~~~~~~~~~~~~~~~~~~~~lvvD~Gggt~dvsv~~~~~g~~~v~~~~~~~~~~GG~~id~~l~~~  240 (453)
                       |++.||++.+.+++||+|||++|+..+....+..++|||+||||+|++++++.+|.+++++..++..+||.+||+.|+++
T Consensus      161 a~~~aGl~~~~li~Ep~Aaa~~y~~~~~~~~~~~~~vlvvD~Gggt~dvsv~~~~~g~~~v~~~~~~~~lGG~~id~~l~~~  240 (554)
T 1yuw_A       161 AGTIAGLNVLRIINEPTAAAIAYGLDKKVGAERNVLIFDLGGGTFDVSILTIAAGIFEVKSTAGDTHLGGEDFDNRMVNH  240 (554)
T ss_dssp               HHHTTTCEEEEEEEHHHHHHHHTTCSTTCSSCEEEEEEEECSSCEEEEEEEEETTEEEEEEEEEETTCSHHHHHHHHHHH
T ss_pred               HHHHcCCCeEEEeCcHHHHHHHHHhhccCCCCcEEEEEEcCCCeEEEEEEEEcCCcEEEEEEeCCCCCCHHHHHHHHHHH


Q ss_pred               HHHHHHHHHcCCCCCcCHHHHHHHHHHHHHHHHHhCCCCceEEEEEeeccCCceeEEEEEHHHHHHHHHHHHHHHHHHHHHH
Q HSPA_1A        241 FVEEFKRKHKKDISQNKRAVRRLRTACERAKRTLSSSTQASLEIDSLFEGIDFYTSITRARFEELCSDLFRSTLEPVEKA  320 (453)
Q Consensus      241 l~~~~~~~~~~~~~~~~~~~~~~~l~~~~e~~K~~ls~~~~~~~~i~i~~~~~g~~~~~~itr~~~~~~~~~~~~~~~~i~~~i~~~  320 (453)
                       +.++++++++.++..+++.+.+|+.+||++|+.|+...+..+.++++++|.++.+.++|++|+++++|+++++.+.|+++
T Consensus      241 l~~~~~~~~~~~d~~~~~~~~~~~l~~~~e~~K~~ls~~~~~~~~i~v~~~~~g~~~~~~itr~~~e~l~~~~~~~~i~~~i~~~  320 (554)
T 1yuw_A       241 FIAEFKRKHKKDISENKRAVRRLRTACERAKRTLSSSTQASIEIDSLYEGIDFYTSITRARFEELNADLFRGTLDPVEKA  320 (554)
T ss_dssp               HHHHHHHHISCCTTSCHHHHHHHHHHHHHHHHHHHTTSSEEEEEEETTCSSSCCEEEEEEHHHHHHHTHHHHHHTTHHHHHH
T ss_pred               HHHHHHHHhCCCccccCHHHHHHHHHHHHHHHhhhcccCceEEEEEeeccCCceEEEEEEEHHHHHHHHHHHHHHHHHHHHHHH


Q ss_pred               HHHcCCCcccCCEEEEECCccccHHHHHHHHHHcCCCCCCCCCChhhHHHHHHHHHHHHHhcCCccccCceEEEEeecce
Q HSPA_1A        321 LRDAKLDKAQIHDLVLVGGSTRIPKVQKLLQDFFNGRDLNKSINPDEAVAYGAAVQAAILMGDKSENVQDLLLLDVAPLS  400 (453)
Q Consensus      321 l~~~~~~~~~~~i~~v~lvGG~s~~p~l~~~l~~~~~~~~~~~~v~~~~~p~~ava~Gaa~~a~~~~~~~~~~~~~~~~~~~~~~~~~~~~~  400 (453)
                       |+.+++.+.+++.|+|+||+|++|+|++.|++.|++.++..+.||++|||+|||++|+.+.+...++++++.+.++++++
T Consensus      321 l~~~~~~~~~~~i~~V~LvGG~s~ip~v~~~l~~~f~~~~~v~~~~~p~~ava~Gaa~~a~~l~~~~~~~~~~~~~~~~~~~~~~~~~~  400 (554)
T 1yuw_A       321 LRDAKLDKSQIHDIVLVGGSTRIPKIQKLLQDFFNGKELNKSINPDEAVAYGAAVQAAILSGDKSENVQDLLLLDVTPLS  400 (554)
T ss_dssp               HHHTTCCGGGCCEEEEESGGGGCHHHHHHHHHHTTTCCCBCCSCTTTHHHHHHHHHHHHTTSCCCCCTTSSCCCCBCSSC
T ss_pred               HHHcCCChhhCcEEEEECCccccChHHHHHHHHHcCCCccccCCCchhHHHHHHHHHHHHHhcCCccccccceEEEEeeeeE


Q ss_pred               EEEEEcCCeEEEEEecCCCCCcceEEEEEEccCCCcEEEEEEEecCCcccCCC
Q HSPA_1A        401 LGLETAGGVMTALIKRNSTIPTKQTQIFTTYSDNQPGVLIQVYEGERAMTKDN  453 (453)
Q Consensus      401 ~gi~~~~~~~~~~~~~~~i~~g~~~p~~~~~~~~~~~~~~~~~~~~~~i~i~~g~~~~~~~~n  453 (453)
                       ||+++.++.+.++||+|+++|.++++.|.+..|+|..+.|.||+|++..+.+|
T Consensus      401 ~gi~~~~~~~~~~li~~g~~ip~~~~~~~~~~~~~~d~q~~~~i~i~~g~~~~~~~~  453 (554)
T 1yuw_A       401 LGIETAGGVMTVLIKRNTTIPTKQTQTFTTYSDNQPGVLIQVYEGERAMTKDN  453 (554)
T ss_dssp               EEEEETTTEEEEEEECTTCBSSEEEEEEEEEESSTTCSEEEEEEEEESSSSBGGGS
T ss_pred               EEEEecCceEEEEEECCCceCceeEEEeeecCCCCceEEEEEEecCcccccC
```

2



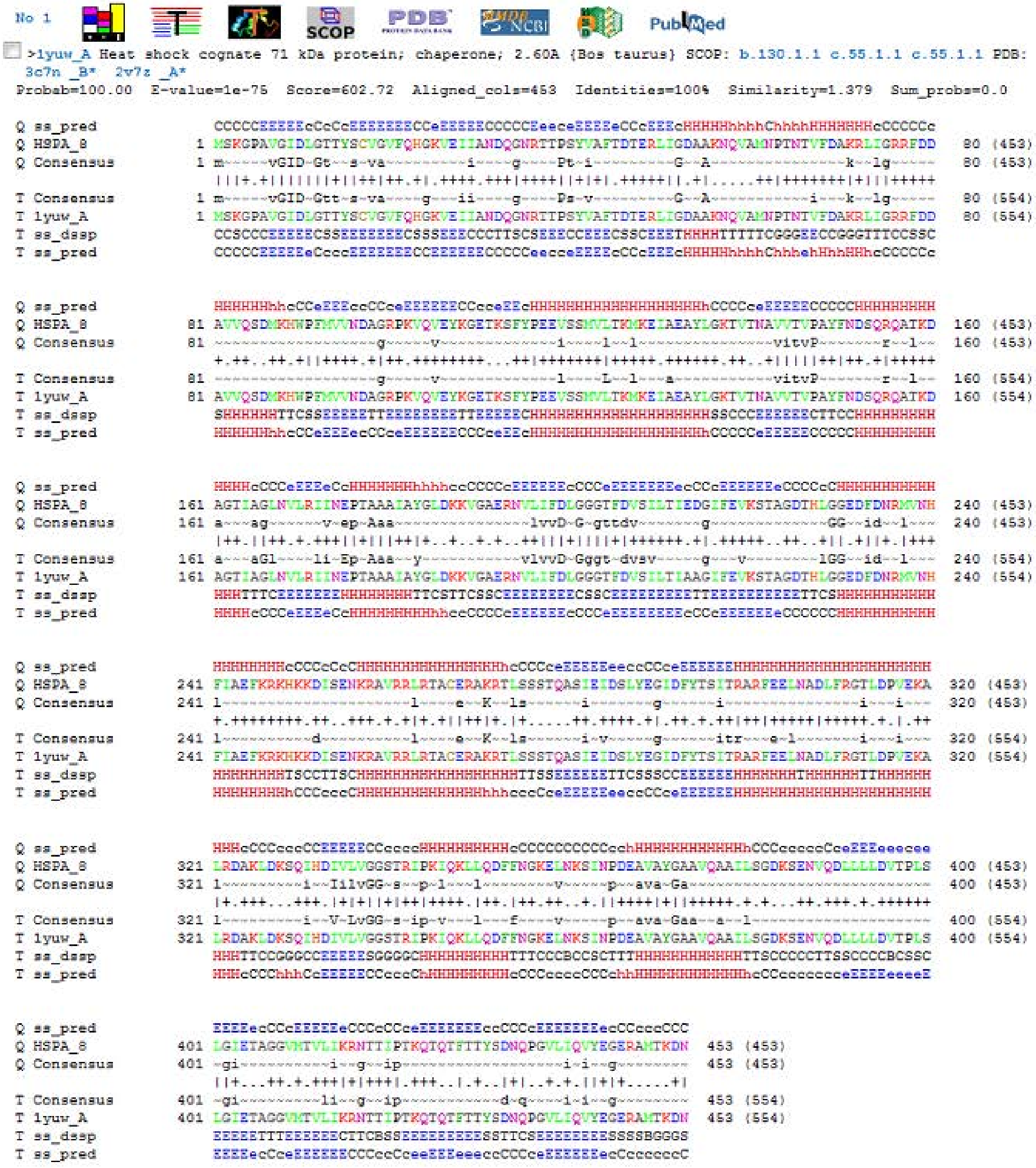
```
No 1
□ >1yuw_A Heat shock cognate 71 kDa protein; chaperone; 2.60A {Bos taurus} SCOP: b.130.1.1 c.55.1.1 c.55.1.1 PDB:
  3c7n _B*  2v7z _A*
Probab=100.00  E-value=1.6e-78  Score=627.14  Aligned_cols=453  Identities=89%  Similarity=1.288  Sum_probs=0.0

Q ss_pred              CCCCCEEEEEeCcccEEEEEEEECCeEEEEECCCCCeecceEEEEeCCeEEECHHHHHhhhhCCceeeehhHHHeCCCCCe
Q HSPA_1B          1 MAKAAAIGIDLGTTYSCVGVFQHGKVEIIANDQGNRTTPSYVAFTDTERLIGDAAKNQVALNPQNTVFDAKRLIGRKFGD   80 (453)
Q Consensus        1 ~~m~~~vGID~Gtt~s~va~~~~~~~~~~ii~~~~g~~~~Ps~v~~~~~~~~~~~G~~A~~~~~~~~~~~~~~~~~~~k~~lg~~~~~   80 (453)
                       |+|+.+|||||||+||++|++.+|.++++.+++|++++||+|+|.+++++||.+|.....++|+++++++|++||+++++
T Consensus        1 m~~~~~vGID~Gtt~s~va~~~~g~~~ii~~~~g~~~~Ps~v~~~~~~~~~~~G~~A~~~~~~~~~~~~~~~i~~~k~~lg~~~~~   80 (554)
T 1yuw_A          1 MSKGPAVGIDLGTTYSCVGVFQHGKVEIIANDQGNRTTPSYVAFTDTERLIGDAAKNQVAMNPTNTVFDAKRLIGRRFDD   80 (554)
T ss_dssp              CCSCCCEEEEECSSEEEEEEEECSSSEEECCCTTSCSEEECCEEECSSCEEETHHHHTTTTTCGGGEECCGGGTTTCCSSC
T ss_pred              CCCCCEEEEEeCcccEEEEEEECCEEEEEECCCCCeecceEEEEcCCcEEEcHHHHHhhhhChhhehHhhHHhcCCCCCe


Q ss_pred              HHHHHHhhcCCeEEEcCCCceEEEEEECCceeEEcHHHHHHHHHHHHHHHHHHHHhCCCCceEEEEECCCCCHHHHHHHHH
Q HSPA_1B         81 PVVQSDMKHWPFQVINDGDKPKVQVSYKGETKAFYPEEISSMVLTKMKEIAEAYLGYPVTNAVITVPAYFNDSQRQATKD  160 (453)
Q Consensus       81 ~~~~~~~~~~~~~~~~~~~g~~~~~~v~~~~~~~~~~~~~~~~~i~~~~L~~l~~~~~~~~~~~~~~~~~~~vitvP~~~~~~~~r~~l~~  160 (453)
                       +.++..++.+||.+++.+|++.++++++++...++|+++++++|+++++.++++++.++.++|||||++|++.||+++++
T Consensus       81 ~~~~~~~~~~~~~~~~~~~g~~~~~~v~~~~~~~~~~~~~~~~l~~~~L~~l~~~a~~~~~~~~~~~~~~~vitvP~~~~~~~~r~~l~~  160 (554)
T 1yuw_A         81 AVVQSDMKHWPFMVVNDAGRPKVQVEYKGETKSFYPEEVSSMVLTKMKEIAEAYLGKTVTNAVVTVPAYFNDSQRQATKD  160 (554)
T ss_dssp              SHHHHHHTTCSSEEEEEETTEEEEEEEEETTEEEEECHHHHHHHHHHHHHHHHHHHHHSSCCCEEEEEECTTCCHHHHHHHHH
T ss_pred              HHHHHHhhcCCeEEEecCCCceEEEEEECCCceEEcHHHHHHHHHHHHHHHHHHHHhCCCCCeEEEEECCCCCHHHHHHHHH


Q ss_pred              HHHHcCCCeEEEecchHHHHHHHHHhhccCCCCcEEEEEEcCCCEEEEEEEEEecCCcEEEEEEeCCCCcCHHHHHHHHHHH
Q HSPA_1B        161 AGVIAGLNVLRIINEPTAAAIAYGLDRTGKGERNVLIFDLGGGTFDVSILTIDDGIFEVKATAGDTHLGGEDFDNRLVNH  240 (453)
Q Consensus      161 a~~~ag~~~~~li~Ep~Aaa~~~~~~~~~~~~~~~~~~~~~~lVvD~Gggt~dvsv~~~~~~~~~~v~~~~~~~~~~~~GG~~id~~l~~~  240 (453)
                       |++.||++.+.+++||+|||++|........+..++|||+||||+|++++++.+|.+++++..++..+||.+||+.|+++
T Consensus      161 a~~~aGl~~~~li~Ep~Aaa~~y~~~~~~~~~~~~~~~vlvvD~Gggt~dvsv~~~~~g~~~v~~~~~~~~~lGG~~id~~l~~~  240 (554)
T 1yuw_A        161 AGTIAGLNVLRIINEPTAAAIAYGLDKKVGAERNVLIFDLGGGTFDVSILTIAAGIFEVKSTAGDTHLGGEDFDNRMVNH  240 (554)
T ss_dssp              HHHTTTCEEEEEEEHHHHHHHHTTCSTTCSSCEEEEEEEEECSSCEEEEEEEEETTEEEEEEEEEETTCSHHHHHHHHHHH
T ss_pred              HHHHeCCCeEEEeCcHHHHHHHHHHhhccCCCCcEEEEEEcCCCeEEEEEEEEcCCcEEEEEEeCCCCCCHHHHHHHHHHH


Q ss_pred              HHHHHHHHhCCCCCeCHHHHHHHHHHHHHHHHHHHHhcCcCceEEEEEeeeeCCceeEEEEeHHHHHHHHHHHHHHHHHHHHHH
Q HSPA_1B        241 FVEEFKRKHKKDISQNKRAVRRLRTACERAKRTLSSSTQASLEIDSLFEGIDFYTSITRARFEELCSDLFRSTLEPVEKA  320 (453)
Q Consensus      241 l~~~~~~~~~~~~~~~~~~~~~~~l~~~~e~~K~~ls~~~~~~~~~~i~i~~~~g~~~~~~~~itr~~~e~~~~~~~~~~~i~~~i~~~  320 (453)
                       +.++++++++.++..+++.+.+|+.+||++|+.||...+..+.++++.+|.++.+.++|++|+++++|+++++.+.|+++
T Consensus      241 l~~~~~~~~~~~~d~~~~~~~~~~~~l~~~~e~~K~~ls~~~~~~~~~~i~v~~~~g~~~~~~~~itr~~~e~l~~~~~~~~~~i~~~i~~~  320 (554)
T 1yuw_A        241 FIAEFKRKHKKDISENKRAVRRLRTACERAKRTLSSSTQASIEIDSLYEGIDFYTSITRARFEELNADLFRGTLDPVEKA  320 (554)
T ss_dssp              HHHHHHHHTSCCTTSCHHHHHHHHHHHHHHHHHHHTTSSEEEEEEETTCSSSCCEEEEEEHHHHHHHTHHHHHHTTHHHHHH
T ss_pred              HHHHHHHHhCCCCcccCHHHHHHHHHHHHHHHHhhhcccCceEEEEEeeccCCceEEEEEEHHHHHHHHHHHHHHHHHHHHHH


Q ss_pred              HHHhCCCCeeeCCEEEEEECCccccHHHHHHHHHHeCCCCCCCCCCchhHHHHHHHHHHHHHHHheCCeeeeeeCceEEEEeecce
Q HSPA_1B        321 LRDAKLDKAQIHDLVLVGGSTRIPKVQKLLQDFFNGRDLNKSINPDEAVAYGAAVQAAILMGDKSENVQDLLLLDVAPLS  400 (453)
Q Consensus      321 l~~~~~~~~~~~~i~~V~LvGG~s~~p~l~~~l~~~~~~~~~~~~v~~~~~~p~~ava~Ga~~~a~~~~~~~~~~~~~~~~~~~~~~~~~~~  400 (453)
                       |+++++.+.+++.|+|+||+|++|+|++.|++.|++.++..+.||+++||+||+++|+.+.+...++.+++.+.+++|++
T Consensus      321 l~~~~~~~~~~~~i~~V~LvGG~s~ip~v~~~l~~~f~~~~~v~~~~~~p~~ava~Gaa~~a~~l~~~~~~~~~~~~~~~~~~~~~~~~   400 (554)
T 1yuw_A        321 LRDAKLDKSQIHDIVLVGGSTRIPKIQKLLQDFFNGKELNKSINPDEAVAYGAAVQAAILSGDKSENVQDLLLLDVTPLS  400 (554)
T ss_dssp              HHHTTCCGGGCCEEEEESGGGGCHHHHHHHHHHTTTCCCBCCSCTTTHHHHHHHHHHHHHTTSCCCCCTTSSCCCCBCSSC
T ss_pred              HHHeCCChhhCcEEEEECCcccChHHHHHHHHHeCCCCcccCCCchhHHHHHHHHHHHHHhcCCcccccceEEEEeeeeE


Q ss_pred              EEEEeeCCCeEEEEEeCCCCCCCcceEEEEEEeeCCCCeEEEEEEEeeCCCcccCCC
Q HSPA_1B        401 LGLETAGGVMTALIKRNSTIPTKQTQIFTTYSDNQPGVLIQVYEGERAMTKDN  453 (453)
Q Consensus      401 ~gi~~~~~~~~~~~~~~~i~~g~~~p~~~~~~~~~~~~~~~~~~~~~~~~~~i~i~~g~~~~~~~~n  453 (453)
                       ||+++.++.+.++||+|+++|.+++..|.+..|+|..+.|.||+|++..+.+|
T Consensus      401 ~gi~~~~~~~~~~~li~~g~~ip~~~~~~~~~~~~~~~d~q~~~~i~i~~g~~~~~~~~~~~   453 (554)
T 1yuw_A        401 LGIETAGGVMTVLIKRNTTIPTKQTQTFTTYSDNQPGVLIQVYEGERAMTKDN  453 (554)
T ss_dssp              EEEEETTTEEEEEECTTCBSSEEEEEEEEEESSTTCSEEEEEEEEESSSSBGGGS
T ss_pred              EEEEecCceEEEEEECCCcccCceeEEEeeeccCCCceEEEEEEecCcccccccC
```

3



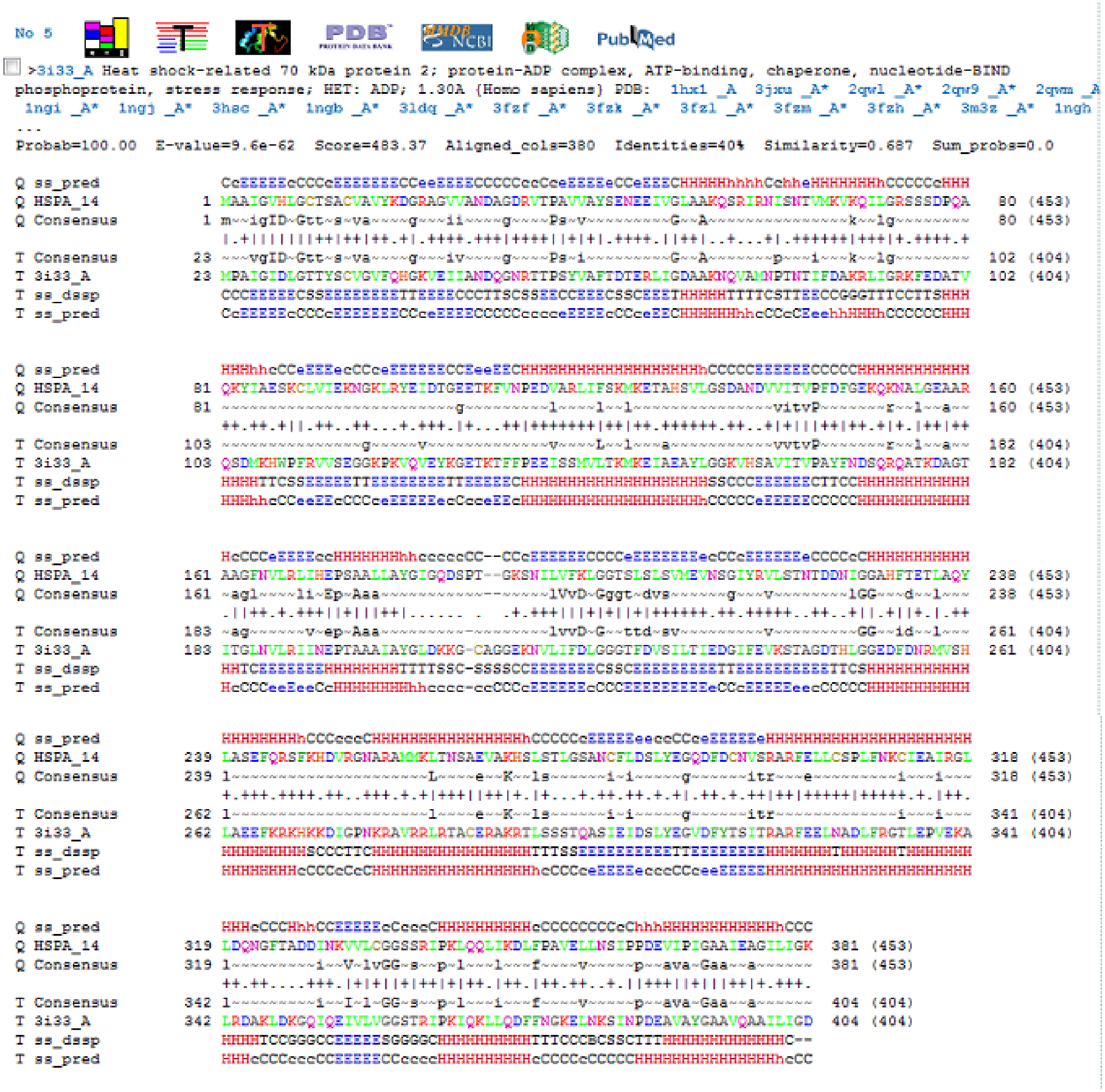
```
No 6

>3qfu_A 78 kDa glucose-regulated protein homolog; HSP70, KAR2, BIP, chaperone; HET: ADP; 1.80A {Saccharomyces
cerevisiae) PDB:  3qfp _A  3qml _A  3ldo _A*  3ldl _A  3ldn _A*  3ldp _A*
Probab=100.00  E-value=8.7e-58  Score=453.42  Aligned_cols=379  Identities=72%  Similarity=1.115  Sum_probs=0.0

Q ss_pred              ccCCEEEEEECCCceEEEEEEEECCceEEEECCCCCeeeceEEEEeCCCcEEECHHHHHhhhCcccHHHHHHHhCCCCCC
Q HSPA_5          26 DVGTVVGIDLGTTYSCVGVFKNGRVEIIANDQGNRITPSYVAFTPEGERLIGDAAKNQLTSNPENTVFDAKRLIGRTWND  105 (453)
Q Consensus       26 ~~~~~vgID~Gt~~t~va~~~~~~~~ii~~~~g~~~~Pt~i~~~~~~~~~~~G~~A~~~~~~~p~~~i~~~k~~l~~~~~~  105 (453)
                       +|+.+|||||||++|+++++.+|.|+++.+++|++++||+|+|  .++++.||++|......+|.+++.++|++|+.++++
T Consensus       16 ~~~~~vgID~Gt~~~~va~~~~g~~~iv~~~~g~~~~pt~i~~~~~~~~~~~G~~A~~~~~~~~~~~~~~~~~~k~~l~~~~~~   94 (394)
T 3qfu_A         16 NYGTVIGIDLGTTYSCVAVMKNGKTEILANEQGNRITPSYVAF-TDDERLIGDAAKNQVAANPQNTIFDIKRLIGLKYND   94 (394)
T ss_dssp             CCCSCEEEEEECSSEEEEEEEECSSCEEECCCTTSCSSEECCEEE-CSSCEEESHHHHHTGGGCGGGEECCGGGTTTCCTTC
T ss_pred              hhceEEEEEeCCCcEEEEEEEECCceEEEECCCCCeeeceEEEE-eCCceEEeHhHHhhhheCcccHHHHHHHhCCCCCC


Q ss_pred              HHHHHHhhcCCeEEEccCCceEEEEEEecCCCceEEcHHHHHHHHHHHHHHHHHHHHhCCCCceEEEEECCCCCHHHHHHHH
Q HSPA_5         106 PSVQQDIKFLPFKVVEKKTKPYIQVDIGGGQTKTFAPEEISAMVLTKMKETAEAYLGKKVTHAVVTVPAYFNDAQRQATK  185 (453)
Q Consensus      106 ~~~~~~~~~~p~~~~~~~~~~~~~~v~~~~g~~~~~~~~~~~~l~~~~l~~l~~~~~~~~~~~~~~vitvP~~~~~~~r~~l~  185 (453)
                     +.+++..+.+|+..++.+++..++++++ |....+++++++++||+++++.++++++.+..++++++|+.|++.+|+.++
T Consensus       95 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~g~~~~~~~~~~~i~~~~l~~l~~~~~~~~~~~~~~~vitvP~~~~~~~r~~l~  173 (394)
T 3qfu_A         95 RSVQKDIKHLPFNVVNKDGKPAVEVSVK-GEKKVFTPEEISGMILGKMKQIAEDYLGTKVTHAVVTVPAYFNDAQRQATK  173 (394)
T ss_dssp             HHHHHHHTTCSSEEEEEETTEEEEEEEESS-SSEEEECHHHHHHHHHHHHHHHHHHHHISCCCEEEEEECTTCCHHHHHHHH
T ss_pred              HHHHHHhhcCCeEEEecCCceEEEEEeC-CcceEEcHHHHHHHHHHHHHHHHHHHHhCCCCCeEEEEECCCCCHHHHHHHH


Q ss_pred              HHHHHcCCCeEEEeCcHHHHHHHhhhhcCCCCcEEEEEEcCCCeEEEEEEEEEeCCcEEEEEecCCCCcCHHHHHHHHHHH
Q HSPA_5         186 DAGTIAGLNVMRIINEPTAAAIAYGLDKREGEKNILVFDLGGGTFDVSLLTIDNGVFEVVATNGDTHLGGEDFDQRVMEH  265 (453)
Q Consensus      186 ~a~~~ag~~~~~~v~ep~Aaa~~~~~~~~~~~~~~vlvvDiG~~ttdvsv~~~~~~~~~~~~~~~~~~~~~~GG~~id~~l~~~  265 (453)
                     ++++.+|++.+.+++||+|++++|......++..++|||+|++|||++++++.++.++++.+.++..+||++||+.|+++
T Consensus      174 ~a~~~ag~~~~~~v~e~~Aaa~~~~~~~~~~~~~~vlvvDiG~~ttd~~v~~~~~~~~~~~~~~~~~~~~~~Gg~~id~~l~~~  253 (394)
T 3qfu_A        174 DAGTIAGLNVLRIVNEPTAAAIAYGLDKSDKEHQIIVYDLGGGTFDVSLLSIENGVFEVQATSGDTHLGGEDFDYKIVRQ  253 (394)
T ss_dssp             HHHHHTTCEEEEEEEEHHHHHHHHTTTTSCSSCEEEEEEEEECSSCEEEEEEEEEETTEEEEEEEEEEETTCSHHHHHHHHHHH
T ss_pred              HHHHHcCCCceEEecCcHHHHHHHHhhcCCCCcEEEEEECCCCceEEEEEEEeCCcEEEEEEeCCCCCCHHHHHHHHHHH


Q ss_pred              HHHHHHHHhCCCccccCHHHHHHHHHHHHHHHHHHHcCCCCeEEEEEeccccCCceEEEEEEHHHHHHHHHHHHHHHHHHHHHHH
Q HSPA_5         266 FIKLYKKKTGKDVRKDNRAVQKLRREVEKAKRALSSQHQARIEIESFYEGEDFSETLTRAKFEELNMDLFRSTMKPVQKV  345 (453)
Q Consensus      266 l~~~~~~~~~~~~~~~~~~~~~l~~~~e~~K~~ls~~~~~~i~~~~~~~~~~~~~~~i~r~~~~~~~~~~~~~~~i~~~i~~~  345 (453)
                     +.+++.++++.++..+++.+.+|+++|+++|+.++...+..++++.+.++.+..+.+++++|+++++|.++++.+.|+++
T Consensus      254 l~~~~~~~~~~~~~~~~~~~~l~~~~e~~K~~l~~~~~~~~i~i~~~~~~~~~~~~~i~~~~~~~~~~~~~~~~~~i~~~i~~~  333 (394)
T 3qfu_A        254 LIKAFKKKHGIDVSDNNKALAKLKREAEKAKRALSSQMSTRIEIDSFVDGIDLSETLTRAKFEELNLDLFKKTLKPVEKV  333 (394)
T ss_dssp             HHHHHHHHHSCCCTTCHHHHHHHHHHHHHHHHHHHTTTCSEEEEEEEEEEETTEEEEEEEEEHHHHHHHHHHHHHHHTHHHHHHH
T ss_pred              HHHHHHHHcCCCccccCHHHHHHHHHHHHHHHHHHHhccCCceEEEEeccccCCceEEEEEEHHHHHHHHHHHHHHHHHHHHHHH


Q ss_pred              HHHcCCCccccCCEEEEEECCccccCHHHHHHHHHHcCCCCcCCCCCHHHHHHHhHHHHHHHHhc
Q HSPA_5         346 LEDSDLKKSDIDEIVLVGGSTRIPKIQQLVKEFFNGKEPSRGINPDEAVAYGAAVQAGVLS  406 (453)
Q Consensus      346 l~~~~~~~~~i~~I~L~GG~s~~p~l~~~i~~~~~~~~~~v~~~~p~~ava~Ga~~~a~~~~  406 (453)
                     ++++++...+++.|+|+||+|++|+|++.|++.||+.+|..+.+|++++|+||+++|++|+
T Consensus      334 l~~~~~~~~~i~~I~l~GG~s~~p~l~~~l~~~~~~~~~~v~~~~p~~ava~Ga~~~a~~l~  394 (394)
T 3qfu_A        334 LQDSGLEKKDVDDIVLVGGSTRIPKVQQLLESYFDGKKASKGINPDEAVAYGAAVQAGVLS  394 (394)
T ss_dssp             HHHHTCCGGGCCEEEEESGGGGSHHHHHHHHHHTTTCCCBCCSCTTTHHHHHHHHHHHHHHC
T ss_pred              HHHcCCCccccCCeEEEecCccccCHHHHHHHHHHcCCCCCCCCCcChhHHHHHHHHHHHHHhhC
```

4



```
No 1
☐ >1yuw_A Heat shock cognate 71 kDa protein; chaperone; 2.60A {Bos taurus} SCOP: b.130.1.1 c.55.1.1 c.55.1.1 PDB:
 3c7n _B*  2v7z _A*
Probab=100.00  E-value=1e-75  Score=602.72  Aligned_cols=453  Identities=100%  Similarity=1.379  Sum_probs=0.0

Q ss_pred                  CCCCCEEEEEcCcCcEEEEEEEECCeEEEEECCCCCEeeceEEEEeCCcEEEcHHHHHhhhhChhhhHHHHHHHcCCCCCc
Q HSPA_8            1 MSKGPAVGIDLGTTYSCVGVFQHGKVEIIANDQGNRTTPSYVAFTDTERLIGDAAKNQVAMNPTNTVFDAKRLIGRRFDD   80 (453)
Q Consensus         1 m~~~~~vGID~Gt~~s~va~~~~~~~~~i~~~~~g~~~~Pt~i~~~~~~~~~~~G~~A~~~~~~~~~~~~~~~k~~lg~~~~~   80 (453)
                        |||+.+|||||||+||++|++.+|.++++.+++|++++||+|+|.+++++||.+|.....++|+++++++|+|||+++++
T Consensus         1 m~~~~~vGID~Gtt~s~va~~~~g~~~ii~~~~g~~~~Ps~v~~~~~~~~~~~G~~A~~~~~~~~~~~~i~~~k~~lg~~~~~   80 (554)
T 1yuw_A            1 MSKGPAVGIDLGTTYSCVGVFQHGKVEIIANDQGNRTTPSYVAFTDTERLIGDAAKNQVAMNPTNTVFDAKRLIGRRFDD   80 (554)
T ss_dssp                  CCSCCCEEEEECSSEEEEEEEECSSSEEECCCTTSCSEEECCEEECSSCEEETHHHHTTTTTCGGGEECCGGGTTTCCSSC
T ss_pred                  CCCCCEEEEEeCcccEEEEEEEECCEEEEEECCCCCeecceEEEEcCCcEEEcHHHHHhhhhChhhehHhhHHhcCCCCCc


Q ss_pred                  HHHHHHhhcCCeEEEccCCceEEEEEEECCcceEEcHHHHHHHHHHHHHHHHHHHHhCCCCceEEEEECCCCCHHHHHHHHH
Q HSPA_8           81 AVVQSDMKHWPFMVVNDAGRPKVQVEYKGETKSFYPEEVSSMVLTKMKEIAEAYLGKTVTNAVVTVPAYFNDSQRQATKD  160 (453)
Q Consensus        81 ~~~~~~~~~~~~~~~~~~~~g~~~~~~v~~~~~~~~~~~~~~~~~i~~~~l~~l~~~~~~~~~~~~~~~~~~~~vitvP~~~~~~~~~r~~l~~  160 (453)
                        +.++..++.+||++++.+|++.++++++++...++|+++++++|+++++.++++++.++..+|||||++|++.+|+++++
T Consensus        81 ~~~~~~~~~~~~~~~~~~~~g~~~~~~v~~~~~~~~~~~~~~~~~l~~~~L~~l~~~a~~~~~~~~~~~~~~~vitvP~~~~~~~~~r~~l~~  160 (554)
T 1yuw_A           81 AVVQSDMKHWPFMVVNDAGRPKVQVEYKGETKSFYPEEVSSMVLTKMKEIAEAYLGKTVTNAVVTVPAYFNDSQRQATKD  160 (554)
T ss_dssp                  SHHHHHHTTCSSEEEEEETTEEEEEEEEETTEEEEECHHHHHHHHHHHHHHHHHHHHSSCCCEEEEEEECTTCCHHHHHHHHH
T ss_pred                  HHHHHHhhcCCeEEEecCCceEEEEEEECCCceEEcHHHHHHHHHHHHHHHHHHHHhCCCCCeEEEEECCCCCHHHHHHHHH


Q ss_pred                  HHHHcCCCeEEEeCcHHHHHHHhhhhccCCCCcEEEEEEcCCCeEEEEEEEEecCCcEEEEEEeCCCCcCHHHHHHHHHHH
Q HSPA_8          161 AGTIAGLNVLRIINEPTAAAIAYGLDKKVGAERNVLIFDLGGGTFDVSILTIEDGIFEVKSTAGDTHLGGEDFDNRMVNH  240 (453)
Q Consensus       161 a~~~ag~~~~~~~v~ep~Aaa~~~~~~~~~~~~~~~~~~~~~~lvvD~G~gttdv~~~~~~~~~g~~~~~~~~~~~~~~~~~~GG~~id~~l~~~  240 (453)
                        |++.||++.+.+++||+|||++|+..+..+.+..++|||+||||+|++++++.+|.+++++..++..+||.+||+.|+++
T Consensus       161 a~~~aGl~~~~li~Ep~Aaa~~y~~~~~~~~~~~~~~~vlvvD~Gggt~dvsv~~~~~~g~~~v~~~~~~~~~~~~lGG~~id~~l~~~  240 (554)
T 1yuw_A          161 AGTIAGLNVLRIINEPTAAAIAYGLDKKVGAERNVLIFDLGGGTFDVSILTIAAGIFEVKSTAGDTHLGGEDFDNRMVNH  240 (554)
T ss_dssp                  HHHTTTCEEEEEEEHHHHHHHHTTCSTTCSSCEEEEEEEEECSSCEEEEEEEEETTEEEEEEEEEETTCSHHHHHHHHHHH
T ss_pred                  HHHHcCCCeEEEeCcHHHHHHHHHHhhccCCCCcEEEEEEcCCCeEEEEEEEEcCCcEEEEEEeCCCCCCHHHHHHHHHHH


Q ss_pred                  HHHHHHHHcCCCCcCcCHHHHHHHHHHHHHHHHHHhcCCCceEEEEEeeccCCceEEEEEEHHHHHHHHHHHHHHHHHHHHHHH
Q HSPA_8          241 FIAEFKRKHKKDISENKRAVRRLRTACERAKRTLSSSTQASIEIDSLYEGIDFYTSITRARFEELNADLFRGTLDPVEKA  320 (453)
Q Consensus       241 l~~~~~~~~~~~~~~~~~~~~~~~~~l~~~~e~~K~~ls~~~~~~~~i~~~~~~~~~~g~~~~~~~~i~~~~~~~~~~~~~~~~~~~~i~~~i~~~  320 (453)
                        +.++++++++.++..+++.+.+|+.+||++|+.|+.....++.++++.+|.++.+.++|++|+++++|+++++.+.|.++
T Consensus       241 l~~~~~~~~~~~~d~~~~~~~~~~~~~l~~~~e~~K~~ls~~~~~~~i~v~~~~~g~~~~~~itr~~~~e~l~~~~~~~~~i~~~i~~~  320 (554)
T 1yuw_A          241 FIAEFKRKHKKDISENKRAVRRLRTACERAKRTLSSSTQASIEIDSLYEGIDFYTSITRARFEELNADLFRGTLDPVEKA  320 (554)
T ss_dssp                  HHHHHHHHTSCCTTSCHHHHHHHHHHHHHHHHHHHTTSSEEEEEETTCSSSCCEEEEEEHHHHHHHTHHHHHHTTHHHHHH
T ss_pred                  HHHHHHHHhCCCccccCHHHHHHHHHHHHHHHhhhcccCceEEEEEeeccCCceEEEEEEHHHHHHHHHHHHHHHHHHHHHHH


Q ss_pred                  HHHcCCCccccCCEEEEECCccccHHHHHHHHHHcCCCCCCCCCCcchHHHHHHHHHHHHHhCCCccccccCceEEEeeecee
Q HSPA_8          321 LRDAKLDKSQIHDIVLVGGSTRIPKIQKLLQDFFNGKELNKSINPDEAVAYGAAVQAAILSGDKSENVQDLLLLDVTPLS  400 (453)
Q Consensus       321 l~~~~~~~~~~~~~~i~~IilvGG~s~~p~l~~~l~~~~~~~~~~~~~v~~~~~~p~~ava~Ga~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~  400 (453)
                        |+.+++...+++.|+|+||+|++|++++.|++.|++.++..+.||++++|+||+++++.+.+...++.+++.+.++++++
T Consensus       321 l~~~~~~~~~~~~~i~~V~LvGG~s~ip~v~~~l~~~f~~~~~v~~~~~~p~~ava~Gaa~~a~~l~~~~~~~~~~~~~~~~~~~~~~~~~~  400 (554)
T 1yuw_A          321 LRDAKLDKSQIHDIVLVGGSTRIPKIQKLLQDFFNGKELNKSINPDEAVAYGAAVQAAILSGDKSENVQDLLLLDVTPLS  400 (554)
T ss_dssp                  HHHTTCCGGGCCEEEEESGGGGCHHHHHHHHHHTTTCCCBCCSCTTTHHHHHHHHHHHHTTSCCCCCTTSSCCCCBCSSC
T ss_pred                  HHHcCCChhhCcEEEEECCccccChHHHHHHHHHcCCCcccccCCCchhHHHHHHHHHHHHhcCCcccccccceEEEEeeeeE


Q ss_pred                  EEEEecCCCcEEEEEeCCCcCCceEEEEEEEccCCCcEEEEEEEecCCccccCCC
Q HSPA_8          401 LGIETAGGVMTVLIKRNTTIPTKQTQTFTTYSDNQPGVLIQVYEGERAMTKDN  453 (453)
Q Consensus       401 ~gi~~~~~~~~~~~~~~~~i~~g~~ip~~~~~~~~~~~~~~~~~~~~~~~~~~~~~i~i~~g~~~~~~~~~~~  453 (453)
                        ||+...++.+.+++|+|+++|.+++..|.+..|+|..+.+.||+|+.....+|
T Consensus       401 ~gi~~~~~~~~~~~~li~~g~~ip~~~~~~~~~~~~~~d~q~~~~~i~i~~g~~~~~~~~~~  453 (554)
T 1yuw_A          401 LGIETAGGVMTVLIKRNTTIPTKQTQTFTTYSDNQPGVLIQVYEGERAMTKDN  453 (554)
T ss_dssp                  EEEEETTTEEEEEEECTTCBSSEEEEEEEEEESSTTCSEEEEEEEEESSSSBGGGS
T ss_pred                  EEEEecCceEEEEEEECCCccCceeEEEeeeccCCCceEEEEEEEecCcccccccC
```

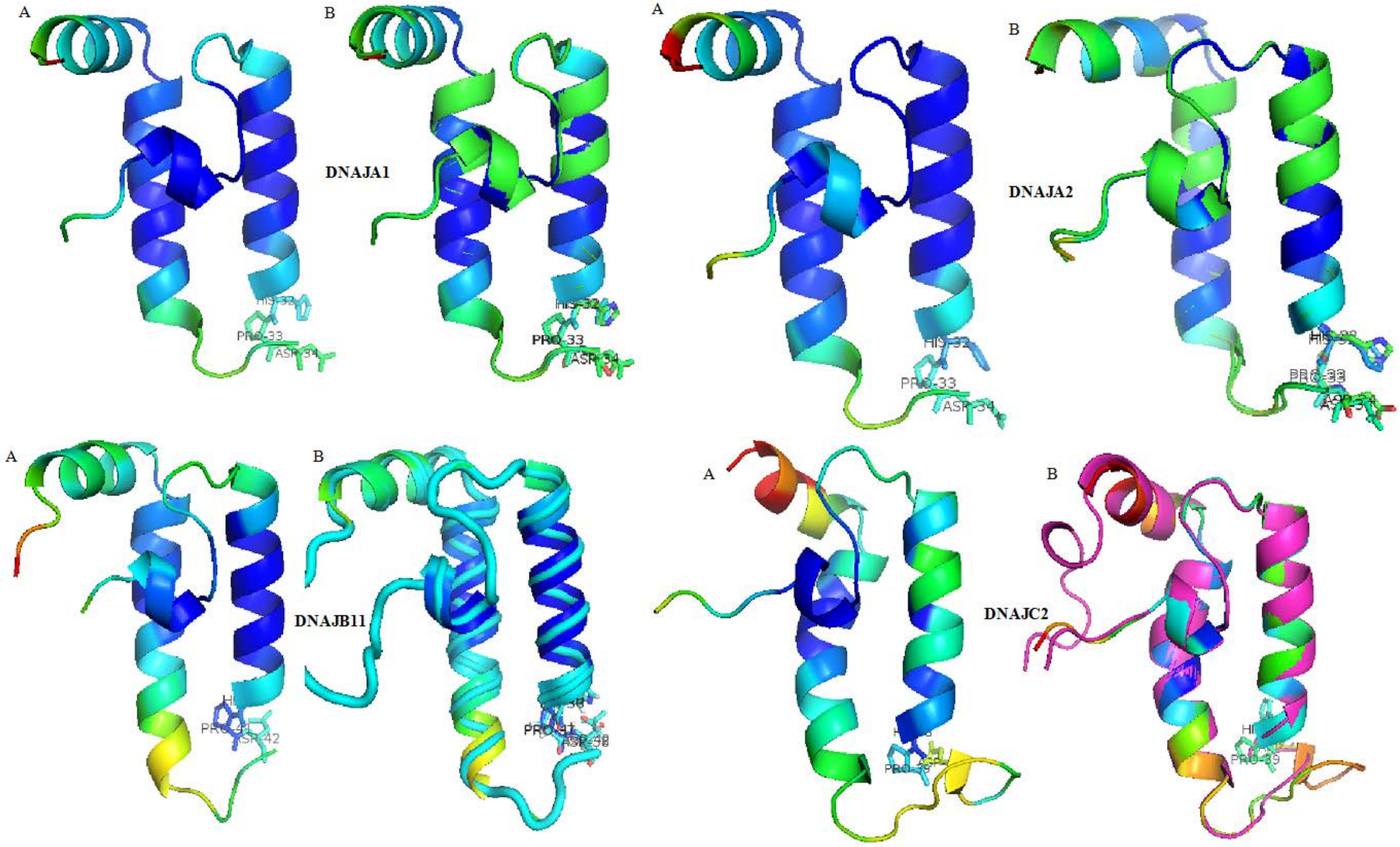
5



```
□ >3i33_A Heat shock-related 70 kDa protein 2; protein-ADP complex, ATP-binding, chaperone, nucleotide-BIND
phosphoprotein, stress response; HET: ADP; 1.30A {Homo sapiens} PDB:  1hx1 _A  3jxu _A*  2qwl _A*  2qw9 _A*  2qwm _A
 1ngi _A*  1ngj _A*  3hsc _A*  1ngb _A*  3ldq _A*  3fzf _A*  3fzk _A*  3fzl _A*  3fzm _A*  3fzh _A*  3m3z _A*  1ngh
...
Probab=100.00  E-value=9.6e-62  Score=483.37  Aligned_cols=380  Identities=40%  Similarity=0.687  Sum_probs=0.0

Q ss_pred              CcEEEEEEeCCCcEEEEEEEECCeeEEEECCCCCcccCceeEEEEeCCeEEECHHHHHhhhhCchheHHHHHHHhCCCCCcHHH
Q HSPA_14         1 MAAIGVHLGCTSACVAVYKDGRAGVVANDAGDRVTPAVVAYSENEEIVGLAAKQSRIRNISNTVMKVKQILGRSSSDPQA   80 (453)
Q Consensus       1 m~~igID~Gtt~s~va~~~~g~~~ii~~~~g~~~~Ps~v~~~~~~~~~~~G~~A~~~~~~~~~~~~~~~~~~k~~lg~~~~~~~~   80 (453)
                    |.+||||||||++|++|++.+|.++++.+++|++++||+|+|.++++.||++|..+...+|.++++++|+++|+.++++.+
T Consensus      23 ~~~vgID~Gtt~s~va~~~~g~~~iv~~~~g~~~~Ps~i~~~~~~~~~~~G~~A~~~~~~~~p~~~i~~~k~~lg~~~~~~~~  102 (404)
T 3i33_A        23 MPAIGIDLGTTYSCVGVFQHGKVEIIANDQGNRTTPSYVAFTDTERLIGDAAKNQVAMNPTNTIFDAKRLIGRKFEDATV  102 (404)
T ss_dssp           CCCEEEEEECSSEEEEEEEEETTEEEEECCCTTSCSSEECCEEECSSCEEETHHHHHTTTTCSTTEECCGGGTTTCCTTSHHH
T ss_pred              CcEEEEEcCCCcEEEEEEEECCceEEEECCCCCcccceEEEEcCCceEECHHHHHHhhcCCcCEeehhHHHhCCCCCCHHH


Q ss_pred              HHHhhcCCeEEEecCCceEEEEEEECCEeeEECHHHHHHHHHHHHHHHHHHHHhCCCCCEEEEEECCCCCHHHHHHHHHHHH
Q HSPA_14        81 QKYIAESKCLVIEKNGKLRYEIDTGEETKFVNPEDVARLIFSKMKETAHSVLGSDANDVVITVPFDFGEKQKNALGEAAR  160 (453)
Q Consensus      81 ~~~~~~~~~~~~~~~~~~~~~~~~~~g~~~~~~~~~~~~l~~~~l~~l~~~~~~~~~~~~~~~~vitvP~~~~~~~r~~l~~a~~  160 (453)
                    ++.++.+||.++..++...+.+++.|+...++|+++++++|++|++.++++++.++..+|+|||++|++.+|+.|++|++
T Consensus     103 ~~~~~~~~~~~~~~~~~~~~~~~~~g~~~~~v~~~~~~~~~~~~~~~v~~~~L~~l~~~a~~~~~~~~~~~vvtvP~~~~~~~r~~l~~a~~  182 (404)
T 3i33_A       103 QSDMKHWPFRVVSEGGKPKVQVEYKGETKTFFPEEISSMVLTKMKEIAEAYLGGKVHSAVITVPAYFNDSQRQATKDAGT  182 (404)
T ss_dssp           HHHHTTCSSEEEEEETTEEEEEEEEETTEEEEEECHHHHHHHHHHHHHHHHHHHHHSSCCCEEEEEEECTTCCHHHHHHHHHHHH
T ss_pred              HHHhhcCCeeEEcCCCceEEEEEecCcceEEcHHHHHHHHHHHHHHHHHHHHHhCCCCCeEEEEECCCCCHHHHHHHHHHHH


Q ss_pred              HcCCCeEEEEccHHHHHHHhhccccccCC--CCcEEEEEECCCCeEEEEEEEecCCCcEEEEEEeCCCCcCHHHHHHHHHHH
Q HSPA_14       161 AAGFNVLRLIHEPSAALLAYGIGQDSPT--GKSNILVFKLGGTSLSLSVMEVNSGIYRVLSTNTDDNIGGAHFTETLAQY  238 (453)
Q Consensus     161 ~agl~~~~li~Ep~Aaa~~~~~~~~~~~~--~~~~~~lVvD~Gggt~dvs~~~~~~~~g~~~v~~~~~~~~~~~lGG~~~d~~l~~~  238 (453)
                    .||++.+.+++||+|||++|......  .  .+.+++|||+|+||+|++++++.++.+++++..++..+||.+||+.|.++
T Consensus     183 ~ag~~~~~~v~ep~Aaa~~~~~~~~~~~~-~~~~~~~~~~lvvD~G~~ttd~sv~~~~~~~~~~~v~~~~~~~~~~~~GG~~id~~l~~~  261 (404)
T 3i33_A       183 ITGLNVLRIINEPTAAAIAYGLDKKG-CAGGEKNVLIFDLGGGTFDVSILTIEDGIFEVKSTAGDTHLGGEDFDNRMVSH  261 (404)
T ss_dssp           HHTCEEEEEEEEHHHHHHHHTTTTSSC-SSSSCCEEEEEEEECSSCEEEEEEEEEETTEEEEEEEEEEEETTCSHHHHHHHHHHH
T ss_pred              HcCCCeeEeeCcHHHHHHHHhhcccc-ccCCCcEEEEEEcCCCEEEEEEEEEeCCcEEEEEeecCCCCCHHHHHHHHHHH


Q ss_pred              HHHHHHHHhCCCccccCHHHHHHHHHHHHHHHHHHhCCCCCcEEEEEeeccCCceEEEEEeHHHHHHHHHHHHHHHHHHHHHHHH
Q HSPA_14       239 LASEFQRSFKHDVRGNARAMMKLTNSAEVAKHSLSTLGSANCFLDSLYEGQDFDCNVSRARFELLCSPLFNKCIEAIRGL  318 (453)
Q Consensus     239 l~~~~~~~~~~~~~~~~~~~~~L~~~~e~~K~~ls~~~~~~~~i~i~~~~~~g~~~~~~~itr~~~e~~~~~~~~~~~~~i~~~i~~~  318 (453)
                    +.+++.++++.++..+++.+.+|+++||++|+.|+...+.++.++.+.+|.++.+.++|++|+++++|+++++.+.|++.
T Consensus     262 l~~~~~~~~~~~~~~~~~~~~~l~~~~e~~K~~ls~~~~~~~~i~i~~~~~~g~~~~~~~itr~~~~~~~~~~~~~~~~~i~~~i~~~  341 (404)
T 3i33_A       262 LAEEFKRKHKKDIGPNKRAVRRLRTACERAKRTLSSSTQASIEIDSLYEGVDFYTSITRARFEELNADLFRGTLEPVEKA  341 (404)
T ss_dssp           HHHHHHHHHSCCCTTCHHHHHHHHHHHHHHHHHHHTTTSSEEEEEEEEEEEETTEEEEEEEEEHHHHHHHTHHHHHHTHHHHHHH
T ss_pred              HHHHHHHHcCCCcCcCHHHHHHHHHHHHHHHHHHHHhcCCCceEEEEeccccCCceeEEEEEHHHHHHHHHHHHHHHHHHHHHHHH


Q ss_pred              HHHcCCCHhhCCEEEEEEcCcccCHHHHHHHHHHcCCCCCCCCcChhhHHHHHHHHHHHHHhCCC
Q HSPA_14       319 LDQNGFTADDINKVVLCGGSSRIPKLQQLIKDLFPAVELLNSIPPDEVIPIGAAIEAGILIGK  381 (453)
Q Consensus     319 l~~~~~~~~~~~i~~V~lvGG~s~~p~l~~~~l~~~f~~~~~~v~~~~~~~p~~ava~Gaa~~a~~~~~~~~  381 (453)
                    ++.++....+++.|+|+||+|++|+|++.|++.|++.++..+.||+++||+|||++|+.+++.
T Consensus     342 l~~~~~~~~~~~i~~I~l~GG~s~~p~l~~~~i~~~f~~~~~~v~~~~~~~p~~ava~Gaa~~a~~~~~~~~  404 (404)
T 3i33_A       342 LRDAKLDKGQIQEIVLVGGSTRIPKIQKLLQDFFNGKELNKSINPDEAVAYGAAVQAAILIGD  404 (404)
T ss_dssp           HHHHTCCGGGCCEEEEESGGGGCHHHHHHHHHHHTTTCCCBCSSCTTTHHHHHHHHHHHHHHC--
T ss_pred              HHHcCCCccccCCEEEEEECCccccHHHHHHHHHHcCCCCcCCCCCCHHHHHHHHHHHHHHHHhcCC
```

Figure 1: Template selection, alignments and secondary structure prediction of HSP70 ATPase_linker region using HHpred. Templates 1, 2, 4 (1YUW) is a crystal structure of HSPA8 from Bos Taurus, template 3 (QFU) is from Saccharomyces cerevisiae, and template 5 (3I33) is from Homo sapiens.

Appendix III: Predicted model structures of selected human HSP40 J domains



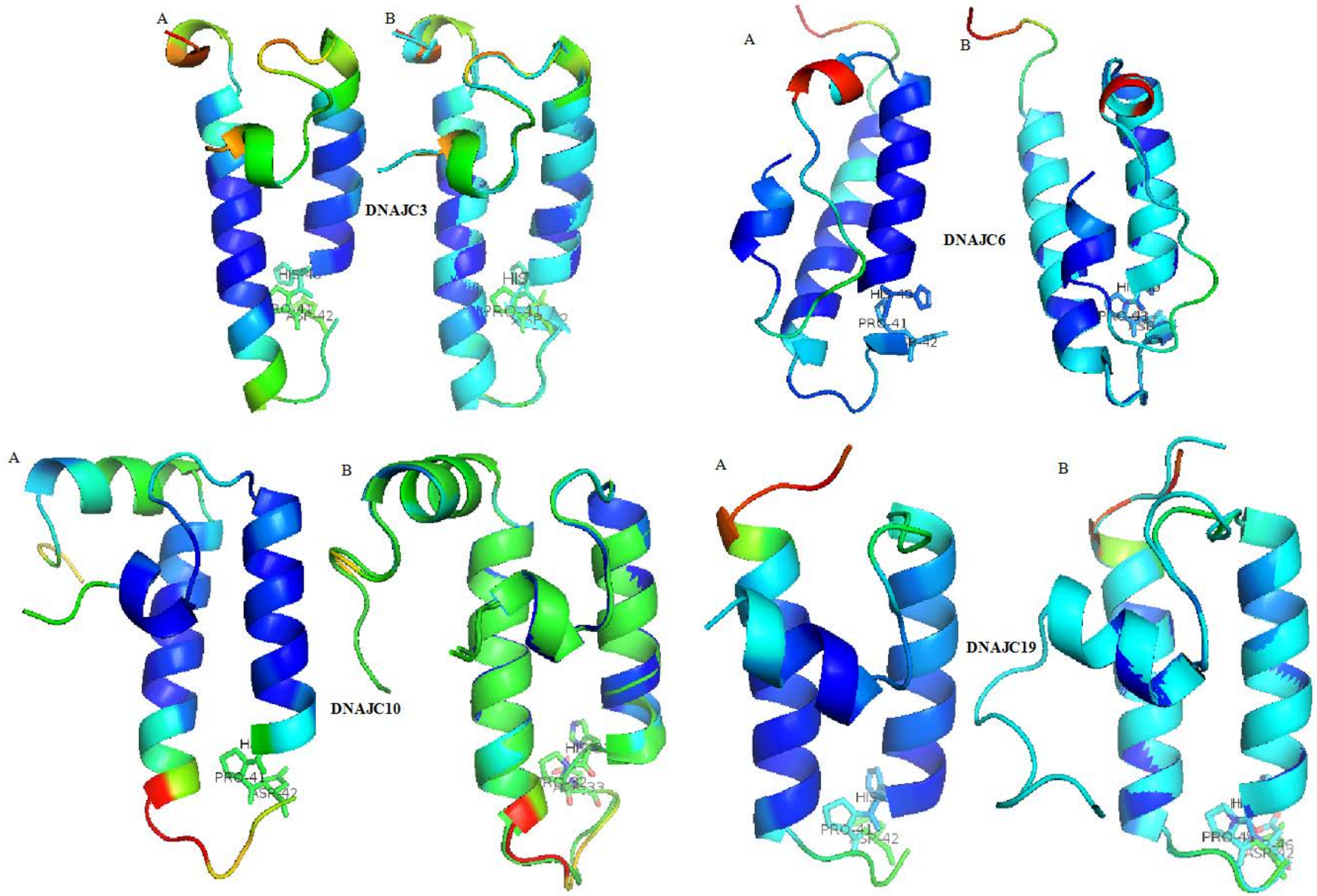
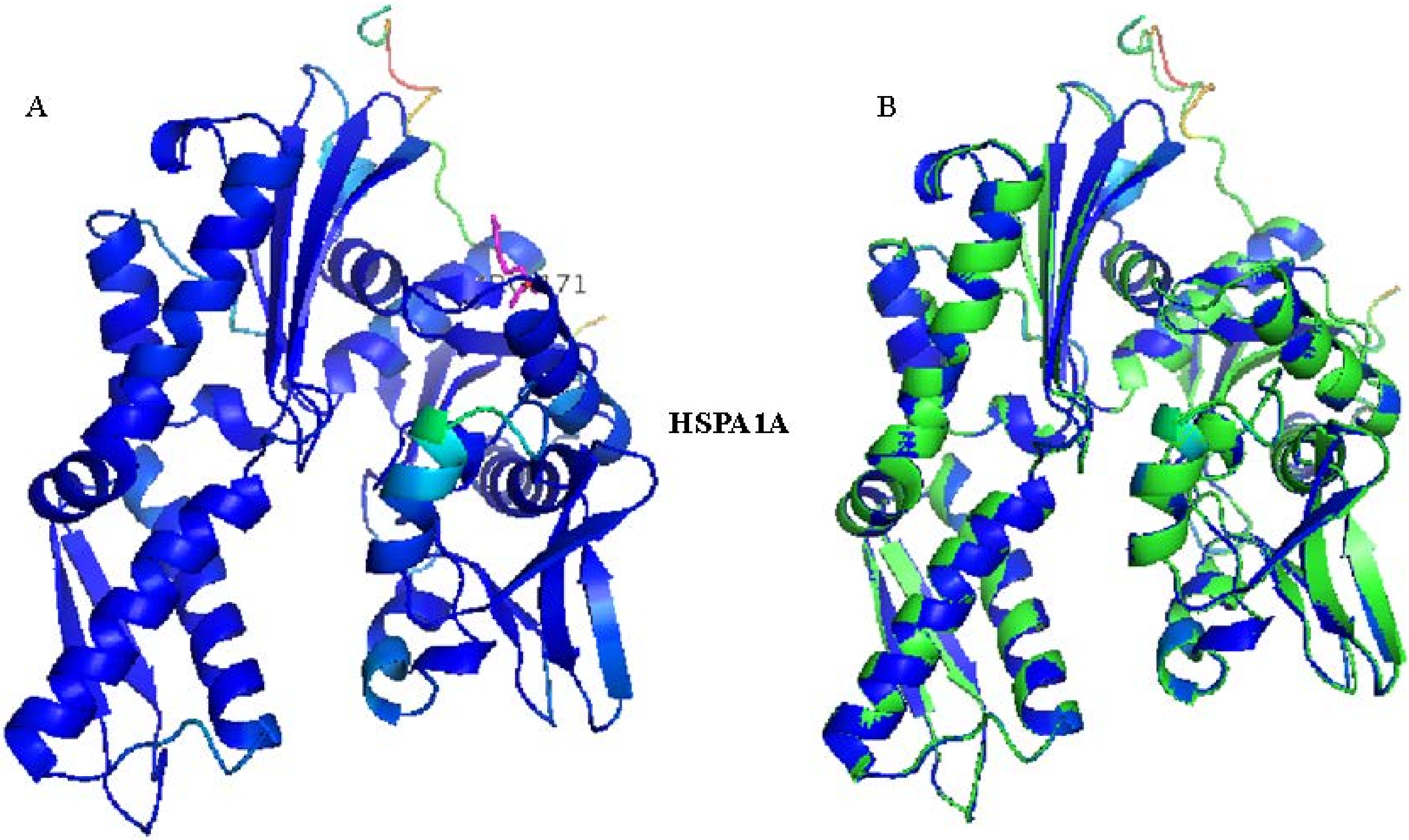
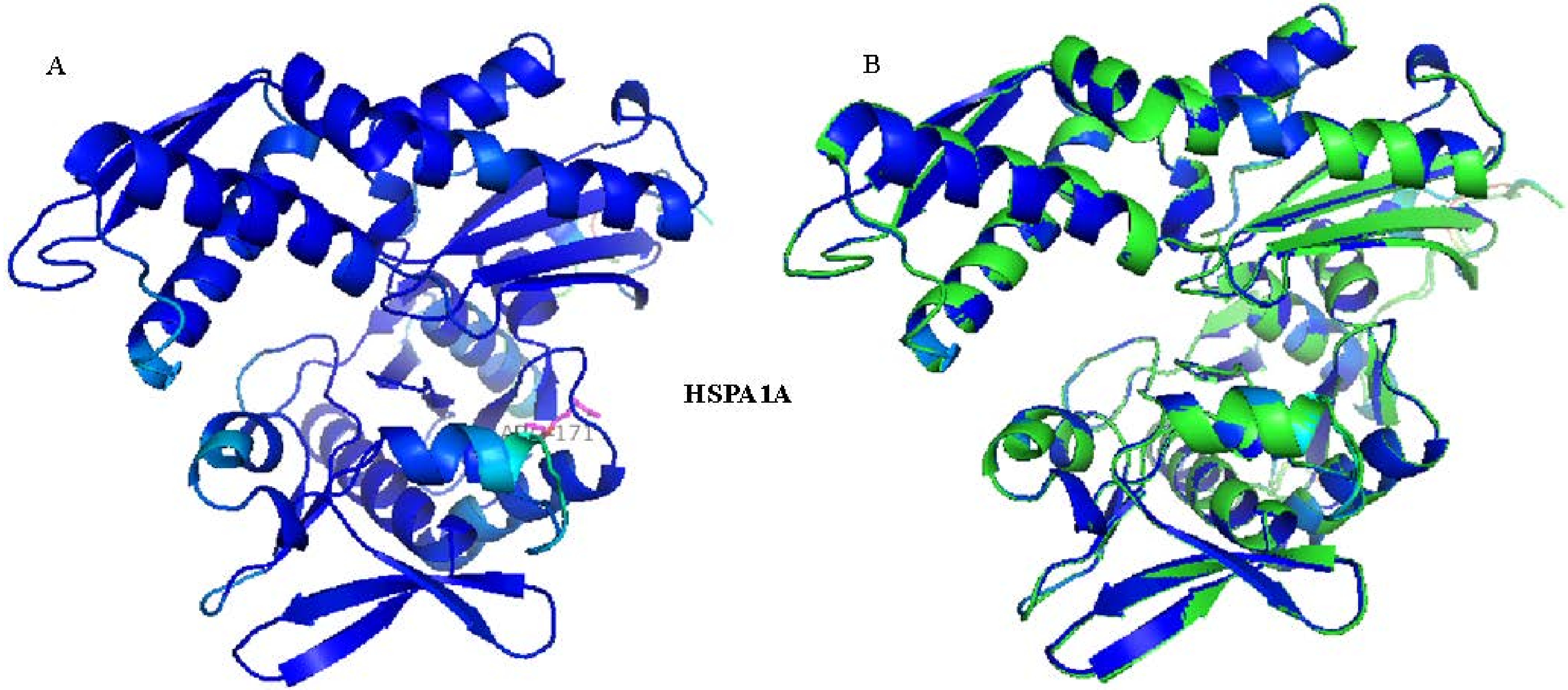
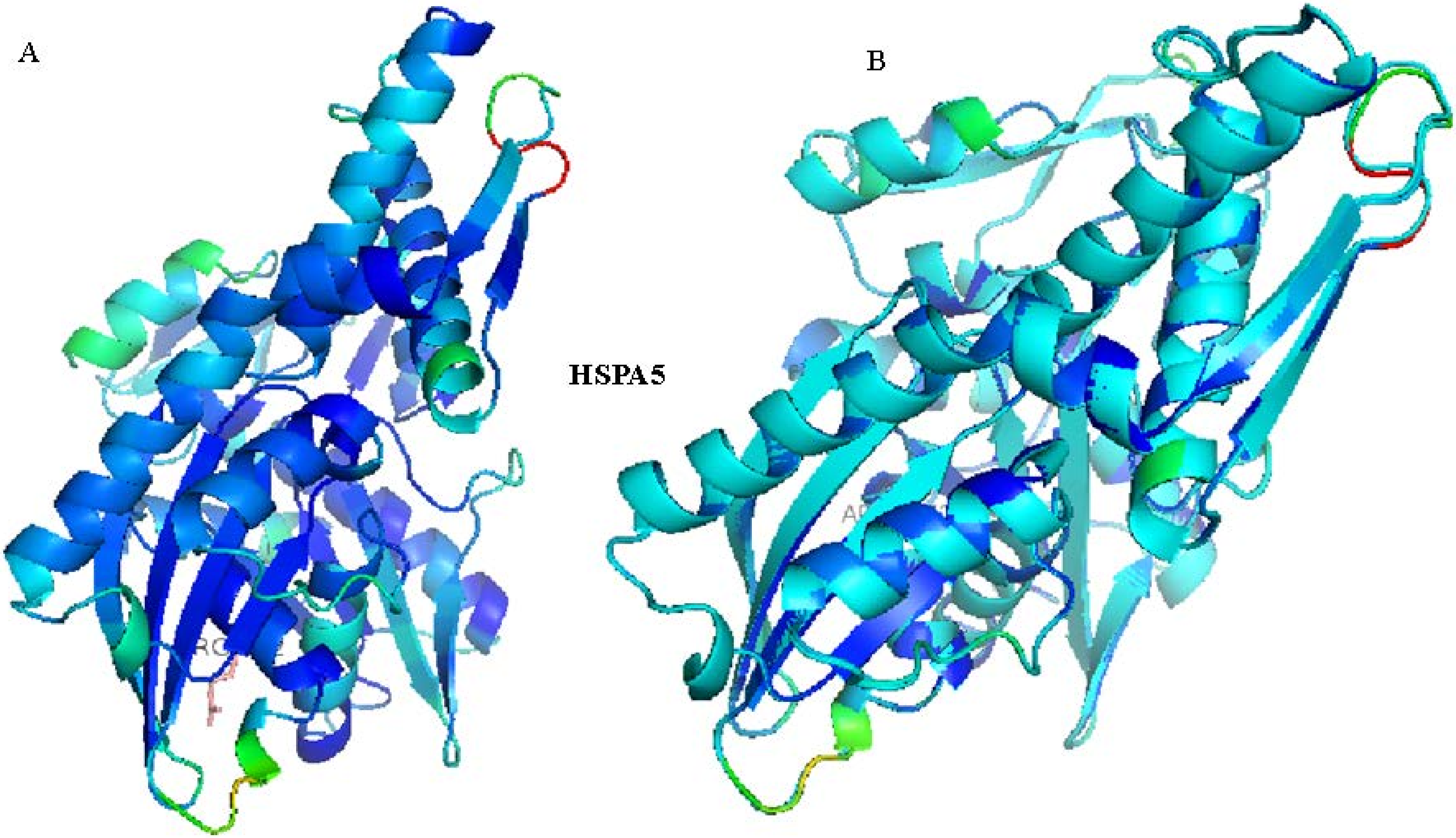


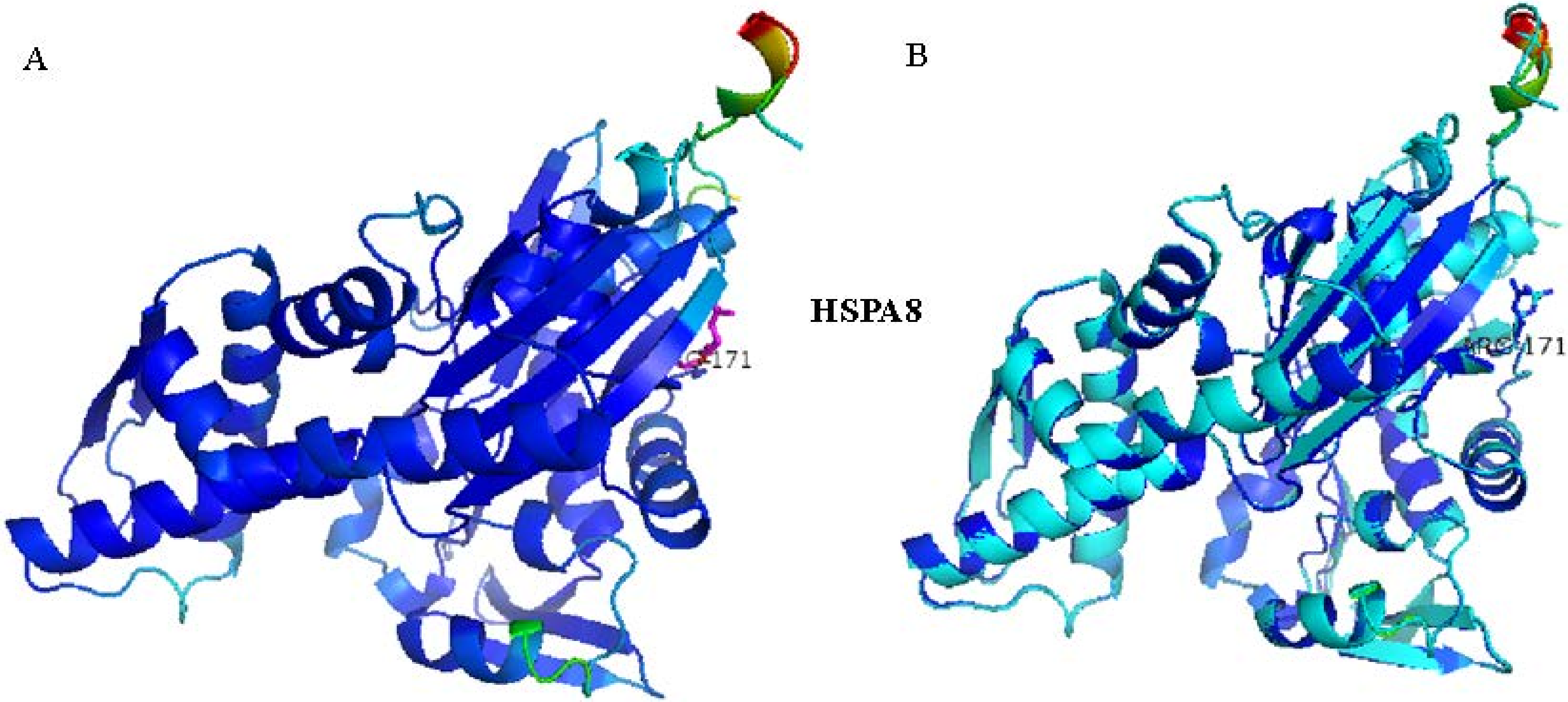
Figure 2: Predicted model structures of HSP40 J domains. (A) Shows the predicted model structure and (B) represents the superposition of the predicted model and the template structure. Models are displayed in cartoon and colored by B-factors. Problematic regions are colored in red while correct and reliable regions are colored green to blue. The HPD motif is depicted in sticks and labeled. Pictures were rendered in PyMol (Delano, 2002).

Appendix IV: Predicted structural models of selected human HSP70 ATPase_linker regions




A
ARG171
HSPA1A
B


A
HSPA5
B


A
B
HSPA8
-171
ARG-171

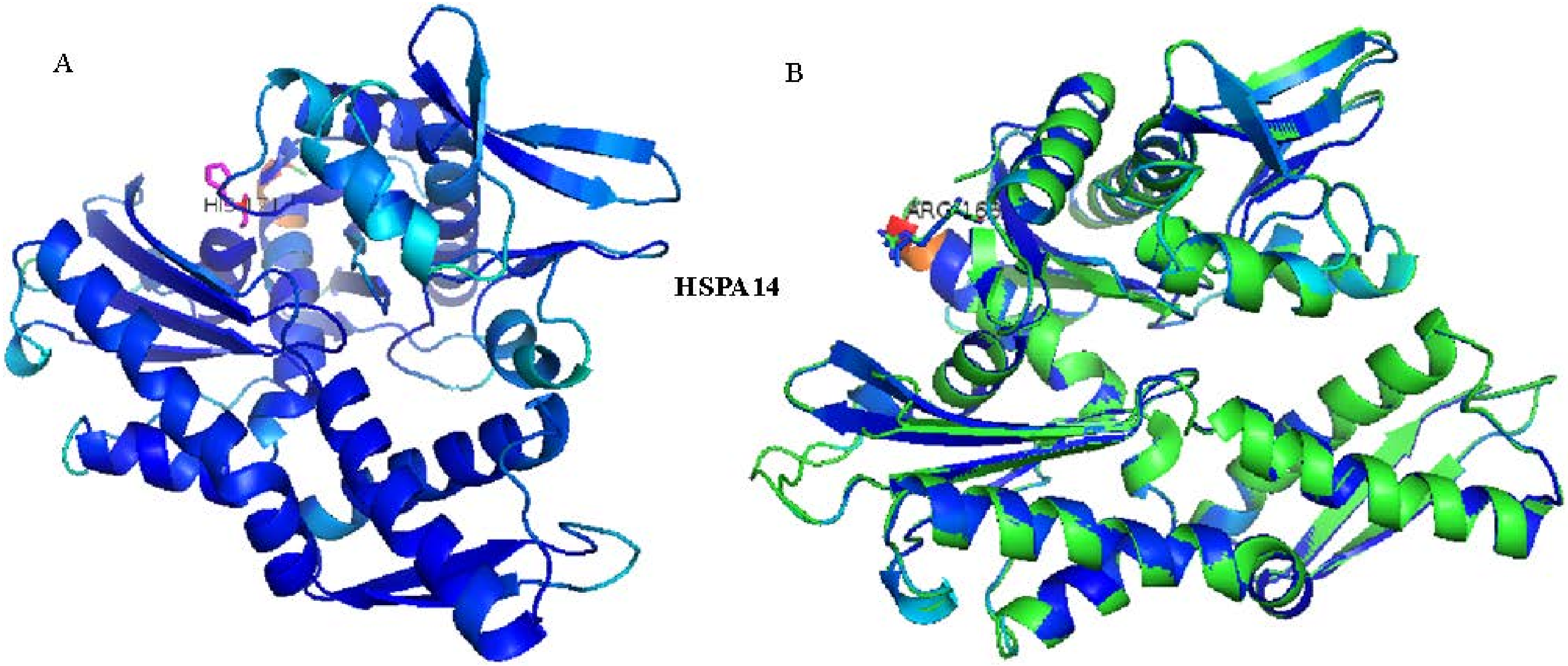
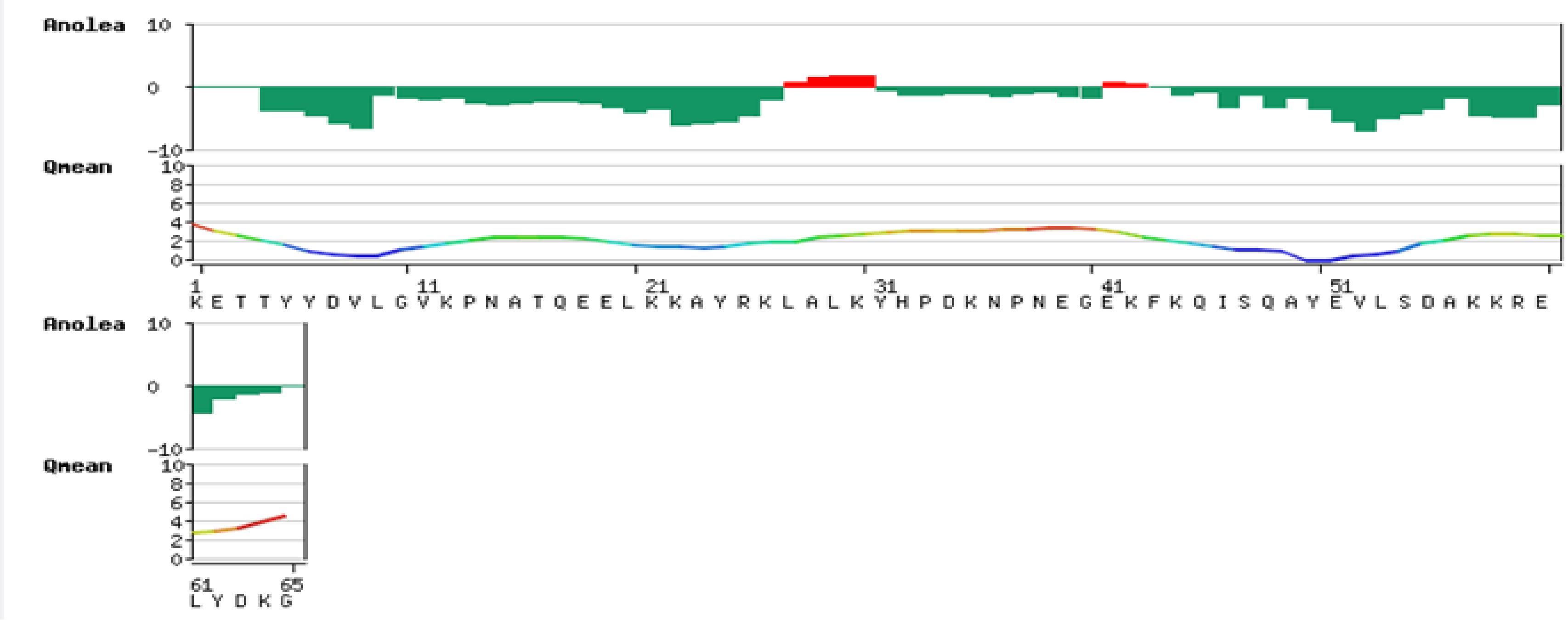


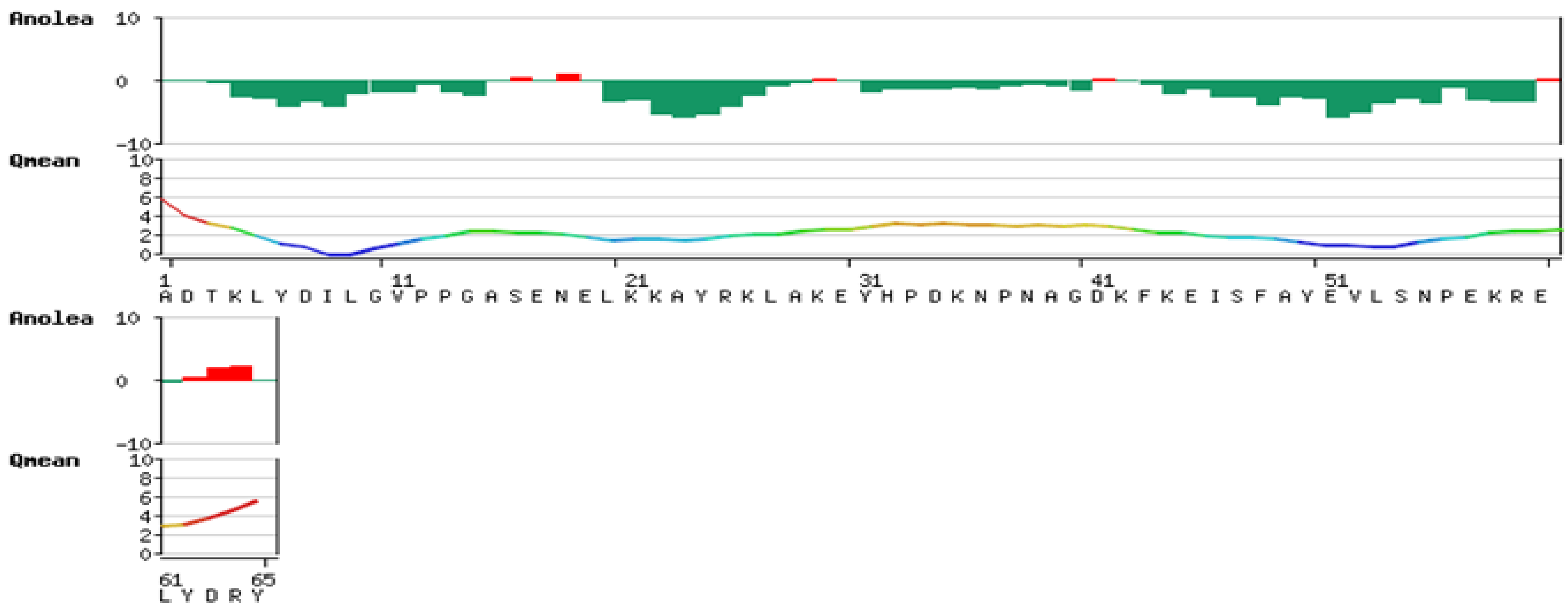
Figure 3: Predicted structures of HSP70 ATPase domain_linker from human using Homology modelling. Proteins are displayed in cartoon and colored by B-factors. The degree of fitness ranges between blue to red color. Problematic regions are colored red. (A) Shows the predicted model and (B) shows the superposition of the predicted model with the template structure. The position of ARG 171 proposed to be important for binding with Hsp40 J domain is highlighted in sticks and labeled accordingly. Pictures are rendered in PyMol (Delano, 2002).

Appendix V: Model quality assessment using Anolea and Qmean evaluations for selected human HSP40 J domains

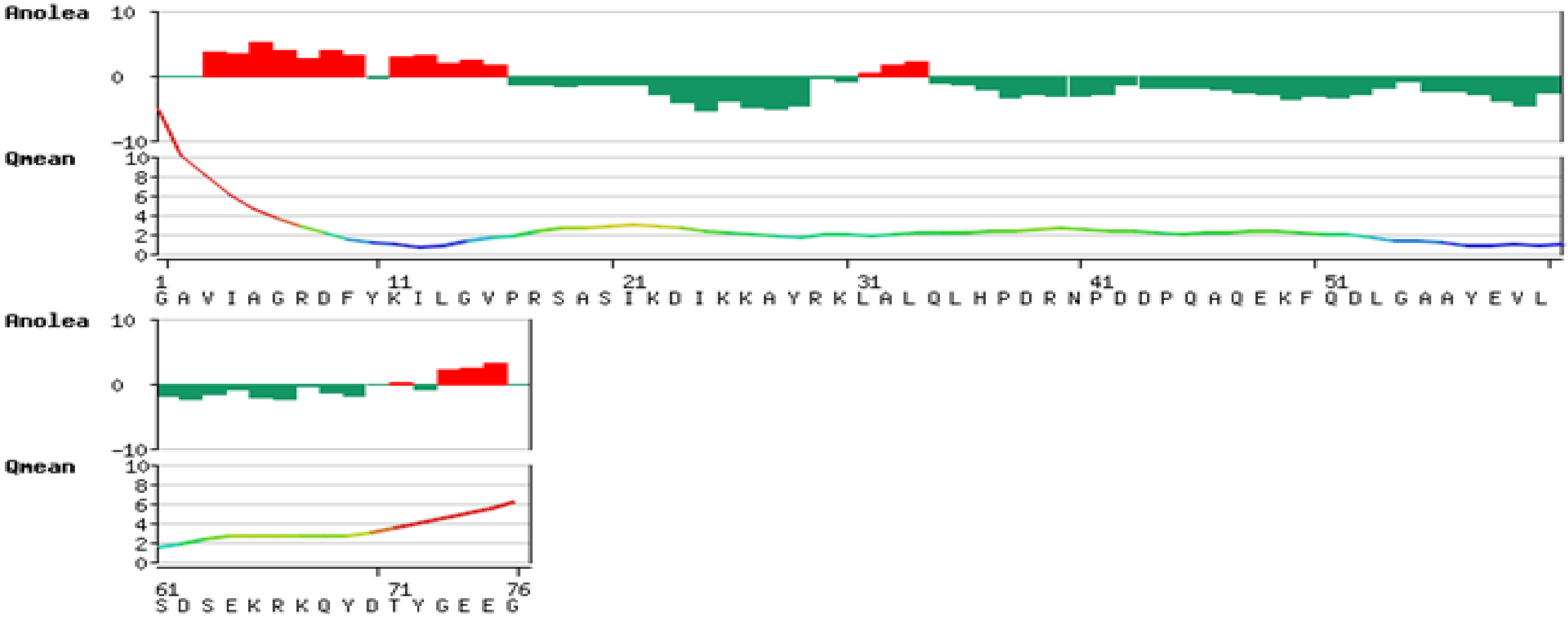
DNAJA1



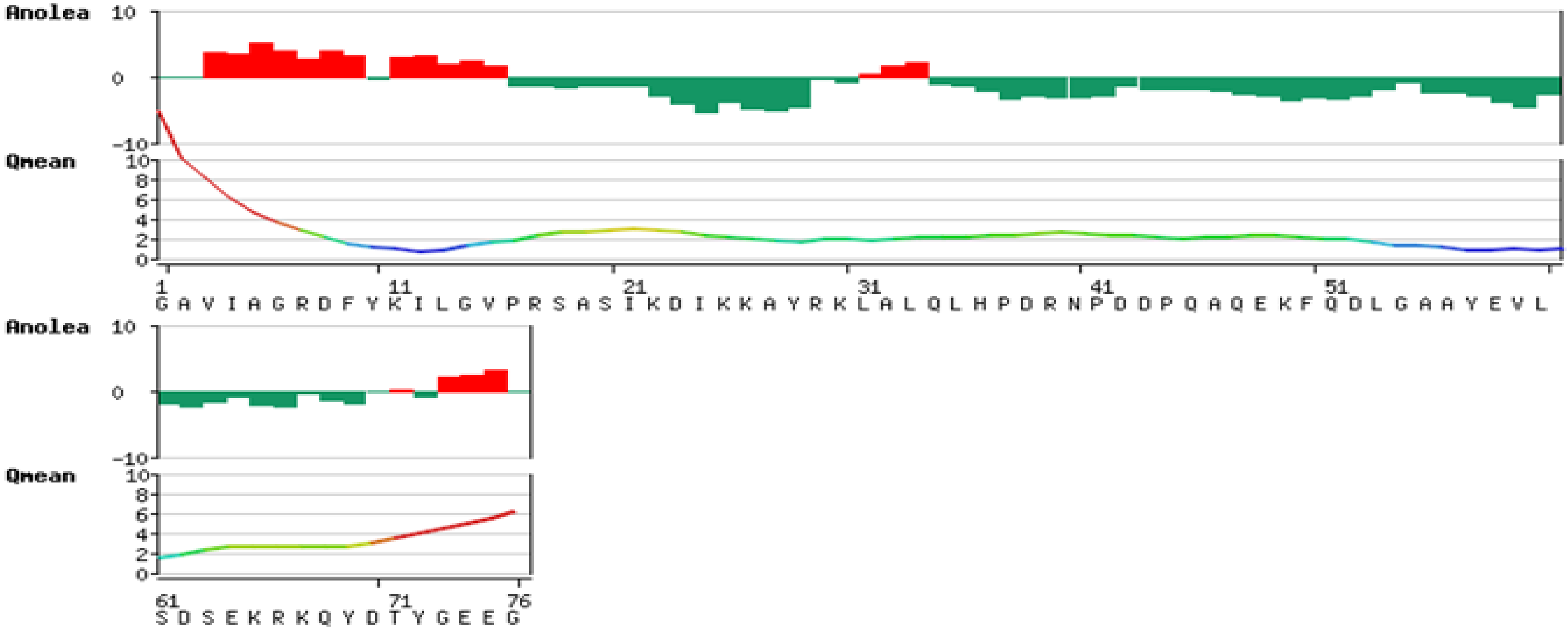
DNAJA2


Anolea
10
0
-10
Qmean
10
8
6
4
2
0
1
11
21
31
41
51
A D T K L Y D I L G V P P G A S E N E L K K A Y R K L A K E Y H P D K N P N A G D K F K E I S F A Y E V L S N P E K R E
Anolea
10
0
-10
Qmean
10
8
6
4
2
0
61
65
L Y D R Y

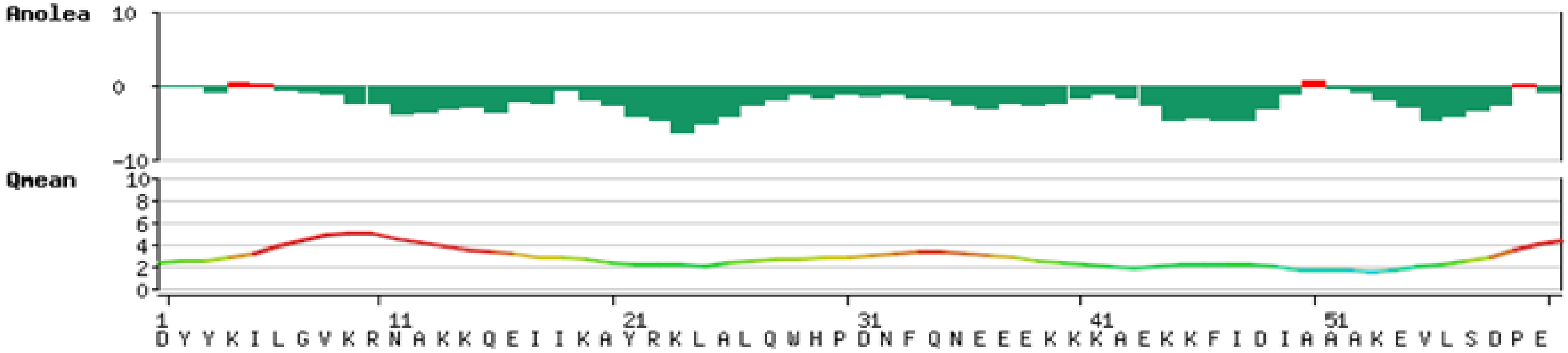
DNAJB11


Anolea
10
0
-10
Qmean
10
8
6
4
2
0
1
11
21
31
41
51
G A V I A G R D F Y K I L G V P R S A S I K D I K K A Y R K L A L Q L H P D R N P D D P Q A Q E K F Q D L G A A Y E V L
Anolea
10
0
-10
Qmean
10
8
6
4
2
0
61
71
76
S D S E K R K Q Y D T Y G E E G

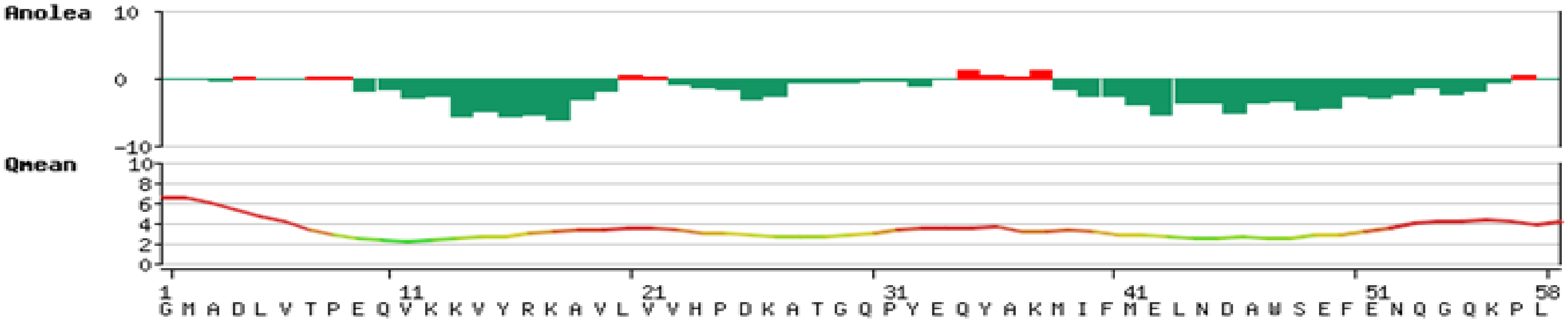
DNAJC2


Anolea
10
0
-10
Qmean
10
8
6
4
2
0
1
11
21
31
41
51
G A V I A G R D F Y K I L G V P R S A S I K D I K K A Y R K L A L Q L H P D R N P D D P Q A Q E K F Q D L G A A Y E V L
Anolea
10
0
-10
Qmean
10
8
6
4
2
0
61
71
76
S D S E K R K Q Y D T Y G E E G

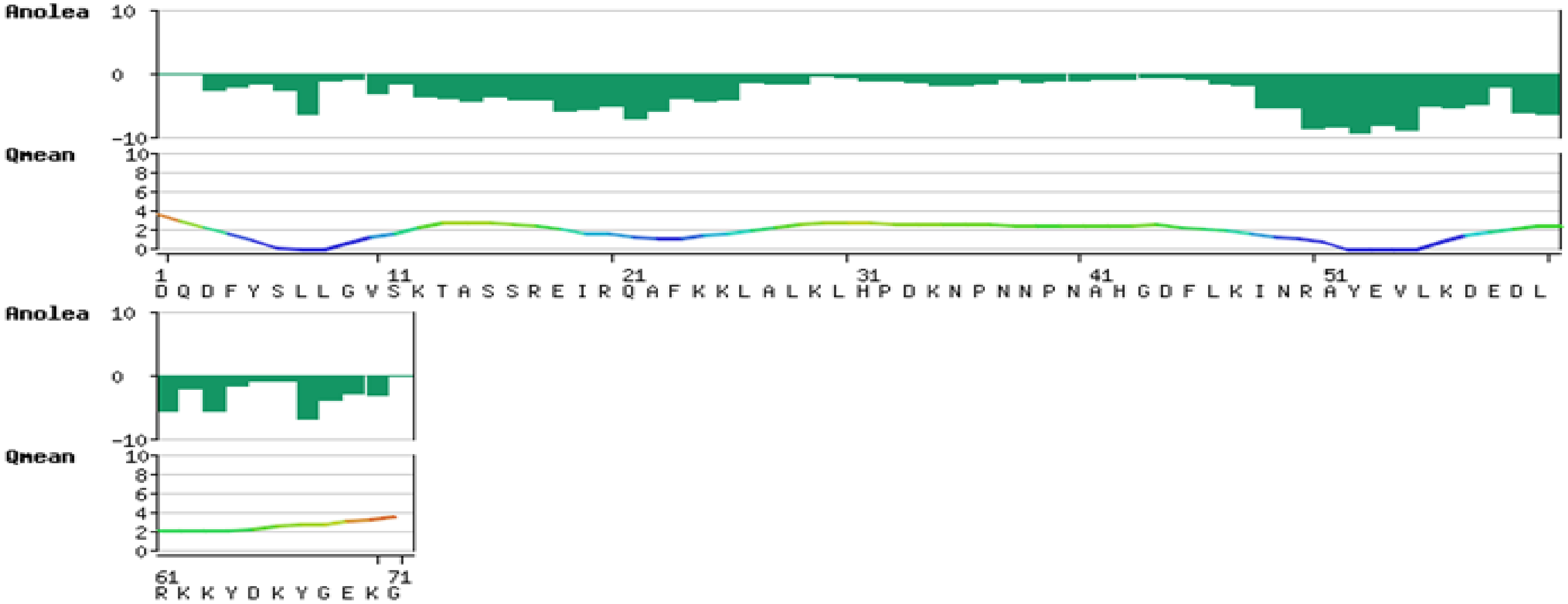
DNAJC3



DNAJC6



DNAJC10


Anolea
10
0
-10
Qmean
10
8
6
4
2
0
1
11
21
31
41
51
D Q D F Y S L L G V S K T A S S R E I R Q A F K K L A L K L H P D K N P N N P N A H G D F L K I N R A Y E V L K D E D L
Anolea
10
0
-10
Qmean
10
8
6
4
2
0
61
71
R K K Y D K Y G E K G

DNAJC19

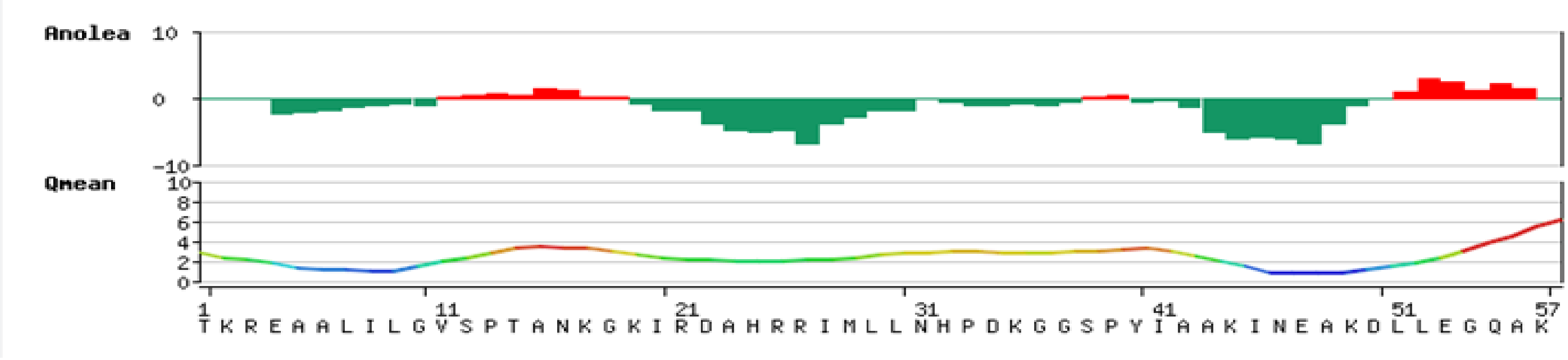
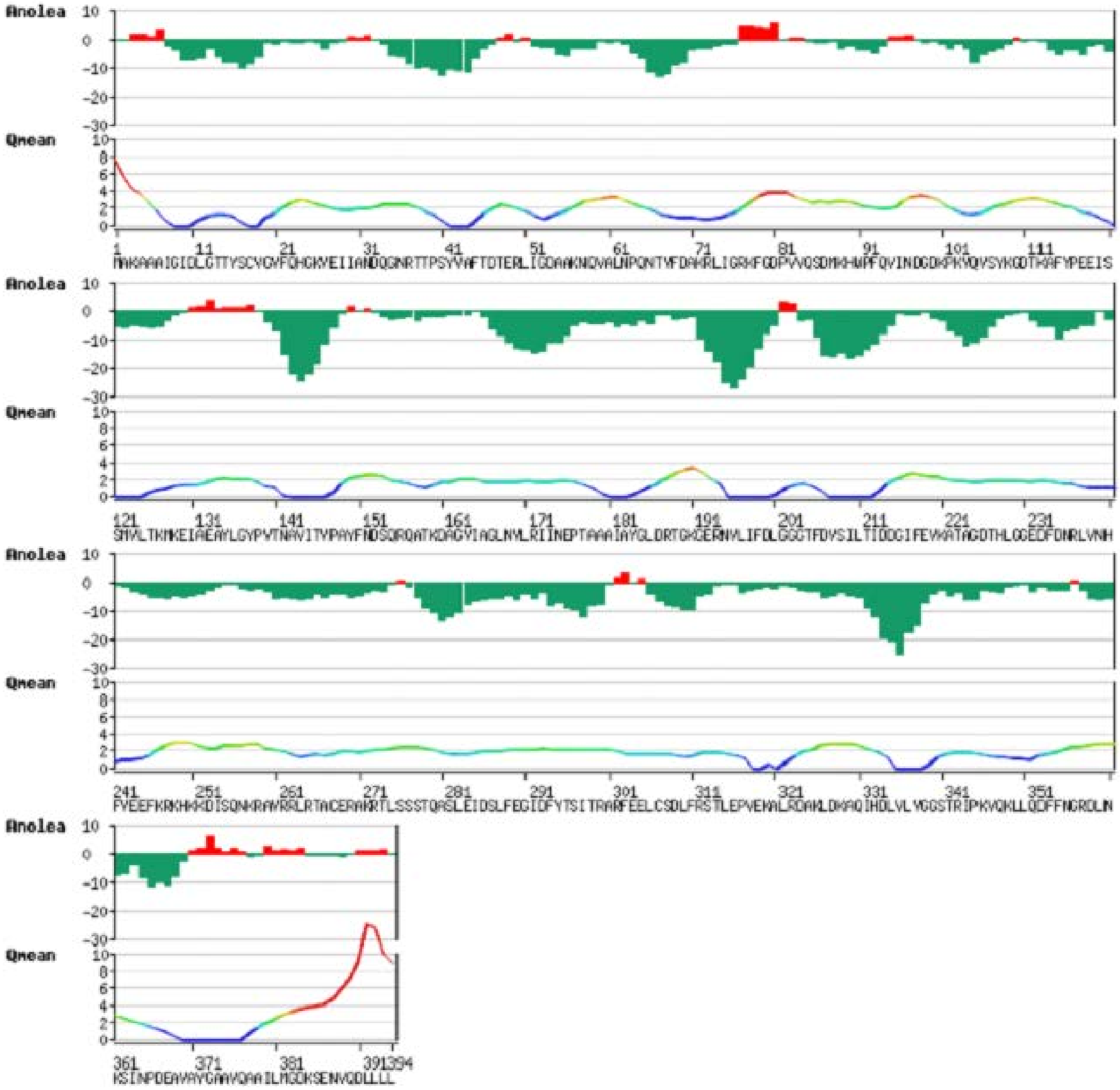


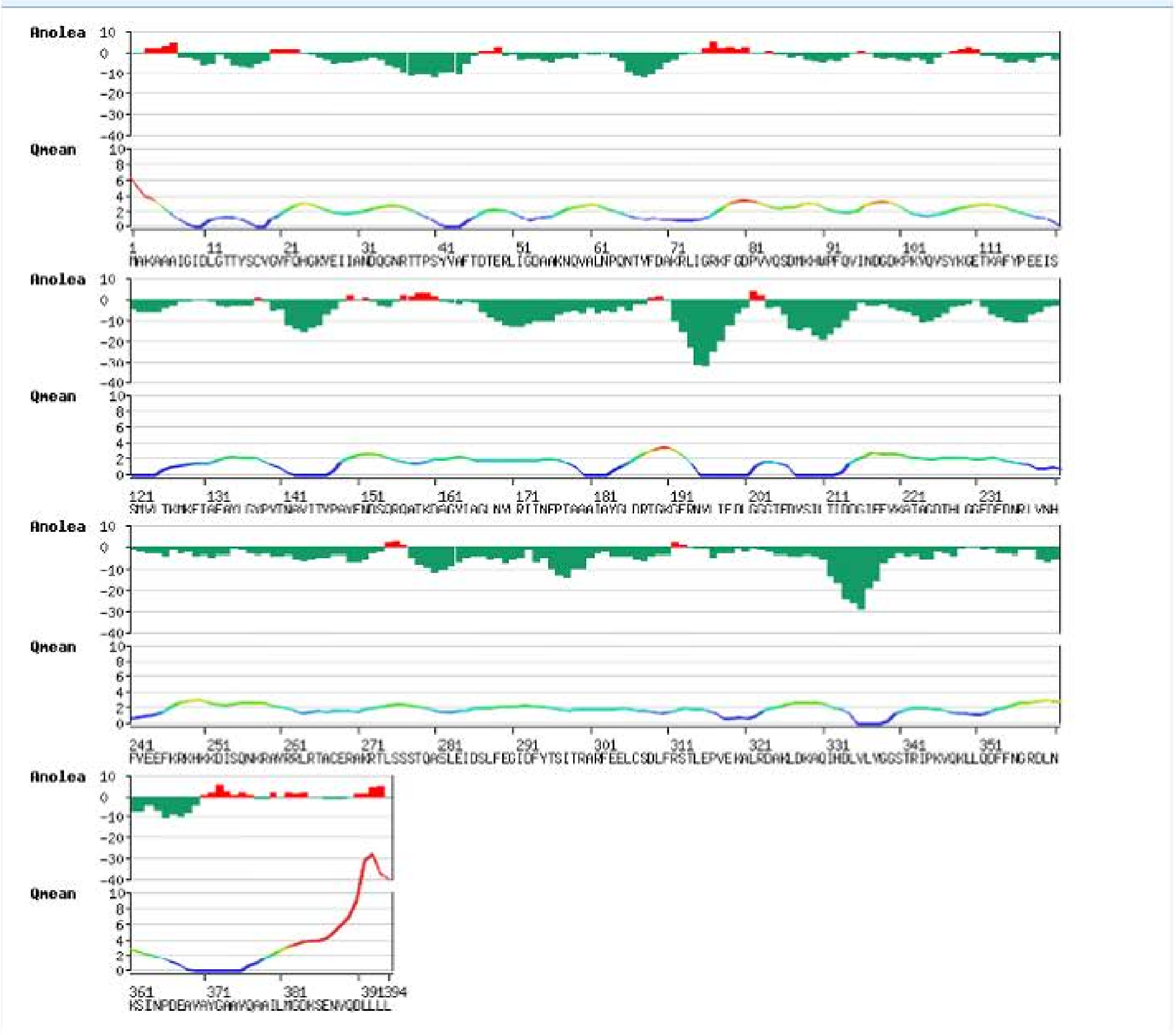
Figure 4: Model quality assessments using Anolea and Qmean evaluations respectively for selected human HSP40 J domains. Problematic residues within the model are colored in red and reliable residues are colored ranging from yellow, green, and blue according to their quality.

Appendix VI: Model quality assessment using Anolea and Qmean evaluations for selected human HSP70 ATPase-linker regions

HSPA1A

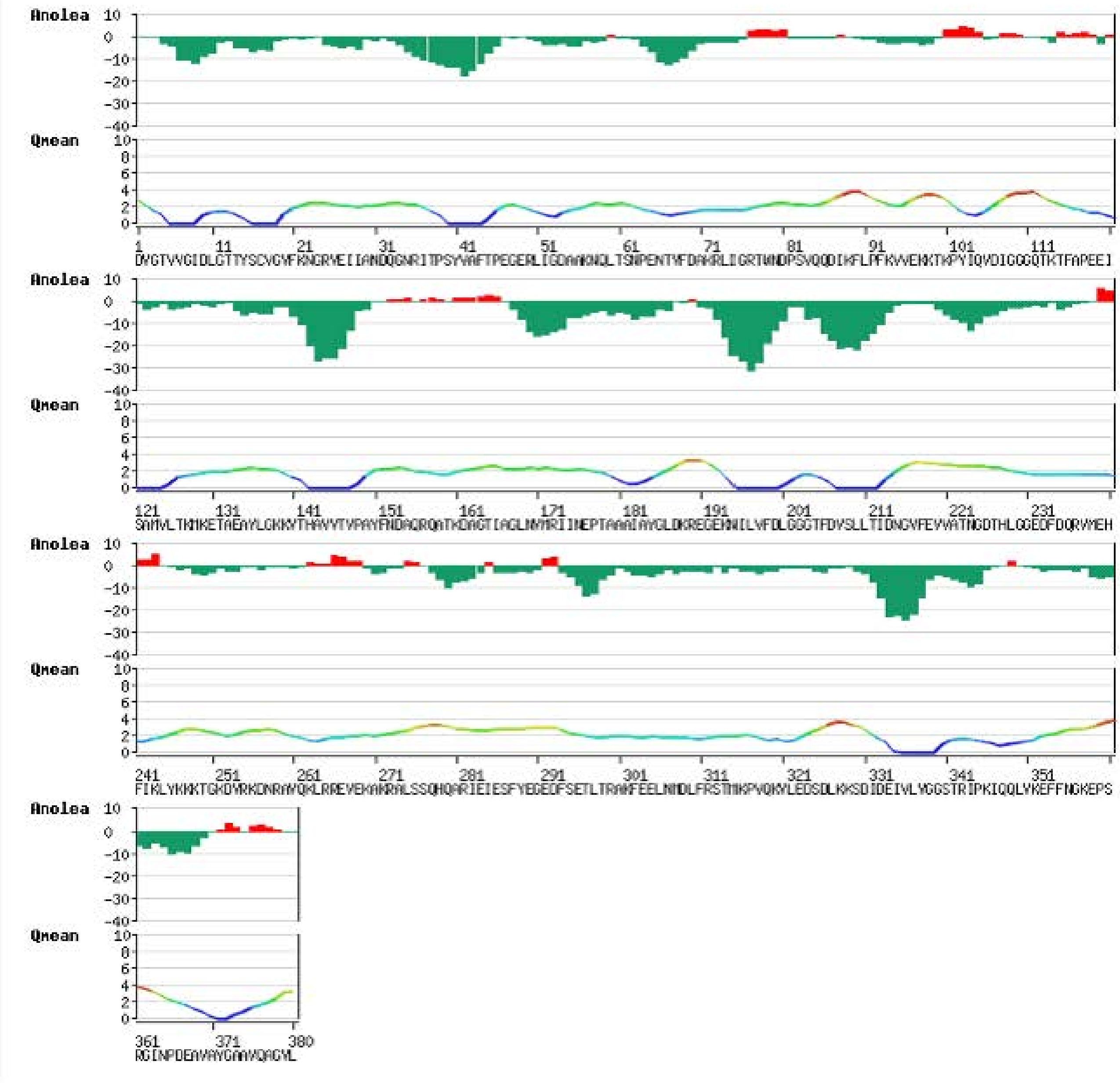


HSPA1B

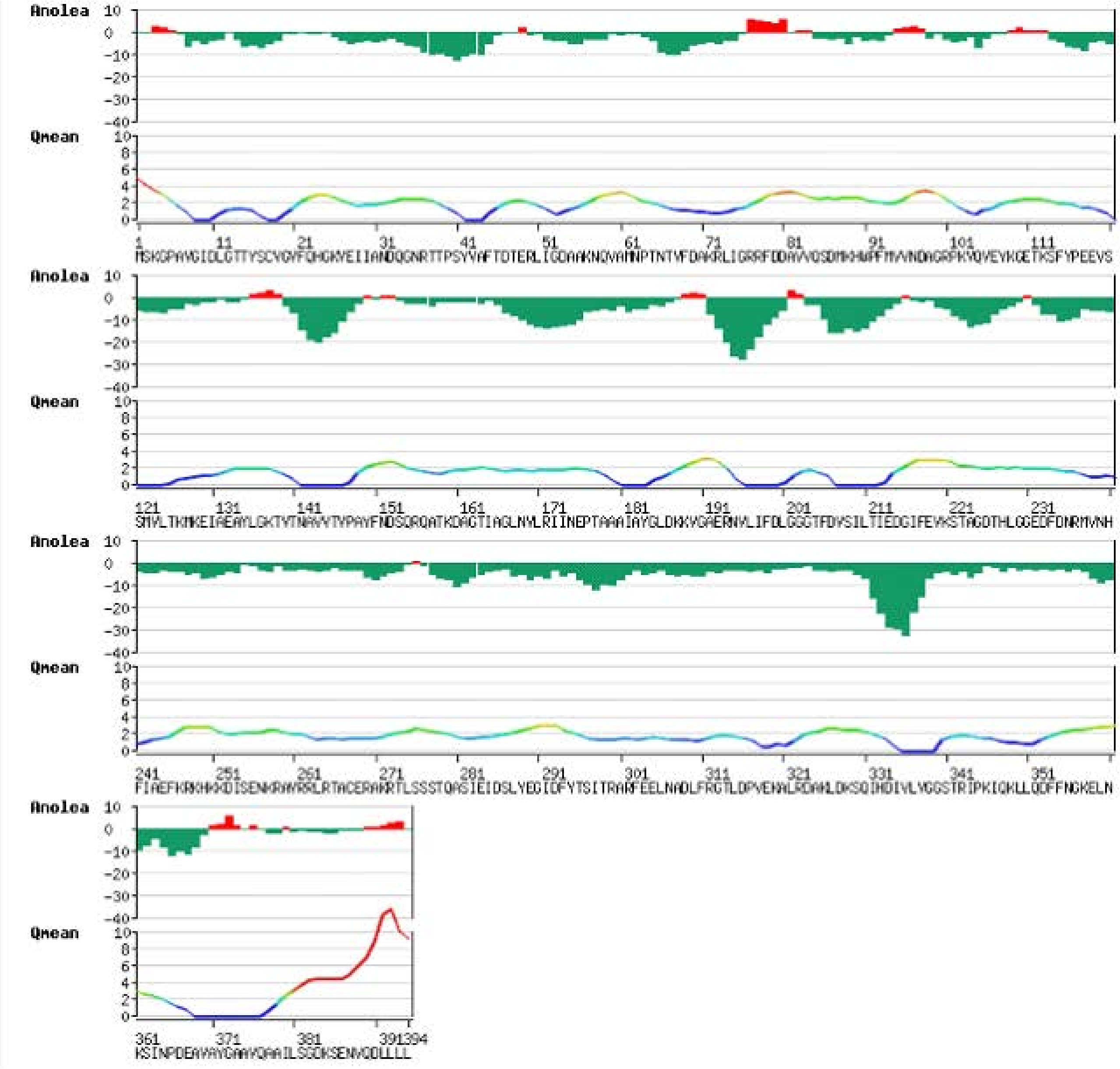


Anolea
10
0
-10
-20
-30
-40
Qmean
10
8
6
4
2
0
1 11 21 31 41 51 61 71 81 91 101 111
MAKAAAIGIDLGTTYSCVGVFQHGKVEIIANDQGNRTTPSYVAFTDTERLIGDAAKNQVALNPQNTVFDAKRLIGRKFGDPVVQSDMKHWPFQVINDGDKPKVQVSYKGETKAFYPEEIS
121 131 141 151 161 171 181 191 201 211 221 231
SMVLTKMKEIAEAYLGYPVTNAVITVPAYFNDSQRQATKDAGVIAGLNVLRIINEPTAAAIAYGLDRTGKGERNVLIFDLGGGTFDVSILTIDDGIFEVKATAGDTHLGGEDFDNRLVNH
241 251 261 271 281 291 301 311 321 331 341 351
FVEEFKRKHKKDISQNKRAVRRLRTACERAKRTLSSSTQASLEIDSLFEGIDFYTSITRARFEELCSDLFRSTLEPVEKALRDAKLDKAQIHDLVLVGGSTRIPKVQKLLQDFFNGRDLN
361 371 381 391394
KSINPDEAVAYGAAVQAAILMGDKSENVQDLLLL

HSPA5


Anolea
10
0
-10
-20
-30
-40
Qmean
10
8
6
4
2
0
1
11
21
31
41
51
61
71
81
91
101
111
DVGTVVGIDLGTTYSCVGVFKNGRVEIIANDQGNRITPSYVAFTPEGERLIGDAAKNQLTSNPENTVFDAKRLIGRTWNDPSVQQDIKFLPFKVVEKKTKPYIQVDIGGGQTKTFAPEEI
Anolea
Qmean
121
131
141
151
161
171
181
191
201
211
221
231
SAMVLTKMKETAEAYLGKKVTHAVVTVPAYFNDAQRQATKDAGTIAGLNVMRIINEPTAAAIAYGLDKREGEKNILVFDLGGGTFDVSLLTIDNGVFEVVATNGDTHLGGEDFDQRVMEH
Anolea
Qmean
241
251
261
271
281
291
301
311
321
331
341
351
FIKLYKKKTGKDVRKDNRAVQKLRREVEKAKRALSSQHQARIEIESFYEGEDFSETLTRAKFEELNMDLFRSTMKPVQKVLEDSDLKKSDIDEIVLVGGSTRIPKIQQLVKEFFNGKEPS
Anolea
Qmean
361
371
380
RGINPDEAVAYGAAVQAGVL

HSPA8


Anolea
10
0
-10
-20
-30
-40
Qmean
10
8
6
4
2
0
1 11 21 31 41 51 61 71 81 91 101 111
MSKGPAVGIDLGTTYSCVGVFQHGKVEIIANDQGNRTTPSYVAFTDTERLIGDAAKNQVAMNPTNTVFDAKRLIGRRFDDAVVQSDMKHWPFMVVNDAGRPKVQVEYKGETKSFYPEEVS
121 131 141 151 161 171 181 191 201 211 221 231
SMVLTKMKEIAEAYLGKTVTNAVVTVPAYFNDSQRQATKDAGTIAGLNVLRIINEPTAAAIAYGLDKKVGAERNVLIFDLGGGTFDVSILTIEDGIFEVKSTAGDTHLGGEDFDNRMVNH
241 251 261 271 281 291 301 311 321 331 341 351
FIAEFKRKHKKDISENKRAVRRLRTACERAKRTLSSSTQASIEIDSLYEGIDFYTSITRARFEELNADLFRGTLDPVEKALRDAKLDKSQIHDIVLVGGSTRIPKIQKLLQDFFNGKELN
361 371 381 391394
KSINPDEAVAYGAAVQAAILSGDKSENVQDLLLL

HSPA14

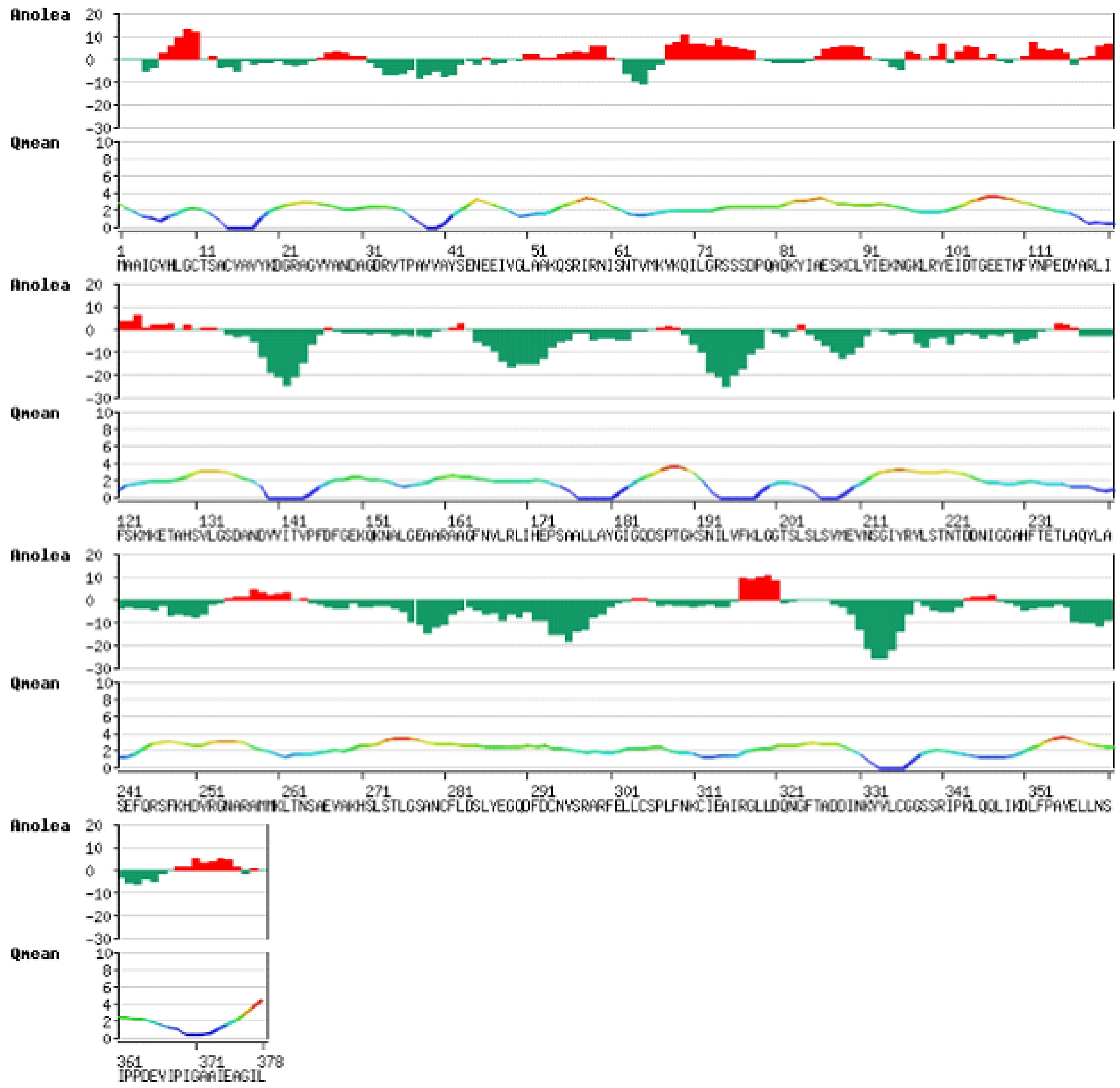


Figure 5: Model quality assessments using Anolea and Qmean evaluations respectively for human HSP70 ATPase-linker region. Problematic residues within the model are colored in red and reliable residues are colored ranging from yellow, green, and blue according to their quality.

CHAPTER FOUR

Appendix I: Predicted complex model structures of ATPase domain_linker region and J domain of HSP70 and HSP40 respectively.

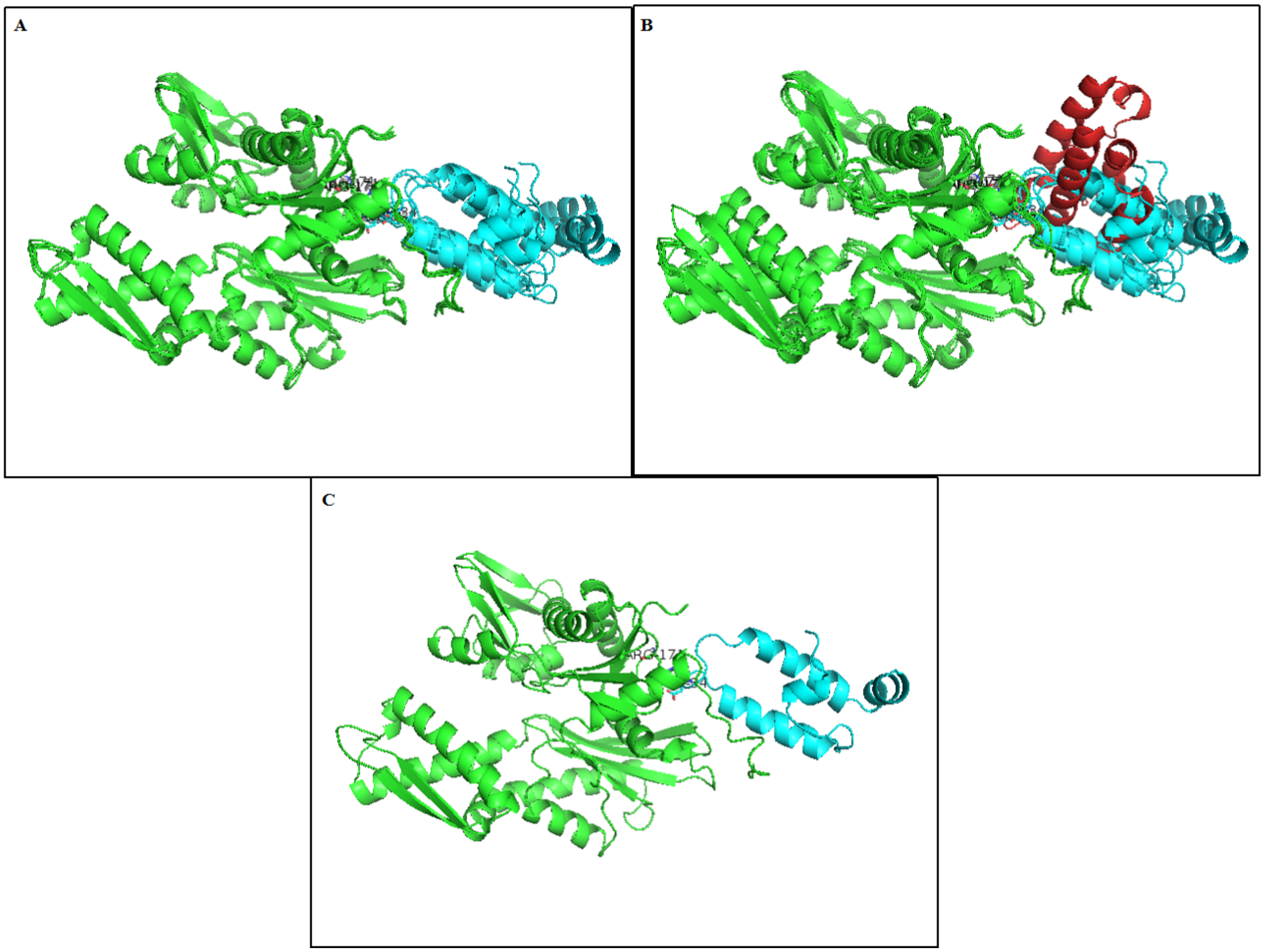


Figure 6: HSPA1B-DNAJA1 complex. Structures are represented in cartoon. ARG 171 and ASP 34; predicted to be involved in HSP70-HSP40 interactions; are shown as sticks and labeled accordingly. (A) The four best predicted complexes by HADDOCK were superposed. ATPase domain_linker colored green and J domain is colored cyan. (B) Superposed structures of 2QWO and complex structures from (A). The ATPase-linker domain of the structures is colored in green. While the J domain in 2QWO is colored in red, that of HSPA1B-DNAJA1 complex is colored cyan. (C) Structure of the best selected complex model with the least energy. Pictures are rendered using PyMol (Delano, 2002).

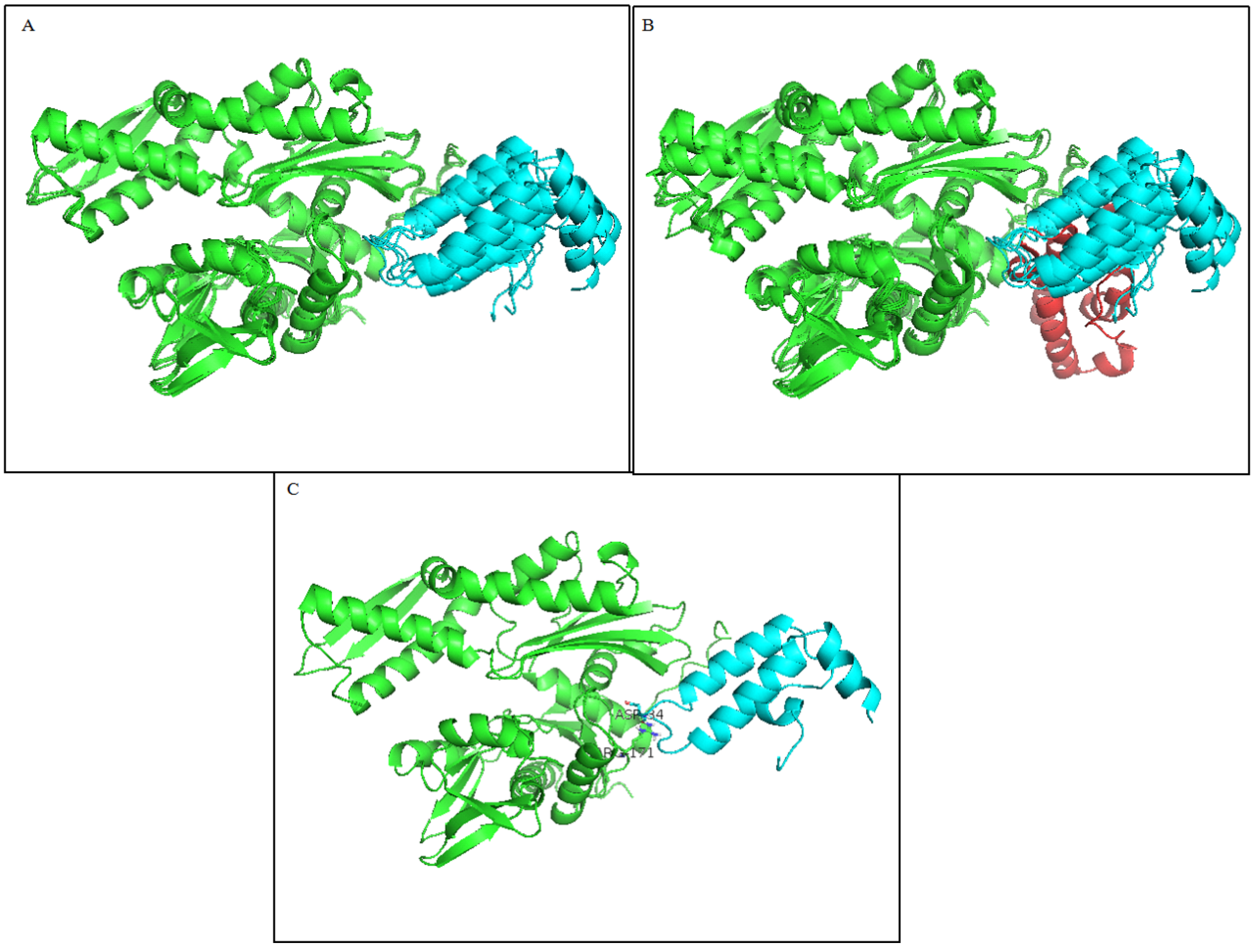


Figure 7: HSPA8-DNAJA2 complex. Structures are represented in cartoon. ARG 171 and ASP 34; predicted to be involved in HSP70-HSP40 interactions; are shown as sticks and labeled accordingly. (A) The four best predicted complexes by HADDOCK were superposed. ATPase domain_linker colored green and J domain is colored cyan. (B) Superposed structures of 2QWO and complex structures from (A). The ATPase-linker domain of the structures is colored in green. While the J domain in 2QWO is colored in red, that of HSPA8-DNAJA2 complex is colored cyan. (C) Structure of the best selected complex model with the least energy. Pictures were rendered using PyMol (Delano, 2002).

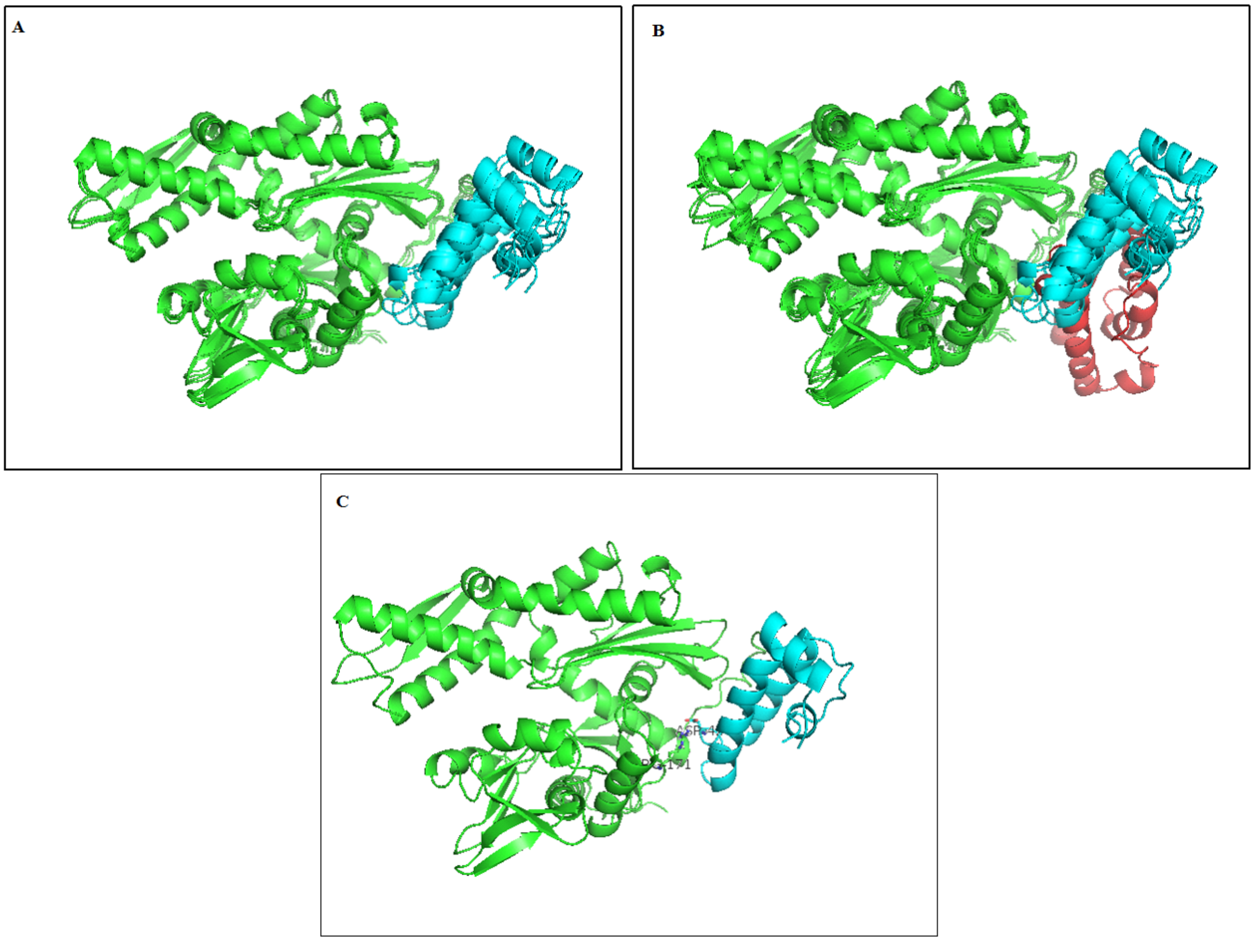


Figure 8: HSPA1A-DNAJB11 complex. Structures are represented in cartoon. ARG 171 and ASP 42; predicted to be involved in Hsp70-Hsp40 interactions; are shown as sticks and labeled accordingly. (A) The four best predicted complexes by HADDOCK were superposed. ATPase domain_linker colored green and J domain is colored cyan. (B) Superposed structures of 2QWO and complex structures from (A). The ATPase-linker domain of the structures is colored in green. While the J domain in 2QWO is colored in red, that of HSPA1A-DNAJB11 complex is colored cyan. (C) Structure of the best selected complex model with the least energy. Pictures were rendered using PyMol (Delano, 2002).

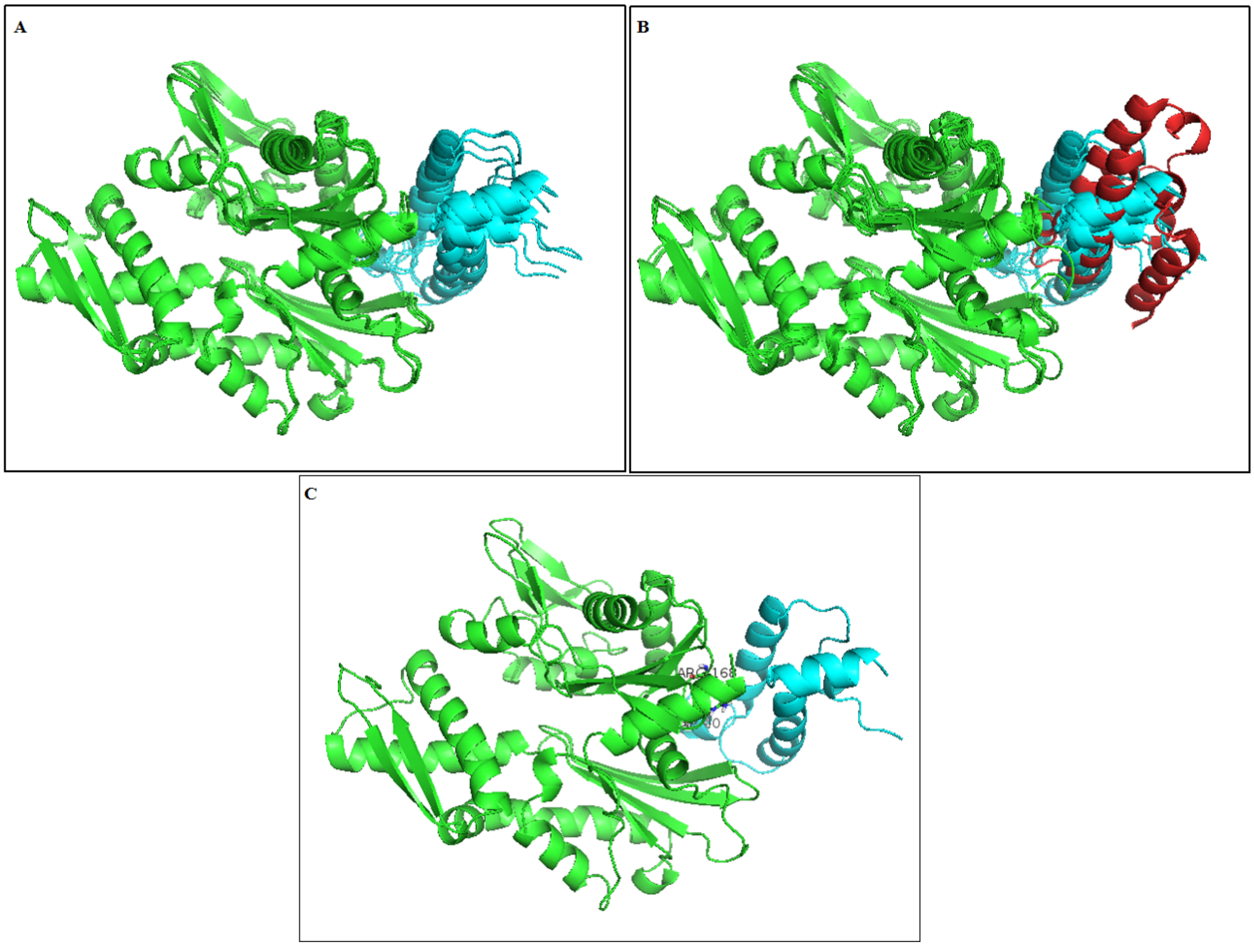


Figure 9: HSPA14-DNAJC2 complex. Structures are represented in cartoon. ARG 171 and ASP 40; predicted to be involved in Hsp70-Hsp40 interactions; are shown as sticks and labeled accordingly. (A) The four best predicted complexes by HADDOCK were superposed. ATPase domain_linker colored green and J domain is colored cyan. (B) Superposed structures of 2QWO and complex structures from (A). The ATPase-linker domain of the structures is colored in green. While the J domain in 2QWO is colored in red, that of HSPA14-DNAJC2 complex is colored cyan. (C) Structure of the best selected complex model with the least energy. Pictures were rendered using PyMol (Delano, 2002).

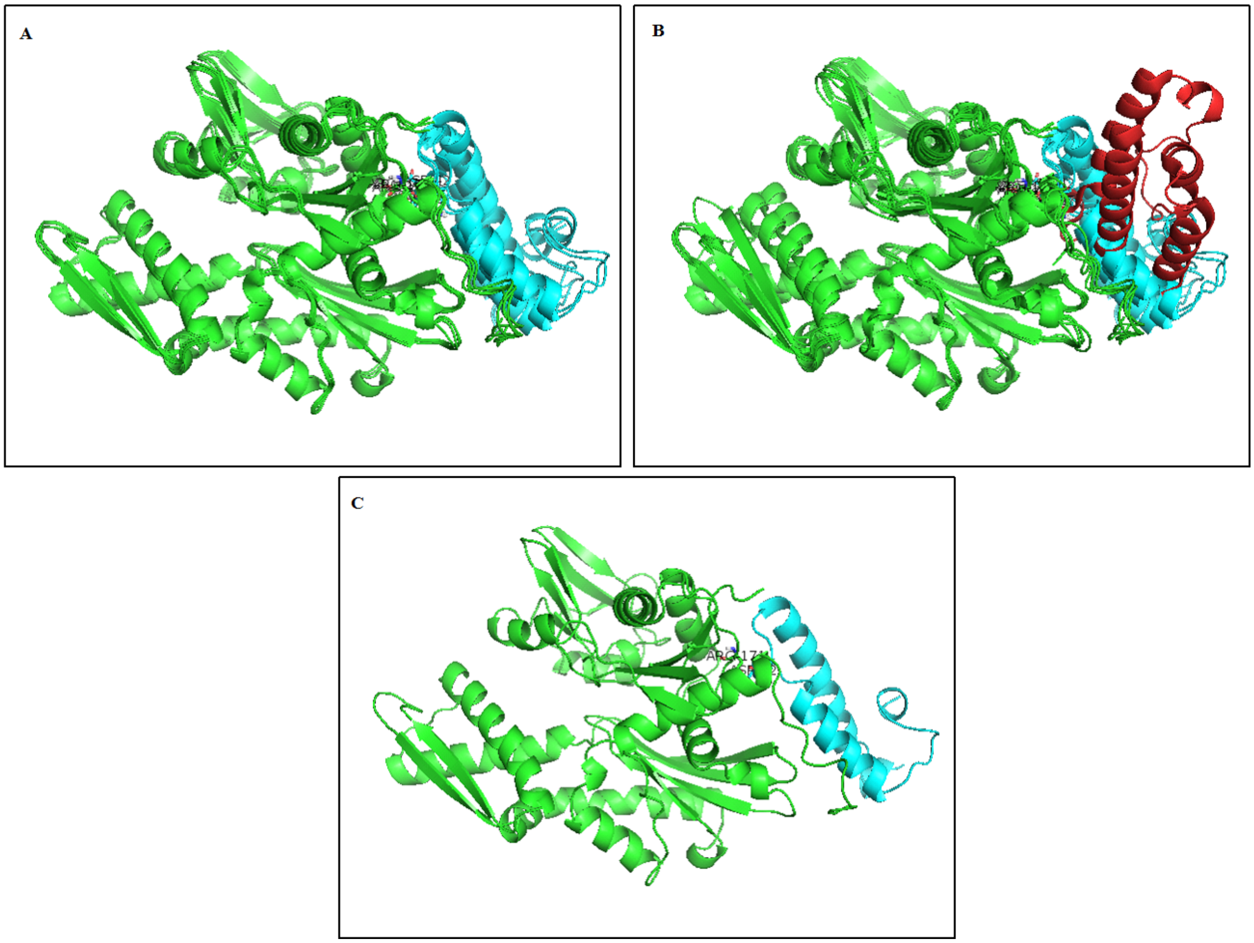


Figure 10: HSPA1A-DNAJC3 complex. Structures are represented in cartoon. ARG 171 and ASP 42; predicted to be involved in Hsp70-Hsp40 interactions; are shown as sticks and labeled accordingly. (A) The four best predicted complexes by HADDOCK were superposed. ATPase domain_linker colored green and J domain is colored cyan. (B) Superposed structures of 2QWO and complex structures from (A). The ATPase-linker domain of the structures is colored in green. While the J domain in 2QWO is colored in red, that of HSPA1A-DNAJC3 complex is colored cyan. (C) Structure of the best selected complex model with the least energy. Pictures were rendered using PyMol (Delano, 2002).

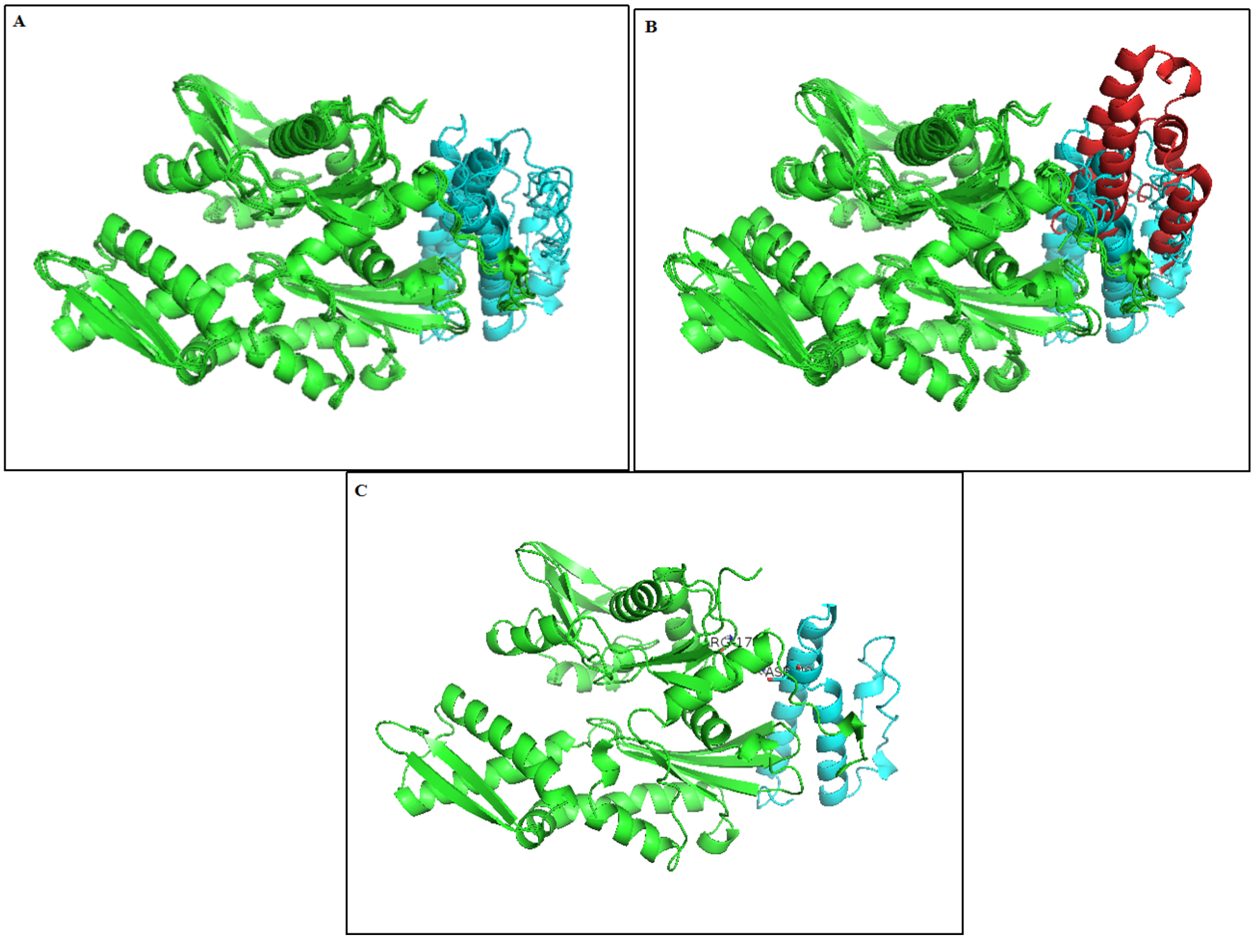


Figure 11: HSPA8-DNAJC6 complex. Structures are represented in cartoon. ARG 171 and ASP 42; predicted to be involved in Hsp70-Hsp40 interactions; are shown as sticks and labeled accordingly. (A) The four best predicted complexes by HADDOCK were superposed. ATPase domain_linker colored green and J domain is colored cyan. (B) Superposed structures of 2QWO and complex structures from (A). The ATPase-linker domain of the structures is colored in green. While the J domain in 2QWO is colored in red, that of HSPA8-DNAJC6 complex is colored cyan. (C) Structure of the best selected complex model with the least energy. Pictures were rendered using PyMol (Delano, 2002).

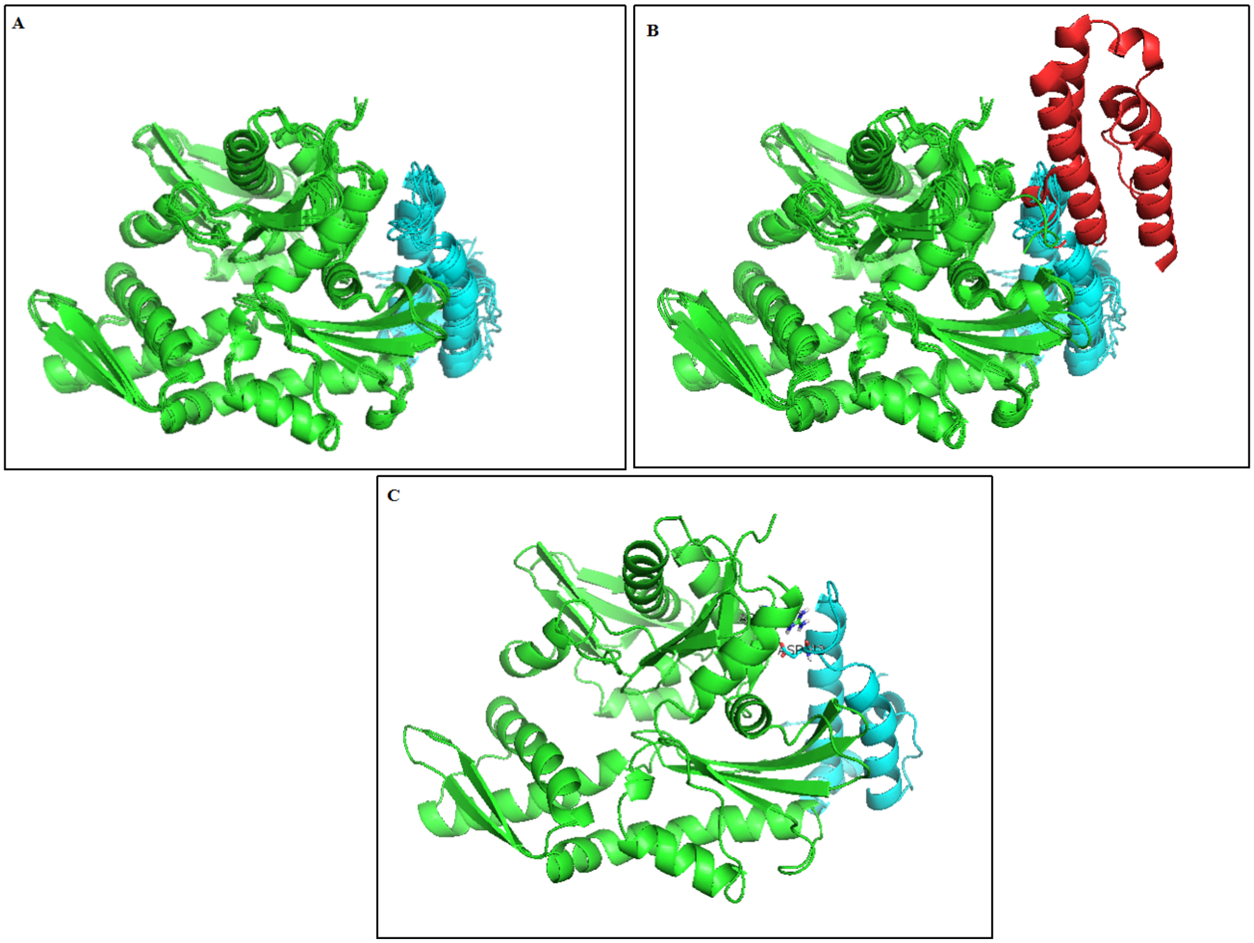


Figure 12: HSPA5-DNAJC10 complex. Structures are represented in cartoon. ARG 171 and ASP 42; predicted to be involved in Hsp70-Hsp40 interactions; are shown as sticks and labeled accordingly. (A) The four best predicted complexes by HADDOCK were superposed. ATPase domain_linker colored green and J domain is colored cyan. (B) Superposed structures of 2QWO and complex structures from (A). The ATPase-linker domain of the structures is colored in green. While the J domain in 2QWO is colored in red, that of HSPA5-DNAJC10 complex is colored cyan. (C) Structure of the best selected complex model with the least energy. Pictures were rendered using PyMol (Delano, 2002).

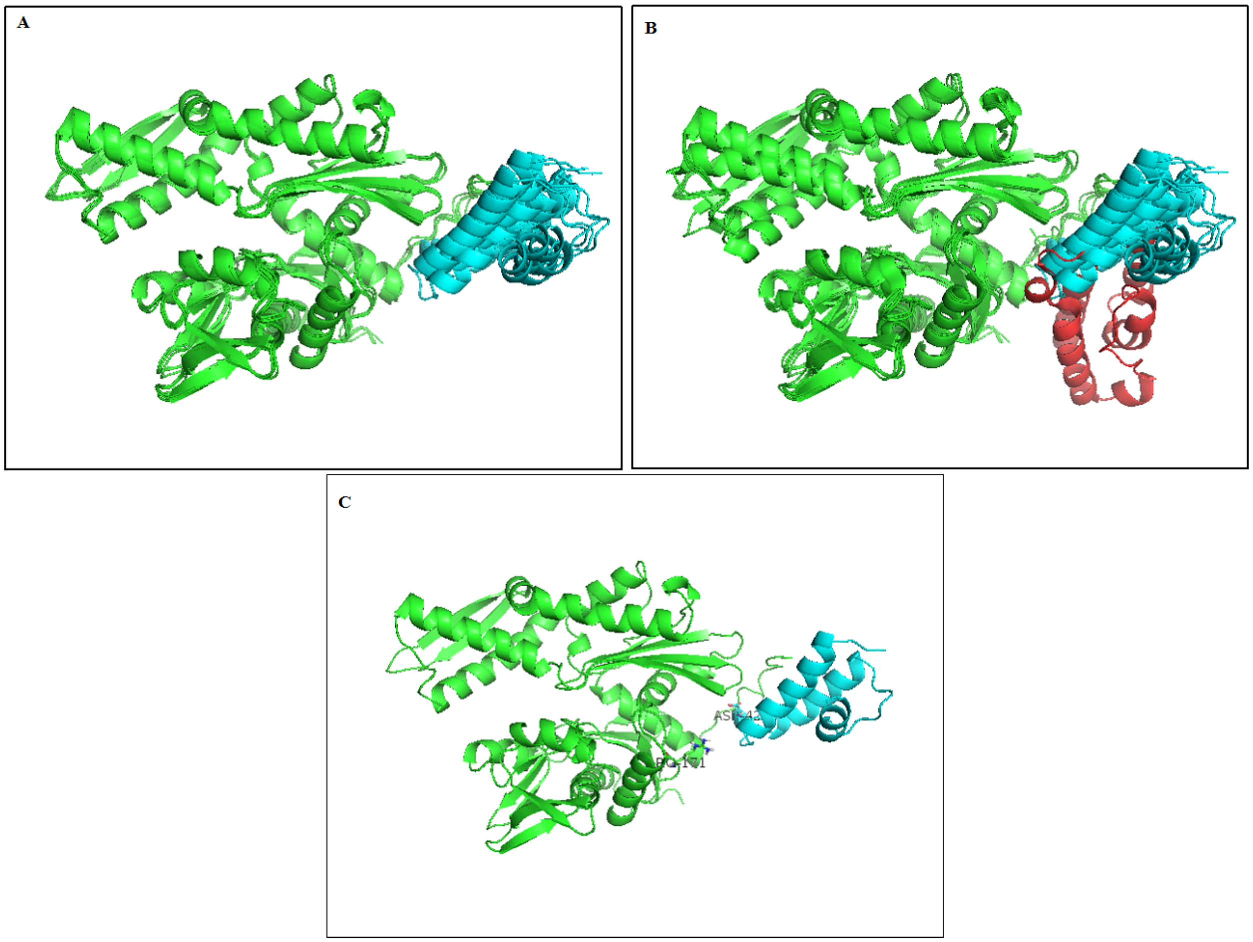


Figure 13: HSPA8-DNAJC19 complex. Structures are represented in cartoon. ARG 171 and ASP 42; predicted to be involved in Hsp70-Hsp40 interactions; are shown as sticks and labeled accordingly. (A) The four best predicted complexes by HADDOCK were superposed. ATPase domain_linker colored green and J domain is colored cyan. (B) Superposed structures of 2QWO and complex structures from (A). The ATPase-linker domain of the structures is colored in green. While the J domain in 2QWO is colored in red, that of HSPA8-DNAJC19 complex is colored cyan. (C) Structure of the best selected complex model with the least energy. Pictures were rendered using PyMol (Delano, 2002).

Appendix II: Clustering and energy scores evaluations for predicted model complexes in HADDOCK

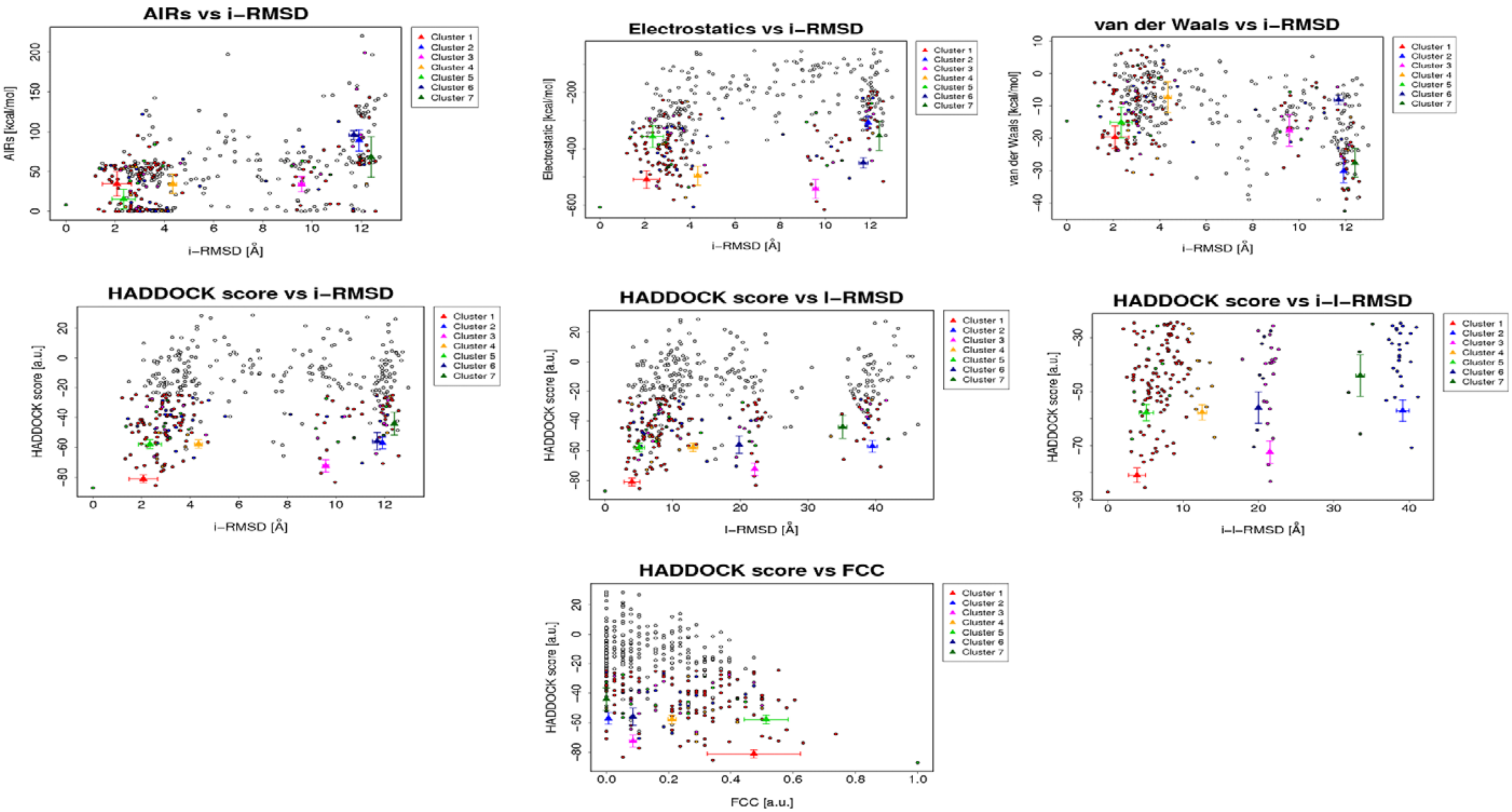


Figure 14: Clustering and energy scores evaluations for HADDOCK complex prediction of HSPA8-DNAJC19. AIRs = Ambiguity interaction Restraints, i-RMSD = interface-root mean square deviation, l-RMSD = ligand-root mean square deviation, i-l-RMSD = interface-ligand-root mean square deviation and FCC = Fraction of Common Contact

Appendix III: Experimental structures of 2WO, 2QWP, 2QWQ and 2QWR

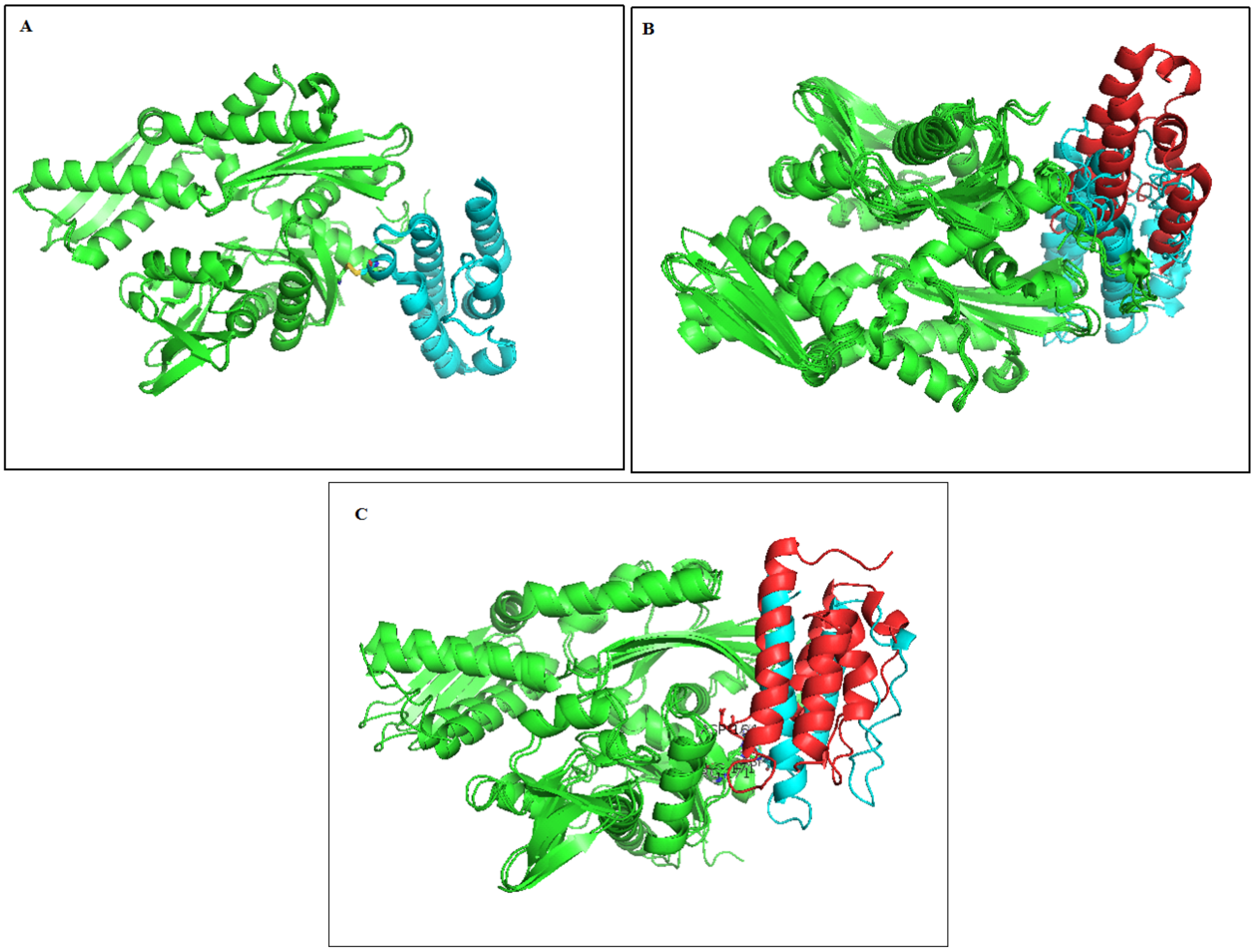


Figure 15: Complex structures of HSP70 ATPase domain_linker and HSP40 J domain. Structures are displayed as cartoon with the ATPase domain_linker in green and J domain in cyan. (A) Superposition of the four experimental crystal complexes (2QWO, 2QWP, 2QWQ, 2QWR) by (Jiang *et al.*, 2007). (B) The four crystal structures from (A) were superposed with the predicted docked complex of HSPA8-DNAJC6 using HADDOCK server. The ATPase domain_linker in all the complexes is colored green while the J domain of the experimental crystal structures where colored cyan and that of the predicted docked complex (HSPA8-DNAJC6) colored red. (C) Superposed complexes of 2QWL and HSPA8-DNAJC6. The ATPase domain_linker are colored green and the J domain in 2QWL was colored cyan and that of HSPA8-DNAJC6 colored red. The structures were rendered using PyMol (Delano, 2002).