

**Structure Modeling of the
Poly(A) Binding Protein C-terminus family**

by Stéphane Coillet-Matillon

Department of Biochemistry
McGill University
Montreal, Quebec, Canada

May 2002

A thesis submitted to the Faculty of Graduate Studies and Research in fulfillment of the
requirements for the degree of Master of Science

©Stéphane Coillet-Matillon, 2002



National Library
of Canada

Bibliothèque nationale
du Canada

Acquisitions and
Bibliographic Services

Acquisitions et
services bibliographiques

395 Wellington Street
Ottawa ON K1A 0N4
Canada

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

ISBN: 0-612-85776-X

Our file Notre référence

ISBN: 0-612-85776-X

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

Canada

Abstract

The solution structure of the C-terminal domain of the yeast Poly(A) Binding Protein (yPABC) was determined by NMR spectroscopy. The tridimensional fold is similar to that previously observed for the human isoform, except for a missing N-terminal alpha helix out of the five helices expected in total. A series of putative binding partners was identified in public databases, based on a consensual interaction motif that we formally defined as being S-X-L-[NS]-X-[ND]-A-X-E-F-X-P. However, no putative binding partner could be found that would match this pattern in *S.Cerevisiae*. Phylogenetic modeling tends to indicate that such an interacting domain might not be a fundamental part, if existing at all, of PABC function.

A tridimensional model of each of the remaining twenty PABC orthologs was generated: the overall fold of five consecutive helices seen in human PABC is well conserved, in contradiction with recently published modeling studies based on the crystal structure of the homologous PABC domain from HYD, which predicts only four helices.

Résumé

La structure en solution du domaine C-terminal de la *Poly(A) Binding Protein* de levure (yPABC) a été déterminée par spectroscopie RMN. La structure tridimensionnelle est proche de celle déjà observée pour la version humaine, à la différence près qu'une hélice N-terminal sur les cinq attendues est manquante. Une série de partenaires d'interaction possibles a été identifiée sur les bases de données publiques en se basant sur le motif consensuel d'interactions dont la nature a été formellement définie comme de type S-X-L-[NS]-X-[ND]-A-X-E-F-X-P. Cependant, aucun partenaire potentiel n'a été trouvé pour yPABC. Des études phylogénétiques tendent à montrer que chez les levures, l'essentialité voire l'existence d'un tel motif est peu probable.

Un modèle tridimensionnel a été généré pour chacun des vingt autres orthologues de PABC connus à ce jour: la structure générale du domaine en cinq hélices consécutives est bien conservée, en contradiction avec des résultats de modélisation récemment publiés et basés sur la structure cristalline d'une protéine homologue, HYD, qui ne leur en accorde que quatre.

Table of Contents

Abstract	2
Résumé	3
Table of Contents	4
List of Figures and Tables	6
Abbreviations	8
Acknowledgements	9
Chapter I: Introduction	10
1.1 Introduction to translation	11
1.1.1 Translation in eukaryotes	11
1.1.2 Initiation step	13
1.1.2.1 <i>The poly(A) tail, 40S subunit recruitment and mRNA stability</i>	16
1.1.2.2 <i>Associated factors</i>	17
1.1.2.3 <i>tRNA binding to 40S, 40S binding to mRNA, scanning and 60S binding</i>	20
1.1.3 Elongation step	21
1.1.4 Translation termination and ribosome recycling	23
1.2 The Poly(A) Binding Protein	27
1.2.1 Generalities	27
1.2.2 C-terminal domain	28
1.3 Objectives	33
Chapter II: Materials and Methods	34
Chapter III: Results and Discussion	41

3.1 Solution structure of the yeast Poly(A) Binding Protein (yPABC)	42
3.1.1 Purification of yPABC	42
3.1.2 Tridimensional structure of yPABC	43
3.2 PABP Interacting Motif 2 (PAM2)	46
3.3 Phylogenetic analysis of the PABP family	52
3.4 Comparative modeling of PABC orthologs_	59
3.4.1 Foreword	59
3.4.2 Comparative modeling	60
3.4.3 Comparative modeling of PABC orthologs	61
Chapter IV: Conclusion	68
4.1 Summary of results	69
4.2 Concluding remarks - Future directions	71
Appendix	73
References	74

List of Figures and Tables

Chapter I		10
Figure 1	Mechanism for translation initiation	15
Table 1	Initiation factors from mammalian, plants and yeast cells	19
Figure 2	NMR structure of human PABC	30
Chapter III		41
Figure 3	12.5% SDS polyacrylamide gel of PABC purification	42
Figure 4	(A) Ribbon drawing of the yPABC domain (B) Closeup view of the hydrophobic core of yPABC	44
Table 2	Φ -BLAST search and sequence alignment of proteins exhibiting the PABP interacting domain 2	48
Figure 5	BLAST search of <i>P.pinus</i> RF3 versus a database of fungal sequences	49
Table 3	Pileup, alignment and conserved residues of homologous PABC	52
Figure 6	(A) Bootstrapped NJ-tree from known PABP C-termini (B) Tree based on the whole sequence of the protein	53
Table 4	Sequence identity between selected PABPs	57
Figure 7	Average model accuracy	60
Figure 8	Homology models of 20 PABC orthologs	63
Figure 9	Closeup view of the conserved salt bridge between Lys38 and Glu45	65

Figure 10 *Left:* Ramachandran plot for *L.major* PABC
 Right: Ramachandran plot for *T.aestivum* PABC

66

List of abbreviations:

5' UnTranslated Region	5' UTR
Cytoplasmic Polyadenylation Factor	CPE
Elongation Factor	EF
eukaryotic Initiation Factor	eIF
Internal Ribosome Entry Site	IRES
Methionine-tRNA	Met-tRNA
Poly(A) Binding Protein	PABP
Poly(A) Binding protein C-terminal domain	PABC
PABP interacting Motif 2	PAM2
PABP Interacting Protein	PAIP
Release Factor	RF
eukaryotic Release Factor	eRF
Ribosome Release Factor	RRF
RNA Recognition Motif	RRM
Root Mean Square Deviation	RMSD

Acknowledgments

I want to thank my supervisor Kalle Gehring for his knowledge, patience and support. I also want to acknowledge those at BRI who made this experience a very fruitful and memorable one indeed, even if it implies at some point that I'll hate figure skating until my very last breath. I'll be more knowledgeable not only because of what I learned on the bench or in front of my computer, but also because of all their input on all sorts of topics, from lab protocols to the intricacies of English grammar, to the detailed history of the Montreal Canadiens. Special thanks go to Alexei Denisov, Irena Ekiel, Greg Finak, Pablo Gutiérrez, Herve Hogue, Guennadi Kozlov, Nadeem Siddiqui, Tara Sprules and Jean-François Trempe.

And of course I want to dedicate this work to my parents, who had to endure (if not support) my roaming the vast world in the past few years.

Chapter I: Introduction

The only good is knowledge and the only evil is ignorance.

-Socrates

1.1 Introduction to translation:

Elucidation of the detailed molecular mechanisms of protein synthesis is essential for understanding translational control. Translating the ribonucleic code into functional proteins is a fundamental process that is subject to many regulatory interactions. Three distinct steps are used to describe the translational process: initiation, elongation and termination. Considerable advances have been made in the past few years in the refinement of our knowledge of the assembly pathways and tridimensional structures of key components of the translation machinery (Wimberly et al., 2000).

The process of translation initiation has been extensively studied since the late 1960s, primarily through biochemical, *in vitro* studies that relied on radiolabeled amino acids and fractionated lysates (Brawerman, 1969) and led to the identification of most of the two hundred or so macromolecular components of the translation apparatus that are now known.

1.1.1 Translation in eukaryotes

Translation in eukaryotic cells is both very close and very different from prokaryote translation. Similarities range from the dissociation of the ribosome into ribosomal subunits, the assembly of a preinitiation complex on the small ribosomal subunit, the common use of the AUG initiation codon, the use of an initiator tRNA^{Met}. However, eukaryotic initiation relies much more on protein-protein and RNA-protein interactions, whereas prokaryotes are more dependent on

RNA-RNA interactions. In eukaryotes, protein translation is spatially and temporally separate from transcription, while these events are linked in prokaryotes. In eukaryotes, additional steps are present as messenger RNAs first synthesized in the nucleus must be spliced, capped, polyadenylated and exported to the cytoplasm through the nuclear pores before they are translated into proteins.

Another major difference lies in the recognition of the initiation codon. Rather than looking for a Shine-Dalgarno sequence and binding through mRNA-rRNA interactions, a scanning model (Kozak and Shatkin, 1978) has been adopted in eukaryotes. A fundamental element is the presence of a m⁷G-cap at the 5' end of the messenger RNA, which acts as an assembly point for the ribosome with the binding of the small 40S subunit. Although this 5' cap has been shown to be non-essential (Gunnery and Mathews, 1995), the greater its accessibility, the higher the efficiency of translation. The 40S subunit then scans downstream until a translation initiation codon is identified, preventing in theory any sort of polycistronic activity (since an mRNA has only one 5' terminus). An alternative, albeit less common initiation mechanism has been identified in eukaryotes, often involving viral RNA, and involves the direct binding of the 40S subunit to an Internal Ribosome Entry Site (IRES) at or just upstream of the initiation codon (Pyronnet et al., 2000; Sachs, 2000). Another difference in mechanisms between pro- and eukaryotic systems is the extra complexity due to the regulation in eukaryotes by numerous phosphoprotein initiation factors (reviewed in (Browning, 1996).

1.1.2 Initiation step

Translation initiation (Fig.1) is a key point in regulating mRNA translation, and is usually divided into three further different steps: (i) How the initiation codon is recognized; (ii) How the initiation factors interact with the initiator tRNA, mRNA and small ribosomal subunit; (iii) How the largest ribosomal unit joins to form a complex capable of elongation.

The nature of the initiation codon is nearly undisputed, being predominantly AUG in mammalian cells (and on rare occasions, GUG and UUG) and exclusively AUG in yeast. Its proximity to the 5' untranslated region (UTR) as well as the sequence surrounding it (at least in mammalian cells) play a dominant role in its selection. Scanning ribosomes that encounter a start codon with a unfavourable environment may pass over it and keep going until they find a more suitable one, which can be interpreted as a mean of regulating a more downstream AUG. This phenomenon, dubbed "leaky scanning", can lead to two proteins being translated, either with one being shorter than the other or, if the reading frames are shifted, with totally different sequences (Kozak, 1991). However, when two AUG codons are placed one after the other with the same favourable environment, it is the first codon that is strictly considered for initiation, indicating that the 40S subunit goes along the mRNA in a linear manner (Kozak, 1995). Non-linear scanning has been postulated to occur in very particular cases, such as heat shocked cells or cell lines infected by adenoviruses (Yueh and Schneider, 1996).

The length and above all secondary structure of the 5' UTR is therefore essential to translational efficiency: shortening and lengthening of this 5' UTR lead to a decrease and an increase of translational efficiency, respectively (Kozak, 1991). Experiments also support the

hypothesis that 40S ribosomal subunit enter at the 5' end of the mRNA, most notably since eukaryotic ribosomes are unable to bind to circular forms of synthetic mRNAs (Kozak, 1979) and that ribosome binding is inhibited when a base-paired structure or a repressor protein blocks the 5' end of the mRNA (Paraskeva et al., 1999).

The methylated cap structure m⁷GpppN (where m is a methyl group, p stands for phosphate and N for any nucleotide type) appended to the 5' end of all cellular mRNA is required for efficient translation but is not absolutely required, as Ohlmann and colleagues showed (Ohlmann et al., 1995). However, ribosomes remain dependent on the presence of a 5' structure when it comes to binding the mRNA (Kozak, 1979). Thus, the methylated cap holds some important functional properties: it affects RNA splicing, transport, stabilization and translation (Horikami et al., 1984; Lo et al., 1998). Its main purpose in the initiation process consists of recruiting a ternary complex eIF4F. Within it, and after eIF4A has unwound any secondary structures in the 5'UTR, the small ribosomal subunit (40S) is directed onto the RNA, mainly through interactions with eIF4G.

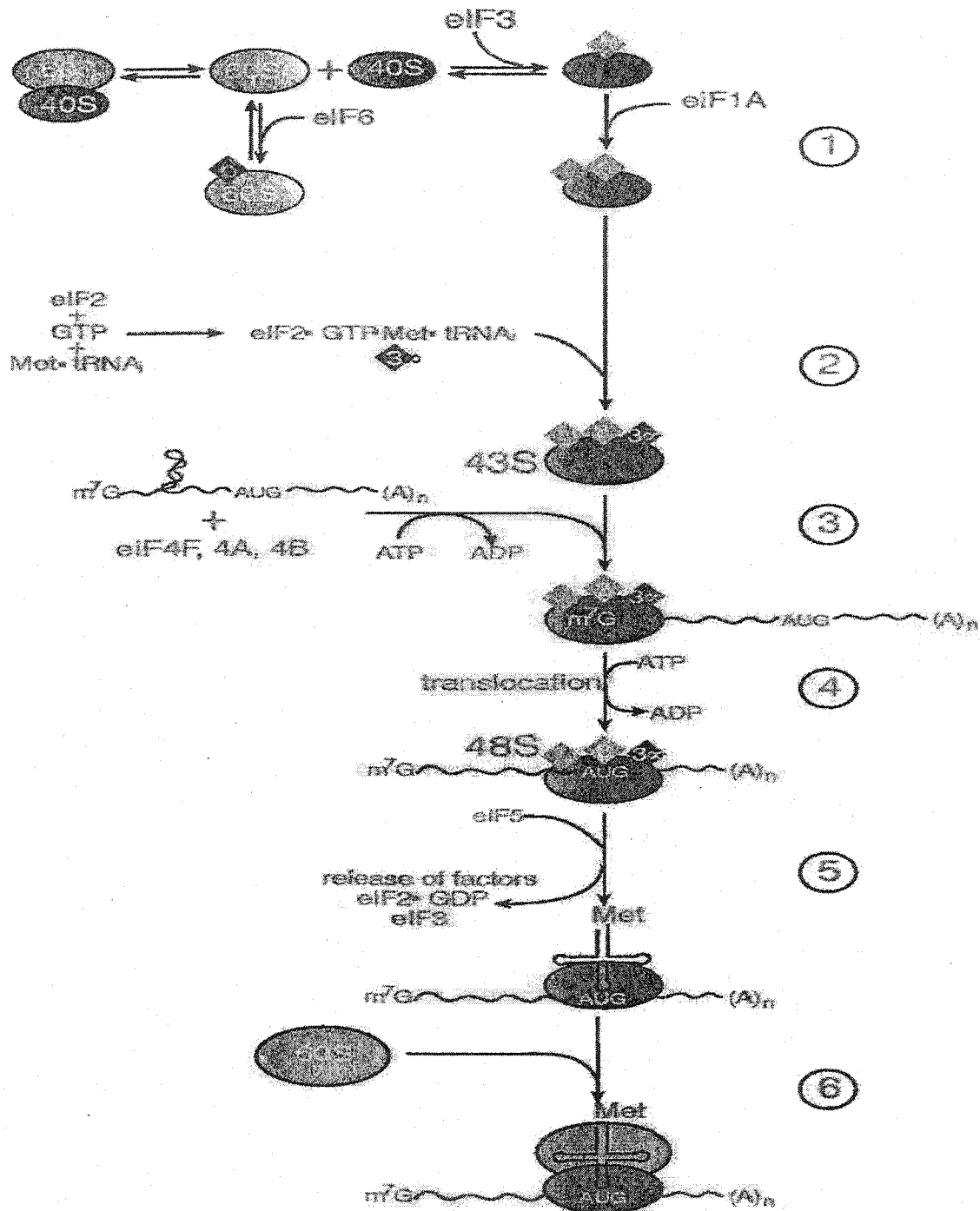


Figure 1: Mechanism for translation initiation. See text for details. Taken from Gingras *et al* (Gingras et al., 1999).

1.1.2.1 The poly(A) tail, 40S subunit recruitment and mRNA stability

To bring things into perspective, a brief review of mRNA polyadenylation might be useful. The majority of eukaryotic mRNAs possess a 3'poly(A) tail, which ranges from 50 to 250 bases in yeast and higher eukaryotes, respectively (Jacobson and Peltz, 1996), and whose main purpose is to help circularize the mRNA in the ribosomal machinery: extensive work on this has been performed by Steitz and coworkers (Moore and Steitz, 2002), allowing us to get a clearer structural idea of what is happening. The length of the tail is not encoded in genes, but is rather added to nascent mRNAs shortly after transcription, through a process involving site-specific cleavage and subsequent poly(A) polymerase-directed addition of nucleotides (Wahle and Keller, 1996). Polyadenylation is regulated by the presence of an AAUAAA sequence as well as a Cytoplasmic Polyadenylation Element (CPE), which interact with a variety of elements, six of which are known to date. Three are considered "core" factors : the poly(A) polymerase, the cleavage and polyadenylation specificity factor and the CPE-binding protein. A more complete description of the 3'-end processing can be found in Wahle (Wahle and Ruegsegger, 1999). Finally, the rate at which the tail is removed correlates well with mRNA decay rates.

In the late 1980s and early 90s, the poly-A tail was shown to be capable of stimulating initiation in amphibian oocytes, and that in yeast the Poly(A) tail Binding Protein (PAB1P) is a 70kDa highly conserved protein with four RNA recognition motifs (RRM). Multiple copies of

PAB1P coat the poly(A) tail, and are required for efficient mRNA translation (Tarun and Sachs, 1996; Tarun and Sachs, 1995). Moreover, a definite synergism exists between the cap structure and poly(A) tail. The poly(A) tail is involved in the binding of the 40S small ribosomal subunit to the mRNA and PAB1P deficiency results in alterations of the assembly and structure of the 60S subunit. Thus, both parts of the ribosome are targeted by the 3' end of the messenger RNA, although through separate processes (Tarun and Sachs, 1995). The influence and role of PAB1P is discussed below.

1.1.2.2 Associated factors

The 40S and 60S ribosomal subunits come from the dissociation of the 80S ribosome, in a process thought to be activated by eukaryotic Initiation Factors eIF3 and eIF1A, which bind to 40S and prevent it from further association with 60S through steric effects. Allosteric changes in the small subunit, upon binding of eIF3, are an alternate explanation (Srivastava et al., 1992). eIF6 has also been shown to bind to the large subunit and prevent its association with the 40S subunit, but its ribosomal status has recently been questioned, as it appears it could have more to do with 60S biogenesis and stability (Si and Maitra, 1999). Other factors of interest are listed in Table 1. eIF1A is a small, stable and essential protein, one of the most conserved of all initiation factors: the mammalian and yeast isoforms can substitute for each other (Wei et al., 1995). Interacting with eIF5B, both the mammalian and yeast proteins may occupy the tRNA binding A site on the 40S ribosome. eIF3 is by far the largest mammalian factor, and contains at least eleven subunits. A

similar complex also exists in plants, and a smaller one in yeast, but no complete structure has been determined so far. Most of the other initiation factors seem to be interacting with eIF3 at some point (Bandyopadhyay and Maitra, 1999; Fletcher et al., 1999). eIF4 is a complex involved in helicase activity (eIF4A), recognition of the 5' cap structure (eIF4E), and bridging of the mRNA and ribosome as well as mRNA circularization through interactions with the Poly(A) Binding Protein (eIF4G) (Gingras et al., 1999).

Table 1: Initiation factors from mammalian, plant and yeast cells^a

Name	Mammals ^b		Plants ^b		Yeast ^b		%ID ^c	Function
	mass (kD)	acc.#	mass (kD)	acc.#	mass (kD)	acc.#		
eIF1	12.6	L26247	12.6	AC005287	12.3	M77514	58	AUG recognition
eIF1A	16.5	L18960	16.6	AC006951	17.4	U11585	65	Met-tRNA binding to 40S subunit
eIF2 α	36.2	J02646	41.6	AF085279	34.7	M25552	58	Affects eIF2B binding by phosphorylation
eIF2 β	39.0	M29536	26.6	AL162351	31.6	M21813	42	Binds to EIF2B, EIF5
eIF2 γ	51.8	L19161	50.9	AC002411	57.9	L04268	71	Binds GTP, Met-tRNA; GTPase
eIF2B α	33.7	U05821	39.8	AC016529	34.0	M23356	42	Nonessential; helps recognize P-eIF2
eIF2B β	39.0	U31880	43.6	AC012395	42.6	L07116	36	Binds GTP, helps recognize P-eIF2
eIF2B γ	50.4	U38253			65.7	X07846		GEF activity
eIF2B δ	57.8	Z48225	29.4	AC012395	70.9	X15658	36	Binds ATP, helps recognize P-eIF2
eIF2Be	80.2	U19511	81.9	AC004238	81.2	L07115	30	GEF activity
eIF3			Multi-subunits (11) complex - see (Hershey et al., 2000) for details					
eIF4AI	44.4	X03039.1	46.7	AB019229	45.1	X12813	65	ATPase, RNA helicase
eIF4AII	46.3	X12507.1	46.8	AC005287	44.6	X12814		ATPase. RNA helicase
eIF4B	69.3	S12566	57.6	AF021805	48.5	X71996	26	Binds RNA; stimulates helicase
eIF4E	25.1	M15353	26.5	AL110123	24.3	M21620	33	Binds m ⁷ G-cap of mRNA
eIF4GI	171.6	AF104913	153.2		107.1	L16923	22	Binds eIF4E, 4A, 3, PABP-RNA
eIF4GII	176.5	AF012072	176.5		103.9	L16924	21	Binds eIF4E, 4A, 3, PABP-RNA
eIF5	48.9	L11651	48.6	AC007576	45.2	L10840	39	Stimulates eIF2 GTPase
eIF5B	139.0	AF078035			97.0	L29389	70	GTPase; promotes junction reaction

^a Taken from (Hershey et al., 2000).^b The masses and accession numbers are from human (or rat), *Arabidopsis thaliana*, and *Saccharomyces cerevisiae* proteins. Only one of the numerous isoforms from the plant proteins was arbitrarily chosen and indicated. A complete listings can be found in (Browning et al., 1996).^c %ID, percent sequence identity shared by yeast and mammalian proteins.

1.1.2.3 tRNA binding to 40S, 40S binding to mRNA, scanning and 60S binding

Protein synthesis is initiated through the binding of a methionine-tRNA to eIF2-GTP (GDP inhibits complex formation). One feature of eIF2 is the ability to differentiate this initiator tRNA from the elongator tRNAs, by recognition of the methionyl residue. This ternary complex binds the 40S small ribosomal subunit to form a preinitiation (or 43S) complex; it is interesting to note that met-tRNA binding occurs prior to mRNA and/or 60S binding (Chaudhuri et al., 1997; Smith and Henshaw, 1975). eIF2 leaves the ribosome after GTP hydrolysis. The m⁷G-cap is recognized by eIF4E as mentioned earlier, thus committing the ribosomal complex to mRNA binding. eIF4F, through its eIF4A helicase, also commits itself to unwinding any secondary structure in the bound messenger RNA. eIF4F is then free to interact with eIF3 (bound to the 40S subunit), bringing the mRNA and ribosome in contact through a protein bridge (Hershey and Merrick, 2000). The ribosome then scans down the length of the mRNA until it finds a start codon in a favourable context. The questions of how the ribosome travels, at what rate, and if it is a strictly energy-dependent process, have not yet been solved. Evidence supporting the scanning mechanism comes from a series of experimental results: (i) if an AUG codon is inserted between the 5' cap and the native initiator codon, the former is used as a start site, and (ii) RNA secondary structure or chemical modification at the 5' end block 40S binding to the mRNA and subsequent localization of the start codon (Hershey and Merrick, 2000). Since the initiator Met-tRNA is already bound to the scanning 40S small ribosomal subunit, recognition of the AUG initiation codon occurs through tRNA-mRNA codon-anticodon

interactions. As explained above, context plays an important secondary role: leaky scanning, where the first AUG encountered is bypassed, can result from a poor environment. A more downstream codon, in a stronger context, will then be selected by the translational machinery. Also, initiation factors do play a particular role, since mutations in eIF1, eIF5 or eIF2 can lead to initiation at UUG codons (Donahue et al., 1988). Upon GTP hydrolysis by eIF5B and the release of all other initiation factors, the 60S large ribosomal subunit can bind to form the 80S ribosome (Pestova et al., 2000).

1.1.3 Elongation step

Protein biosynthesis has been well studied and the subject of many books (Hill et al., 1990; Nierhaus et al., 1993; Söll and RajBandhary, 1995). In the recent years, the prokaryotic elongation cycle has been favoured in biochemical studies. The structure of the major elongation factor EF-Tu, as well as that of ribosomal subunits (50S, 30S and even the full 70S ribosome) and many individual protein subunits have been elucidated (Ramakrishnan, 2002). These give us a good picture of what is happening during the process of protein synthesis (Cate et al., 1999; Krab and Parmeggiani, 1998; Wimberly et al., 1999). Fortunately, it is possible to extrapolate much about eukaryotic elongation from what is already known in prokaryotes. With the exception of Elongation Factor 3 (EF3) found in yeast (Kovalchuk et al., 1998), all eukaryotic factors are related to prokaryotic homologs: EF-1 α (eEF-1A), EF-1 $\beta\gamma$ (eEF-1B) and EF-2 correspond to the bacterial EF-Tu, EF-Ts and EF-G (Riis et al., 1990). Although EF-1A and EF-2 appear to be well conserved, the eukaryotic factors are about 30%

longer than their prokaryotic counterparts (Cavallius et al., 1993). Electron cryomicroscopy has also revealed the 25Å structure of a complete mammalian ribosome (Dube et al., 1998). Amino acids are sequentially added onto the growing peptide chain based on the order given by the template mRNA. During elongation, a ternary complex, comprising EF-1A, GTP and acylated tRNA is formed and enters the ribosome's A-site, binding the messenger RNA through codon-anticodon interactions. Upon recognition, GTP hydrolysis leads to EF-1A•GDP release. Peptidyl transferase immediately catalyzes the formation of a peptide bond between the amino-acids on the P and A sites. EF-1A is regulated both through phosphorylation that impairs its activity, and through regulation of its expression level. GTP hydrolysis is favored by the guanine nucleotide exchange factor EF-1B, which is thought to be complexed with EF-1A. Surprisingly, none of its constituents is in anyway phylogenetically related to EF-Ts: no sequence homology can be found, although they clearly fill the same function. Once the peptide bond between A and P residues is created, eEF-2 intervenes and shifts the peptidyl-tRNA out of the A site and towards the P site in a GTP-dependent manner, while the deacylated tRNA goes to an exit (E) site. By displacing the acylated tRNA on to the P site, eEF-2 also favors a shift of the mRNA by exactly three codons. EF-3, which is only found in yeast, is less well studied but seems to contribute to the release of the deacylated tRNA from the ribosomal E site, using ATP as a substrate.

1.1.4 Translation termination and ribosome recycling

Compared to translation initiation, very little is known about the termination of protein synthesis and the mechanisms that lead to the subsequent reinitiation through the recycling of the ribosomal machinery. Even though it is well accepted that in eukaryotes 61 of the 64 codons code for amino acids and that the three remaining codons (UAA, UAG and UGA) are nonsense or stop codons, this rule is not universal among all living organisms (Osawa et al., 1992): AUA (isoleucine) and UGA (stop) code for methionine and tryptophan, respectively, in human mitochondria (Barrell et al., 1979). However, the general rules of termination are unaffected by those local evolutionary changes and can be freely discussed.

The termination of protein synthesis takes place in response to the presence of an antisense codon in the A site. A heterodimer made up of two distinct types of polypeptide release factors (RF) is required to properly end elongation: it contains a class I, codon-specific factor (eRF1), and a non-specific class II factor (eRF3) that binds GDP/GTP and stimulates class I factors activity. In eukaryotes, eRF1 recognizes all three codons, but this function is devoluted to two distinct factors, RF1 and RF2, in bacteria (Nakamura et al., 2000). Recent evidence in prokaryotes shows that specific recognition of anticodons occurs through interaction with a specific tripeptide within RF1 and RF2 (Ito et al., 2000). Song *et al.* (Song et al., 2000) solved the crystal structure of human eRF1, and revealed that the overall shape and dimensions is similar to a tRNA, with domains 1, 2 and 3 of eRF1 corresponding respectively to the anticodon, aminoacyl acceptor and T- stems. Based on results

from prokaryotes and the high sequence conservation, it was postulated that a Thr-Ser-Ala tripeptide in domain 1 plays the role of the anticodon. eRF1 enters the A site upon identification of the stop signal, and through interaction of G¹⁸⁵ of the conserved GGQ motif (located in a region equivalent to the aminoacyl location in regular tRNA (Song et al., 2000)), triggers peptidyl-tRNA hydrolysis by the ribosome, releasing the nascent polypeptide chain (Drugeon et al., 1997). The universal conservation of the GGQ motif in prokaryotes and eukaryotes likely reflects a similar mechanism.

The rate of peptidyl release can be greatly enhanced by the binding of eRF1-GTP-eRF3 onto the A site; eRF3, a G-protein, triggers GTP hydrolysis to enhance the rate of peptidyl release. The eRF3 C-terminal domain shows a strong sequence similarity to elongation factor EF-1 α and prokaryotic elongation factor EF-G, implying their functions may be similar. To date, there is no archaeobacterial equivalent of RF3 known and eRF1 is still active *in vivo* when the eRF3-interacting region is deleted. eRF3 could act either in activating peptidyl release by eRF1, or perhaps by working on eRF1 ejection and recycling in a RF3 manner, or even in proofreading termination, since eRF1 might have problems discriminating against sense codons by having such a broad specificity as three antisense codons (Nakamura et al., 2000). Moreover, eRF3 has been shown to interact with the C-terminus of PAB1P (Hoshino et al., 1999; Kozlov et al., 2001). In yeast, the eRF3 gene was first identified as a nonsense suppressor (SUP35 and SUP2) (Hawthorne and Leupold, 1974). When overexpressed, yeast eRF3 leads to a non-mendellian inheritance which is due to protein aggregation [*psi*+] (Santoso et al., 2000). This phenomenon has been extensively studied as a yeast

model system for prions. More biochemical evidence will be needed before the breadth of functions and processes in which eRF3 is implicated will be elucidated.

While the ribosomes that reach a stop codon may undergo a conformational change to expose the A site to RF interaction, the mRNA sequence downstream of the antisense codon is also important for termination efficiency. The nucleotide context is not random, particularly for genes that are highly expressed. The identity of the base immediately following the stop codon (+1) is most strongly biased, but the bias can reach as far as position +6 (Namy et al., 2001). The stop codon should therefore be considered at least like a tetranucleotide, the nature of the fourth nucleotide determining the efficiency of termination: from most efficient to least efficient termination, it was found to be $G > U, A > C$ in UGA and UAA, while for the UAG codon it was $U, A > C > G$ (Bonetti et al., 1995). Strong termination is particularly important in genes that are highly expressed. On the other hand, viral proteins that require extended synthesis have a use for codons in a weak context, as it allows for expression of different proteins from the same RNA strand (Bertram et al., 2001). Upstream context is also important, although less so in eukaryotes than prokaryotes. A number of non-essential *trans*-acting factors, like for instance the *UPF* genes, can also have an influence on translation termination efficiency, and are reviewed in Czaplinski (Czaplinski et al., 1999).

The steps following peptide hydrolysis were investigated mostly in *E.coli* and could differ in eukaryotic organisms, particularly since bacterial transcripts are polycytronic, thus requiring a different mechanism for ribosome dissociation and recycling. Upon release of the nascent polypeptide chain, the ribosome is left in complex with deacylated tRNA (in the P site), eRF1/eRF3

(in the A site), and the mRNA strand. The cell still has to perform efficient recycling of the ribosome in order to complete new rounds of protein synthesis. The GTPase activity of eRF3 is thought to favor the dissociation of the eRF1/eRF3 complex from the A site. It has also been postulated that eRF3 could act as some sort of hybrid between RF3 and the Ribosome Releasing Factor (RRF), whose purpose it is to help dissociating the complex, in order to perform ribosome dislocation (Buckingham et al., 1997). *In vivo*, following termination and in the absence of RRF, prokaryote 70S ribosomes remain on the mRNA and scan until they find a new initiation codon anywhere between 17 and 45 nucleotides downstream (Janosi et al., 1998). Circularization of the mRNA strand during eukaryotic initiation might also play a role in recycling translation factors (Wells et al., 1998).

1.2 The Poly(A) Binding Protein

1.2.1 Generalities

Poly(A) Binding Proteins in the PABP1 family are approximately 70 KDa in mass and comprised of four highly conserved RNA recognition motifs followed by a long unstructured region and a C-terminal domain (PABC). The PABP multigene family is distributed on several loci in the human genome. Numerous studies have been accomplished since PABP's discovery (Blobel, 1973) and its purification (Baer and Kornberg, 1983). At least three functional isoforms of poly(A) binding protein exist in humans: PABP1 is ubiquitously expressed, most abundant and the best studied form. An inducible form, iPABP, which exhibits 77% identity with PABP1, has been found in T cells and platelets; a third isoform has been found and is testis-specific (Feral et al., 2001). An unrelated protein, PABP2, is found in the nucleus (Wahle, 1991) and mediates the recognition of a nascent poly(A) tail during elongation. The RRM motif is found in proteins of both bacterial and eukaryotic origin, indicating that the motif is an ancient and fundamental element of protein structure. Individual RRM motifs are composed of two tandem RNP motifs folded into a beta-sandwich of 4 beta sheets and 2 alpha helices.

PABP has been found in significant excess (about three fold) over cytoplasmic RNA (Gorlach et al., 1994), but it is also thought to participate in mRNP maturation in the nucleus

(Afonina et al., 1998). Multiple copies of the protein assemble on poly(A) tail via the RRM's to form a repetitive, higher order complex with an RNA footprint of 27 nucleotides per PABP molecule (Baer and Kornberg, 1983). All four RRM's exhibit distinct affinities and specificities for the polyadenylated tract: domains 1 and 2, whose 3D structure in complex with RNA was solved (Deo et al., 1999), have been shown to be equal to the whole protein in respect to RNA affinity; domains 3 and 4 have an approximately ten fold-reduced affinity for poly(A), and cooperate with the C-terminal region to form a complex on poly(A) with multiple copies of PABP (Kuhn and Pieler, 1996). The observation that RRM1 and 3 share homology at 11 of 13 RNA-binding residues and RRM2 and 4 share 8 of 9 residues is reconciled with this difference in affinity for poly(A) by noting that the linker between RRM3 and RRM4 is longer (9 vs 26 residues), thus making binding less efficient than that between RRM1 and RRM2 (Deo et al., 1999).

1.2.2 C-terminal domain

The C-terminal quarter of PABP (PABC) became a center of attention when it was demonstrated that the four RRM's alone were not sufficient for normal growth (Kessler and Sachs, 1998). This C-terminal region holds a strongly conserved domain of 68-72 residues, whose sequence identity ranges from 54% to 40 and 38% between mammalian, plant and yeast isoforms, respectively. One of the most divergent features in the various PABC sequences lies with a two amino acid insertion at position 34 and 35 of the PABC domain in yeast and *C.elegans*, although

the biological significance of this sequence divergence is not known. In yeast, PAB1 is an essential gene whose deletion leads to inhibition of translation initiation, poly(A) shortening and delay in the onset of mRNA decay (Amrani et al., 1997; Sachs, 1990; Sachs and Davis, 1989), but those effects can be suppressed by mutations that alter the 60S subunit of the ribosome, as well as those that inhibit mRNA decay (Hatfield et al., 1996; Sachs and Davis, 1989; Wyers et al., 2000). Deletion of the C-terminal motif in yeast containing a PABP-MS2 fusion protein instead of wild-type PABP is lethal (Coller et al., 1998).

In addition to poly(A) binding proteins, PABC domains also occur in a subclass of ubiquitin E3 protein ligases that contain a HECT (Homologous to E6-AP C-Terminus) domain. The function of the PABC domain in these ubiquitin ligases is unknown. The crystal and solution structure of a PABC domains in a ubiquitin ligase homolog and human PABP1 were recently solved (Deo et al., 2001; Kozlov et al., 2001). The PABC domain from PABP1 is a well defined, alpha-helical globular domain made up of five helices, the global shape similar to that of an arrowhead (Fig.2). Within the arrow is a well-defined hydrophobic core, part of which serves for peptide recognition through stacking interactions.

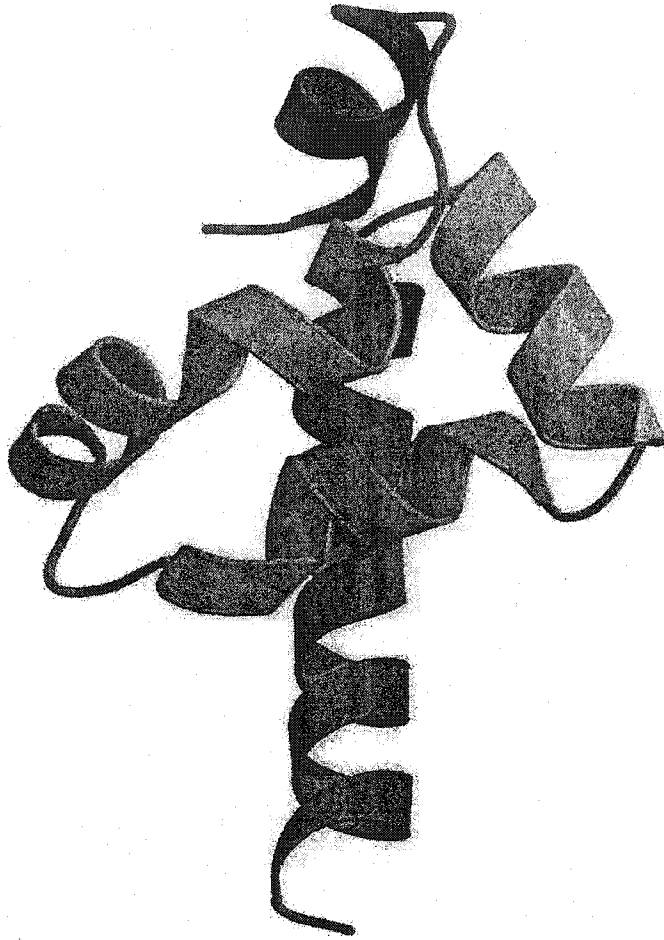


Figure 2: NMR structure of human PABC. The arrow-like shape is clearly visible, with helix 1 (blue) as the tip of the arrow, and helix 5 (red) as its shaft. Taken from (Kozlov et al., 2001)

There is a number of proteins known to interact with PABC domains. Among these are Pbp1p (Mangus et al., 1998), which promotes proper polyadenylation, initiation factor eIF4G (Tarun et al., 1997), polypeptide release factor eRF3/GSPT (Hoshino et al., 1999), Paip1, a mammalian protein that shows some areas of homology to eIF4G, and Paip2 (both of which are unique to metazoans), cleavage factor IA (Minvielle-Sebastia et al., 1997), α CP mRNA stability protein (Wang and Kiledjian, 2000), as well as rotavirus nonstructural protein 3 (NSP3) (Deo et al., 2002) and potyvirus RNA-dependent RNA polymerase (Wang et al., 2000). The range of functions of these proteins is pretty wide: from premessenger RNA 3' formation and polyadenylation (CFIA and Pbp1p), to endoribonuclease regulation and mRNA stability (α CP and eRF3), mRNA circularization and translational activation (eIF4G and Paip1) or repression (Paip2). Both Paip1 and Paip2 interact with PABP1, but Paip2 competes with Paip1 for binding to PABP (Khaleghpour et al., 2001). Paip1 has been shown to contain two distinct binding sites for PABP: PAM2 (for PABP interacting Motif 2) is a large C-terminal acidic amino acid rich region that interacts with RRM1 and RRM2 from PABP, while PAM1 is a 12-15 amino acid stretch residing in the N-terminus portion of Paip1 and that is specific to PABP's C-terminal domain (Roy et al., 2002). Interaction of PABC with Paip1 leads to a 40-fold increase in eIF4F cap-binding, showing a role for PABP in mRNA circularization (Luo and Goss, 2001). PABP thus contributes to mRNA stability in a poly(A) independent manner (Coller et al., 1998), probably through its interaction with eIF4G-eIF4F that contributes to "protect" the 5'-cap from

degradation. Besides, PABP also has its own task of delaying mRNA decay, by inhibiting indirectly the assembly of a 3'-to-5' exonuclease (Ford et al., 1997).

1.3 Objectives:

The Poly(A) Binding Protein is an important part of the translational machinery and mRNA biogenesis in eukaryotes; its C-terminal domain only now starts to be the focus of attention, mostly through the role it plays in mRNA circularization. However, the mechanism(s) by which proteins interact with PABC is or are not yet fully understood. The tridimensional solution structure of yeast PABC was determined by NMR spectroscopy and compared to that of human PABC. To further characterize the interactions at the center of which PABC is found, we went searching for all the putative proteins that held a twelve residue-long PABP interactive motif, and undertook a detailed phylogenetic study of all PABC orthologs in order to understand and highlight differences that appeared between lower and higher eukaryotes, most notably yeasts. Finally, comparative protein structure modeling was used to generate high-quality model structures for twenty PABP C-terminal domain orthologs. The general fold of the modeled proteins is comparable to that observed in human PABC, in contrast to previously published modeling studies based on human HYD, a PABC ortholog.

Chapter II: Materials and Methods

Perl: the swiss-army chainsaw of UNIX tools.

-Henry Spencer

Buffers used in protein purification

- *PBS buffer*: 50 mM Na_2HPO_4 (EM Science, Gibbstown, NJ), 150 mM NaCl (EM Science), pH 7.25
- *Lysate buffer*: PBS buffer 1x, 1 mM Phenylmethylsulfonyl fluoride (Sigma Co., St Louis, MO), 10 mg/ml Lysozyme (ICN Biomedicals, Aurora, OH), 1.4 mM β -mercaptoethanol (EM Science) pH 7.0
- *PreScission protease buffer*: 50 mM Tris (GibcoBRL, Grand Island, NY), 150 mM NaCl, 1 mM disodium ethylenediamine tetraacetate (Fisher Biotech, Fair Lawn, NJ), 1 mM Dithiothreitol (BioVectra, Italy), pH 7.0
- *NMR buffer*: 50 mM Tris, 300 mM NaCl, pH 7.0
- *M9 (minimal) media for isotopically labelled samples*: 6 g of Na_2HPO_4 , 3 g of KH_2PO_4 , 0.5 g of NaCl, 0.5 g of $^{15}\text{N-NH}_4\text{Cl}$ (Isotec, Miamisburg, OH) and 2 g of $^{13}\text{C}_6$ -glucose (Cambridge Isotope Laboratory, Andover, MA) for one liter of media.

PAB1P construct

The yeast Poly(A) Binding Protein C-terminal domain, residues 491 to 577 (gi417441) was subcloned from *S. cerevisiae* genomic DNA (kind gift from Dr. Malcom Whiteway), PCR amplified by Nadeem Siddiqui, cloned into vector pGEX-6P-1 (Amersham Pharmacia Biotech)

using BamHI and EcoRI restriction sites, and the PAB1P construct was expressed in *E. coli* BL21 Gold (DE3) (Stratagene).

Host expression and sample preparation

An overnight culture was prepared from a single colony of transformed host containing the PAB1P construct. One liter of 1x Luria Broth (LB) or M9 Medium was supplemented with 100ug/ml of ampicillin (Fisher), and inoculated with 20 ml of overnight culture. The culture was vigorously shaken and grown at 37°C until $OD_{600} = \sim 0.8$. The growth temperature was reduced to room temperature and IPTG was added for a final concentration of 0.5 mM to induce protein expression. The culture was left for three hours to allow for overexpression of the fusion protein. The cells were harvested by centrifugation and stored at -20°C.

Cells were resuspended in 10 ml of lysate buffer for each liter of culture and sonicated repeatedly to ensure proper destruction of cell walls. Lysed cells were then centrifuged for 25 minutes at 4°C in a Beckman JA-17 rotor, at 21,500g. The supernatant was filtered with Whatman paper and 2 ml of glutathione Sepharose 4B (Amersham Pharmacia) were added. The mix was left for one hour at 4°C, with occasional gentle mixing. Cell filtrate was then removed, and the beads were washed with 30 ml of 1x PBS buffer, and then 20 ml of PreScission protease buffer. 50U (2.5U per mg of protein) PreScission protease (Amersham Pharmacia) were added and the column was stored overnight at 4°C, so that the N-terminal GST tag be cleaved and yield a 92 residue protein fragment, consisting of the 87 C-terminal residues of PAB1P plus a five

residue (Gly-Pro-Leu-Gly-Ser) N-terminal extension. The flowthrough containing the cleaved protein was collected and an additional 0.5 ml glutathione Sepharose was added to collect any remaining GST-tag. After 30 min at 4°C, the flowthrough was collected and protein concentration measured at 280 nm.

NMR experiments and structure calculations

To conduct NMR experiments, the sample was resuspended in NMR buffer. Backbone assignments and solution structure were completed by Guennadi Kozlov. Experiments and procedures were conducted as previously described in (Kozlov et al., 2001).

Sequence comparison of PABC domains and phylogenetics

PABC sequences were obtained from an iterative Ψ -BLASTP search (Altschul et al., 1997) of the NCBI non-redundant database with the yPABC sequence (gi417441, residues 490-563) as query. Iterations were run until no new sequence was reported, and to ensure that all possible homologs were found the initial default expected maximal E-value was increased to 1000, while the default starting inclusion threshold for defining the position-specific score matrix (PSSM) to the next iteration was increased by two orders of magnitude to 1 and decreased by a factor of 10 for each new iteration. Unique sequences from distinct orthologous proteins were then analyzed by ClustalX v1.81 (Thompson et al., 1997) to generate an alignment and bootstrapped (1,000 replicates) Neighbor Joining (NJ) tree in order to reflect the degree of

difference and branching between the various orthologs. Alignments were built using either the PABC domains (1-72) or the whole PABP protein, and unrooted trees were drawn using TreeView v1.6.5.

To simplify comparisons between different PABC domains, a numbering scheme was adopted which is anchored on the highly conserved KITGMLLE motif common to all PABC domains. The PABC domain was defined to begin 37 residues prior to the conserved K38 (or 39 residues prior in the case of *S. cerevisiae* and *C. elegans* PABC). The two additional residues that are unique to the *Saccharomyces* and *Caenorhabditis* proteins were numbered residue 33a and 33b.

Comparative structure modeling

The comparative modeling program Modeller v6.1 (Sali and Blundell, 1993) was used for structural modeling. Modeller computes 3D structure of proteins by satisfaction of spatial restraints based on an alignment of known related proteins. A 3D model is obtained by optimization of a molecular probability density function (pdf).

The amino acid sequences of all orthologous proteins were aligned against a combination of template sequences from human (PDB code 1G9L), yeast (PDB 1IFW) and/or HYD (PDB 1I2T) PABC or PABC-like experimentally determined structures, the two former coming from NMR data (Kozlov et al., 2001), and the latter from X-ray crystallography (Deo et al., 2001). Alignments were manually optimized to improve the quality of the automatically generated

alignments. In cases where identity with one of the templates was 90%, modeling was made only against this template, as additional templates of lower resemblance would actually lower the overall accuracy of the modeling. Twenty successive modelizations were run for each target protein on a Silicon Graphics Indigo² 10000 Impact, and the best model was picked as the one with the lowest (most optimized) pdf.

As a mean of comparison, an alternate model was generated using SWISS-MODEL v3.5¹, using the NMR structure of human PABC as a template. No manual optimization was available for this program and default parameters were used.

Evaluation of generated models

All generated models were submitted to DALI² (Holm and Sander, 1993) for comparison with hPABC, and PROCHECK (Laskowski et al., 1993) was used to check for the overall stereochemistry and generate the Ramachandran plot of Fig. 10.

¹ available on-line at <http://www.expasy.org/swissmod/SWISS-MODEL.html>

² <http://www2.ebi.ac.uk/dali/>

PABC binding consensus sequence

Based on the conserved motif (S-X-L-N-X-N-A-X-E-F-X-P) described in (Kozlov et al., 2001), a Φ -BLAST (Zhang et al., 1998) was run successively using poly(A) binding interacting proteins (PAIP) 1 and 2 as templates. A second set of iterative searches was then run based on proteins that were found in the during the first BLAST, and that exhibited a more loosely consensual binding motif. Iterations were run using the same parameters as described earlier (E-value of 10,000 to compensate for the shortness of the sequence) until no new sequences of interest could be identified.

Scripts

Scripts written to facilitate file editing and data mining were done in Perl v5.005_02.

Chapter III: Results and Discussion

It doesn't matter whether the cat is black or white, as long as it catches mice.

-Deng Xiaoping

3.1 Yeast Poly(A) Binding Protein (yPABC)

3.1.1 Purification of yPABC

As shown on Figure 3, the purification process of the yeast PABC from PAB1P (residues 491-577) was successful, leading to a final quantity of about 3 mg of protein per liter of LB medium and 2mg for M9 medium, as measured by absorbance in the 280 nm region. As expected, the purified protein's MW is in the 10kDa range (10434 Da theoretical weight). The C-terminal fragment from *Saccharomyces cerevisiae* PABC was also prepared as an isotopically labeled, recombinant protein fragment for complete structural studies by NMR spectroscopy at 500 and 750 MHz.

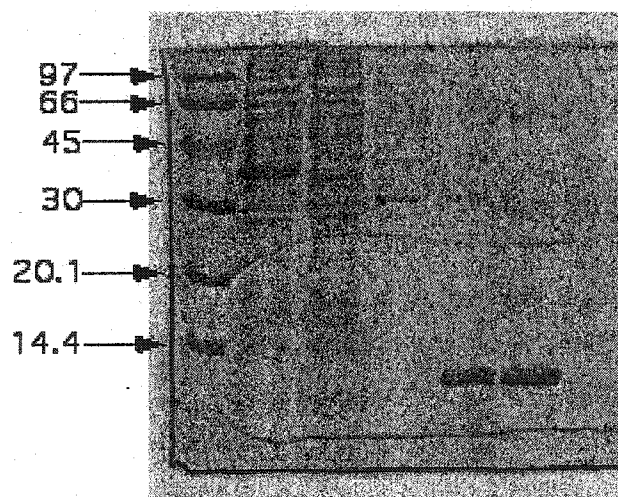


Figure 3: yPABC purification procedure. 12.5% SDS-PAGE. Lanes: 1) Molecular weight markers; 2) Lysed cells supernatant; 3) GST Column flowthrough; 4) PBS Buffer wash; 5) Eluted, concentrated cleaved protein.

3.1.2 Tridimensional structure of yPABC

The yPABC domain gave excellent quality spectra and its 3D structure could be determined (Fig 4 A) in our laboratory by Dr Kozlov (Kozlov et al., 2002). The final value of the backbone Root Mean Square Deviation (RMSD) for the 30 accepted structures (using residual dipolar couplings) is 0.34Å. The overall structure is very similar to that of the human C-terminus fragment. One notable difference is that when compared to the human isoform the first N-terminal helix is missing, which gives the protein a fold more similar to the crystal structure of HYD (Deo et al., 2001); this can be easily explained by the lack of sequence conservation in this N-terminal region. Yeast PABC also contains a two residue insertion (Asn30a and Glu30b) which partly extends the N-terminal side of helix 3.

The folded domain of yPABC includes approximately 65 residues (res. 502-567 of PABP1) that resembles a fairly compact cluster of four helices that are numbered from 2 to 5 (to facilitate comparison with the human domain), with a well-defined hydrophobic core consisting of Leu20, Gly21, Leu24, Val28, Ala36, Ile39, Ile43, Leu46, Val51, Phe52, Leu54, Leu55, Phe61, Tyr65, Ala68 and Tyr72 (Fig 4 B). yPABC differs from structures previously determined in the presence of a strong 60° bend halfway down the fifth helix, most likely due to the change Val65 → Tyr65 and the steric hindrance coming from the tyrosine side chain. The C-terminus of helix 5 is nearly anti-parallel to helix 3, while it is nearly perpendicular to it in HYD and hPABC. This bend in helix 5 also helps bringing two more aromatic residues (Tyr65 and Tyr72)

closer to the hydrophobic core, along with F61 which replaces the otherwise nearly perfectly conserved L61.

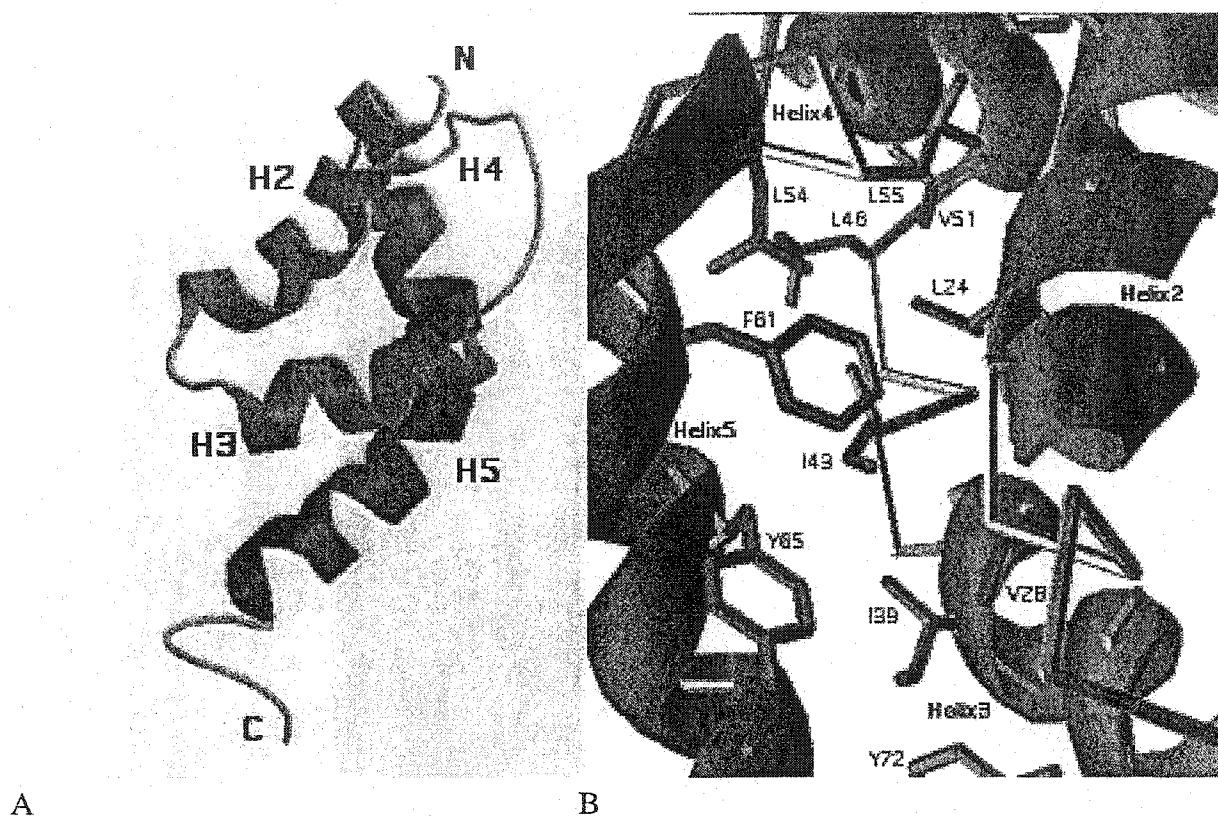


Figure 4: (A) Ribbon drawing of the yPABC domain, where the four helices are labeled according to the hPABC nomenclature. (B) A close-up view of the hydrophobic core of yPABC. Helices are indicated, as well as the amino acids, in one-letter code. Figures generated using RASMOL (Sayle and Milner-White, 1995), MOLSCRIPT (Kraulis, 1991), RENDER (Merritt and Murphy, 1994) and SWISS PDBVIEWER (Guex and Peitsch, 1997).

Comparison of the structural fold of yeast PABC with that of its human and hyperplastic disc counterparts using DALI (Holm and Sander, 1993) confirms the first impression that HYD and yPABC are indeed structurally more similar. The Z-score is a measure of compatibility

between the predicted energies calculated from the structures. The higher the Z-score, the better the fit is between two proteins. When comparing hPABC and HYD to their yeast homolog, the Z-score is nearly three times higher for HYD than for hPABC. Nevertheless, though the two C-termini are different, higher eukaryotic PABP have been shown to be able to substitute for the yeast protein and allow for cell viability (Chekanova et al., 2001), implying that most of the protein's activity does not reside in this part.

3.2 Interaction with the Poly(A) Binding Protein interacting Motif (PAM2)

Interactions of human Poly(A) Binding Protein with two PABP Interacting Proteins (Paip1 and Paip2) have been described (Gray et al., 2000; Khaleghpour et al., 2001; Roy et al., 2002). The PABC-like domain of HYD has been shown to bind Paip1 (Deo et al., 2001). The binding specificity of Paip2 with the C-terminal fragment of human PABP was also assessed by NMR (Kozlov et al., 2001): addition of unlabeled Paip2 to ^{15}N -labeled PABC shifted about half of the ^{15}N - ^1H correlation peaks, indicating the formation of a Paip2-PABC complex. When a peptide derived from the 16-residue long portion of Paip2 that has been identified by deletion mutagenesis studies to be responsible of interactions with PABC (Khaleghpour et al., 2001) was added to PABC, the same spectral changes resulted. Experiments with peptides from two other PABP-binding proteins (Paip1 and a human release factor hRF3) gave similar results.

Binding of Paip occurs on a surface between helices 3 and 5 and extending towards helix 2. There is a deep hydrophobic pocket between helices 2 and 3, bounded by Phe22, Ile25, Ala33, and residues 34-38 of helix 3. The presence of numerous residues with small but significant chemical shift changes indicates the possibility that intramolecular rearrangements occur upon binding (Kozlov et al., 2001). Not surprisingly, this series of experiments showed that the stretches of PABC that were most implicated in binding were also amongst the most conserved ones between all PABC isoforms, particularly the universally conserved KITGMLLE motif in

helix 3. The 12-residue long Paip2 peptide that is minimal for binding to occur can be represented with residue 1 being near helix 5 and the aromatic end of the peptide inside the pocket, near the loop preceding helix 2 (Kozlov et al., 2001).

in the wake of the overall structure determination, similar experiments to detect peptide binding by yeast PABC were carried out by Dr Kozlov. Interactions were detected with Paip2 and hRF3, but none occurred with Paip1, despite the strong similarity between the two sequences, notably the presence of a highly conserved alanine at the middle of C-terminal phenylalanine motif (F-X-P) that is required for binding (Kozlov et al., 2002). Based on the sequences of Paip1 and Paip2, a PHI-BLAST (Zhang et al., 1998) search was undertaken using the non-redundant database at the National Center for BioInformatics (NCBI), and the results of the search are indicated in table 2. The original pattern was modified to actually retrieve all Paip1 and Paip2 entries in the Swiss-Prot/TrEMBL database, and the final motif [S-X-L-[NS]-X-[ND]-A-X-E-F-X-P] has been sent and accepted at the PROSITE pattern database and should be part of the next release of the updated databank.

Paip1 is a mammalian protein that shows some areas of homology to eIF4G, yet no protein could be found in yeast that exhibited its peculiar PABC interacting motif. Similarly, no Paip2-like protein could be retrieved in fungi. Amongst all the proteins found and shown in Table 2, the only one that is also occurring in yeast is RF3. *Saccharomyces cerevisiae* RF3, also known as SUP35, was shown a few years ago to be responsible for the maintenance of a non-Mendelian determinant that confers suppression of nonsense mutations through a prion-like self-

Name	Organism	Length	Starts		Sequence	gi number
hPaip1	Human	480	127-	lm	SKLSVNAPEFYF	sgyss 3046900
mPaip1	Mouse	338	18-	lm	SKLSANAPEFYF	sgyss 4704764
hPaip2	Human	127	107-	vk	SNLNPNAKEFVP	gvkyg 11416373
mPaip2	Mouse	124	107-	vk	SNLNPNAKEFVP	gvky- 12838033
dPaip2	Drosophila	124	103-	ek	SVLNPMADFEFVP	rchvi 7299698
Ataxin2	Human	1312	911-	rk	STLNPNAKEFNP	rsfsq 4506795
SCA2	Human	914	831-	rk	STLNPNAKEFNP	rsfsq 1770390
Ataxin2	Mouse	1285	883-	rk	STLNPNAKEFNP	rsfsq 6677857
RNA-BP	A.Thaliana	320	64-	ei	SHLNPMKEFVP	sflaq 9758532
RNA-BP	A.Thaliana	829	3-	mf	KKLNPEAEFFFP	sykm 8671859
RNA-BP	Rice	289	40-	ll	SKLNPAEEFVP	ssaav 8096600
CG8963	Drosophila	559	71-	da	SNLQATATIFIP	sfsyn 7302843
SCP	Drosophila	1106	165-	sa	SKLQATAPEFVP	nfakl 730843
Tob1	Mouse	363	283-	qt	SALSPNAKEFIF	prmqg 6678397
Tob1	Human	345	265-	kt	SALSPNAKEFIF	prmqg 5032187
Tob2	Human	344	131-	ik	SSFNPDAQVFVP	igsqd 7706739
			251-	pq	SQLSPNAKEFVY	ngggs
Tob2	Mouse	345	130-	ik	SSFNPDAQVFVP	igsqd 10048440
			252-	pq	SQLSPNAKEFVY	ngggs
Blackjack	Grasshopper	1547	140-	me	SRLNPDAKEFVP	sflaq 7511753
MAP 205	Drosophila	1185	234-	nh	SQLNPNAVAFVP	gvgsq 126746
mei-218	Drosophila	1186	533-	mg	SVLNSNAKEFHP	tgesh 7293319
hRF3	Human	628	49-	fs	RKLNVNAKPFVP	nvhaa 8922424
rRF3	Rabbit	588	18-	fs	RQLNVNAKPFVP	nvhaa 7077174
xRF3	Bullfrog	614	69-	sm	AALNVNAKPFVP	nvnav 2134158
mRF3	Mouse	597	20-	fs	SQLNIHAKPFVP	svsaa 3461882
cRF3	C.Elegans	573	2-	-m	SGWNVNASSFVP	nanar 3878044
dRF3	Drosophila	619	12-	kf	STLNVNAVEFVP	sfsyn 7297950
pRF3	P.Pinus	741	34-	vq	SYIHTAQAQFVP	saqpy 3236
ppRF3	P.Pastoris	315	18-	eq	RFNPNSASSFQP	sfnpq 6478794
caRF3	C.Albicans	715	36-	qp	YFNPNQAQAFVP	tggyq 13676380
cmRF3	C.Maltosa	712	27-	qn	YNNPNAAQSFVP	tggyq 2582369
spRF3	S.Pombe	662	20-	qt	SKLSMSAKAPTF	TPkaa 3819704
dhRF3	D.Hansenii	701	42-	qn	NFNANSAPTFIP	sgqqg 15080702
MDC16.14	A.Thaliana	596	459-	ek	STLNPNAKEFKL	NPkak 11994373
CG6441	Drosophila	482	21-	re	YSLNPNAVEFVP	sfrqk 7297259
L5134.7	Leishmania	665	487-	gg	SMVRPVATSPTP	hfqls 7769961
Hyp.Prot.	S.Pombe	1621	1330-	qk	VALNPSPAPEFVP	dstps 7522319
Unk. Prot.	A.Thaliana	557	20-	kp	TTLNPFAAEFVP	ftlrs 3075391
deubiquitinase	Human	813	96-	is	STLNPQAPEFIP	gctas 1136438
Ub. hydrolase	Mouse	793	81-	is	STLNPQAPEFIP	gctts 6678493
ERD15	A.Thaliana	163	9-	gr	STLNPDAPLFIP	sfvrq 15227351
ERD15	Tomato	156	9-	gr	STLNPDAPLFVP	sfvrq 8489786
Hyp.Prot.	A.Thaliana	255	10-	rl	STLNPDAFVDFP	vefre 15236461
Consensus					SxL[NS]x[ND]AxEFxP	

Table 2: PHI-BLAST search and sequence alignment of proteins exhibiting the PABP interacting Motif 2, with the consensual motif at the bottom. Within the alignment, conserved positions are shown in light gray.

This table can also be found on-line at <http://www.bri.nrc.ca/mcgnmr/pabp/pabc.html>.

propagating aggregation mechanism (phenotypically designated as [psi+]) (Ter-Avanesyan et al., 1994). The region of interaction is part of a N-terminal portion of Sup35 that is conserved among distantly related yeast species (Kushnirov et al., 1990; Santoso et al., 2000). A few yeast species indeed did exhibit a paip-like sequence, the first one we found being that of *Pichia pinus*. However, when a yeast-specific BLAST search is run based on *Pichia*'s RF3, two types of homologous RF3s are retrieved (Fig. 5): a few hits are that of closely related proteins that match *Pichia* RF3 on its whole length, but the vast majority of the proteins that are highly homologous (with E-values $\ll 1$) only starts matching past the 160 first residues of those PAM2 containing protein, and correspond mostly to EF1-like proteins, that catalyze the GTP-dependent binding of amino-acyl-tRNA to ribosomes.

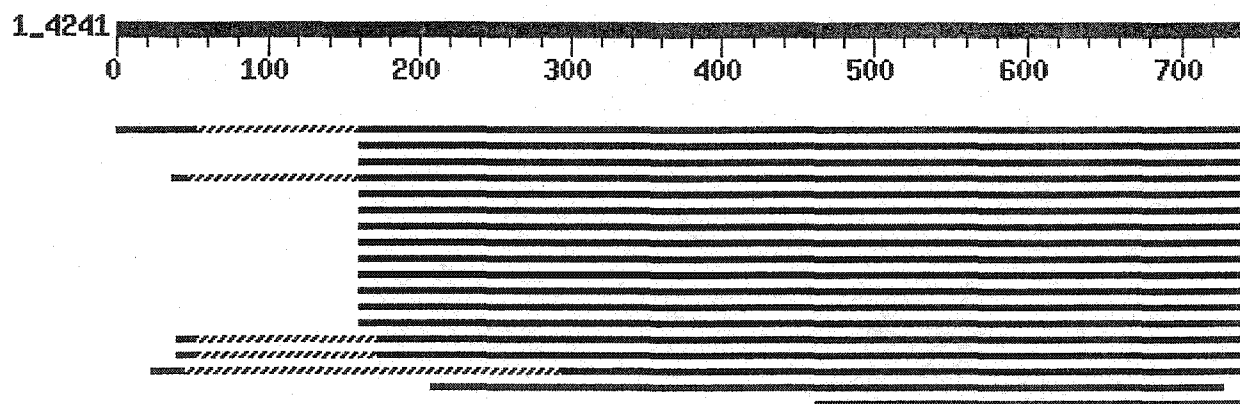


Figure 5: Excerpt of a BLAST search of the *Pichia pinus* RF3 (741 aa) versus a database of fungal sequences. Hits (bars) are all from fungal RF3s. Hatched parts are regions of low complexity, plain lines have a similarity score above 200. The putative PABC-binding domain of pRF3 is found between residues 34 to 46.

Most importantly, absolutely no homologous motif could be retrieved in *S.cerevisiae*, despite the fact that the whole genome of this organism has been sequenced; this should be contrasted with the fact that the *S. pombe* genome has also just been completed (Wood et al., 2002), and that we did indeed find what seems to be a PAM2 motif in some of its RF3-like proteins. Kushnirov *et al.* (Kushnirov et al., 1990) had already noticed something similar when comparing SUP35 from *P.pinus* and *S.cerevisiae*, but could not come up with an explanation for this strong difference in the N-terminal region, that contrasted so much with the high homology on the rest of the protein. Also, when looked at more closely, no real phyletic distribution of those "have" and "have not" organisms could be found in the fungal subkingdom: although the protein sequences are highly homologous, no family or subgroups could be clearly isolated that could be said to exhibit (or not) in a significant portion of its members the extra RF3-like domain in its EF1-like proteins. Of course, considering that only two yeasts have had so far their genome sequenced, this situation is bound to evolve as more light will be shed as other genome sequencing projects are progressing. Alternatively, and based on the slightly divergent fold of *S.cerevisiae* PABC, it could be possible that the presence or absence of helix 1 and/or the lack of sequence conservation in the loops between helices 1-2 and helices 2-3 (the former being implicated in Paip 2 binding in hPABC as seen earlier) play a role on ligand selectivity. The conserved motif could then be different, particularly at the F-X-P C-terminus that is inserted into the deep of the hydrophobic core. Unfortunately, various searches in the N-terminal region of *S.cerevisiae*

SUP35 and other yeast RF3s could not provide a consistent pattern that could be considered as even remotely homologous to the current PAM2 motif.

3.3 Phylogenetic analysis of the PABP family

Poly(A) Binding Proteins are eukaryotic proteins made up of four N-terminal RRM domains and a highly conserved C-terminal sequence 64-72 amino-acids in length (Table 3). An iterated Ψ -blast search with human and yeast PABPs as a search template led to the identification of about 100 isoforms of the protein spread over 22 different eukaryotic organisms (8 from the animal kingdom, 8 plants, 3 parasites, 3 fungi). Paralogs (genes originating from a gene duplication event within an organism) were filtered out so as to show only orthologs (found in separate species but which are derived from a common ancestor) as the taxa in the two trees shown in Fig. 6, and also to compensate for the overrepresentation that would arise for organisms whose genome has been completely sequenced over those that have not. The general appearance of both trees shown in Figure 6 is similar: it clearly appears that three subfamilies can be clearly distinguished. A first group encompasses PABs of animal origin (top branch), with an overall identity of 75% or more. Tissue-specific (testies) or inducible (activated platelets or T cells) isoforms have also been described (Feral et al., 2001; Houng et al., 1997) in humans. Paralogs are duplicated genes within an organism, usually answering to the need for subcellular specialization: this interpretation explains the fact that the human isoforms including PABP show a slightly lower identity level between themselves than when compared to isoforms in other vertebrates.

Ap	MPALASALASAPPEERVMIGEOLYPLVDRLH--DHACKVTGMLLEMDQPEVLHLIDLLRQIKAKVAEAMDVL
At	ISKLASDLALASPDKHPRLGCHLYPLVEQOEP--ANAACKVTGMLLEMDQAEILHLLSPEALKAKVSEALDVL
En	EPLTASMLASAPPOEKQMLGERLFLIQAMHP--TLACKITGMLLEIDNSELHMLLESPELSRSKDEAVAM
Bt	EPHLSAMLAAPPOEKQMLGERIYALIEKLYPGHKDACKITGMLLEIDNSELIMMLQDSILFRSKDEASVL
Ce	QPLIASALAAAAPPOQKMMIGERLYPCVAELOP--DLACKITGMLLEMDNAELIMLLESHEALVSKVDEAIAVL
Cr	VGALASALANATPDQORTMIGENLYPLVEQLEP--DNAACKVTGMLLEMDQTEVHLLLESPEALKAKVAEAMEVL
Cs	ITALASALANAPADQORTMIGENLYPLVDQLEH--DHAACKVTGMLLEMDQTEVHLLLESPEALKAKVAEAMDVL
Dc	EKLIASLIANAKPOEKQILGERLYPMIEFMHA--NLACKITGMLLEIDENSELHMTEDDEALKAKVEEAVAM
Dm	EPLTASMLASAPPOEKQMLGERLFLIQAMHP--TLACKITGMLLEIDNSELHMLLESPELSRSKDEAVAM
Hs	GVHTAQAASAAPPOQKMLGHALYPKIQATOP--ELACKITGMLLEMDNTETLGLLRMITR--LCAPRSTKPLAF
Lm	PEITPQELLESMSPOHRAALGDRLFLKYETIPP--DVAPKTTGMLLEMDKPKKEAYELLNDOKRLEERVTEALCML
Mc	VGILATILANATPEQORLLIGENLYPLVEQLEP--EMAACKVTGMLLEMDQTEVHLLLESPEALKAKVAEAMEVL
Mm	EPLTASMLASAPPOEKQMLGERLFLIQAMHP--SLACKITGMLLEIDNSELHMLLESPELSRSKDEAVAM
Nt	VGALATALANSSPTERTMIGENLYPLVEQLEP--ETAACKVTGMLLEMDQTEVHLLLESPEALKAKVAEAMEVL
Pm	EPLTASMLAAAPPEOKQMLGERLFLIHGMYP--TLACKITGMLLEIDNSELHMTLESPELSRAKVEEAVAM
Rn	EPLTASMLASAPPOEKQMLGERLFLIQAMHP--SLACKITGMLLEIDNSELHMLLESPELSRSKDEAVAM
Sc	FERNANINNOFYQOKORQALGEOLYKKVSAKTSNEFAACKITGMLIDLPFOVFPPLESDELFEQHYKEASAAY
Sp	ERFIAADLAAPPEESRKQVIGLILYKVFVREE--KLSCKITGMLLEMDPNSELLEIDDSAINERNVNAIGML
Ta	IGALASALANSPEETORMMIGENLYPLVDQLEH--DQAAKVTGMLLEMDQTEVHLLLESPEALKAKVAEAMEVL
Tb	GQNLSTVLASMTPDQQNVLGERLYNYIVRNIP--SFAACKVTGMLLEMDENSEILNLDNHSLLDITKVQALDVL
Tc	GQNLSTVLANLTPQQRNVLGERLYNHIVAINP--AAAACKVTGMLLEMDNGEILNLDTEGLIDAKVQALEVL
Xl	EPLTASMLAAAPPOEKQMLGERLFLIQAMHP--TLACKITGMLLEIDNSELHMLLESPELSRSKDEAVAM
DH	TDNATPESLNDHLSVHLQIGERLYPKIHSINO--THAPKITGMLLEIPTPOLLSVISSEETIFOKNEATEIT
HH	ASEGNPSDDPEPLPAHROALGERLYRQAMQP--AFASKITGMLLELSPAQILLIASEDSLRARVDEAMELI
R100	ASEGNPSDDPDPLPAHROALGERLYRQAMQP--AFASKITGMLLELSPAQILLIASEDSLRARVDEAMELI

Table 3: Alignment and conserved residues of homologous PABC sequences identified from 22 different organisms and 3 HYD PABC-like sequences. The sequence of 74 residues contains 58 conserved residues with 50% identity. For readability purposes, paralogs are not shown. Abbreviations, organisms (vulgar name if applicable) and accession numbers by alphabetical order: Ap, *Anemia Phyllitidis*, CAA81127; At, *Arabidopsis thaliana* (cress), AAF43230; Ce, *Caenorhabditis elegans* (worm), CAA21572; Cr, *Chlamydomonas reinhardtii*, AAC39368; Cs, *Cucumis sativus* (cucumber), AAF53202; Dc, *Daucus carota* (carrot), AF349964; Dm, *Drosophila melanogaster* (fruit fly), P21187; DH, *Drosophila melanogaster* HYD, AF252698; En, *Emericella Nidulans*, AAB16848; HH, *Homo sapiens* (human) HYD, AAF88143; Hs, *Homo sapiens*, P11940; Lm, *Leishmania major*, AAC64372; Mc, *Mesembryanthemum crystallinum* (ice plant), AAB61594; Mm, *Mus musculus* (mouse), CAA46522; Nt, *Nicotiana tabacum* (tobacco plant), AAF66823; Pm, *Petromyzon marinus* (sea lamprey), AAB8849; R100, *Rattus Norvegicus* (Norwegian rat) 100kDa protein, Q62671; Rn, *Rattus Norvegicus*, CAC21554; Sc, *Saccharomyces cerevisiae* (baker's yeast), P04147; Sp, *Schizosaccharomyces pombe* (fission yeast), T38950; Ta, *Triticum Aestivum* (wheat), AAB38974; Tb, *Trypanosoma brucei*, AAD13337; Tc, *Trypanosoma cruzi*, AAC46487; Xl, *Xenopus laevis* (frog), CAA40721.

Figure 6: (A) Bootstrapped NJ-tree from known Poly(A) Binding Proteins C-termini from 22 different organisms and HYD PABC-like sequences from three others. (B) Tree based on the whole sequence from the 22 Poly(A) Binding Proteins. Abbreviations are similar to those used in Table 3.

Trypanozoma cruzi is grouped with plant homologues, at a position clearly away from the two other main sub-groups. This grouping is supported by recent work involving several plant and trypanosomal genes that hints at a distant phylogenetic link between the euglenozoan lineage (to which trypanozomids belong) and plants (Baldauf et al., 2000).

A third, more divergent sub-group contains the ubiquitin ligase proteins of the hyperplastic discs (HYD) family, as well as the *S. cerevisiae* and *S.pombe* PABC isoforms. Although the relationship between those ligases and poly-A binding proteins has not been yet clearly established, alignment of human HYD (table 3) shows a clear homology with other PABC proteins (45% identity on average), excepted on the N-terminal side, where a significant structural difference has been shown to occur with the loss of the first alpha helix which is present in human PABC (Deo et al., 2001; Kozlov et al., 2001). Interestingly, the equally divergent sequence of *S.cerevisiae* PABC on the N-terminal side leads to the absence of the first alpha helix. This could conceivably reflect a functional evolution of the PAB protein driven by the need to meet an increasingly complex variety of ligands interacting with the protein in higher eukaryotic lines. Ubiquitin ligases on the other hand, having a somewhat different function, could function in the absence of helix 1, due to the lack of requirement for interactions with more complex (or specific) substrate proteins.

The two trees shown in figure 6 were generated using either the PABC domain or the whole protein sequence -including unfolded or weakly conserved regions like the spacer arm

between PABC and the four N-terminal RRM domains. It is interesting to note that the latter case, despite lower overall sequence identity and homology (Table 4), gives a taxonomically more accurate grouping into families and subgroups. In Figure 6.B, *Leishmania major* PAB groups logically with other parasitic trypanozomids, while *Emericella nidulans* moves from the edge of the animal branch to go with other fungi *S.pombe* and *S.cerevisiae*. Even more interestingly, PABP from *Chlamydomonas reinhardtii*, a unicellular green algae, rearranges itself within the fungal branch of the tree: *Chlamydomonas* is often referred to as a "green yeast", mainly because it is unicellular, grows quickly, forms colonies on plates, can grow haploid or diploid and is easy to transform (Lefebvre and Silflow, 1999). Trees generated using maximum parsimony gave similar results, and so did a tree built from the RRM motifs only (data not shown). When comparing sequence identities in Table 4, the general pattern is that PABC domains are more conserved than RRM domains. The same result is found between paralogs, as RRM domains are also found to be more divergent from one another than the C-termini when comparing the inducible and ubiquitous human isoforms. However, in the particular case of yeasts, *S.cerevisiae* and *S.pombe* PABPs have a lower homology with higher eukaryotic PABCs than there is between the RRM domains. The RRM family is an ancient motif that is found in a variety of proteins of bacterial, fungal, plant and animal origin, suggesting that the RRM is an important structural feature of protein-RNA interaction. We therefore suggest that the RRM part of the protein results from an early duplication event that gave rise to orthologs that evolved in parallel, but separately. It is only much later that the C-terminal domain appeared on the protein,

	HS	PM	XL	SC	TC	TA	AT	SP	EN	CR
HS		77	88	51	34	51	50	57	50	48
PM			80	53	36	56	53	60	54	51
XL				59	34	55	56	61	53	54
SC					34	51	50	61	54	51
RRM Domains					TC		37	34	39	40
TC						TA	59	52	48	51
TA							AT	54	47	49
AT								SP	61	55
SP									EN	51
EN										CR
CR										

	HS	PM	XL	SC	TC	TA	AT	SP	EN	CR
HS										
PM	90									
XL	99	91								
SC	40	37	40							
TC	50	50	50	36						
TA	54	55	54	38	55					
AT	54	55	54	35	53	66				
SP	50	51	51	41	44	48	45			
EN	71	67	73	46	58	70	66	76		
CR	66	65	68	38	51	61	57	56	71	

Table 4: Sequence identity between selected PABPs. Top right: RRM domains. Bottom left: conserved C-termini only were BLASTed against each other. Except for yeasts, homology between RRM domains is lower than between PABC domains. AT, *Arabidopsis thaliana*; CR, *Chlamydomonas reinhardtii*; EN, *Emericella nidulans*; HS, *Homo sapiens*; PM, *Petromyza marinus*; SC, *Saccharomyces cerevisiae*; SP, *Schizosaccharomyces pombe*; TA, *triticum aestivum*; TC, *trypanozoma cruzi*; XL, *Xenopus laevis*.

when a stronger specialization of the protein evolved, explaining the evolutionary difference between the two parts of PABP. Conversely, in yeast, lower conservation of the PABC domain may reflect the lesser need for a specialization of the protein, since unicellular eukaryotes are subject to less complex interactions. Hence, the looser evolutionary pressure for strong sequence conservation, and the lack of PAM2-bearing proteins in *S.cerevisiae*. In a yeast two-hybrid

study, Mangus *et al.* (Mangus et al., 1998) found a number of proteins interacting with the C-terminus of PAB1P. Focusing on Pbp1p, they could not, however, identify a single, twelve residue long, binding motif or interacting region as described in chapter 3.2. Interestingly, their protein had a weak homology to ataxin (and none with paip1 or 2) and it was not in the region that holds the PAM2 motif; interactions with PABP are thus likely to take place through a very different mechanism, closer to that used by eIF4G during mRNA circularization for instance, rather than to that of the higher eukaryotic Paip1 protein. Additionally, a recent publication (Cosson et al., 2002) tried to map the interactions of *S.cerevisiae* RF3/SUP35 with PABC: the strongest interacting region in SUP35 was between residues 1 to 239, while removing the last twenty residues of PABC had a negative effect on binding. This goes in contradiction with a short PAM2 motif at the N-terminus of SUP35 that would regulate interactions with PABC, and also with the central role played by Lys35 in PABC binding of peptides.

3.4 Comparative modeling of PABC orthologs

3.4.1 Foreword

Comparative modeling predicts the three-dimensional structure of a given protein sequence (target) based primarily on its alignment to one or more proteins of known structure (templates) (Marti-Renom et al., 2000). The number of protein sequences that can be modeled and the accuracy of prediction are increasing steadily because of the growth in the number of known protein structures and because of improvements in the modeling algorithms.

Although steady progress has been made in *ab initio* fold prediction, comparative structure modeling remains the most accurate method of guessing reliably at a protein's structure, particularly in cases where identities are fairly high (>30%) (Schoonman et al., 1998). Thus, literature has started to appear in the past couple of years that relies only on modeled folds and virtual docking experiments (Al-Lazikani et al., 2001; Brody et al., 1999; Xu et al., 2001). The overall accuracy spans a wide range, from low resolution models with only a correct fold to models comparable to medium resolution structures as determined by X-ray crystallography or NMR data ((Sanchez and Sali, 1997) and Figure 7). Moreover, the structure of proteins in a family are more conserved than their sequences; thus, if sequence homology can be detected,

structural identity can largely be assumed. It has been speculated that within ten years at least one example of most structural folds will be known, making comparative modeling applicable to most protein sequences (Sanchez and Sali, 1997). It is with this in mind that IBM recently launched its "Blue Gene" project, a massively parallel petaflop (10^{15} floating-point operations per second) supercomputer that is expected to correctly predict protein folding (Allen et al., 2001) within a year's worth of calculations.

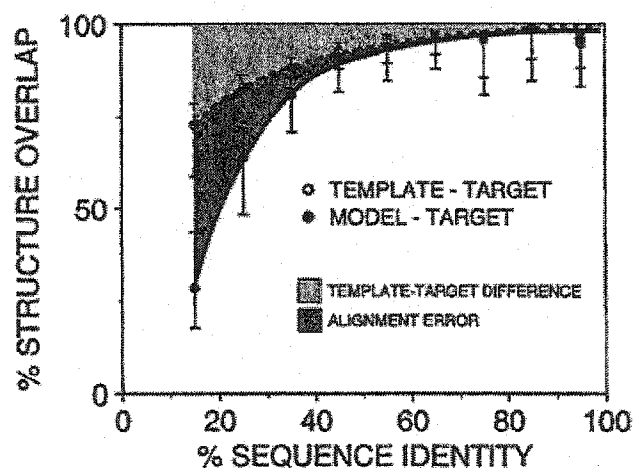


Figure 7: Average model accuracy as a function of the template–target sequence similarity, as calculated with MODELLER. When multiple sequence and structure information is used, and the alignments are edited by hand, the models can be significantly more accurate. The higher the sequence identity, the more significant the accuracy of modeling. Percentage structure overlap is defined as the fraction of structurally equivalent residues. Two residues were considered equivalent when their C atoms are within 3.5 Å of each other upon rigid-body, least-squares superposition of the two structures. Taken from (Marti-Renom et al., 2000).

3.4.2 Comparative Modeling


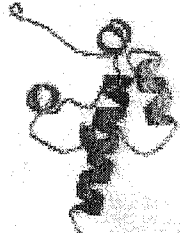
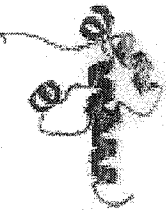

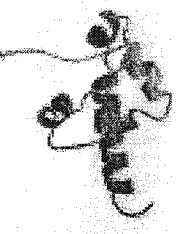

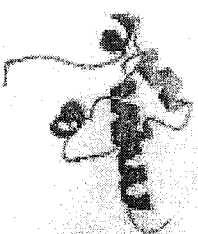

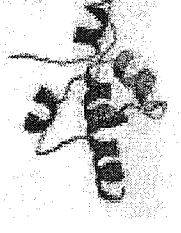
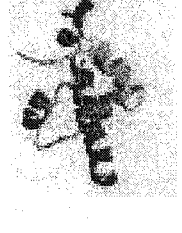
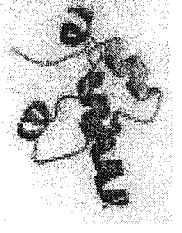



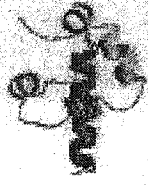
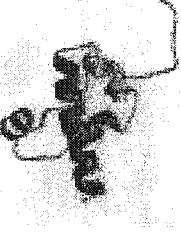
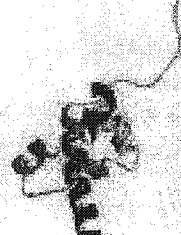
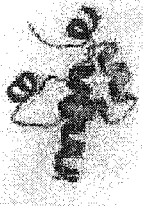

All current comparative modeling methods currently consist of four sequential steps: template selection, template-target alignment, model building and model evaluation. Template

selection was made easy by the fact that three PABC sequences have been solved, all of them with medium (40%) to good (>95%) identity to the remaining isoforms, which facilitated the alignment of target and template sequences. Modeling itself, using the MODELLER program (Sali and Blundell, 1993), then relies on the satisfaction of so-called spatial restraints, that are obtained by assuming that the corresponding distances and angles between aligned residues in the template and target are similar. The model is then derived by minimizing the violations of all the restraints (Marti-Renom et al., 2000). Thus, given a correct initial alignment of sequences, it has been shown that when the evaluation of the template-target similarity is based on the template-target alignment used for modeling, the final model is generally closer to the target structure than any of the templates (Sanchez and Sali, 1997).

An interesting way to monitor progresses made by modeling softwares from around the world is to follow the biannual meetings on Critical Assessment of Techniques for Protein Structure Prediction (CASP), where modelers are challenged to solve the structure of a set of proteins that are simultaneously solved by X-ray crystallography or NMR. Round V of CASP is scheduled for december 2002 and results from CASP IV can be found in (Venclovas et al., 2001).

3.4.3 Comparative modeling of PABC orthologs

Deo *et al.* used this modeling based on spatial restraints technique to perform homology modeling based on the HYD PABC-like crystal structure they had solved (Deo et al., 2001); however, they relied on a fairly distant (45%) and single structure in order to generalize their findings and extend them to the poly(A) binding protein's C-terminal domain. They concluded that the whole PABC family of proteins was made of the same bundle of four helices as their HYD template was. This assertion was partly proved wrong by the simultaneous publication of the human structure of PABC (Kozlov et al., 2001), which is made of five helices, and also by the fact that the template they had used was that of an HYD domain 64 residue-long (residues 16-76 of the conserved PABC domain). Sixteen N-terminal residues were "missing", and from the structure published by Kozlov *et al.* (Kozlov et al., 2001), those were exactly where helix 1 was actually located. However, the N-terminal region from HYD-P is sufficiently divergent from the other PABC protein to lead us to think that no extra structural information could have been derived from the crystal structure, had it been that of a larger sequence. With this in mind, the structure of all 20 extant PABP orthologs were modeled with MODELLER, using the NMR coordinates of hPABC and those of the newly solved yeast structure as templates. The only exception was in the case of mammalian orthologs that all had an identity >95% to hPABC: in this case, the use of two templates, with one being very close and the other being mediumly distant, can prove counter-productive for the accuracy of the final model. The proteins were therefore modeled against hPABC only. A ribbon drawing of each of the models generated for each protein (with the lowest probability density function) is presented in Fig. 8.

A N I M A L S				
	<i>Drosophila melanogaster</i>	<i>Petromyzon marinus</i>	<i>Bos taurus</i>	<i>Caenorhabditis elegans</i>
				
	<i>Xenopus laevis</i>	<i>Mus Musculus</i>	<i>Rattus norvegicus</i>	
Y E A S T				
	<i>Emericella nidulans</i>	<i>Schyzosaccharomyces pombe</i>	<i>Chlamydomonas reinhardtii</i>	
P L A N T S / P A R A S I T E S				
	<i>Nicotiana tabaccum</i>	<i>Arabidospis thaliana</i>	<i>Daucus carota</i>	<i>Cucumis sativus</i>
				
	<i>Trypanosoma brucei</i>	<i>Trypanosoma cruzi</i>	<i>Mesembryanthemum crystallinum</i>	<i>Triticum aestivum</i>

(see next page for legend)

Figure 8: Homology models of 20 PABC orthologs as predicted using MODELLER (Sali and Blundell, 1993). The most conserved feature is the relative length of the C-terminal helix (red), as well as the characteristic arrow-like shape of the whole domain. Figures generated using RASMOL (Sayle and Milner-White, 1995), MOLSCRIPT (Kraulis, 1991) and RENDER (Merritt and Murphy, 1994).

Contrary to the claim by Deo and colleagues, the models obtained and modeled from both yPABC and hPABC are much closer to the human structure than to the hyperplastic disc protein, in that all 20 models are predicted to contain five consecutive helices arranged in a arrowhead-like shape, with no major difference in the core region defined by helices 3,4 and 5 of the human domain. Helix one (residues 5-9) is conserved in all models, including the yeast *Schyzosaccharomyces pombe*, as could be predicted from the high sequence homology between all the N-terminal regions of all the proteins, except for the yPABC and the various HYD isoforms. Another generally conserved feature is the presence of a salt bridge between Lys38 and Glu45 (Fig. 9). As for differences and variations, an extended helix 2 in *L. major* probably results from the absence of the conserved Pro26 (changed to Leu) that otherwise acts as a destabilizing factor for the helical conformation in the other molecules. Also, the loss of the LLH stretch and the mutation Leu52→Phe52 could explain the shorter helix 4 in the *L. major* domain.

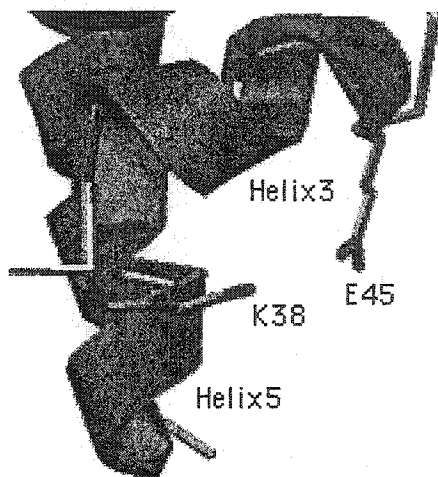


Figure 9: Close up view of the conserved salt bridge between Lys38 and Glu45 from the modeled structure of *Anemia phyllitidis*. Figure generated using SwissPDB Viewer (Guex and Peitsch, 1997).

For most models and residues, the ϕ/ψ angles fall in the most favoured or additionally allowed regions of the Ramachandran plot (Fig. 10). Of the 20 modeled structures, seven ended up having a residue in a disallowed region: Ala37, which is a mutation from the conserved Gly37 at the end of the loop preceding helix 3, was in disallowed regions for *A. thaliana*, *C. sativus*, *D. carota*, *M. Crystallinum*, *T. aestivum*, *T. brucei* and *T. cruzi*.

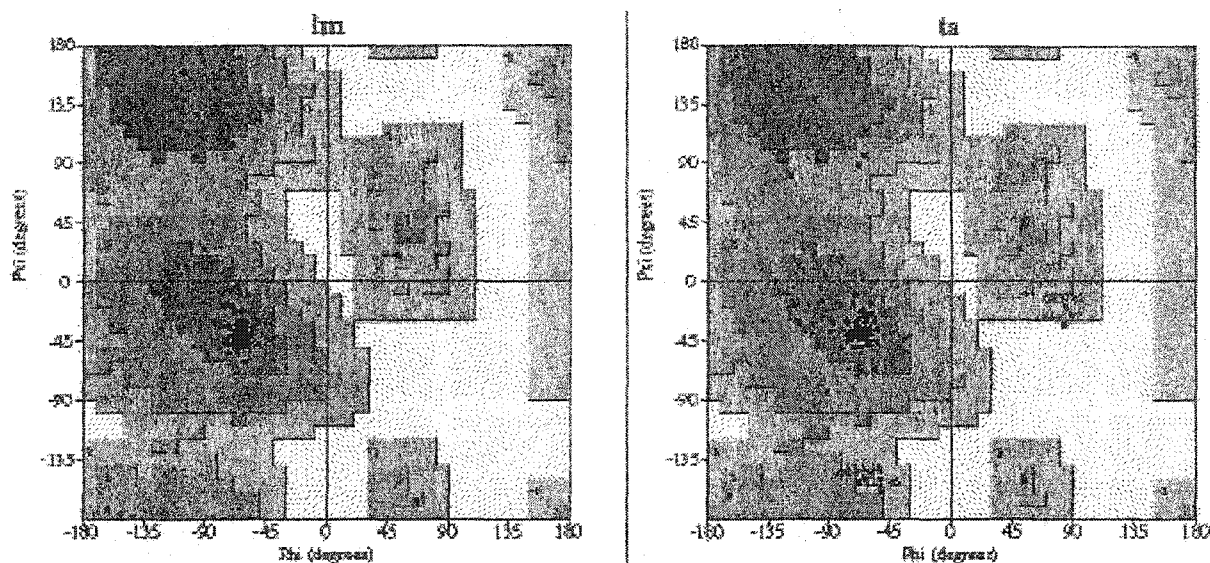


Figure 10: Ramachandran plot for the lowest pdf value of two modeled structures. Left: *L. major*. PABP residues are at 92.3 % in the most favoured regions and 7.7% in additionally allowed regions. Right: *T. aestivum*. 82.6% of the residues are in the most favoured regions, 10.1% in additionally allowed regions, 5.8% in generously allowed regions and 1.4% (1 residue) in disallowed regions.

All those proteins have also a non-conserved residue instead of the expected Leu35, which could explain the incoherent torsion angles applied two residues later in the loop. Also, loops that are longer than five residues have proved to be a very problematic area for modeling systems (Martin et al., 1997). Models generated using as a template the experimentally solved structures of hPABC, yPABC, HYD or a combination of the three, as well as a similar model building using the on-line server SwissModel (Guex and Peitsch, 1997) gave similar results (data not shown). We also decided to run a prediction of the structure of HYD-P, using the full-length of the PABC-like domain. The templates used were alternative combinations of HYD and hPABC or yPABC, as well as the three proteins simultaneously aligned. Neither MODELLER nor

SwissModel predicted that the missing N-terminus is host to an additional helix (results not shown).

During CASP IV (December 2000), prediction of a protein structure with about 51% identity with the template used (*T.cruzi* PABC has 50% identity with hPABC) gave a Root Mean Square Deviation (RMSD) of 1.09 Å³; as a mean of comparison, the NMR structure of human PABC that was published by Kozlov *et al.* had a RMS Deviation of 0.48 Å, and that of yPABC an R.M.S.D. of 0.34 Å. The structure of the C-terminal domain of the *Trypanosoma* Poly(A) Binding Protein is currently being investigated by NMR in our laboratory: it should soon be possible to check the accuracy of the models presented here. It has already been confirmed that tcPABC harbours indeed five helices (Nadeem Siddiqui, personal communication). However, from the convergence curve presented in Figure 7, it is reasonable to assume that models derived from mammalian proteins that exhibit a sequence identity >95% should be of a much higher quality, possibly under 1 Å RMSD.

³ See <http://predictioncenter.llnl.gov/> for a complete list of results.

Chapter IV: Concluding remarks

*To be or not to be, that is the question:
Whether 'tis nobler in the mind to suffer
The slings and arrows of outrageous fortune,
Or to take arms against a sea of troubles
And by opposing end them?*

-William Shakespeare, Hamlet, act III, sc.1

4.1 Summary of results

Yeast PABC was overexpressed as a GST fusion protein, cleaved and purified to homogeneity, for a final yield of 3 mg per liter of LB medium. From the three-dimensional structure of the yeast PABC solved in the laboratory by heteronuclear and multi-dimensional NMR spectroscopy, it was observed that despite the loss of the first alpha helix that is present in residues 5-9 of the human isoform, the overall fold of the protein is well conserved, and resembles a compact bundle of four alpha helices, arranged in a arrowhead-like shape.

A putative consensual binding site for translational proteins interacting with PABC was investigated and was established to be of the form S-X-L-[NS]-X-[ND]-A-X-E-F-X-P in order to match all of Paip 1 and 2 isoforms in a non-redundant database search. A host of proteins were identified that also held this conserved motif. Intriguingly, of all the putative proteins that were retrieved from a search in the non-redundant NCBI database, only release factor 3 (RF3) proteins are present in both lower and higher eukaryotes (all of yeast, vegetals and animals). Most importantly, no RF3 could be found in *S.cerevisiae* that contain a PABC binding site. A series of Φ -BLASTs and phylogenetic analyses showed that yeast RF3-like proteins could be divided into two very distinct groups, based on the loss (or gain?) of a 160 residues long N-terminal domain: longer proteins have a PAM2 conserved motif located in the first third of this extra region, whereas the shorter proteins are more to EF-1 alpha-like. No obvious pattern in the phyletic

distribution of those proteins could be detected: this indicates that eRF3 and eEF1- α have very similar functions, and that they may have very similar roles in lower eukaryotes. Both eEF1- α and eRF3 bind to the ribosomal machinery through the A site and bind to GDP/GTP, but eEF1- α requires eEF-2 for GTP hydrolysis and nucleotide exchange, while eRF3 is a G-protein in itself (see Chapter 1.1.3 and 1.1.4). No explanation has so far been offered in the literature as to why these two proteins are so alike.

The phylogeny of the Poly(A) Binding Protein C-terminal domain was established. It was shown that orthologs could be separated into three distinct groups (and two sub-groups), all related to their respective domains of the eukaryotic kingdoms: yeasts, plants/parasites and animals. Phylogeny is best established when the proteins were considered as a whole rather than by segments, even though the apparition of the RRM domains and the PABC domain are two totally distinct events in the evolutionary history of the protein, and despite the presence of a long, unconserved spacer between the four consecutive RNA-binding motifs and the C-terminus. The average identity in an all-against-all sequence comparison is about 55%, but yeast PABC identity with other proteins is more around 40% and has a greater similarity of sequence with the homologous domain of the hyperplastic disc protein. It is also only in yeast that sequence conservation is weaker at the C-terminus than in the RRM region, reflecting some likely lesser need for the PAM2 interaction mechanism in lower eukaryotes.

A three-dimensional comparative model of each of the 20 PABC orthologs was generated in order to confirm the overall conservation of the PABC fold. The NMR coordinates of both

yeast and human PABC were used as templates for alignment, and modeling was done to satisfy spatial restraints. All models showed a remarkable conservation of the protein fold, including the conservation of the first N-terminal helix, that is absent from both yeast and HYD experimental structures.

The whole length of the PABC-like domain of the hyperplastic disc protein was modeled using the experimental data of yeast and human PABC and HYD, and two distinct modeling programs confirmed that a longer, more complete PABC-like domain of HYD still harboured only four helices as previously published.

4.2 Concluding remarks - Future directions

The absence of a putative binding protein for yPABC that would harbor the PAM2 consensus sequence probably means that because of the uniquely different sequence and structure of the *S. cerevisiae* PAB C-terminal domain, the motif of the corresponding binding partner must be slightly different or different enough to be bypassed during a pattern-based search of a yeast proteome database. However, that what used to be grouped as being all yeast SUP35 can in fact be more accurately subdivided between RF3-like and EF1-like proteins (the latter being 160 residues shorter and lacking an apparent PAM2 motif) hints at a possible evolutionary divergence in the binding partners and interaction mechanisms of PABC, in the sense that interactions of

PABP through this domain are not as much of an essential part of the protein's function as in higher eukaryotic proteins -in mRNA circularization, for instance. Future work should concentrate on elucidating the relationship between hydrophobic core sequence and structural conservation, thus leading to the prediction of a new conserved motif for a yeast binding partner. Conversely, the elucidation of the function(s) of PABC could lead the way to new interacting proteins to look at. Examining their sequence and binding requirements would then pave the way for finding the sequence of the "other" yeast PAM2 sequence, if it exists.

APPENDIX:

Poly(A) Binding Protein Orthologs:

<http://www.bri.nrc.ca/mcgnmr/pabp/pabp.html>

List of proteins with a putative Poly(A) Binding Protein interacting Motif (PAM2):

<http://www.bri.nrc.ca/mcgnmr/pabp/pabc.html>

Modeller Home Page:

<http://guitar.rockefeller.edu/modeller/modeller.html>

Swiss Model:

<http://www.expasy.org/swissmod/SWISS-MODEL.html>

Dali:

<http://www2.ebi.ac.uk/dali/>

Protein Structure Prediction Centre:

<http://predictioncenter.llnl.gov/>

References:

- Afonina, E., Stauber, R., and Pavlakis, G. N. (1998). The human poly(A)-binding protein 1 shuttles between the nucleus and the cytoplasm. *J Biol Chem* 273, 13015-21.
- Al-Lazikani, B., Sheinerman, F. B., and Honig, B. (2001). Combining multiple structure and sequence alignments to improve sequence detection and alignment: application to the SH2 domains of Janus kinases. *Proc Natl Acad Sci U S A* 98, 14796-801.
- Allen, F., Almasi, G., Andreoni, W., Beece, B., Berne, B. J., Bright, A., Brunheroto, J., Cascaval, C., Castanos, J., Coteus, P., Crumley, P., Curioni, A., Denneau, M., Donath, W., Eleftheriou, M., Fitch, B., Fleischer, B., Georgiou, C. J., Germain, R., Giampapa, M., Gresh, D., Gupta, M., Haring, R., Ho, H., Hochschild, P., Hummel, S., Jonas, T., Lieber, D., Martyna, G., Maturu, K., Moreira, J., Newns, D., Newton, M., Philhower, R., Picunko, T., Pitera, J., Pitman, M., Rand, R., Royyuru, A., Salapura, V., Sanomiya, A., Shah, R., Sham, Y., Singh, S., Snir, S., Suits, F., Swetz, R., Swope, W. C., Vishnumurthy, N., Ward, T. J. C., Warren, H., and R., Z. (2001). Blue Gene: A vision for protein science using a petaflop supercomputer. *IBM Systems Journal* 40, 310-327.
- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25, 3389-402.
- Amrani, N., Minet, M., Le Gouar, M., Lacroute, F., and Wyers, F. (1997). Yeast Pab1 interacts with Rna15 and participates in the control of the poly(A) tail length in vitro. *Mol Cell Biol* 17, 3694-701.
- Baer, B. W., and Kornberg, R. D. (1983). The protein responsible for the repeating structure of cytoplasmic poly(A)-ribonucleoprotein. *J Cell Biol* 96, 717-21.
- Baldauf, S. L., Roger, A. J., Wenk-Siefert, I., and Doolittle, W. F. (2000). A kingdom-level phylogeny of eukaryotes based on combined protein data. *Science* 290, 972-7.
- Bandyopadhyay, A., and Maitra, U. (1999). Cloning and characterization of the p42 subunit of mammalian translation initiation factor 3 (eIF3): demonstration that eIF3 interacts with eIF5 in mammalian cells. *Nucleic Acids Res* 27, 1331-7.

- Barrell, B. G., Bankier, A. T., and Drouin, J. (1979). A different genetic code in human mitochondria. *Nature* 282, 189-94.
- Bertram, G., Innes, S., Minella, O., Richardson, J., and Stansfield, I. (2001). Endless possibilities: translation termination and stop codon recognition. *Microbiology* 147, 255-69.
- Blobel, G. (1973). A protein of molecular weight 78,000 bound to the polyadenylate region of eukaryotic messenger RNAs. *Proc Natl Acad Sci U S A* 70, 924-8.
- Bonetti, B., Fu, L., Moon, J., and Bedwell, D. M. (1995). The efficiency of translation termination is determined by a synergistic interplay between upstream and downstream sequences in *Saccharomyces cerevisiae*. *J Mol Biol* 251, 334-45.
- Brawerman, G. (1969). Role of initiation factors in the translation of messenger RNA. *Cold Spring Harb Symp Quant Biol* 34, 307-12.
- Brody, S. S., Gough, S. P., and Kannangara, C. G. (1999). Predicted structure and fold recognition for the glutamyl tRNA reductase family of proteins. *Proteins* 37, 485-93.
- Browning, K. S. (1996). The plant translational apparatus. *Plant Mol Biol* 32, 107-44.
- Buckingham, R. H., Grentzmann, G., and Kisselev, L. (1997). Polypeptide chain release factors. *Mol Microbiol* 24, 449-56.
- Cate, J. H., Yusupov, M. M., Yusupova, G. Z., Earnest, T. N., and Noller, H. F. (1999). X-ray crystal structures of 70S ribosome functional complexes. *Science* 285, 2095-104.
- Cavallius, J., Zoll, W., Chakraborty, K., and Merrick, W. C. (1993). Characterization of yeast EF-1 alpha: non-conservation of post- translational modifications. *Biochim Biophys Acta* 1163, 75-80.
- Chaudhuri, J., Si, K., and Maitra, U. (1997). Function of eukaryotic translation initiation factor 1A (eIF1A) (formerly called eIF-4C) in initiation of protein synthesis. *J Biol Chem* 272, 7883-91.
- Chekanova, J. A., Shaw, R. J., and Belostotsky, D. A. (2001). Analysis of an essential requirement for the poly(A) binding protein function using cross-species complementation. *Curr Biol* 11, 1207-14.

Coller, J. M., Gray, N. K., and Wickens, M. P. (1998). mRNA stabilization by poly(A) binding protein is independent of poly(A) and requires translation. *Genes Dev* 12, 3226-35.

Cosson, B., Couturier, A., Chabelskaya, S., Kiktev, D., Inge-Vechtomov, S., Philippe, M., and Zhouravleva, G. (2002). Poly(A)-Binding Protein Acts in Translation Termination via Eukaryotic Release Factor 3 Interaction and Does Not Influence [PSI(+)] Propagation. *Mol Cell Biol* 22, 3301-15.

Czaplinski, K., Ruiz-Echevarria, M. J., Gonzalez, C. I., and Peltz, S. W. (1999). Should we kill the messenger? The role of the surveillance complex in translation termination and mRNA turnover. *Bioessays* 21, 685-96.

Deo, R. C., Bonanno, J. B., Sonenberg, N., and Burley, S. K. (1999). Recognition of polyadenylate RNA by the poly(A)-binding protein. *Cell* 98, 835-45.

Deo, R. C., Groot, C. M., Rajashankar, K. R., and Burley, S. K. (2002). Recognition of the rotavirus mRNA 3' consensus by an asymmetric NSP3 homodimer. *Cell* 108, 71-81.

Deo, R. C., Sonenberg, N., and Burley, S. K. (2001). X-ray structure of the human hyperplastic discs protein: an ortholog of the C-terminal domain of poly(A)-binding protein. *Proc Natl Acad Sci U S A* 98, 4414-9.

Donahue, T. F., Cigan, A. M., Pabich, E. K., and Valavicius, B. C. (1988). Mutations at a Zn(II) finger motif in the yeast eIF-2 beta gene alter ribosomal start-site selection during the scanning process. *Cell* 54, 621-32.

Drugeon, G., Jean-Jean, O., Frolova, L., Le Goff, X., Philippe, M., Kisselev, L., and Haenni, A. L. (1997). Eukaryotic release factor 1 (eRF1) abolishes readthrough and competes with suppressor tRNAs at all three termination codons in messenger RNA. *Nucleic Acids Res* 25, 2254-8.

Dube, P., Wieske, M., Stark, H., Schatz, M., Stahl, J., Zemlin, F., Lutsch, G., and van Heel, M. (1998). The 80S rat liver ribosome at 25 Å resolution by electron cryomicroscopy and angular reconstitution. *Structure* 6, 389-99.

Feral, C., Guellaen, G., and Pawlak, A. (2001). Human testis expresses a specific poly(A)-binding protein. *Nucleic Acids Res* 29, 1872-1883.

- Fletcher, C. M., Pestova, T. V., Hellen, C. U., and Wagner, G. (1999). Structure and interactions of the translation initiation factor eIF1. *Embo J* 18, 2631-7.
- Ford, L. P., Bagga, P. S., and Wilusz, J. (1997). The poly(A) tail inhibits the assembly of a 3'-to-5' exonuclease in an in vitro RNA stability system. *Mol Cell Biol* 17, 398-406.
- Gingras, A. C., Raught, B., and Sonenberg, N. (1999). eIF4 initiation factors: effectors of mRNA recruitment to ribosomes and regulators of translation. *Annu Rev Biochem* 68, 913-63.
- Gorlach, M., Burd, C. G., and Dreyfuss, G. (1994). The mRNA poly(A)-binding protein: localization, abundance, and RNA-binding specificity. *Exp Cell Res* 211, 400-7.
- Gray, N. K., Collier, J. M., Dickson, K. S., and Wickens, M. (2000). Multiple portions of poly(A)-binding protein stimulate translation in vivo. *Embo J* 19, 4723-33.
- Guex, N., and Peitsch, M. C. (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* 18, 2714-23.
- Gunnery, S., and Mathews, M. B. (1995). Functional mRNA can be generated by RNA polymerase III. *Mol Cell Biol* 15, 3597-607.
- Hatfield, L., Beelman, C. A., Stevens, A., and Parker, R. (1996). Mutations in trans-acting factors affecting mRNA decapping in *Saccharomyces cerevisiae*. *Mol Cell Biol* 16, 5830-8.
- Hawthorne, D. C., and Leupold, U. (1974). Suppressors in yeast. *Curr Top Microbiol Immunol* 64, 1-47.
- Hershey, J. W., and Merrick, W. C. (2000). The Pathway and Mechanism of Initiation of Protein Synthesis. In *Translational Control of Gene Expression*, J. W. B. H. a. M. B. M. N. Sonenberg, ed. (Cold Spring Harbour, N.Y.: ColdSpring Harbour Laboratory Press), pp. 33-88.
- Hill, W. E., Dahlberg, A., Garrett, R. A., Moore, P. B., Schlessinger, D., and Warner, J. R. (1990). *The ribosome: Structure, function and evolution* (Washington, D.C.: ASM Press).
- Holm, L., and Sander, C. (1993). Protein structure comparison by alignment of distance matrices. *J Mol Biol* 233, 123-38.

Horikami, S. M., De Ferra, F., and Moyer, S. A. (1984). Characterization of the infections of permissive and nonpermissive cells by host range mutants of vesicular stomatitis virus defective in RNA methylation. *Virology* 138, 1-15.

Hoshino, S., Imai, M., Kobayashi, T., Uchida, N., and Katada, T. (1999). The eukaryotic polypeptide chain releasing factor (eRF3/GSPT) carrying the translation termination signal to the 3'-Poly(A) tail of mRNA. Direct association of erf3/GSPT with polyadenylate-binding protein. *J Biol Chem* 274, 16677-80.

Houng, A. K., Maggini, L., Clement, C. Y., and Reed, G. L. (1997). Identification and structure of activated-platelet protein-1, a protein with RNA-binding domain motifs that is expressed by activated platelets. *Eur J Biochem* 243, 209-18.

Ito, K., Uno, M., and Nakamura, Y. (2000). A tripeptide 'anticodon' deciphers stop codons in messenger RNA. *Nature* 403, 680-4.

Jacobson, A., and Peltz, S. W. (1996). Interrelationships of the pathways of mRNA decay and translation in eukaryotic cells. *Annu Rev Biochem* 65, 693-739.

Janosi, L., Mottagui-Tabar, S., Isaksson, L. A., Sekine, Y., Ohtsubo, E., Zhang, S., Goon, S., Nelken, S., Shuda, M., and Kaji, A. (1998). Evidence for in vivo ribosome recycling, the fourth step in protein biosynthesis. *Embo J* 17, 1141-51.

Kessler, S. H., and Sachs, A. B. (1998). RNA recognition motif 2 of yeast Pab1p is required for its functional interaction with eukaryotic translation initiation factor 4G. *Mol Cell Biol* 18, 51-7.

Khaleghpour, K., Kahvejian, A., De Crescenzo, G., Roy, G., Svitkin, Y. V., Imataka, H., O'Connor-McCourt, M., and Sonenberg, N. (2001). Dual interactions of the translational repressor Paip2 with poly(A) binding protein. *Mol Cell Biol* 21, 5200-13.

Khaleghpour, K., Svitkin, Y. V., Craig, A. W., DeMaria, C. T., Deo, R. C., Burley, S. K., and Sonenberg, N. (2001). Translational repression by a novel partner of human poly(A) binding protein, Paip2. *Mol Cell* 7, 205-16.

Kovalchuke, O., Kambampati, R., Pladies, E., and Chakraborty, K. (1998). Competition and cooperation amongst yeast elongation factors. *Eur J Biochem* 258, 986-93.

Kozak, M. (1995). Adherence to the first-AUG rule when a second AUG codon follows closely upon the first. *Proc Natl Acad Sci U S A* 92, 7134.

Kozak, M. (1991). Effects of long 5' leader sequences on initiation by eukaryotic ribosomes in vitro. *Gene Expr 1*, 117-25.

Kozak, M. (1979). Inability of circular mRNA to attach to eukaryotic ribosomes. *Nature 280*, 82-85.

Kozak, M. (1991). A short leader sequence impairs the fidelity of initiation by eukaryotic ribosomes. *Gene Expr 1*, 111-5.

Kozak, M., and Shatkin, A. J. (1978). Migration of 40 S ribosomal subunits on messenger RNA in the presence of edeine. *J Biol Chem 253*, 6568-77.

Kozlov, G., Siddiqui, N., Coillet-Matillon, S., Trempe, J. F., Ekiel, I., Sprules, T., and Gehring, K. (2002). Solution structure of the orphan PABC domain from *Saccharomyces cerevisiae* poly(A) binding protein. *J Biol Chem 8*, 8.

Kozlov, G., Siddiqui, N., Coillet-Matillon, S., Trempe, J. F., Ekiel, I., Sprules, T., and Gehring, K. (2002). Solution structure of the orphan PABC domain from *Saccharomyces cerevisiae* poly(A)-binding protein. *J Biol Chem 277*, 22822-8.

Kozlov, G., Trempe, J. F., Khaleghpour, K., Kahvejian, A., Ekiel, I., and Gehring, K. (2001). Structure and function of the C-terminal PABC domain of human poly(A)- binding protein. *Proc Natl Acad Sci U S A 98*, 4409-13.

Krab, I. M., and Parmeggiani, A. (1998). EF-Tu, a GTPase odyssey. *Biochim Biophys Acta 1443*, 1-22.

Kraulis, P. J. (1991). Similarity of protein G and ubiquitin. *Science 254*, 581-2.

Kuhn, U., and Pieler, T. (1996). *Xenopus* poly(A) binding protein: functional domains in RNA binding and protein-protein interaction. *J Mol Biol 256*, 20-30.

Kushnirov, V. V., Ter-Avanesyan, M. D., Didichenko, S. A., Smirnov, V. N., Chernoff, Y. O., Derkach, I. L., Novikova, O. N., Inge-Vechtormov, S. G., Neistat, M. A., and Tolstorukov, II (1990). Divergence and conservation of SUP2 (SUP35) gene of yeast *Pichia pinus* and *Saccharomyces cerevisiae*. *Yeast 6*, 461-72.

Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* 26, 283-291.

Lefebvre, P. A., and Silflow, C. D. (1999). *Chlamydomonas*: the cell and its genomes. *Genetics* 151, 9-14.

Lo, H. J., Huang, H. K., and Donahue, T. F. (1998). RNA polymerase I-promoted HIS4 expression yields uncapped, polyadenylated mRNA that is unstable and inefficiently translated in *Saccharomyces cerevisiae*. *Mol Cell Biol* 18, 665-75.

Luo, Y., and Goss, D. J. (2001). Homeostasis in mRNA initiation: wheat germ poly(A)-binding protein lowers the activation energy barrier to initiation complex formation. *J Biol Chem* 276, 43083-6.

Mangus, D. A., Amrani, N., and Jacobson, A. (1998). Pbp1p, a factor interacting with *Saccharomyces cerevisiae* poly(A)- binding protein, regulates polyadenylation. *Mol Cell Biol* 18, 7383-96.

Marti-Renom, M. A., Stuart, A. C., Fiser, A., Sanchez, R., Melo, F., and Sali, A. (2000). Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct* 29, 291-325.

Martin, A. C., MacArthur, M. W., and Thornton, J. M. (1997). Assessment of comparative modeling in CASP2. *Proteins Suppl*, 14-28.

Merritt, E. A., and Murphy, M. E. P. (1994). ***Raster3D* Version 2.0. A program for photorealistic molecular graphics.** *Acta Cryst D* 50, 869-873 .

Minvielle-Sebastia, L., Preker, P. J., Wiederkehr, T., Strahm, Y., and Keller, W. (1997). The major yeast poly(A)-binding protein is associated with cleavage factor IA and functions in premessenger RNA 3'-end formation. *Proc Natl Acad Sci U S A* 94, 7897-902.

Moore, P. B., and Steitz, T. A. (2002). The involvement of RNA in ribosome function. *Nature* 418, 229-35.

Nakamura, Y., Ito, K., and Ehrenberg, M. (2000). Mimicry grasps reality in translation termination. *Cell* 101, 349-52.

Namy, O., Hatin, I., and Rousset, J. P. (2001). Impact of the six nucleotides downstream of the stop codon on translation termination. *EMBO Rep* 2, 787-93.

Nierhaus, K. H., Franceschi, F., A.R., S., Erdman, V. A., and Wittmann-Liebold, B. (1993). *The translational apparatus: Structure, function, regulation, evolution* (New York, NY: Plenum Press).

Ohlmann, T., Rau, M., Morley, S. J., and Pain, V. M. (1995). Proteolytic cleavage of initiation factor eIF-4 gamma in the reticulocyte lysate inhibits translation of capped mRNAs but enhances that of uncapped mRNAs. *Nucleic Acids Res* 23, 334-40.

Osawa, S., Jukes, T. H., Watanabe, K., and Muto, A. (1992). Recent evidence for evolution of the genetic code. *Microbiol Rev* 56, 229-64.

Paraskeva, E., Gray, N. K., Schlager, B., Wehr, K., and Hentze, M. W. (1999). Ribosomal pausing and scanning arrest as mechanisms of translational regulation from cap-distal iron-responsive elements. *Mol Cell Biol* 19, 807-16.

Pestova, T. V., Lomakin, I. B., Lee, J. H., Choi, S. K., Dever, T. E., and Hellen, C. U. (2000). The joining of ribosomal subunits in eukaryotes requires eIF5B. *Nature* 403, 332-5.

Pyronnet, S., Pradayrol, L., and Sonenberg, N. (2000). A cell cycle-dependent internal ribosome entry site. *Mol Cell* 5, 607-16.

Ramakrishnan, V. (2002). Ribosome structure and the mechanism of translation. *Cell* 108, 557-72.

Riis, B., Rattan, S. I., Clark, B. F., and Merrick, W. C. (1990). Eukaryotic protein elongation factors. *Trends Biochem Sci* 15, 420-4.

Roy, G., De Crescenzo, G., Khaleghpour, K., Kahvejian, A., O'Connor-McCourt, M., and Sonenberg, N. (2002). Paip1 interacts with PABP through two independent binding motifs. *Mol. Cell. Biol* *In press*.

Sachs, A. (1990). The role of poly(A) in the translation and stability of mRNA. *Curr Opin Cell Biol* 2, 1092-8.

Sachs, A. B. (2000). Cell cycle-dependent translation initiation: IRES elements prevail. *Cell* 101, 243-5.

- Sachs, A. B., and Davis, R. W. (1989). The poly(A) binding protein is required for poly(A) shortening and 60S ribosomal subunit-dependent translation initiation. *Cell* 58, 857-67.
- Sali, A., and Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 234, 779-815.
- Sanchez, R., and Sali, A. (1997). Advances in comparative protein-structure modelling. *Curr Opin Struct Biol* 7, 206-14.
- Sanchez, R., and Sali, A. (1997). Evaluation of comparative protein structure modeling by MODELLER-3. *Proteins Suppl*, 50-8.
- Santoso, A., Chien, P., Osherovich, L. Z., and Weissman, J. S. (2000). Molecular basis of a yeast prion species barrier. *Cell* 100, 277-88.
- Sayle, R. A., and Milner-White, E. J. (1995). RASMOL: biomolecular graphics for all. *Trends Biochem Sci* 20, 374.
- Schoonman, M. J., Knegt, R. M., and Grootenhuys, P. D. (1998). Practical evaluation of comparative modelling and threading methods. *Comput Chem* 22, 369-75.
- Si, K., and Maitra, U. (1999). The *Saccharomyces cerevisiae* homologue of mammalian translation initiation factor 6 does not function as a translation initiation factor. *Mol Cell Biol* 19, 1416-26.
- Smith, K. E., and Henshaw, E. C. (1975). Binding of Met-tRNA-f to native and derived 40S ribosomal subunits. *Biochemistry* 14, 1060-7.
- Söll, D., and RajBandhary, U. L. (1995). *tRNA: Structure, biosynthesis and function* (Washington, D.C.: ASM Press).
- Song, H., Mugnier, P., Das, A. K., Webb, H. M., Evans, D. R., Tuite, M. F., Hemmings, B. A., and Barford, D. (2000). The crystal structure of human eukaryotic release factor eRF1--mechanism of stop codon recognition and peptidyl-tRNA hydrolysis. *Cell* 100, 311-21.
- Srivastava, S., Verschoor, A., and Frank, J. (1992). Eukaryotic initiation factor 3 does not prevent association through physical blockage of the ribosomal subunit-subunit interface. *J Mol Biol* 226, 301-4.

Tarun, S. Z., Jr., and Sachs, A. B. (1996). Association of the yeast poly(A) tail binding protein with translation initiation factor eIF-4G. *Embo J* 15, 7168-77.

Tarun, S. Z., Jr., and Sachs, A. B. (1995). A common function for mRNA 5' and 3' ends in translation initiation in yeast. *Genes Dev* 9, 2997-3007.

Tarun, S. Z., Jr., Wells, S. E., Deardorff, J. A., and Sachs, A. B. (1997). Translation initiation factor eIF4G mediates in vitro poly(A) tail- dependent translation. *Proc Natl Acad Sci U S A* 94, 9046-51.

Ter-Avanesyan, M. D., Dagkesamanskaya, A. R., Kushnirov, V. V., and Smirnov, V. N. (1994). The SUP35 omnipotent suppressor gene is involved in the maintenance of the non-Mendelian determinant [psi+] in the yeast *Saccharomyces cerevisiae*. *Genetics* 137, 671-6.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25, 4876-82.

Venclovas, Zemla, A., Fidelis, K., and Moul, J. (2001). Comparison of performance in successive CASP experiments. *Proteins* 45, 163-70.

Wahle, E. (1991). A novel poly(A)-binding protein acts as a specificity factor in the second phase of messenger RNA polyadenylation. *Cell* 66, 759-68.

Wahle, E., and Keller, W. (1996). The biochemistry of polyadenylation. *Trends Biochem Sci* 21, 247-50.

Wahle, E., and Ruegsegger, U. (1999). 3'-End processing of pre-mRNA in eukaryotes. *FEMS Microbiol Rev* 23, 277-95.

Wang, X., Ullah, Z., and Grumet, R. (2000). Interaction between zucchini yellow mosaic potyvirus RNA-dependent RNA polymerase and host poly-(A) binding protein. *Virology* 275, 433-43.

Wang, Z., and Kiledjian, M. (2000). The poly(A)-binding protein and an mRNA stability protein jointly regulate an endoribonuclease activity. *Mol Cell Biol* 20, 6334-41.

Wei, C. L., Kainuma, M., and Hershey, J. W. (1995). Characterization of yeast translation initiation factor 1A and cloning of its essential gene. *J Biol Chem* 270, 22788-94.

Wells, S. E., Hillner, P. E., Vale, R. D., and Sachs, A. B. (1998). Circularization of mRNA by eukaryotic translation initiation factors. *Mol Cell* 2, 135-40.

Wimberly, B. T., Brodersen, D. E., Clemons, W. M., Jr., Morgan-Warren, R. J., Carter, A. P., Vonnrhein, C., Hartsch, T., and Ramakrishnan, V. (2000). Structure of the 30S ribosomal subunit. *Nature* 407, 327-39.

Wimberly, B. T., Guymon, R., McCutcheon, J. P., White, S. W., and Ramakrishnan, V. (1999). A detailed view of a ribosomal active site: the structure of the L11- RNA complex. *Cell* 97, 491-502.

Wood, V., Gwilliam, R., Rajandream, M. A., Lyne, M., Lyne, R., Stewart, A., Sgouros, J., Peat, N., Hayles, J., Baker, S., Basham, D., Bowman, S., Brooks, K., Brown, D., Brown, S., Chillingworth, T., Churcher, C., Collins, M., Connor, R., Cronin, A., Davis, P., Feltwell, T., Fraser, A., Gentles, S., Goble, A., Hamlin, N., Harris, D., Hidalgo, J., Hodgson, G., Holroyd, S., Hornsby, T., Howarth, S., Huckle, E. J., Hunt, S., Jagels, K., James, K., Jones, L., Jones, M., Leather, S., McDonald, S., McLean, J., Mooney, P., Moule, S., Mungall, K., Murphy, L., Niblett, D., Odell, C., Oliver, K., O'Neil, S., Pearson, D., Quail, M. A., Rabinowitsch, E., Rutherford, K., Rutter, S., Saunders, D., Seeger, K., Sharp, S., Skelton, J., Simmonds, M., Squares, R., Squares, S., Stevens, K., Taylor, K., Taylor, R. G., Tivey, A., Walsh, S., Warren, T., Whitehead, S., Woodward, J., Volckaert, G., Aert, R., Robben, J., Grymonprez, B., Weltjens, I., Vanstreels, E., Rieger, M., Schafer, M., Muller-Auer, S., Gabel, C., Fuchs, M., Fritz, C., Holzer, E., Moestl, D., Hilbert, H., Borzym, K., Langer, I., Beck, A., Lehrach, H., Reinhardt, R., Pohl, T. M., Eger, P., Zimmermann, W., Wedler, H., Wambutt, R., Purnelle, B., Goffeau, A., Cadieu, E., Dreano, S., Gloux, S., Lelaure, V., et al. (2002). The genome sequence of *Schizosaccharomyces pombe*. *Nature* 415, 871-80.

Wyers, F., Minet, M., Dufour, M. E., Vo, L. T., and Lacroute, F. (2000). Deletion of the PAT1 gene affects translation initiation and suppresses a PAB1 gene deletion in yeast. *Mol Cell Biol* 20, 3538-49.

Xu, D., Baburaj, K., Peterson, C. B., and Xu, Y. (2001). Model for the three-dimensional structure of vitronectin: predictions for the multi-domain protein from threading and docking. *Proteins* 44, 312-20.

Yueh, A., and Schneider, R. J. (1996). Selective translation initiation by ribosome jumping in adenovirus- infected and heat-shocked cells. *Genes Dev* 10, 1557-67.

Zhang, Z., Schaffer, A. A., Miller, W., Madden, T. L., Lipman, D. J., Koonin, E. V., and Altschul, S. F. (1998). Protein sequence similarity searches using patterns as seeds. *Nucleic Acids Res* 26, 3986-90.