

Universidade Federal de Santa Catarina
Centro Tecnológico
Departamento de Engenharia Química e Engenharia de
Alimentos
Programa de Pós-Graduação em Engenharia Química

Gisele Serpa

**Genômica do X-Frágil:
Elementos de Regulação do Gene
*FMR1***

Tese apresentada ao Programa de Pós-Graduação
em Engenharia Química da
Universidade Federal de Santa Catarina como
requisito parcial para obtenção do grau de
Doutor em Engenharia Química

Orientador: Prof. Dr. Luismar Marques Porto

**Florianópolis, SC
Agosto 2008**

Gisele Serpa

**Genômica do X-Frágil:
Elementos de Regulação do Gene *FMR1***

Esta tese foi julgada e aprovada para a obtenção de grau de Doutor em Engenharia Química no Programa de Pós-Graduação em Engenharia Química da Universidade Federal de Santa Catarina
Área de Concentração:
Processos Químicos e Biotecnológicos.
Linha de Pesquisa em
Engenharia Genômica e Engenharia Biomédica

Prof. Dr. Luismar Marques Porto
Orientador

Prof. Dr. Leonel Teixeira Pinto
Coordenador do Programa de Pós-Graduação em Engenharia Química

Banca Examinadora:

Prof. Dr. Leonel Teixeira Pinto
Presidente

Prof. Dr. André Oliveira de Souza
Lima
Membro Externo (UNIVALI)

Prof^a Dr^a Angélica Francesca Maris
Membro Externo (UNOESC)

Prof^a Dr^a Carla Inês Taska
Membro Interno (BQA/CCB/UFSC)

Prof^a Dr^a Cíntia Soares
Membro Interno (Pós-ENQ/UFSC)

Florianópolis – SC, Agosto de 2008

Serpa, Gisele

Genômica do X-Frágil: Elementos de Regulação do Gene FMR1. 90 p.

Tese (Doutorado) – Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Engenharia Química.

1. Síndrome do X-Frágil – 2. Gene FMR1 – 3. Fatores de Transcrição –
4. Genômica Comparativa

Este trabalho é parte integrante das pesquisas realizadas pelo Grupo de Engenharia Genômica, e foi desenvolvido no Laboratório de Tecnologias Integradas (InteLAB), do Departamento de Engenharia Química e Engenharia de Alimentos, e Laboratório de Neuroquímica, do Departamento de Bioquímica da Universidade Federal de Santa Catarina, e no Department of Biochemistry and Molecular Medicine, University of California, School of Medicine, Davis, Califórnia, Estados Unidos.

AGRADECIMENTOS

Gostaria de deixar registrado aqui o meu mais profundo agradecimento àqueles que de alguma forma contribuíram para esta conquista, em especial:

A Deus, o companheiro fiel das horas fáceis e difíceis.

À minha mãe Ingelora e à minha irmã Patrícia, que sempre estiveram presentes, mesmo que em pensamento.

Ao meu marido Ramon, que entrou em minha vida durante esta empreitada e esteve ao meu lado nos momentos bons e ruins.

Aos meus avós Irmgard e Herbert Schmidt, meus anjos da guarda, que continuam vivos no meu coração e nos meus pensamentos.

Ao Prof Dr Luismar Marques Porto, pela paciência e competência na orientação deste trabalho.

À amiga Flora Tassone, ao casal Paul e Randi Hagerman e aos integrantes de sua equipe de pesquisa do Departamento de Bioquímica e Medicina Molecular da Universidade da Califórnia em Davis, pela acolhida e orientação durante o estágio de doutorado.

À Prof^a Dr^a Carla Inês Tasca e aos integrantes de sua equipe no Laboratório de Neuroquímica do Departamento de Bioquímica da UFSC, em especial à Tharine Aparecida Dal-Cim pela ajuda em alguns procedimentos experimentais.

À Ingrid Tremel Barbato, pela amizade e ajuda em todos os momentos.

Às amigas Mariana, Tânia, Maria Cristina e Pricila, pela troca de informações e estímulos constantes.

Aos amigos do Intelab: Fernanda, Claudimir, Itamar, Derce, Sidnei, Carlos Rambo e em especial à Julia e à Patrícia, pelo companheirismo e ajuda valiosa.

Ao CNPq e à CAPES, pelo apoio financeiro.

RESUMO

A Síndrome do X Frágil (SXF) é a forma de retardo mental herdado mais comum encontrada, afetando um entre 4000 homens e uma entre 8000 mulheres. Ela é causada devido a alterações do nível de transcrição e tradução do gene *FMR1*, que nos humanos e em vários outros mamíferos se encontra na região Xq27.3 do cromossomo X humano. Este gene codifica a proteína FMRP, cuja falta de expressão está associada ao retardo mental e a outros fenótipos característicos da SXF. A função da proteína FMRP no organismo ainda não é totalmente conhecida, porém acredita-se que ela esteja intimamente ligada ao transporte de mRNA do núcleo para o citoplasma dos neurônios e na sua condução até os ribossomos, onde faz parte dos complexos de ribonucleoproteínas (RNP) e controla a tradução de determinados mRNA. Diante deste contexto, este trabalho tem o objetivo de contribuir para a compreensão dos elementos reguladores da expressão do gene *FMR1* e também incrementar os estudos sobre as funções desempenhadas pela proteína FMRP. Uma análise da região à montante do códon de início do gene *FMR1*, utilizando técnicas de genômica comparativa, *gel shift* e espectrometria de massa mostrou potenciais sítios de interação para proteínas de transcrição. Foram identificadas várias proteínas que interagiram com os segmentos de DNA conservados estudados *in vitro*, com especial significância para as proteínas Pur- α , Pur- β e as ribonucleoproteínas heterogêneas A2/B1 (hnRNP A2/hnRNP B1). O papel da Pur- α e Pur- β como elemento na regulação da transcrição do gene *FMR1* nunca foi demonstrado, sendo esse o primeiro trabalho que propõe esta relação. No estudo sobre a função da proteína FMRP utilizou-se análise de fluxos metabólicos para calcular os fluxos intracelulares dos complexos envolvidos na regulação da tradução de mRNA nos dendritos neuronais, onde a FMRP é uma proteína chave. O uso de modelos estequiométricos das reações ou interações e a aplicação de balanços de massa permitiram a estimativa de alguns fluxos importantes no papel biológico da FMRP em nível celular. Esses resultados poderão auxiliar no diagnóstico e tratamento da Síndrome do X-Frágil e doenças relacionadas.

ABSTRACT

Fragile X Syndrome (FXS) is the most common form of inherited mental retardation with the incidence of 1:4000 in males and 1:8000 in females. The syndrome occurs as an effect of the lack of expression of the fragile X mental retardation protein (FMRP) codified by the *FMR1* gene located in the q27.3 region of the X chromosome. Absence of the FMRP results in mental retardation and other characteristics of the Fragile X phenotypes. Some FMRP functions in the organism are still unknown; however, structural evidences of the protein suggest that it is involved in nuclear export, cytoplasmic transport, and/or translation control of target mRNA. In this work, we combined *in silico* comparative genomics tools with proteomic analysis to investigate regulatory elements in the upstream region of the *FMR1* gene transcription start site. We identified proteins that interact with studied DNA conserved sequences, particularly Pur- α , Pur- β and heterogeneous nuclear ribonucleoproteins A2/B1 (hnRNP A2/hnRNP B1). To our acknowledge this is the first work that correlate those proteins to the *FMR1* gene transcription. We used Metabolic Flux Analysis (MFA) to estimate turnovers of molecules and complexes involved in mRNA transport to dendrites and its translation at synapses. Stoichiometric models of biochemical reactions or interactions and mass balances may be useful in diagnosis and future treatment of the Fragile X Syndrome and related diseases.

GLOSSÁRIO

AMPA (receptor)

Receptor de glutamato.

Camundongo Knockout

Camundongo geneticamente modificado com um ou mais genes silenciados.

DAG

Diacilglicerol. Produto da clivagem do inositol fosfolipídio devido à estimulação de receptores neuronais que ativam a fosfolipase C. Ativa a proteína quinase C (PKC).

DNA

Ácido desoxirribonucléico. É o ácido nucléico que contém a informação genética para o desenvolvimento celular e de um organismo.

DNase footprinting

Técnica de biologia molecular que detecta a interação DNA-proteína usando o fato de que quando ligada ao DNA a proteína freqüentemente protege o DNA da clivagem enzimática.

Electrophoretic Mobility Shift Assay (EMSA) ou Gel Shift

Técnica de biologia molecular em que se estuda a regulação gênica através das interações DNA-proteína. Esta técnica é baseada na observação de que complexos DNA-proteína migram mais lentamente do que moléculas de DNA livre quando submetidas a uma eletroforese em gel de agarose ou poliacrilamida. Devido à baixa taxa de migração observa-se um retardamento das bandas que contém os complexos DNA-proteína.

Enhancer

Pequena região do DNA onde se ligam proteínas de regulação da transcrição.

FMRP

Fragile X Mental Retardation Protein. Proteína relacionada à Síndrome do X-Frágil.

FT

Fator de Transcrição. Proteína que se liga ao DNA e regula a transcrição de um determinado gene.

FXR1

Proteína homóloga à FMRP, codificada pelo gene *FXR1*.

FXR2

Proteína homóloga à FMRP, codificada pelo gene *FXR2*.

Gel Super Shift

Técnica de biologia molecular usada para caracterizar o fator de transcrição que atua na regulação de um dado gene. O mesmo que Gel Shift, porém, neste caso utilizam-se anticorpos específicos contra determinados fatores e os complexos são formados por DNA-proteína-anticorpo.

Immunoblotting ou Western blot

Técnica para detecção de proteínas específicas em uma amostra, através da interação proteína-anticorpo.

IP3

Trifosfato de inositol. Produto da clivagem do inositol fosfolipídio devido à estimulação de receptores neuronais que ativam a fosfolipase C. Difunde para o citosol e abre os canais de Ca^{2+} dos reservatórios do retículo endoplasmático.

LTD

Long Term Depression. Depressão de Longa Duração.

LTP

Long Term Potenciation. Potenciação de Longa Duração.

Max

Fator de transcrição. Liga-se a uma seqüência específica de DNA chamada E-box.

mGLUR

Receptor metabotrópico de glutamato.

NMDA (receptor)

Receptor ionotrópico de glutamato.

Nrf-1

Nuclear respiratory factor-1. Fator de transcrição. Liga-se a seqüências específicas da região reguladora de determinados genes e regula a sua transcrição.

Phylogenetic footprinting

Técnica para identificação de sítios de interação para fatores de transcrição que sejam bem conservados entre seqüências não codificantes, ortólogas de múltiplas espécies.

PKC

Proteína Quinase C.

Primers

Iniciadores. Seqüência de ácidos nucléicos que servem para iniciar a replicação do DNA.

SP-1

Fator de transcrição. Liga-se a seqüências específicas da região reguladora de determinados genes e regula a sua transcrição.

Splicing

Modificação que ocorre com os mRNA em eucariotos para a retirada de seqüências não codificantes (íntrons) existentes dentro do gene, chamados éxons.

Splicing alternativo

Processo que ocorre nos eucariotos no qual o *splicing* do pré-mRNA de um dado gene pode levar a diferentes mRNA e conseqüentemente a diferentes proteínas, devido a saída de diferentes íntrons.

SXF

Síndrome do X-Frágil.

TFBS

Sítios de interação para fatores de transcrição.

USF1 e USF2

Fatores de transcrição. Ligam-se às seqüências específicas da região reguladora de determinados genes e regulam a sua transcrição.

ÍNDICE

1. Introdução	1
1.1. <i>Objetivos</i>	3
2. Revisão Bibliográfica	5
2.1. <i>A Síndrome do X-Frágil e sua caracterização</i>	5
2.2. <i>FMRP e suas proteínas homólogas</i>	7
2.3. <i>Regulação da expressão gênica</i>	11
2.3.1. <i>Genômica comparativa no estudo da regulação</i>	11
2.3.2. <i>Análise gênica dos elementos de regulação da transcrição</i>	12
2.3.3. <i>Fatores de regulação da transcrição do gene FMR1</i>	17
2.4. <i>Purificação e Caracterização de Proteínas de Ligação ao DNA</i>	19
2.5. <i>FMRP versus plasticidade sináptica e aprendizado</i>	23
3. Materiais e Métodos	30
3.1. <i>Seqüências de DNA Genômico</i>	31
3.2. <i>Determinação in silico das regiões conservadas do DNA que contêm sítios de interação para fatores de transcrição</i>	31
3.3. <i>Animais de laboratório</i>	32
3.4. <i>Extração de proteínas nucleares e citoplasmáticas</i>	32
3.5. <i>Extração de DNA genômico</i>	33
3.6. <i>Amplificação e purificação do fragmento de DNA de interesse</i>	33
3.7. <i>Gel shift para determinação da ligação DNA-proteína de regiões reguladoras para o gene FMR1</i>	34
3.8. <i>Recuperação de proteínas nucleares através de precipitação por afinidade</i>	35
3.9. <i>Separação e identificação das proteínas nucleares purificadas</i>	36
3.10. <i>Imunodeteção para confirmação da proteína Pur-α:</i>	36
3.11. <i>Análise de Fluxos Metabólicos</i>	36

4. Resultados e Discussão	39
4.1. <i>Determinação de possíveis sítios para fatores de regulação da transcrição do gene FMR1</i>	39
4.2. <i>Interação DNA-proteína nas seqüências conservadas</i>	47
4.3. <i>Purificação e identificação das proteínas nucleares que se ligam especificamente aos segmentos de DNA conservados</i>	50
4.4. <i>Análise de fluxos metabólicos no estudo do modelo funcional teórico para a proteína FMRP</i>	59
5. Conclusões e sugestões para Trabalhos futuros	67
5.1. <i>Sugestões para trabalhos futuros</i>	68
6. Referências Bibliográficas	68

1. INTRODUÇÃO

Os recentes avanços nas ciências genômicas têm motivado a introdução de ferramentas de engenharia na análise de processos biológicos em geral. Atualmente, os processos biológicos de interesse para a engenharia podem incluir desde a produção de bioprodutos por vias fermentativas até a compreensão de doenças humanas herdadas ou adquiridas.

Visando contribuir para o avanço das pesquisas genômicas dentro do contexto da engenharia, de tal forma a estreitar os laços entre as ciências médicas, a biologia molecular e a engenharia de processos biológicos, este trabalho tem o objetivo de contribuir para a compreensão dos processos reguladores da expressão do gene *FMR1* e também incrementar os estudos sobre as funções desempenhadas pela proteína FMRP. Utilizou-se para isso ferramentas computacionais de genômica comparativa, técnicas de biologia molecular e análises de espectrometria de massa, a fim de identificar algumas proteínas reguladoras da transcrição do gene e abordagens de processo aliadas a programas computacionais para a determinação de um modelo matemático que represente a atuação da proteína FMRP nos neurônios.

A falta de expressão da FMRP em humanos e em outros animais causa a Síndrome do X-Frágil (SXF), que é a forma de retardo mental herdado mais comum encontrada. Indivíduos afetados por esta síndrome possuem comportamento mental variável que pode ir desde uma dificuldade no aprendizado até o retardo mental mais severo.

A expressão do gene *FMR1* é dependente da situação fisiológica e do estágio de desenvolvimento do organismo. Como em outros genes, essa expressão é seletivamente ativada devido à ligação de determinados Fatores de Transcrição (FT) em seqüências específicas do DNA. A ligação de diferentes combinações de FT em diferentes locais da região reguladora pode modular a transcrição do gene em tecidos específicos, ou em diferentes estágios de desenvolvimento.

Como parte dos esforços para compreender a regulação da expressão do gene *FMR1*, este trabalho estudou a regulação da transcrição através do uso de ferramentas de genômica comparativa (*phylogenetic footprinting*) para identificar possíveis sítios de ligação de fatores de transcrição, em regiões específicas à montante do gene.

Através do uso das ferramentas de bioinformática foram encontradas mais de vinte regiões do DNA, conservadas ao longo da evolução, que apresentavam potencial para serem regiões reguladoras. Dentre estas, escolheu-se cinco que foram testadas experimentalmente através da técnica de *gel shift*. Os resultados mostraram que todas as cinco regiões escolhidas eram reconhecidas por proteínas nucleares, já que seus respectivos oligonucleotídeos sofriam um retardo na migração em gel poliacrilamida quando combinados com as proteínas. Duas destas seqüências de DNA foram utilizadas como ligantes de afinidade para a purificação de proteínas nucleares que reconheçam sítios naquelas seqüências. Utilizou-se precipitação com esferas metálicas cobertas pelos ligantes para selecionar estas proteínas. As proteínas recuperadas por afinidade foram separadas por eletroforese SDS-PAGE e submetidas à análise de espectrometria de massa para sua identificação.

Dentre as proteínas apontadas pela espectrometria de massa podemos citar as proteínas ativadoras da transcrição Pur- α e Pur- β , as ribonucleoproteínas heterogêneas A2/B1 (hnRNP A2/hnRNP B1), além de tubulina e actina, entre outras. De todas as proteínas identificadas estudou-se mais profundamente a existência da proteína Pur- α através de experimentos de imunodeteção utilizando um anticorpo anti-Pur- α , que confirmaram a presença desta proteína no conjunto de proteínas isoladas pelo sistema de purificação por afinidade às seqüências estudadas. Esta proteína vem sendo relacionada com o mRNA transcrito do gene *FMR1* como umas das proteínas que podem estar envolvidas no transporte do mRNA ou à regulação da tradução das proteínas FMRP.

Juntamente com os estudos de genômica comparativa que visam contribuir para a compreensão da regulação da expressão do gene *FMR1*. Buscou-se, também neste trabalho, incrementar os conhecimentos referentes às funções da proteína FMRP codificada pelo gene.

A função da proteína FMRP no organismo ainda não é totalmente conhecida, porém, evidências estruturais sugerem que ela esteja intimamente ligada ao transporte de mRNA do núcleo para o citoplasma e na sua condução até os ribossomos, onde faz parte dos complexos de ribonucleoproteínas (RNP).

A FMRP parece também atuar no controle da tradução de alguns mRNA, principalmente nos neurônios, ao nível dos dendritos e sinapses. Esta função coloca a FMRP como uma proteína chave no processo de maturação neuronal e do desenvolvimento cognitivo dos indivíduos.

Com a modelagem funcional das atividades da proteína FMRP através da técnica de Análise de Fluxos Metabólicos (AFM), baseando-se em dados encontrados na literatura mais recente, foi possível estabelecer um modelo que descreve as etapas de atuação da proteína, desde sua expressão até sua atuação como moduladora da tradução.

A análise de fluxos metabólicos é uma ferramenta para determinação dos fluxos (velocidades de reação) em vias metabólicas, onde fluxos intracelulares são calculados usando modelos estequiométricos das reações intracelulares e aplicação de balanços de massa.

Este trabalho propõe o uso desta técnica na análise dos fluxos dos complexos envolvidos na regulação da tradução de mRNA nos dendritos neuronais, onde a FMRP é uma proteína chave. Leva-se, no entanto, em consideração que, ao invés de reações enzimáticas de reagentes sendo transformados em produtos, têm-se uma série de interações bioquímicas entre as moléculas envolvidas (mRNA, proteínas, ribossomos, etc.) gerando complexos protéicos e riboprotéicos.

Acredita-se que a modelagem matemática da atuação da FMRP será útil, no futuro, na compreensão da sua função. O modelo proposto é puramente teórico e foi estabelecido através da utilização de dados retirados da literatura atual.

1.1. Objetivos

Este trabalho tem o objetivo de contribuir para uma melhor compreensão dos processos reguladores da expressão de genes eucariotos, principalmente o gene *FMR1*, responsável pela Síndrome do X-Frágil, e também incrementar os estudos sobre as funções desempenhadas pela proteína FMRP, codificada por este gene. Para atingir estes objetivos, este trabalho foi dividido em três etapas:

A primeira é referente ao estudo da regulação da transcrição do gene *FMR1* através de possíveis sítios de interação para fatores de transcrição em regiões específicas à montante do gene *FMR1*. Nesta etapa foi utilizada a ferramenta de genômica comparativa, chamada *phylogenetic footprinting*.

A segunda etapa consistiu da confirmação e caracterização experimental de alguns sítios de ligação encontrados, incluindo a amplificação e a purificação dos oligonucleotídeos referentes às seqüências de DNA alvo, a purificação das proteínas nucleares e citoplasmáticas de tecidos neuronais, o estudo da interação DNA-proteínas de transcrição através da técnica de *gel shift* e a identificação destas proteínas.

A terceira etapa foi a determinação de um modelo matemático funcional teórico da FMRP que descreve as etapas de interação da proteína, desde sua expressão até sua atuação como moduladora da tradução de determinados mRNA nos dendritos e a modelagem da sua atuação através da Análise de Fluxos Metabólicos, com dados extraídos da literatura.

2. REVISÃO BIBLIOGRÁFICA

2.1. A Síndrome do X-Frágil e sua caracterização

A Síndrome do X-Frágil (SXF) é a forma de retardo mental herdado mais comum encontrada, afetando 1 entre 4000 homens e 1 entre 8000 mulheres (Berry-Kravis *et al.*, 2002; Jin e Warren, 2003). O banco de dados de doenças genéticas humanas herdadas (OMIN) tem catalogado mais de 1500 entradas contendo o termo “*mental retardation*”, as quais normalmente estão associadas com déficit no quociente de inteligência (QI). Segundo pesquisas americanas, de 2 a 3% da população em geral apresenta um QI abaixo de 70 e 0,3% pode ser considerado severamente retardada (Castellvi-Bel e Mila, 2001). Indivíduos com SXF possuem um comportamento mental variável, que pode ir desde uma dificuldade no aprendizado até o retardo mental mais severo. Algumas características comportamentais também estão muito presentes nos afetados pela síndrome, como atraso no desenvolvimento da fala, hiperatividade em meninos, timidez em meninas, dificuldade de atenção, comportamento repetitivo, maneirismos de mão e dificuldade de interação social. Fisicamente os portadores podem ser caracterizados por apresentarem olhos grandes, face alongada, testa proeminente, macroorquidismo, entre outras características fenotípicas (Hagerman e Hagerman, 2001).

A SXF resulta da amplificação da repetição dos trinucleotídeos CGG, e sua conseqüente metilação, localizados na região transcrita mas não traduzida 5' (5' UTR) do gene *FMR1* que se encontra na região Xq27.3 do cromossomo X humano. O número de repetições CGG é polimórfico na maioria da população, encontrando-se entre 6 a 40 com uma média por volta de 30. O nível de mutação deste gene caracteriza três diferentes grupos de afetados pela síndrome. Os mais comumente descritos na literatura são os pré-mutados, que apresentam de 60 a 200 repetições de CGG e os indivíduos completamente mutados ou afetados, que apresentam mais de 200 repetições de CGG (Hagerman e Hagerman, 2002a), e um grupo que tem sido pouco estudado, chamado “zona-gray”, que justamente leva este nome por se encontrar numa faixa muito próxima do normal em termos de mutação (de 40 a 60 repetições de CGG)(Loesch *et al.*, 2007). A falta de caracterização e diagnóstico adequados deste grupo dificulta a identificação, sobretudo de crianças “zona-gray” que, normalmente apresentam problemas de cognição e linguagem em fase escolar. Esses indivíduos quando não propriamente diagnosticados são muitas vezes submetidos a tratamentos inadequados, agravando ainda mais o seu quadro

clínico. Os estudos mais freqüentes são realizados com os pacientes totalmente mutados e pré-mutados, e tais estudos vêm mostrando que os pacientes pré-mutados apresentam anomalias típicas deste tipo de mutação (Hagerman e Hagerman, 2001; Greco *et al.*, 2002; Hagerman e Hagerman, 2002a; 2002b; Primerano *et al.*, 2002). Os pacientes totalmente mutados apresentam um déficit de transcrição de praticamente 100% do gene *FMR1*, causado pela expansão das repetições CGG, da região 5' promotora, seguida de metilação das citosinas. Esta expansão silencia o gene de forma irreversível nos pacientes. Já os pacientes pré-mutados apresentam um aumento no nível de transcrição em até cinco vezes o número de transcritos, o que sugere que a expansão em até 200 repetições potencialize o processo de transcrição do gene, porém este aumento não se reflete na tradução destes mRNA, o que implica que deve haver algum outro meio de controle da expressão para a proteína FMRP (Tassone *et al.*, 2000; Garber *et al.*, 2006).

Estudos de transmissão familiar (Nolin *et al.*, 1996) mostram que mulheres portadoras repassam a sua mutação aos filhos, e possivelmente às filhas, sendo que esta mutação pode se manter estável ou evoluir (expandir) para mais de 200 repetições de CGG o que caracteriza um indivíduo totalmente mutado. Da mesma forma homens portadores passam sua mutação para suas filhas que podem repassar para seus filhos, e assim a Síndrome necessita ser estudada em todos os membros da família. Devido aos estudos mais recentes sobre os indivíduos pré-mutados pode-se agora dar suporte a estes indivíduos quanto aos fenótipos que eles apresentam e que até pouco tempo atrás não estavam relacionados à Síndrome do X-Frágil (Greco *et al.*, 2002; Hagerman e Hagerman, 2002a).

Com base neste quadro geral, nos últimos anos, a atenção de alguns pesquisadores tem sido direcionada a entender não apenas a SXF, a partir do estudo de caso de pacientes totalmente mutados, mas também dos pré-mutados e mais recentemente dos indivíduos zona gray. Estes últimos apresentam baixo número de repetições CGG e há pouco tempo eram considerados pacientes normais; porém pelo fato de apresentarem fenótipos comportamentais muito semelhantes aos portadores da SXF, revelam-se uma importante fonte de informações para a sua compreensão.

Visando o entendimento dos diferentes tipos de mutações da Síndrome do X-Frágil, deve-se estender o diagnóstico e a caracterização dos pacientes. Além da determinação do número de repetições CGG, outra medida que já é empregada com esta finalidade é a quantificação do percentual de expressão da proteína codificada

pelo gene *FMR1*, chamada FMRP, nos linfócitos dos pacientes. Esta quantificação é realizada com técnicas de imunocitoquímica (Willemsen *et al.*, 1997), utilizando-se anticorpos monoclonais anti-FMRP, e nos permite determinar o percentual de linfócitos que expressam a proteína. Com esta técnica pode-se observar que os indivíduos totalmente mutados (CGG)_{n>200} apresentam um nível de expressão da proteína muito baixo ou nulo (0 a 10%) enquanto os pacientes pré-mutados, zona gray e normais apresentam índices que podem variar de 60 a 100% (Tassone *et al.*, 1999). Estes resultados evidenciam que a ausência de proteína é um fator significativo na avaliação dos indivíduos totalmente mutados, que na sua maioria apresentam retardo mental muito evidente. Porém, não deixam muito claro qual o grau de importância da sua redução nos outros indivíduos. Esta avaliação também não é clara nas pacientes mulheres, que mesmo totalmente mutadas, podem apresentar um nível de expressão da proteína de até 70% das células (Tassone *et al.*, 1999), já que as mulheres possuem dois cromossomos X e a mutação normalmente está presente em apenas um deles, permitindo que um ou outro esteja ativo, expressando a proteína em algumas células, mas não em outras (Fenômeno conhecido como mosaicismo). Estes diferentes grupos de indivíduos que são afetados de formas diferentes pela síndrome instigam os pesquisadores a dedicar-se para a melhor compreensão tanto dos mecanismos de regulação da expressão do gene *FMR1* quanto dos mecanismos biológicos nos quais a proteína FMRP está envolvida.

2.2. FMRP e suas proteínas homólogas

A proteína FMRP (Fragile-X Mental Retardation Protein) é codificada pelo gene *FMR1* que possui 38 kb e está localizado na região Xq27.3 do cromossomo X. O gene *FMR1* é dividido em 17 éxons que quando transcritos podem gerar mRNA diferentes, devido a *splicings* alternativos do mesmo gene (Huang *et al.*, 1996). Estudos mostram que o mRNA da proteína FMRP é sujeito a diferentes *splicings* gerando, por consequência, diferentes isoformas, de 70 a 80 kDa, que são conhecidas coletivamente como proteínas FMRP, que muito provavelmente possuem funções específicas em tecidos diferentes (Huang *et al.*, 1996).

A FMRP contém domínios específicos de ligação com RNA, chamados domínios KH, que são regiões presentes numa variedade de proteínas que têm a capacidade de se ligar a ácidos nucleicos, principalmente encontrados nas proteínas que formam o complexo de ribonucleoproteínas associadas a RNA mensageiros. A FMRP possui ainda o domínio RGG que também está associado com

a capacidade de ligação com RNA (Bardoni *et al.*, 2001; Oostra, 2002; Schaeffer *et al.*, 2003). Outras seqüências existentes na estrutura primária da proteína FMRP são as regiões de sinalização e localização nuclear (NSL) e a região de exportação nuclear (NES) (Eberhart *et al.*, 1996), tipicamente encontradas em proteínas que têm a função de auxiliar no transporte de outras proteínas ou RNA no núcleo e sua exportação para o citoplasma.

No genoma humano bem como no de outros mamíferos existem duas cópias homólogas do gene *FMR1*, os genes parólogos *FXR1* e *FXR2*. A proteína FXR1 é codificada pelo gene *FXR1* localizado no cromossomo 3 na região 3q28. Esta proteína possui três isoformas conhecidas (a,b,c) resultantes de *splicings* alternativos do mRNA transcrito. A proteína FXR2 é muito similar à proteína FMRP, apresentando 60% de identidade e é codificada pelo gene *FXR2*, localizado no cromossomo 17 na região 17p13.1. A Figura 2.1 mostra o alinhamento entre as seqüências das proteínas FMRP, FXR1 e FXR2 obtido a partir do programa ClustalW (Thompson *et al.*, 1994).

Estas proteínas homólogas apresentam uma estrutura muito similar à FMRP. Elas possuem similaridade entre 73 e 90% na primeira metade da proteína, e também os dois domínios KH e o RGG Box, como pode ser observado na Figura 2.1. Os domínios de localização (NLS) e exportação do núcleo (NES) também são similares à da FMRP (Siomi *et al.*, 1995; Zhang *et al.*, 1995; Bardoni *et al.*, 2001; Kirkpatrick *et al.*, 2001).

Estudos *in vitro* e *in vivo* revelaram que as proteínas FMRP, FXR1 e FXR2 interagem formando homo e hetero-dímeros, o que sugere que essa oligomerização pode estar relacionada com alguma função biológica (Zhang *et al.*, 1995). A dimerização parece ser possível devido a um domínio identificado na região N-terminal das proteínas, que também pode estar associado com a ligação ao RNA (Adinolfi *et al.*, 2003).

A FXR1 e a FXR2, assim como a FMRP, possuem a capacidade de se ligarem a RNA, e foram encontradas associadas a partículas de ribonucleoproteínas (mRNP) (Ceman *et al.*, 1999). No entanto, estudos mostram que algumas funções atribuídas à FMRP, como a regulação da tradução de determinados RNA, localmente nos dendritos (Laggerbauer *et al.*, 2001), não são realizadas pelas proteínas homólogas, o que sugere que provavelmente estas proteínas tenham uma função diferenciada.


```

Isoform_a_FXR1 -----MAELTVEVRGSGNGAFYKGF IKDVHEDSLTVVFENNWPQERQVFPFNEVRLP 50
Isoform_c_FXR1 -----
Isoform_b_FXR1 -----MAELTVEVRGSGNGAFYKGF IKDVHEDSLTVVFENNWPQERQVFPFNEVRLP 50
FXR2 MGGLASGGDVEPLPVEVRGSGNGAFYKGFVKDVHEDSVT IFFENNWQSERQIFPGDVRLP 60
FMRP -----MEELVVEVRGSGNGAFYKAFVKDVHEDSITVAFENNWPQDRQIFPHDVRFP 50

Isoform_a_FXR1 PPPDIKKEISEGDEVEVYSRANDQPCGWLLAKVRMMKGEFYVIEYAACDATYNEIVTFE 110
Isoform_c_FXR1 -----MMKGEFYVIEYAACDATYNEIVTFE 25
Isoform_b_FXR1 PPPDIKKEISEGDEVEVYSRANDQPCGWLLAKVRMMKGEFYVIEYAACDATYNEIVTFE 110
FXR2 PPADYNKEITEGDEVEVYSRANEQPCGWLLARVRMMKGEFYVIEYAACDATYNEIVTLE 120
FMRP PPVGYNKDINESDEVEVYSRANEKPCGWLLAKVRMIKGEFYVIEYAACDATYNEIVTLE 110
*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*

Isoform_a_FXR1 RLRPVNQKNTVKKNTFFKCTVDVPEDLREACANENAHKDFKKAVGACRIFYHPETTQLMI 170
Isoform_c_FXR1 RLRPVNQKNTVKKNTFFKCTVDVPEDLREACANENAHKDFKKAVGACRIFYHPETTQLMI 85
Isoform_b_FXR1 RLRPVNQKNTVKKNTFFKCTVDVPEDLREACANENAHKDFKKAVGACRIFYHPETTQLMI 170
FXR2 RLRPVNPNPLATKGSFFKVTMAVPEDLREACSNENNVHKEFKKALGANCIFLNITNSGLFI 180
FMRP RLRSVNPNKPKTDFHKIKLDVPEDLRQMCAGEAAHKDFKKAVGAFSVTYDPENYQLVI 170
*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*

Isoform_a_FXR1 LSASEATVKRVNII LSDMHLRSIRTKLMLMSRNEEATKHLECTKQLAAAFHEEFVREDIM 230
Isoform_c_FXR1 LSASEATVKRVNII LSDMHLRSIRTKLMLMSRNEEATKHLECTKQLAAAFHEEFVREDIM 145
Isoform_b_FXR1 LSASEATVKRVNII LSDMHLRSIRTKLMLMSRNEEATKHLECTKQLAAAFHEEFVREDIM 230
FXR2 LSTTEAPVKRASLLGDMHFRSLRKLMLMSRNEEATKHLECTSKQLAAAFQEEFTVREDIM 240
FMRP LSINEVTSKRAHMLIDMHFRSLRKLMLMSRNEEATKHLECTSKQLAAAFQEEFTVREDIM 230
*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*

KH
Isoform_a_FXR1 GLAIGTHGSNIQQAARKVPGVTAI ELD E D T G T F R I Y G S A D A V K K A R G F L E F V E D F I Q V P R 290
Isoform_c_FXR1 GLAIGTHGSNIQQAARKVPGVTAI ELD E D T G T F R I Y G S A D A V K K A R G F L E F V E D F I Q V P R 205
Isoform_b_FXR1 GLAIGTHGSNIQQAARKVPGVTAI ELD E D T G T F R I Y G S A D A V K K A R G F L E F V E D F I Q V P R 290
FXR2 GLAIGTHGANIQQAARKVPGVTAI ELD E E T C T F R I Y G T P E A C R Q A R S Y L E F S E D S V Q V P R 300
FMRP GLAIGTHGANIQQAARKVPGVTAI D L D E D T C T F H I Y G E D Q D A V K K A R S F L E F A E D V I Q V P R 290
*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*

KH-1
Isoform_a_FXR1 NLVGKVI G K N G K V I Q E I V D K S G V V R V R I E G D N E N K L P R E D ----- 330
Isoform_c_FXR1 NLVGKVI G K N G K V I Q E I V D K S G V V R V R I E G D N E N K L P R E D ----- 245
Isoform_b_FXR1 NLVGKVI G K N G K V I Q E I V D K S G V V R V R I E G D N E N K L P R E D ----- 330
FXR2 NLVGKVI G K N G K V I Q E I V D K S G V V R V R V E G D N D K N P R E E ----- 340
FMRP NLVGKVI G K N G K L I Q E I V D K S G V V R V R I E A E N K N V Q E E I M P P N S L P S N N S R V G P N A P 350
*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*

Isoform_a_FXR1 -----GMVFFVFGTKESI 344
Isoform_c_FXR1 -----GMVFFVFGTKESI 259
Isoform_b_FXR1 -----GMVFFVFGTKESI 344
FXR2 -----GMVFFVFGTREN I 354
FMRP EEK K H L D I K E N S T H F S Q P N S T K V Q R V L V A S S V V A G E S Q K P E L K A W Q G M V P F V G T K D S I 410
*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*

Isoform_a_FXR1 GNVQVLL E Y H I A Y L K E V E Q L R M E R L Q I D E Q L R Q I G S ----- 380
Isoform_c_FXR1 GNVQVLL E Y H I A Y L K E V E Q L R M E R L Q I D E Q L R Q I G S ----- 295
Isoform_b_FXR1 GNVQVLL E Y H I A Y L K E V E Q L R M E R L Q I D E Q L R Q I G S ----- 380
FXR2 S N A Q A L L E Y H L S Y L Q E V E Q L R L R L Q I D E Q L R Q I G L G F R P P G S G R G S G G S D K A G Y S T D E S 414
FMRP A N A T V L L D Y H L N Y L K E V D Q L R L R L Q I D E Q L R Q I G A S S R P P P N R T D K E K S Y V T D D Q G M G 470
*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*

Isoform_a_FXR1 -----RSYSGRGRGRRGPNY T S G Y G T N S E L S N P S E T E S E R K D E L S D W S L A - G E 427
Isoform_c_FXR1 -----RSYSGRGRGRRGPNY T S G Y G T N S E L S N P S E T E S E R K D E L S D W S L A - G E 342
Isoform_b_FXR1 -----RSYSGRGRGRRGPNY T S G Y G T N S E L S N P S E T E S E R K D E L S D W S L A - G E 427
FXR2 S S S L H A T R T Y G G S Y G G R G R G R R T G - - G P A Y G P S S D V T A S E T E S E K R E E P - N R A G P - G D 470
FMRP R G S - - - - - R P Y R N R G H G R R G P - - G Y T S G T N S E A S N A S E T E S D H R D E L S D W S L A P T E 519
*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*

Isoform_a_FXR1 DDRDSRHQRDSRRRPPGGRGRSVSGGRGRGGPRGGKSSISSVLKDPDSNPYSLLDNTESDQ 487
Isoform_c_FXR1 DDRDSRHQRDSRRRPPGGRGRSVSGGRGRGGPRGGKSSISSVLKDPDSNPYSLLDNTESDQ 402
Isoform_b_FXR1 DDRDSRHQRDSRRRPPGGRGRSVSGGRGRGGPRGGKSSISSVLKDPDSNPYSLLDNTESDQ 487
FXR2 R D P P T R G E S R R R P T G C R G R G P P A P - R P T S R Y N S S I S S V L K D P D S N P Y S L L D T S E P E P 529
FMRP E E R E S F L R R G D G R R R G G G R G - Q G R G R G G G - - - - - F K G N D - - - - - D H S R T D - 560
*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*

RGG box
Isoform_a_FXR1 T A D T D A S E S H H S T N R R R R S R R R R T D E D A V L M D G M T E S D T A S V N E N G L V T V A D Y I S R A E S Q 547
Isoform_c_FXR1 T A D T D A S E S H H S T N R R R R S R R R R T D E D A V L M D G M T E S D T A S V N E N G L V T V A D Y I S R A E S Q 462
Isoform_b_FXR1 T A D T D A S E S H H S T N R R R R S R R R R T D E D A V L M D G M T E S D T A S V N E N G L G K R C D - - - - - 539
FXR2 P V D S E P G E P P A S A R R R R S R R R R T D E D R T V M D G G L E S D G P N M T E N G L E D E S R P Q R R N R S R 589
FMRP - - - - - N R P R N P R E A K R T T D G S L Q I R V D C N N E R - S V H T K T L Q N T S E G S R L R T G 608
*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*

Isoform_a_FXR1 SRQR--NLPRETLAKNKK-----EMAKDVIEEHGHPSEKAI NGPTSASGDDISKLQRT 597
Isoform_c_FXR1 SRQR--NLPRETLAKNKK-----EMAKDVIEEHGHPSEKAI NGPTSASGDDISKLQRT 512
Isoform_b_FXR1 -----
FXR2 RRRNRGNRTDGSISGDRQPVTVADYISRAESQSRQSAPLERTKPS E D S L S G Q K G S V S K L 649
FMRP KDRN-----QKKEKPDSDVGGQQ----- 625

Isoform_a_FXR1 PGEEKINTLKEENTQEA AVLNGVS 621
Isoform_c_FXR1 PGEEKINTLKEENTQEA AVLNGVS 536
Isoform_b_FXR1 -----
FXR2 PKGFSENGELSAPLELGS MVNGVS 673
FMRP -----PLVNGVP 632

```

Figura 2.1: Alinhamento entre as seqüências das proteínas FMRP, FXR1 e FXR2 humanas utilizando programa ClustalW. As seqüências grifadas correspondem a alguns domínios também conservados. (*) todas as proteínas alinhadas possuem o mesmo aminoácido na posição indicada; (;) substituição de algum aminoácido por outro do mesmo grupo; (.) substituição por aminoácido de outro grupo.

Estudos sobre os níveis de expressão das proteínas FXR1, FXR2 e FMRP mostraram que em cerebelos humanos adultos normais há um alto nível de expressão das três proteínas, que se localizam principalmente no citoplasma. Já nas células de córtex humano de adultos normais as três proteínas não estão apenas localizadas no citoplasma, mas sim fortemente presentes nos dendritos. Como esperado, em células de pacientes afetados pela síndrome não foi observada expressão da proteína FMRP, porém os níveis de FXR1 e FXR2 permaneceram inalterados (Tamanini *et al.*, 1997). Neste mesmo estudo observou-se que os níveis de FXR1, FXR2 e FMRP em células cerebrais de fetos humanos são comparáveis com os resultados para células de indivíduos adultos, porém a proteína FXR1 foi identificada em concentrações altas também nos núcleos celulares de alguns neurônios.

Diversos outros organismos possuem uma ou mais proteínas homólogas à FMRP humana. Dentre os organismos que possuem o seu genoma seqüenciado, que estão depositados no Banco de Dados do NCBI (GenBank), encontramos homólogos ao gene *FMR1* em *Pan troglodites* (chimpanzé), *Canis lupus familiaris* (cachorro), *Bos taurus* (boi), *Mus musculus* (camundongo), *Rattus norvegicus* (rato), *Gallus gallus* (galinha) e *Danio rerio* (peixe Paulistinha). No banco de dados de seqüências de transcritos (UniGene), encontramos homólogos em *Equus caballus* (cavalo), *Sus scrofa* (porco), *Xenopus tropicalis* (rã), *Xenopus laevis* (rã africana), *Aplysia californica* (lesma aquática), *Drosophila melanogaster* (mosca das frutas), *Bombyx mori* (mariposa), entre outros. O conhecimento destas seqüências homólogas à FMRP em outros organismos é muito importante para os estudos genômicos.

A existência de proteínas homólogas à FMRP em ratos ou camundongos, por exemplo, abre uma oportunidade para a criação de animais modelos para a Síndrome do X-Frágil, ou seja, animais manipulados geneticamente, que não expressem o gene homólogo ao *FMR1*. Isso possibilita o estudo e caracterização desta doença sem a necessidade de envolver humanos.

Nesta área existem já disponíveis alguns animais *knockout* para o gene *FMR1* e que apresentam dificuldades de aprendizado, hiperatividade e macroorquidismo, características que mimetizam a síndrome em humanos (Bakker *et al.*, 1994; Slegtenhorst-Eegdeman *et al.*, 1998; Paradee *et al.*, 1999; D'agata *et al.*, 2002; Ishizuka *et al.*, 2002; Pan *et al.*, 2004; Van 'T Padje *et al.*, 2005; Hanson e Madison, 2007; Hayashi *et al.*, 2007; Price *et al.*, 2007; Centonze *et al.*, 2008).

Camundongos, moscas das frutas e peixes zebra têm sido utilizados como animais modelo para o estudo da Síndrome do X-Frágil. Estes estudos buscam compreender os fundamentos genéticos e bioquímicos da síndrome bem como os padrões comportamentais dos animais afetados e fazer uma correlação com o que é observado em humanos. Outros estudos visam buscar possíveis vias de cura ou amenização dos sintomas da síndrome, como o caso apresentado por Hayashi e colaboradores (2007), que através do uso de camundongos modelo observaram que uma perda na expressão da proteína PAK (p21-activated kinase) leva a amenização do fenótipo característico ao da Síndrome do X-Frágil, em camundongos *FMR1 knockout*. Estes autores utilizaram camundongos *FMR1 knockout* e cruzaram com outros com atividade da PAK inibida e comparam os dendritos e estruturas cerebrais destes animais com animais dos dois genótipos. Os autores demonstraram que existe uma interação direta entre a PAK e a FMRP e sugerem que a via de sinalização da PAK, incluindo a sua atividade catalítica, pode ser um novo ponto de intervenção para a Síndrome do X-Frágil.

A comparação dos comportamentos e mecanismos biológicos entre humanos e alguns animais tem sido utilizada nas mais diversas áreas da biologia, como anatomia, bioquímica, farmacologia, imunologia e biologia celular e tem se mostrado fundamental na compreensão de diversos fenômenos. Da mesma forma a genômica, uma área interdisciplinar nova e crescente, pode utilizar-se de ferramentas de comparação para esclarecer muitas questões ligadas à função gênica e à regulação da expressão de genes.

O alinhamento entre seqüências codificantes e não codificantes de genomas indicam a existência de genes homólogos e também regiões reguladoras. Estes alinhamentos podem revelar regiões conservadas ao longo da evolução que sejam de grande importância para a regulação da expressão gênica. O alinhamento entre as proteínas codificadas por genes homólogos pode revelar a existência de domínios relacionados com funções específicas destas proteínas.

2.3. Regulação da expressão gênica

2.3.1. Genômica comparativa no estudo da regulação

O crescente número de seqüências genômicas disponíveis (82.853.685 seqüências depositadas no GenBank até fevereiro de 2008) e principalmente a corrida em prol do seqüenciamento de genomas completos de organismos procariotos (1.803 genomas até junho de 2008) e eucariotos (344 genomas), levam a

um aumento do interesse e da necessidade de estudos relativos à compreensão destas seqüências, sejam elas codificantes ou não-codificantes.

Nas mais diversas áreas, como agricultura, medicina e engenharia, a genômica vem apresentando explicações e alternativas para diversos fenótipos apresentados por diferentes organismos. Nestes campos da pesquisa há muito se comparam fenotipicamente as espécies visando à compreensão de muitos males. Agora, com o advento das técnicas de comparação de seqüências genômicas, pode-se levar a comparação entre as espécies para o nível molecular.

A genômica comparativa contribui para o direcionamento de muitas questões fundamentais da genômica funcional, que incluem o número de genes codificados em um organismo, suas funções e regras de expressão e a análise de vias metabólicas e reguladoras.

As comparações entre seqüências genômicas têm mostrado um alto nível de conservação de genes e proteínas entre as espécies, embora muitas vezes entre espécies bem distantes filogeneticamente. Esse alto nível de conservação tem sido foco de estudos visando tanto a análise de regiões codificantes, como de regiões não-codificantes (regiões reguladoras), que são a chave para o entendimento do processo de controle da expressão dos genes (Dubchak e Frazer, 2003; Frazer *et al.*, 2003; Pennacchio, 2003; Thomas *et al.*, 2003; Ahituv *et al.*, 2004; Lee *et al.*, 2005; Ganapathi *et al.*, 2007).

2.3.2. Análise gênica dos elementos de regulação da transcrição

A expressão dos genes é controlada pela ligação de determinados **fatores de transcrição** (FT) em seqüências específicas e não codificantes do DNA. A expressão destes fatores, e conseqüentemente a sua atuação, é dependente da situação fisiológica e do estágio de desenvolvimento do organismo. Assim, alguns genes são expressos apenas nos momentos apropriados.

Para a compreensão completa da função do gene é necessária a identificação dos fatores de transcrição que regulam a sua expressão, e também regiões do DNA, nas quais estes fatores se ligam, chamadas de **sítios de ligação de fatores de transcrição** (TFBS - Transcription Factor Binding Sites).

Diferentes combinações de fatores de transcrição ligados a diferentes locais da região reguladora podem modular a transcrição do gene em tecidos específicos, ou em determinados estágios de desenvolvimento. Estas proteínas regulam a transcrição através de sua ligação no seu sítio específico, também chamado de

enhancer, e subsequente ligação à **região promotora**, favorecendo o acoplamento da RNA polimerase e sua posterior transcrição. A Figura 2.2 apresenta um esquema ilustrativo da atuação dos fatores de transcrição na regulação da expressão do gene.

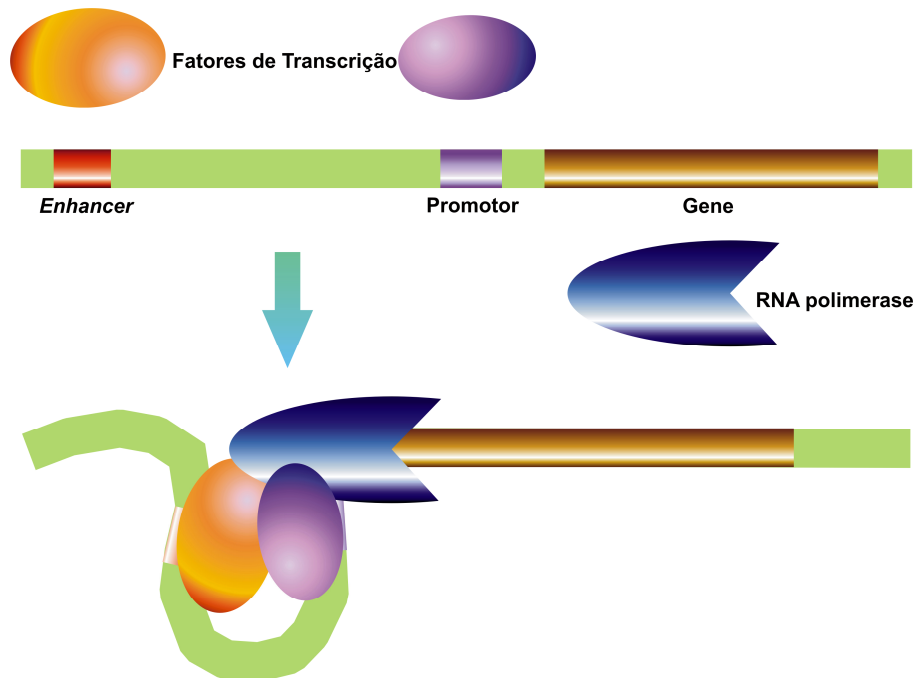


Figura 2.2: Esquema ilustrativo da atuação dos fatores de transcrição na regulação da expressão do gene.

As regiões reguladoras dos genes nos eucariotos podem estar localizadas próximas aos genes, nas suas extremidades 5' e 3', no interior do gene, em regiões intrônicas, ou ainda em regiões muito distantes do gene há milhares de pares de bases de distância à montante do início da transcrição.

Tradicionalmente, a especificidade de ligação dos FT em sítios do DNA é determinada experimentalmente com técnicas como a *DNase I Footprinting* (Weisbrod e Weintraub, 1979) e *Electrophoretic Mobility Shift Assay* (EMSA), também conhecida como gel shift (Carey, 1991; Hellman e Fried, 2007). Estas técnicas permitem a identificação e caracterização *in vitro* dos TFBS. No entanto, estudos de grandes porções dos genomas são ainda complicados e dispendiosos; por esse motivo, técnicas alternativas, baseadas em programas computacionais, têm sido desenvolvidas visando identificar regiões reguladoras dos genes de genomas inteiros.

Bases de dados como TRANSFAC, COMPEL e TRRD (Kolchanov *et al.*, 2000; Kel-Margoulis *et al.*, 2002; Fogel *et al.*, 2005) que contêm um grande número de

TFBS já identificados e caracterizados experimentalmente, quando associadas aos programas computacionais desenvolvidos para comparação gênica tornam possível prever elementos de regulação existentes em regiões conservadas, em grandes seqüências de DNA, entre duas ou mais espécies (Dermitzakis e Clark, 2002; Levy e Hannenhalli, 2002; Dubchak e Frazer, 2003).

Os estudos de genômica comparativa utilizando-se genomas de organismos já seqüenciados como o *Mus musculus* e a *Drosophila melanogaster*, entre outros; podem revelar dados importantes sobre os diversos genes e sobre as regiões conservadas ao longo da evolução onde se encontram possíveis sítios de interação para fatores de transcrição, que regulam a expressão desses genes (Berman *et al.*, 2002; Lenhard *et al.*, 2003). Vários autores têm utilizado genômica comparativa na identificação de regiões não codificantes e codificantes do DNA de várias espécies, conservadas ao longo da evolução (Nandi *et al.*, 2002; Couronne *et al.*, 2003; Lenhard *et al.*, 2003; Pennacchio, 2003; Ahituv *et al.*, 2004; Nobrega e Pennacchio, 2004). Uma das ferramentas para identificação de regiões reguladoras é o *phylogenetic footprinting* (Blanchette e Tompa, 2002), um termo inspirado na técnica de *DNAse footprinting*.

Phylogenetic footprinting é uma ferramenta que faz uso do alinhamento de seqüências de DNA de várias espécies para identificar regiões conservadas e que contenham pequenas seqüências (6 a 12 pb) correspondentes àquelas seqüências de consenso já conhecidas como sítios de ligação para determinados fatores de transcrição.

Alguns programas já foram desenvolvidos especialmente para a identificação de TFBS através desta técnica e estão disponíveis na Web, como por exemplo, o programa VISTA (<http://genome.lbl.gov/vista/index.shtml>) (Frazer *et al.*, 2004), com a ferramenta rVISTA (Loots *et al.*, 2002; Loots e Ovcharenko, 2004), o programa ConSite (<http://phylofoot.org>) (Lenhard *et al.*, 2003), entre outros (Blanchette e Tompa, 2002).

O programa Vista, que é desenvolvido e hospedado pela Genomics Division of Lawrence Berkeley National Laboratory (Berkeley – EUA), é composto de uma série de ferramentas computacionais para genômica comparativa, que permite aos usuários o alinhamento global de grandes seqüências genômicas de vertebrados e outros grupos de organismos.

Através destes alinhamentos é possível posteriormente realizar várias análises comparativas entre estas seqüências. O programa disponibiliza um navegador (*browser*) que permite a navegação do usuário através de todo o

alinhamento que é apresentado numa forma gráfica de fácil compreensão. A Figura 2.3 apresenta o navegador genômico do Vista, com um alinhamento entre as seqüências codificantes e não codificantes do gene *FMR1* humano e seu homólogo nos ratos.

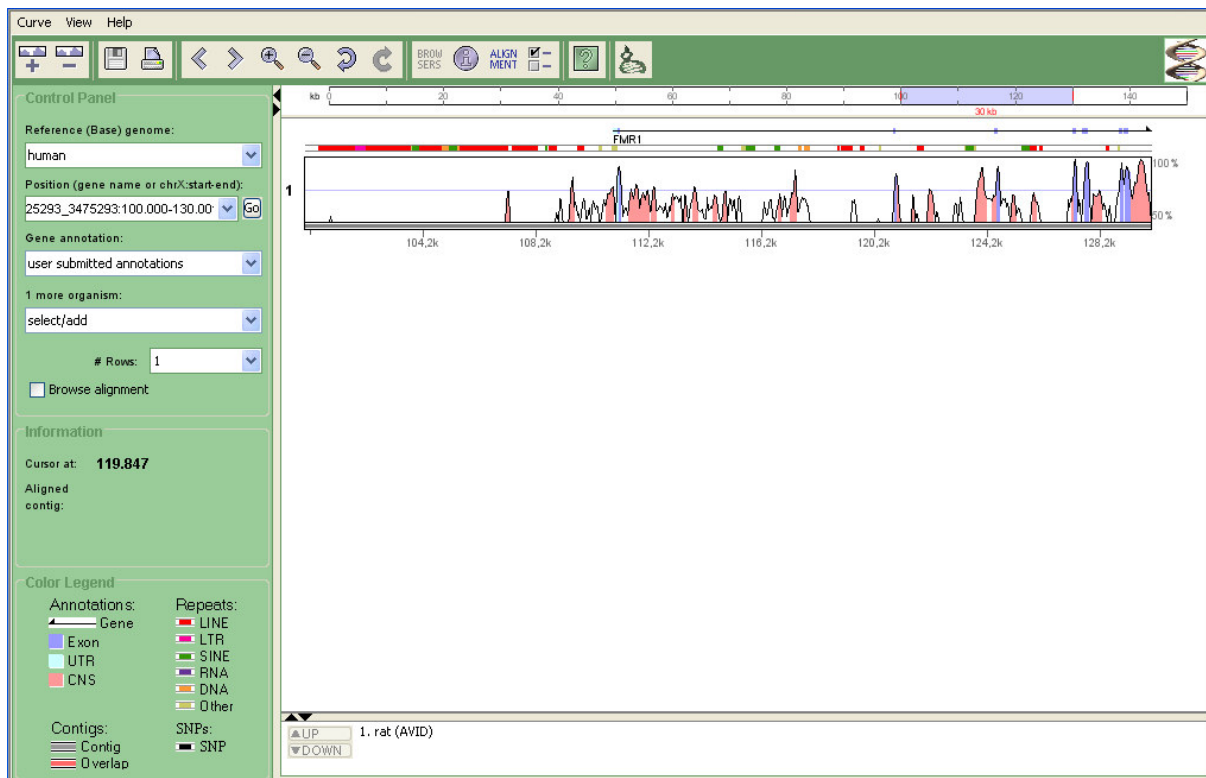


Figura 2.3: Navegador do Programa Vista de alinhamento global e genômica comparativa. Os gráficos de visualização mostram as seqüências conservadas entre homens e ratos, para a região imediatamente à montante do gene *FMR1* e parte de sua região codificante, baseado no alinhamento global realizado pelo programa Vista. O nível de conservação (eixo y) é apresentado em função da seqüência humana (eixo x). Regiões conservadas com mais de 70%/100 pb são destacadas sob a curva, com rosa indicando regiões conservadas não codificantes e azul indicando regiões conservadas codificantes (éxons).

Dentre as ferramentas disponíveis no navegador do programa Vista, encontra-se o servidor rVista (regulatory Vista), que combina a pesquisa no maior banco de dados de fatores de transcrição – TRANSFAC™ Professional, da Biobase (Wolfenbüttel, Germany) – com a análise comparativa entre as seqüências. Através deste processo é possível identificar, nas regiões não codificantes, seqüências candidatas a regiões de regulação da transcrição.

A análise das regiões não codificantes buscando apenas a presença de sítios de interação para fatores de transcrição conhecidos, produz uma imensa quantidade de falsos positivos. A combinação deste procedimento com a análise

comparativa obtida com os alinhamentos, diminui muito estes falsos resultados (Loots *et al.*, 2002).

A análise de possíveis sítios de ligação de FT realizada pelo rVista, segue quatro passos: (i) localização de possíveis TFBS presentes em cada uma das seqüências submetidas, utilizando as matrizes disponíveis na base de dados TRANSFAC; (ii) identificação dos TFBS localizados nas regiões pareadas pelo alinhamento; (iii) seleção dos TFBS localizados em regiões de alta conservação nas seqüências de DNA estudadas e (iv) disponibilização gráfica dos TFBS e das regiões de alta conservação entre as seqüências de DNA (Loots e Ovcharenko, 2004). A Figura 2.4 apresenta um resultado típico do programa rVista, para umas das regiões conservadas entre humanos e ratos à montante do gene *FMR1*.

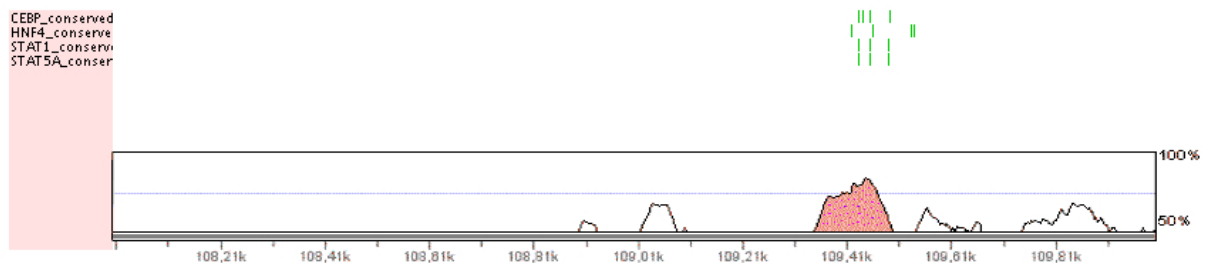


Figura 2.4: Resultado típico do programa rVista para uma região não codificante, conservada entre duas espécies. São apresentados apenas os possíveis sítios de ligação que se encontram nesta região conservada.

Utilizando o programa Vista pode-se comparar a seqüência reguladora de genes com seqüências homólogas em diferentes espécies. Os resultados indicarão as regiões conservadas entre as seqüências homólogas destas espécies e a seqüência de interesse. A partir destes dados obtidos computacionalmente projeta-se os experimentos *in vitro* e/ou *in vivo*, que confirmem as evidências computacionais. Esse procedimento diminui muito as chances de execução de experimentos desnecessários, uma vez que já são conhecidas previamente as seqüências de deverão ser experimentadas no laboratório.

O experimento clássico para estudos de interação entre ácidos nucléicos e proteínas é o procedimento de *Electrophoretic Mobility Shift Assay* (EMSA) ou *Gel Shift*. Com este procedimento pode-se determinar se uma proteína ou uma mistura de proteínas é capaz ou não de se ligar a uma dada seqüência de DNA ou RNA (Woo *et al.*, 2002; Hellman e Fried, 2007). Em alguns casos pode-se inclusive indicar se mais de uma proteína estão envolvidas no complexo de ligação. *Gel shift* é uma técnica baseada em separação por eletroforese onde uma mistura de proteínas e

DNA é aplicada em um gel de poliacrilamida ou agarose e como resultado analisa-se a velocidade de migração dos complexos formados (Carey, 1991). Parte-se do princípio de que a velocidade com que diferentes moléculas migram dentro do gel depende da sua massa e carga. Sendo assim, se houver a formação de complexos DNA-proteína, estes devem migrar mais lentamente do que o segmento de DNA livre. O controle do experimento se dá correndo paralelo aos exemplos testados uma alíquota do DNA alvo livre de proteína (controle negativo). Para possibilitar a visualização e identificação dos complexos e do DNA livre, este normalmente é marcado radioativamente, ou com pequenas moléculas como digoxigenina ou biotina.

Uma vez caracterizados os segmentos de DNA, através de experimentos *in silico* e *in vitro*, parte-se para a etapa de caracterização das proteínas que se ligam a este DNA.

2.3.3. Fatores de regulação da transcrição do gene *FMR1*

No caso do gene *FMR1*, alguns estudos foram realizados sobre a atuação de proteínas de transcrição no controle da expressão do gene (Beilina *et al.*, 2004; Lim, Booker e Fallon, 2005). Estudos *in vitro* da região promotora do gene *FMR1* (Drouin *et al.*, 1997; Kumari e Usdin, 2001) mostraram algumas proteínas reguladoras que possuem sítios de interação nesta região, incluindo dois sítios **Sp1**, um sítio **Nrf-1** e um sítio **E-box**, onde se ligam **USF1** e **USF2** (Smith *et al.*, 2004). Smith e colaboradores (2004) apresentaram evidências *in vivo* da ligação destes fatores de transcrição na região promotora do gene *FMR1* e mostraram que os fatores Nrf-1 e Sp1 são **ativadores da transcrição** quando utilizados individualmente, e funcionam de forma sinérgica quando associados, aumentando a expressão do gene em mais de 250 vezes. Já os fatores USF e Max funcionam como **repressores da expressão gênica**. Neste mesmo estudo os autores comprovaram que a ativação da expressão do gene promovida pelos fatores Nrf-1 e Sp1 é afetada pela *metilação das citosinas da região promotora*, característica típica dos indivíduos afetados pela Síndrome do X-Frágil, sendo que o fator Sp1 parece ser menos sensível do que o fator Nrf-1.

Em outro estudo, Smith e colaboradores (2006) sugerem que membros da família **CREB/ATF** de fatores de transcrição ligam-se ao E-box existente à montante do gene *FMR1* e não o fator USF, conforme sugeriam outros pesquisadores. No mesmo estudo, os autores afirmam que CREB e Nrf-2 atuam

sinergicamente contribuindo positivamente para a expressão do gene *FMR1* (Smith *et al.*, 2006).

Lim e colaboradores (2005), em seus estudos *in vivo*, identificaram sítios de interação para a proteína **AP-2 α** na região promotora do gene *FMR1* e concluíram que esta proteína se associa a este sítio e regula seletivamente a transcrição do gene durante o desenvolvimento embrionário.

A seqüência de repetições CGG presente à montante do gene *FMR1* vem sendo estudada como possível sítio de ligação para proteínas que se ligam especificamente ao DNA e que podem influenciar as taxas de transcrição, visto que este tipo de repetição é freqüente nas regiões reguladoras de muitos genes de mamíferos

Deissler e colaboradores (1996) isolaram uma proteína de 20 kDa, chamada **p20-CGGBP**, que liga-se especificamente à duplas-fitas de DNA contendo repetições CGG. Segundo os autores, a proteína requer mais de oito repetições para ligar-se propriamente e mudanças na primeira base a cada duas repetições leva a perda da capacidade de ligação da proteína. A proteína também não se liga a repetições que tenham as citosinas parcial ou totalmente metiladas. Este estudo revela a importância destas repetições CGG, muito provavelmente na regulação da sua transcrição. Nos estudos seguintes o grupo de pesquisadores observou que a proteína p20-CGGBP quando super-expressa na célula pode causar a inibição da expressão do gene *FMR1* (Muller-Hartmann *et al.*, 2000).

Em um estudo mais recente, Orlov e colaboradores (2007) identificaram outra proteína de ligação às repetições CGG, chamada de **p56-HGCCBP** (Human (GCC)_n-Binding Protein), que se liga especificamente às repetições CGG do gene *FMR1* e regula negativamente sua expressão.

Prasad e Singh (2008) estudaram a variação da expressão da proteína FMRP e da ligação de fatores de transcrição na região promotora do gene *FMR1*, com a idade do indivíduo, no caso de camundongos. Os autores constataram que há uma diminuição no nível de expressão da proteína com a idade, e também alteração das proteínas que se ligam a esta região promotora ao longo do tempo.

Estudos sobre a conformação da cromatina na região promotora do gene *FMR1* mostraram que em pacientes normais a região promotora do gene é o centro de um domínio com baixa interação intersegmentos, o que não ocorre com os pacientes afetados (Gheldof *et al.*, 2006).

Esses trabalhos elucidaram alguns pontos pertinentes, principalmente quanto à caracterização da região promotora do gene *FMR1*. Porém, foram investigados apenas alguns fatores de transcrição e uma região muito limitada à montante do gene, até aproximadamente 350 pb. Visto que a região reguladora dos genes, principalmente em mamíferos, pode localizar-se a muitos kpb de distância do gene, a exemplo de outros trabalhos já realizados para a determinação de regiões reguladoras e TFBS de outros genes (Grehan *et al.*, 2001), torna-se necessária à investigação destes outros possíveis sítios de ligação para fatores de transcrição, numa região mais ampla ao redor e dentro do gene. É necessária também a caracterização das proteínas que se ligam a estes sítios e que podem ser parte fundamental na regulação da expressão do gene *FMR1* e, conseqüentemente, na regulação da Síndrome do X-Frágil.

2.4. Purificação e Caracterização de Proteínas de Ligação ao DNA

O estudo e caracterização das proteínas que se ligam às regiões específicas do DNA e podem influenciar na taxa de transcrição dos genes localizados próximos ou até mesmo distantes destas regiões é uma das linhas de pesquisa que mais tem tido atenção nesta fase pós-genômica. Os procedimentos utilizados em proteômica, por exemplo, incluem: eletroforeses, *Western blot*, espectrometria de massa, entre outras técnicas. Nesta etapa, é necessário inicialmente extrair as proteínas dos sistemas vivos: culturas celulares, tecidos, órgãos, etc, normalmente utilizando-se solventes e seqüências de centrifugação. Após a obtenção do extrato que contém as proteínas a serem utilizadas, normalmente proteínas nucleares, seguem-se as etapas de purificação e identificação, por eletroforese e espectrometria de massa das proteínas.

A purificação das proteínas que interagem com o DNA, no entanto, mostra-se um grande desafio para os pesquisadores. Os extratos obtidos a partir dos sistemas vivos são sempre muito diluídos, em pequena quantidade e formados de uma imensa quantidade de diferentes proteínas, que incluem proteínas estruturais, de transporte, enzimas, fatores de transcrição, entre outras.

As técnicas cromatográficas convencionais para purificação de proteínas tais como cromatografia de troca iônica, cromatografia de interação hidrofóbica ou cromatografia de pseudoafinidade (IMAC, cromatografia com aminoácidos imobilizados, por exemplo), tornam-se muito inespecíficas, levando a um baixo rendimento e um baixo grau de pureza. No caso de purificação de fatores de transcrição, normalmente utiliza-se cromatografia por afinidade ao DNA. Nestes

casos utilizam-se matrizes estáticas, como gel ou membranas, como suporte para a seqüência de DNA onde se ligam as proteínas de interesse (Lin *et al.*, 2003). As proteínas são obtidas após sua eluição com tampões contendo altas concentrações de sal ou baixos pHs, dependendo do sistema. Normalmente, nas cromatografias, as proteínas obtidas encontram-se altamente diluídas, devida a necessidade de utilização de grandes quantidades de tampão de eluição.

Outra técnica que vem sendo muito utilizada por sua praticidade e versatilidade é a precipitação por afinidade a fragmentos de DNA imobilizados em esferas metálicas (Gabrielsen *et al.*, 1989; Nordhoff *et al.*, 1999). Neste caso, ao invés de se utilizar uma matriz estática contida numa coluna de cromatografia, utilizam-se micro-esferas metálicas que podem ser separadas através de forças magnéticas, conforme apresentado na Figura 2.5. Neste caso a proteína liga-se ao DNA imobilizado nas esferas e o complexo é precipitado com o auxílio de um magneto, podendo ser lavado para eliminação das proteínas fracamente aderidas. A seguir, procede-se a etapa de eluição para obtenção das proteínas que estavam ligadas ao DNA. Com esta técnica consegue-se a eluição das proteínas de interesse com menores quantidades de tampão de eluição, o que leva a obtenção de soluções mais concentradas em proteínas.

Uma vez purificadas, as proteínas de interesse podem ser identificadas a partir de eletroforese uni ou bi-dimensional ou então submetidas a análises de espectrometria de massa. Normalmente, os dois processos são necessários para uma boa identificação. Primeiramente submete-se a amostra a uma eletroforese, a fim de separar as proteínas que se ligaram ao DNA e em seguida as bandas, devidamente cortadas do gel, sofrem digestão por tripsina seguida de análise por espectrometria de massa. Em alguns casos submete-se a solução de proteínas purificadas diretamente à digestão por tripsina e análise de espectrometria de massa.

A espectrometria de massa é uma técnica na qual as moléculas estudadas são ionizadas e a sua razão massa/carga é medida para determinar sua massa exata. Se a massa é medida com precisão pode-se determinar a composição da molécula. No caso das proteínas, a sua seqüência de aminoácidos pode ser determinada (Baldwin, 2004).

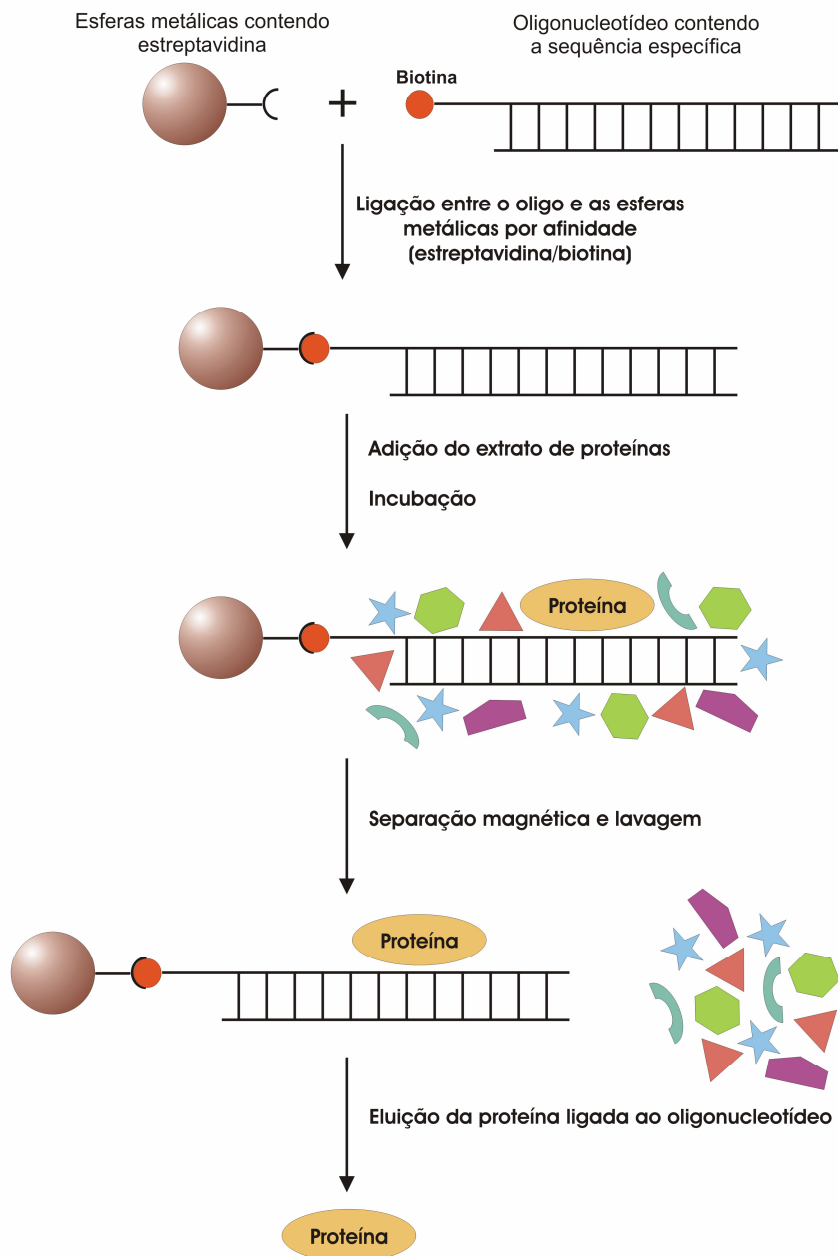


Figura 2.5: Esquema de purificação de proteínas por afinidade a seqüências específicas de DNA.

A identificação de proteínas normalmente é baseada na análise de peptídeos gerados a partir da digestão proteolítica da cadeia peptídica. A enzima mais utilizada é a tripsina, que hidrolisa a proteína na parte C-terminal dos aminoácidos lisina e arginina, com exceção dos casos onde o aminoácido subsequente é a prolina. Esta digestão leva em consideração de que todo peptídeo formado, menos a parte C-terminal da proteína, tem no mínimo dois sítios para protonação: o grupo amino da parte N-terminal e o resíduo C-terminal. Assim, os peptídeos são ionizados e detectados como íons positivos. (Baldwin, 2004).

O método de espectrometria de massa possui uma alta sensibilidade e baixo limite de detecção, permitindo facilmente analisar e identificar amostras com baixa concentração de proteínas (Lin *et al.*, 2003).

Os novos programas de bioinformática também vêm sofrendo grandes melhorias e as previsões de seqüências de proteínas deduzidas a partir das seqüências genômicas têm se tornado muito confiáveis. Utilizando-se destes avanços, muitos algoritmos têm sido criados para relacionar os dados de espectrometria de massa obtidos experimentalmente com seqüências de proteínas conhecidas (Yates, 1998). Deste modo, proteínas isoladas podem ser relacionadas a seqüências de proteínas conhecidas através dos resultados obtidos a partir da análise de espectrometria de massa dos peptídeos obtidos a partir da sua digestão proteolítica, como se fosse uma “impressão digital” daquela proteína. Estes resultados podem ser armazenados em grandes bancos de dados disponíveis na Web e utilizados para a identificação de amostras complexas de proteínas, através da técnica chamada *Tandem Mass Spectrometry* (Lin *et al.*, 2003).

A Espectrometria de Massa em Tandem, isto é, em seqüência, é utilizada para analisar os peptídeos derivados da digestão de uma mistura de proteínas. Os peptídeos resultantes são separados através de cromatografia e introduzidos em um espectrômetro de massa tipo tandem, que possui dois analisadores combinados. Este tipo de espectrômetro consegue selecionar os íons dos peptídeos a partir de uma mistura de íons. A partir daí, o padrão de fragmentação derivado de cada peptídeo, indica a sua seqüência de aminoácidos. Este padrão é comparado com outros padrões disponíveis nos bancos de dados de resultados de espectrometria de massa em tandem, sem que haja a necessidade de primeiramente interpretar a seqüência a partir do espectro.

Muitos algoritmos foram desenvolvidos com o objetivo de automatizar o processo de identificação de proteínas a partir dos dados de Espectrometria de Massa em Tandem como, por exemplo: Mascot (www.matrixscience.com), MS-Tag (www.prospector.ucsf.edu), Pepsea (www.pepsea.protana.com), SEQUEST (www.fields.scripps.edu/sequest), Sonar (www.proteometrics.com), X! Tandem (www.thegpm.org/tandem/), entre outros.

Existem muitas vantagens no uso da espectrometria de massa em tandem para a identificação de proteínas, entre elas: um alto nível de certeza na identificação, já que o método leva a obtenção de vários fragmentos muito específicos da informação, que caracterizam a proteína, e o fato de permitir a identificação de proteínas a partir de uma mistura (Lin *et al.*, 2003).

2.5. FMRP versus plasticidade sináptica e aprendizado

A função da proteína FMRP no organismo ainda não é totalmente conhecida, porém, sabe-se que ela possui domínios típicos de proteínas de ligação com mRNA (Adinolfi *et al.*, 2003; Chen *et al.*, 2003) e outros domínios de sinalização de localização e exportação nuclear (Oostra, 2002; Jin e Warren, 2003), o que sugere que esta proteína esteja intimamente ligada ao transporte de mRNA do núcleo para o citoplasma e na sua condução até os ribossomos, onde faz parte dos complexos de ribonucleoproteínas (RNP) (Ceman *et al.*, 1999).

A capacidade de ligação com mRNA e a participação da FMRP no transporte de outras proteínas dentro da célula foram estudadas por Otero e colaboradores (2002), que demonstraram o transporte da FMRP, associada a complexos protéicos, dentro de células tipo PC12 de ratos. Estas células foram transfectadas com um sistema de indução de expressão para a FMRP-GFP e, utilizando microscopia de fluorescência, foi possível registrar o percurso da FMRP-GFP desde sua tradução no corpo da célula até seu destino, a extremidade dos dendritos, onde provavelmente atua na regulação da tradução de mRNA específicos. Os autores observaram que a FMRP faz parte de grânulos (partículas RNP) que se movem em direção aos dendritos. Juntamente com a FMRP foram co-localizados nestes grânulos mRNA, FXR1, quinesina e subunidades ribossomais.

A proteína FMRP, além de ser sintetizada no corpo dos neurônios, também pode ser traduzida localmente nos dendritos. Weiler e colaboradores (1997) (Weiler *et al.*, 1997) demonstraram que a proteína responsável pela Síndrome do X-Frágil é traduzida nos dendritos em resposta à atividade sináptica em sinaptoneurosomas. Os autores demonstraram que o mRNA codificante da FMRP associou-se com complexos translacionais de 1 a 2 minutos depois da estimulação do receptor metabotrópico de glutamato (mGLUR1), observando um aumento na expressão de FMRP.

O transporte de alguns mRNA do núcleo para as partes mais distantes dos dendritos e a sua tradução próxima das regiões sinápticas é um fenômeno intimamente ligado à plasticidade sináptica, que requer a síntese local de determinadas proteínas em resposta à atividade sináptica (Weiler e Greenough, 1993; Huber *et al.*, 2002). Para que a síntese destas proteínas ocorra no dendrito é necessário que os mRNA que as codificam estejam disponíveis para a tradução no local desejado.

Apenas um pequeno grupo de mRNA é transportado ao longo dos neurônios até os dendritos, e estudos indicam que a seleção desses mRNA parece ser feita pela

sua interação com a proteína FMRP (Schaeffer *et al.*, 2001; Antar *et al.*, 2004). A FMRP é uma proteína de ligação com o mRNA que se liga com alta afinidade àqueles que possuem uma conformação espacial específica formada por 4 guaninas (G quartet) (Schaeffer *et al.*, 2003). Vários mRNA, que possuem essa formação contendo as guaninas e apresentam afinidade pela FMRP foram identificados, e observou-se que vários deles tiveram sua expressão alterada em pacientes afetados pela Síndrome do X-Frágil (Brown *et al.*, 2001).

Brown e colaboradores (2001) utilizaram-se de ensaios de *microarrays* (ou microarranjos) para identificar que 3,9% (aproximadamente 430 mensageiros) dos mRNA transcritos no cérebro de camundongos encontram-se associados à FMRP. Neste caso a FMRP estava fazendo parte dos complexos de ribonucleoproteínas associadas a mRNA (mRNP). Dados semelhantes quanto à capacidade de ligação da proteína FMRP com mRNA já haviam sido divulgados para ensaios *in vitro* em sistemas livres de células (Brown *et al.*, 1998). Um ponto divergente em relação aos dados obtidos *in vitro* e *in vivo* foi o fato de que *in vivo* o mRNA do gene *FMR1*, que codifica a proteína FMRP, não foi encontrado associado a ela, como era esperado devido aos resultados *in vitro* apresentados pelo mesmo grupo em 1998, embora tenha sido identificado como um dos mRNA expressos naquelas células. Brown e sua equipe, nesse mesmo trabalho de 2001, observaram que se for comparado o número de mRNA associados aos complexos mRNP de células de linfoblastos humanos sadios e de pacientes afetados pela Síndrome do X-Frágil, nota-se que o número de alguns mRNA associados aumenta enquanto o de outros diminui nos pacientes afetados (aproximadamente 50%/50%), enquanto outros ainda permaneceram inalterados.

A FMRP parece ser responsável pelo controle da tradução de alguns mRNA, principalmente nos neurônios ao nível dos dendritos e sinapses (Weiler e Greenough, 1999; Greenough *et al.*, 2001; Laggerbauer *et al.*, 2001; Li *et al.*, 2001; Zalfa *et al.*, 2003; Lu *et al.*, 2004; Weiler *et al.*, 2004).

Li e colaboradores (2001) e Laggerbauer e colaboradores (2001) utilizaram ensaios *in vitro* para demonstrar que a proteína FMRP produzida de forma recombinante, inibiu a tradução de determinados mRNA.

Lu e colaboradores (2004) estudaram a influência da FMRP no controle da tradução da proteína 1B associada à microtúbulos (MAP1B), que é uma proteína associada ao desenvolvimento dos dendritos e sinapses. Neste trabalho, os autores observaram que a ausência da FMRP nas células neuronais de indivíduos

X-Frágil resulta em uma expressão elevada da proteína MAP1B devido à perda da repressão da tradução do seu mRNA que supostamente é realizada pela FMRP.

No entanto, estudos realizados com animais modelo, camundongos *knockout* que não expressam o gene *FMR1*, por Jianxue Li e colaboradores (2002), mostraram que esses animais apresentam uma plasticidade sináptica reduzida na região do córtex cerebral, acompanhada de uma redução da expressão da proteína GluR1, uma das subunidades formadoras do receptor ionotrófico de glutamato – AMPA, nas sinapses. Esse estudo sugere que a falta da proteína FMRP nestes camundongos inibiu, ou melhor, não favoreceu a correta expressão da proteína GluR1 que também está relacionada com a plasticidade sináptica, na região do córtex cerebral. O mesmo estudo, no entanto, mostrou que não houve alteração no nível de plasticidade e expressão de GluR1 nas regiões do hipocampo e cerebelo.

A ligação da FMRP com a regulação da atividade dos receptores AMPA também foi estudada por Bear e colaboradores (2004). Segundo estes autores a deficiência de FMRP, característica dos pacientes portadores da Síndrome do X-Frágil, aumenta a depressão de longa duração (LTD – *Long-Term Depression*), na região do hipocampo cerebral. A depressão de longa duração consiste de um conjunto de reações bioquímicas que são desencadeadas nos dendritos devido à atividade sináptica, assim como a potenciação de longa duração (LTP – *Long-Term Potentiation*). O mecanismo da LTP parece ser muito importante em recém-nascidos, uma vez que favorece o nascimento de novos botões sinápticos. Por outro lado, a LTD parece estar relacionada com a eliminação de sinapses, de forma orientada. Estes dois mecanismos, trabalhando em conjunto, contribuem para o aprendizado e para a formação da memória durante a vida (Bear *et al.*, 2004).

A LTD, no hipocampo e no cerebelo, pode ser induzida pela estimulação dos receptores metabotrópicos de glutamato chamados mGLURs (mGLUR1 e mGLUR5) que possuem localizações subcelulares distintas.

Segundo alguns autores, a não expressão ou inativação do receptor mGLUR causa um déficit na LTD. O déficit de LTD pode ser expresso como a diminuição no número de receptores AMPA pós-sinápticos (Bear *et al.*, 2004). A LTD induzida pela ativação do receptor mGLUR (mGLUR-LTD) requer a rápida tradução de mRNA preexistentes nos dendritos pós-sinápticos (Huber *et al.*, 2000). Outra forma de indução da LTD é a ativação do receptor NMDA (NMDA-LTD), porém esta forma de indução da LTD não requer a rápida tradução de mRNA (Sawtell *et al.*, 1999).

Em 2002, Huber e colaboradores propuseram um modelo que visa explicar como a Síndrome do X-Frágil atua para alterar a plasticidade sináptica e,

conseqüentemente, a capacidade de aprendizado e memorização. Este grupo, do qual também fazia parte o pesquisador Mark F. Bear utilizou camundongos modelos *knockout* para o gene *FMR1* para estudar a LTD induzida pela ativação do mGLUR1 e pela ativação do NMDA, no hipocampo. Os resultados mostraram que a LTD-NMDA é normal nos animais modelo enquanto a LTD-mGLUR é aumentada em animais *FMR1-knockout*. Com base em seus resultados e em resultados prévios de outros autores, Huber e colaboradores (2002) propuseram o modelo apresentado na Figura 2.6.

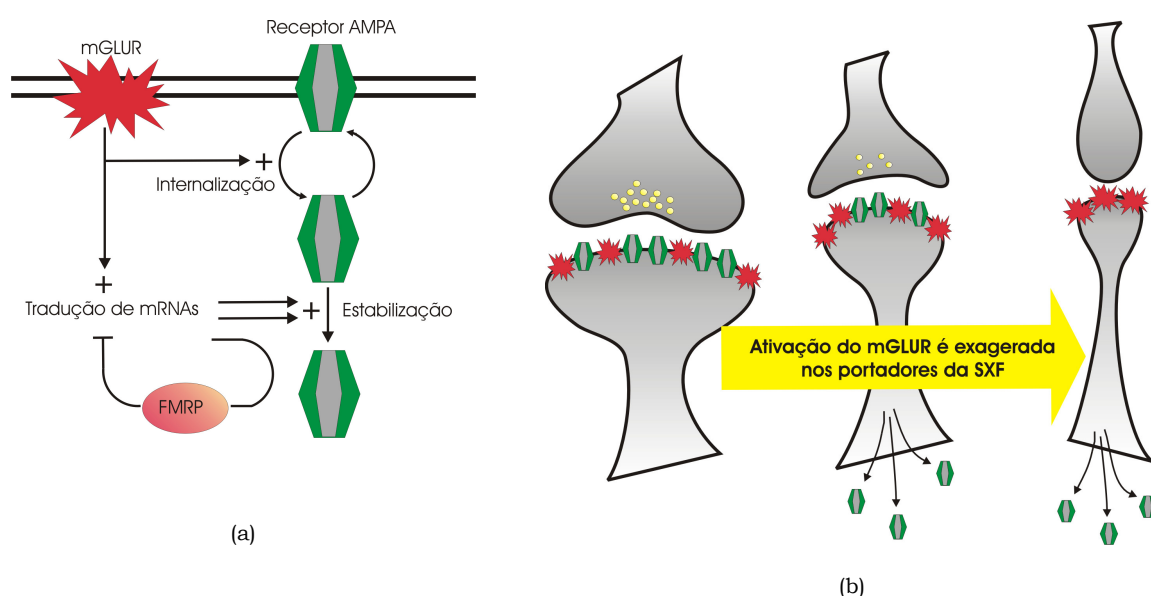


Figura 2.6: Modelo de atuação da FMRP na alteração da LTD em indivíduos afetados pela Síndrome do X-Frágil. (a) A ativação da mGLUR estimula a internalização dos receptores AMPA. A estabilização desta modificação requer a tradução de mRNA, que é regulada negativamente pela FMRP que é sintetizada em resposta à ativação da mGLUR. (b) Na ausência de FMRP a magnitude da LTD é aumentada levando a internalização excessiva de receptores e conseqüente perda da capacidade sináptica, conduzindo ao enfraquecimento das interações sinápticas e ao alongamento dos dendritos, como já observado por alguns autores em pacientes afetados pela SXF. (Adaptado de Bear *et al.*, 2004).

Segundo este modelo, a ativação da mGLUR estimula a internalização dos receptores AMPA, e a estabilização desta modificação requer a tradução de mRNA. Os autores propõem que esta tradução seja regulada negativamente pela FMRP que é sintetizada em resposta à ativação da mGLUR (Weiler e Greenough, 1999). Esta proposta é consistente com os resultados obtidos por outros autores que também constataram que a FMRP regula negativamente a tradução de alguns mRNA (Laggerbauer *et al.*, 2001; Li *et al.*, 2001). Sendo assim, na ausência de FMRP a magnitude da LTD é aumentada, ocorrendo uma internalização excessiva dos receptores AMPA, no hipocampo. Isto levaria a uma dificuldade da maturação sináptica e conseqüente perda da capacidade de aprendizado e memorização

Ceman e colaboradores (2003) propuseram que a regulação da atuação da FMRP na inibição da tradução de alguns mRNA está relacionada com a sua fosforilação. Os grânulos compostos pelos mRNA, FMRP e outras proteínas, localizam-se nos dendritos esperando por um sinal sináptico que indique a necessidade de tradução dos respectivos mRNA. Em resposta à ativação dos receptores mGLURs, que são acoplados à cascata do fosfatidilinositol (PIP2), ocorre a estimulação de uma proteína efetora chamada fosfolipase C, que induz a formação de dois importantes segundos-mensageiros: o diacilglicerol (DAG) e o inositol-1,4,5-trifosfato (IP₃). O IP₃ regula a liberação das reservas de cálcio intracelulares armazenadas no retículo endoplasmático. O cálcio liberado interage com numerosas proteínas celulares, incluindo a proteína quinase C (PKC). A DAG, por sua vez, também ativa a PKC que está envolvida em vários processos celulares, incluindo a expressão gênica (Gallagher *et al.*, 2004).

Embora não exista nenhuma evidência direta de que a PKC fosforile a FMRP, alguns autores sugerem que a ativação da PKC desencadeia uma cascata enzimática com vários intermediários, sendo que um deles sinaliza para que o grânulo, contendo o mRNA de interesse, libere a FMRP, permitindo assim a síntese protéica (Antar *et al.*, 2004; Weiler *et al.*, 2004).

Compilando todos os dados apresentados até aqui, a Figura 2.7 mostra, de forma sintética, um modelo biológico para a atuação da proteína FMRP em células neuronais, como moduladora da tradução de alguns mRNA específicos devido à sinalização proveniente da excitação do receptor metabotrópico de glutamato (mGLUR1). Este modelo está sendo aqui proposto a partir das evidências experimentais encontradas na literatura, obtidas de células de pacientes afetados e com a utilização de camundongos *knockout* (Jin e Warren, 2003; Antar *et al.*, 2004; Bear *et al.*, 2004; Weiler *et al.*, 2004; Aschrafi *et al.*, 2005).

De acordo com esse modelo, a FMRP atua na localização e exportação de mRNA do núcleo, bem como participa da regulação da tradução destes mRNA no corpo e nos dendritos dos neurônios, incluindo seu próprio mRNA (Bear *et al.*, 2004).

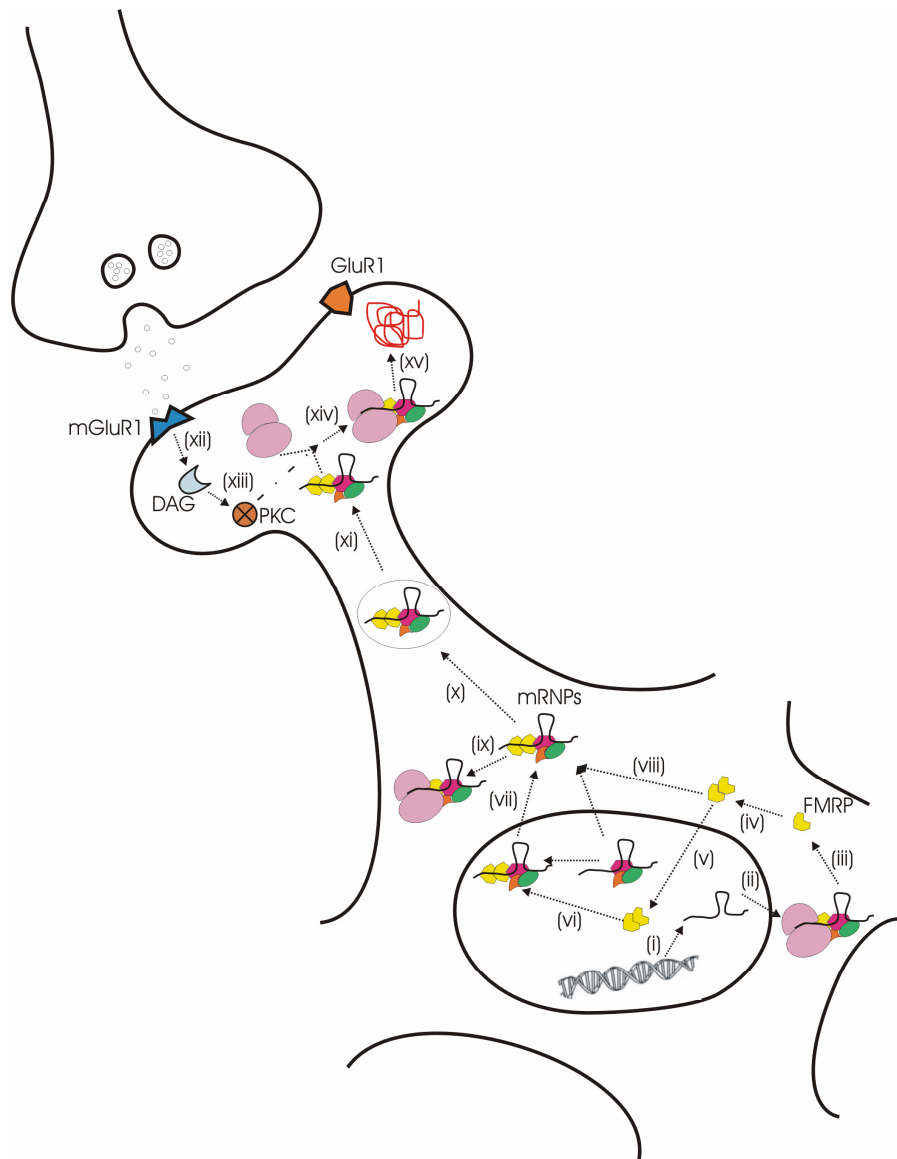


Figura 2.7: Modelo biológico proposto para o transporte de proteínas e regulação da expressão de mRNA pela FMRP em células neuronais. (i) transcrição do gene *FMR1*; (ii) transporte do mRNA transcrito para o citoplasma; (iii) tradução do mRNA; (iv) a FMRP dimeriza no citoplasma; (v) devido ao seu sinal de localização nuclear (NLS) a proteína entra no núcleo do neurônio; (vi) e (viii) a FMRP dimerizada liga-se aos complexos de ribonucleopartículas mensageiras (mRNP); (vii) a proteína promove a exportação deste complexo de proteínas e mRNA no núcleo devido ao seu sinal de exportação do núcleo (NES); (ix) uma vez no citoplasma, a FMRP associada ao complexo mRNP pode associar-se a ribossomos e regular a tradução de proteínas no corpo celular dos neurônios; (x) e (xi) alternativamente, o complexo FMRP+mRNP pode associar-se aos sistemas de transporte celular até os dendritos; (xii) em resposta a um estímulo sináptico no receptor mGluR1, (xiv) desencadeia-se a cascata de ativação da PKC; (xiii) a PKC causa a inativação, possivelmente por fosforilação, da proteína FMRP, que atua na repressão da tradução do mRNA associado ao complexo FMRP+mRNP, (xiv) permitindo a sua associação com os ribossomos dos dendritos; (xv) sem a repressão causada pela FMRP o mRNA é traduzido na proteína, o que é importante para a estrutura e plasticidade sináptica, como por exemplo o receptor GLUR1 (Jin e Warren, 2003; Antar *et al.*, 2004; Weiler *et al.*, 2004).

Em relação a este modelo biológico, muitas dúvidas ainda precisam ser desvendadas, principalmente quanto aos dados quantitativos. Estes dados podem ser determinados tanto experimentalmente, através de cultura de células, quanto

através de modelagem matemática dos fluxos de proteínas e outros fatores que possam afetar seu mecanismo de atuação. Algumas medições podem ser realizadas, porém, para se compreender melhor a participação efetiva de todos os componentes envolvidos na regulação da tradução de determinados mRNA em locais específicos dos neurônios, e em resposta a estímulos específicos, o ideal é conhecer a participação de todos os fatores envolvidos.

Uma abordagem possível e já utilizada na modelagem matemática e análise de vias metabólicas é a técnica conhecida como Análise de Fluxos Metabólicos (AFM) (Stephanopoulos *et al.*, 1998; Chatziioannou *et al.*, 2003; Ozkan *et al.*, 2005). A AFM utiliza alguns dados de fluxos experimentais e, através de balanços de massa e a produção da respectiva matriz estequiométrica, pode estimar valores para o fluxo de partículas ou proteínas que participam do processo regulador, e que são de difícil, senão impossível, determinação experimental, nas condições atuais.

A Análise de Fluxos Metabólicos é uma técnica muito eficiente e tem se mostrado aplicável a diferentes estudos *in silico*, isto é, computacionais, de vias metabólicas. No entanto, o uso desta abordagem no estudo de vias reguladoras é um procedimento novo, e espera-se que possa fornecer informações relevantes sobre o estado homeostático celular.

Chatziioannou e colaboradores (2003), por exemplo, aplicaram essa metodologia como uma alternativa na elucidação do metabolismo do neurotransmissor glutamato, onde, a partir de poucos dados experimentais foi possível determinar o fluxo de todos os componentes da via metabólica de produção, liberação e captação do glutamato em neurônios e células gliais.

Além de tornar possível a estimativa de valores, este método permite ainda a correção de dados experimentais e ainda o uso do modelo para simular possíveis alterações na regulação, como por exemplo, a deleção de genes, que resultaria na inexistência de alguma das proteínas reguladoras, ou a super-expressão dessas proteínas. Essas alterações, de outra forma, só poderiam ser feitas através de técnicas de engenharia genética em organismos modelo.

No entanto, para que a modelagem seja a mais fiel possível, um maior número de fatores (proteínas) envolvidos, desde a regulação da expressão da proteína FMRP até a sua atuação como proteína reguladora no transporte e tradução de mRNA dentro dos neurônios, devem ser caracterizados através de determinações experimentais.

3. MATERIAIS E MÉTODOS

Um esquema geral da metodologia empregada nos estudos genômicos desenvolvidos neste trabalho e apresentado na Figura 3.1.

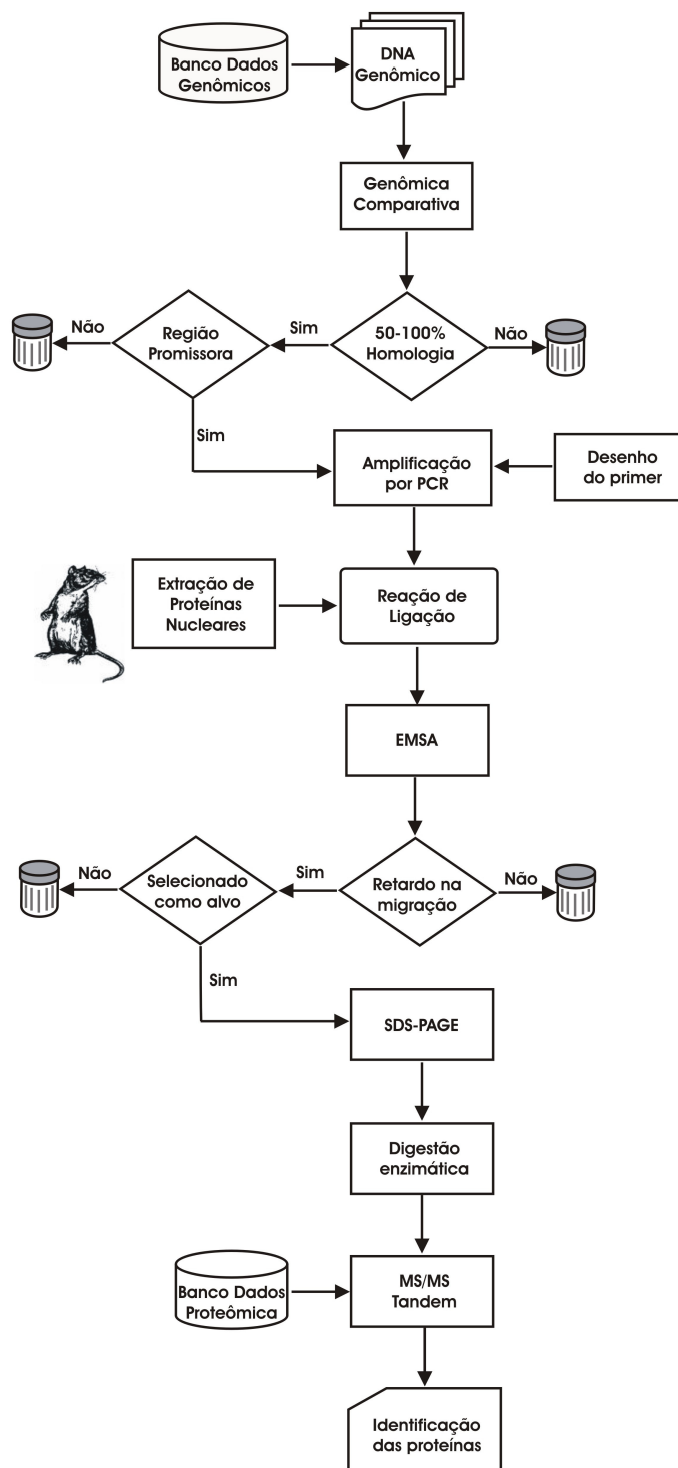


Figura 3.1: Resumo da metodologia utilizada para a determinação dos sítios de interação para proteínas de transcrição do gene *FMR1*

3.1. Sequências de DNA Genômico

As seqüências de DNA estudadas visando a determinação de possíveis regiões conservadas contendo TFBS foram obtidas a partir do GenBank – NCBI. Foram estudadas as seguintes espécies para comparação com a seqüência reguladora humana (contig NT_011681.15:11838:3325293-3475293), *Pan troglodytes* (contig NW_001252165.1:988399-1138399), *Canis familiaris* (contig NW_879563.1:67539336-67689336), *Mus musculus* (contig NT_039706.7:13405245-13555245) e *Rattus norvegicus* (contig NW_048052.2:13525688-13675688). Para este estudo foram utilizadas seqüências que compreendem toda a região codificante do gene *FMR1* e a região não codificante à montante do gene, onde podem estar situados os possíveis TFBS.

3.2. Determinação *in silico* das regiões conservadas do DNA que contêm sítios de interação para fatores de transcrição

A determinação dos possíveis TFBS foi feita através do programa mVISTA para genômica comparativa (<http://genome.lbl.gov/vista>). O mVista é um programa para comparação de seqüências de DNA de duas ou mais espécies (Mayor *et al.*, 2000; Frazer *et al.*, 2004).

Na busca por possíveis sítios de interação para proteínas reguladoras da transcrição do gene *FMR1*, estudou-se toda a seqüência de nucleotídeos à montante do gene, até a região codificante anterior na mesma fita de DNA, que corresponde a uma proteína hipotética (LOC100133053). Esta primeira região estudada compreendeu *mais de um milhão de pares de bases* (pb) de nucleotídeos.

Esta seqüência de DNA à montante do gene acrescida da seqüência codificante do gene *FMR1* foi alinhada globalmente com as seqüências à montante e codificante dos homólogos deste gene nas espécies *Pan troglodytes*, *Canis familiaris*, *Mus musculus* e *Rattus norvegicus*, também incluindo toda a região não codificante existente entre o gene *FMR1* destas espécies até a região codificante anterior na mesma fita. Estes primeiros alinhamentos revelaram que a porção da seqüência humana mais próxima ao gene apresentava mais regiões conservadas nas quatro outras espécies. Esta região próxima que apresentava maior grau de conservação (75% no mínimo) compreendia aproximadamente 150.000 pares de bases de nucleotídeos, no DNA humano, incluindo o próprio gene. Focou-se então os estudos nesta porção de 150.000 pb. A análise dos possíveis TFBS foi realizada utilizando o programa rVista (regulatory Vista) (Loots *et al.*, 2002). O programa

rVista combina o alinhamento obtido pelo mVista, que identifica as regiões conservadas entre as seqüências, com uma pesquisa no principal banco de dados de sítios de ligação para fatores de transcrição, o TRANSFAC Professional (<http://helixweb.nih.gov/transfac/>), para localizar nas regiões conservadas os possíveis sítios de ligação para fatores de transcrição conhecidos. Um esquema da metodologia utilizada é apresentado na Figura 3.2.

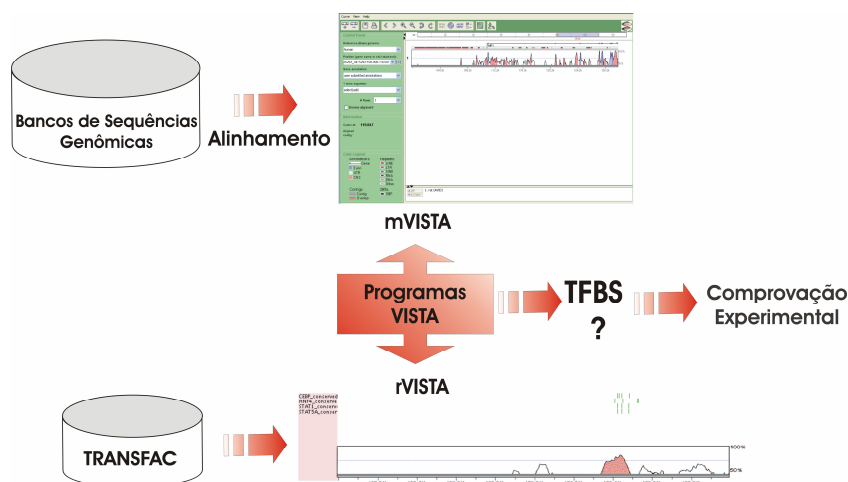


Figura 3.2: Resumo da metodologia para determinação *in silico* das regiões conservadas do DNA que podem conter sítios de interação para fatores de transcrição.

3.3. Animais de laboratório

Ratos Wistar adultos (60-90 dias) saudáveis, mantidos com quantidade adequada de comida e água, foram obtidos do Biotério Central da Universidade Federal de Santa Catarina. Os experimentos foram realizados segundo protocolos aprovados pelo Comitê de Ética para Pesquisas com Animais da Universidade.

3.4. Extração de proteínas nucleares e citoplasmáticas

As proteínas citoplasmáticas e nucleares foram extraídas a partir de porções de 100 mg de córtex de ratos Wistar utilizando o protocolo e o *kit* de extração CellLyctic NuCLEAR (Sigma, EUA). As proteínas citoplasmáticas foram recuperadas, após lise das células com tampão de lise (10 mM HEPES, pH 7,9, 1,5 mM MgCl₂, 10 mM KCl, 1 mM DTT e 0,5 mM PMSF), seguido de centrifugação por 20 minutos a 10.000×g. O precipitado foi ressuscitado em tampão de extração (20 mM HEPES, pH 7,9, 1,5 mM MgCl₂, 0,42 M NaCl, 0,2 mM EDTA, 25% v/v glicerol, 1 mM DTT e 0,5 mM PMSF) e agitadas por 30 minutos. As proteínas nucleares foram obtidas, após centrifugação a 20.000×g por 5 minutos, no sobrenadante, que foi transferido

para um tubo limpo, aliquotado, congelado e estocado a -20°C . A concentração de proteínas foi determinada pelo Método de Lowry (Lowry *et al.*, 1951). Nos casos em que a concentração salina dos tampões usados era muito alta, as frações protéicas foram dessalinizadas por filtração em gel utilizando-se colunas Sephadex PD-10 (Amersham Biosciences). As proteínas extraídas foram visualizadas por Eletroforese SDS-PAGE em gel gradiente de acrilamida, de 4–15% (Biorad, EUA) e coloração com solução de nitrato de prata

3.5. Extração de DNA genômico

O DNA genômico foi recuperado a partir de tecido cerebral de ratos Wistar adultos normais. O DNA foi recuperado utilizando o *kit* GenomicPrep™ Cells and Tissue DNA Isolation (GE Healthcare, EUA), conforme instruções do fabricante. Aproximadamente 10 a 20 mg de tecido cerebral foi homogeneizado em “tampão de lise” e incubado por 12 horas a 55°C , após adição de proteinase K (aprox. 0,5 mg/mL). Tratou-se a solução com RNAase por 60 min. As proteínas foram precipitadas pela adição de “solução de precipitação” e centrifugação a $16.000\times g$ por 15 minutos. O sobrenadante foi transferido para um tubo contendo isopropanol e mantido por 12 horas no freezer. A solução contendo o DNA precipitado foi centrifugada a $16.000\times g$ por 15 min. O sobrenadante foi descartado e o DNA presente no precipitado foi seco em estufa (37°C) por 60 minutos. O precipitado foi hidratado com a “solução de hidratação” e armazenado em geladeira ($2-8^{\circ}\text{C}$).

3.6. Amplificação e purificação do fragmento de DNA de interesse

As seqüências reguladoras estudadas foram amplificadas a partir do DNA genômico em uma reação de cadeia de polimerase (PCR) utilizando-se os seguintes iniciadores, ou *primers*, determinados para cada região:

Seqüência 1

5'GCCCTGAGCTTTCAAAGTTAAAGC3'

5'GAGTACTTTAGGCTTCCAGTTCCTCA3'

Seqüência 2

5'AGATTAAAACCAGTTCCTCCAGCCTC3'

5' TGCTTTCACCTGGTAAATGGTAATT3'

Seqüência 3a

5'AGGGATATTGGATGAACAACAGGG3'

5'GCAAGGTTGCATGGCTGTTTT3'

Seqüência 3b

5'TACCACAGTAAAATCATGGGAAAGC3'

5'CCCTTGGATAGTTTTCACTGCAGT3'

Seqüência 4

5'AGCCAGATACTTTATGTCCACCCA3'

5'GTGTCCTTCTACCAGCAGCATCA3'

Seqüência 5

5'ACTATTGTAGGCAGAAAGGCAGAGC3'

5'TGTTTTCTTCCCCAAGTCCTGC3'

Cada amplificação foi realizada em tampão contendo diferentes concentrações de cloreto de magnésio, de acordo com otimização preliminar da amplificação, e também diferentes temperaturas de anelamento, conforme tabela a seguir:

Tabela 3.1: Temperaturas de anelamento e concentrações de MgCl₂ utilizados para cada uma das amplificações realizadas:

Seqüência	Concentração de MgCl₂ (mM)	Temperatura de anelamento (°C)
Seqüência 1	1,5	59
Seqüência 2	1,5	56
Seqüência 3a	2,5	55
Seqüência 3b	2,0	50
Seqüência 4	2,0	59
Seqüência 5	2,0	50

Os produtos do PCR foram separados por eletroforese em gel de agarose (1%) e corados com brometo de etídio. As bandas correspondentes aos oligonucleotídeos (oligos) de interesse foram identificadas de acordo com o marcador de peso molecular para DNA, Hi-Lo all purpose (Bionexus, EUA). O DNA foi recuperado, a partir do gel de agarose, utilizando-se o *kit* QIAquick Gel Extraction Kit (Qiagen, EUA).

Para utilização como ligante de afinidade na etapa de purificação de proteínas nucleares, um *primer* de cada par usado para amplificação das seqüências foi adquirido na forma biotilada, a fim de obter o produto do PCR ligado a biotina.

3.7. Gel shift para determinação da ligação DNA-proteína de regiões reguladoras para o gene *FMR1*

Os produtos obtidos com as reações de PCR foram marcados com Digoxigenina (DIG) e os ensaios de ligação de proteínas reguladoras ao segmento de DNA foram realizados utilizando-se o *kit* DIG Gel Shift Kit 2nd Generation (Roche Applied Science, Suíça). Os produtos amplificados e marcados, contendo os segmentos de interesse, foram utilizados nas reações de *gel shift*, conforme instruções do fabricante. Cada reação de ligação foi realizada com 30 fmol do produto da amplificação marcado com DIG e um ou dois microgramas de proteínas nucleares totais na presença do tampão de ligação (20 mM HEPES pH 7,6, 1mM EDTA, 10 mM (NH₄)₂SO₄, 1 mM DTT, 0,2% Tween 20, 30 mM KCl) em temperatura ambiente. Competidores específicos e não-específicos foram usados nas reações de ligação para reduzir os resultados falso-positivos. Como competidores específicos

foram utilizados os próprios produtos da amplificação em estudo (100× concentrado) sem a marcação com DIG e como competidor inespecífico foi utilizado o oligonucleotídeo poli-AT (100×).

Os complexos proteína-DNA foram separados por eletroforese nativa em gel de poliacrilamida 5% em tampão Tris-borato (TBE) 0,5×. As bandas resultantes da eletroforese foram transferidas para membranas de Nylon® positivamente carregadas (Roche, Suíça). A detecção foi feita por quimioluminescência dos complexos DIG-DNA-Proteína reconhecidos por anticorpo anti-DIG marcado com fosfatase alcalina, de acordo com o fabricante (Roche, Suíça).

3.8. Recuperação de proteínas nucleares através de precipitação por afinidade

Aproximadamente 40 pmol dos produtos da amplificação biotilados, dupla fita, foram imobilizados em 200 µg de esferas metálicas Dynabeads M-280 recobertas com streptavidina (Invitrogen, EUA), conforme protocolo do fabricante. Um volume de 60 µl de solução de proteínas nucleares (75 µg de proteínas totais) foi diluído para um total de 90 µl com Tampão de Ligação (20 mM HEPES pH 7,6, 1 mM EDTA, 10 mM (NH₄)₂SO₄, 1 mM DTT, 0,2% Tween 20 e 30 mM KCl), contendo também 1 µg/µl de competidor não específico poli-(AT)₂₅. Esta solução foi incubada, a 4°C, por 10 minutos antes de ser adicionada ao frasco contendo 200 µg de Dynabeads com ou sem o DNA alvo imobilizado. A mistura foi incubada por 15 minutos, à temperatura ambiente, sob agitação branda. Depois de uma separação magnética as esferas foram lavadas duas vezes com o mesmo Tampão de Ligação. A primeira lavagem continha 0,1 µg/µl do oligonucleotídeo poli-(AT)₂₅, que atuou como competidor não-específico para eluição das proteínas ligadas inespecificamente ao DNA. Após as lavagens as esferas foram lavadas duas vezes com Tampão de Eluição (20 mM HEPES pH 7,6, 1 mM EDTA, 10 mM (NH₄)₂SO₄, 1 mM DTT, 0,2% Tween 20 e 1 M KCl). O procedimento descrito foi realizado repetidas vezes para se obter uma quantidade significativa do produto da amplificação desejado. A solução total de proteínas eluídas foi dessalinizada utilizando colunas de filtração em gel Protein Desalting Spin (Pierce, EUA), mantidas em Tampão de Ligação e congeladas. As proteínas purificadas foram separadas por Eletroforese SDS-PAGE em gel gradiente 4–15% (Biorad, EUA) e coradas com Coomassie Brilliant Blue R-250, e as bandas distintas foram recortadas do gel.

3.9. Separação e identificação das proteínas nucleares purificadas

As proteínas contidas nas bandas recortadas do gel de poliacrilamida foram submetidas à digestão enzimática com tripsina e os peptídeos resultantes foram analisados por espectrometria de massa (Tandem Mass Spectroscopy), no Laboratório de Proteômica da Universidade da Califórnia em Davis, CA, EUA. O tratamento dos dados foi realizado com o uso do servidor da WEB chamado The Global Proteome Machine (<http://gpmdb.thegpm.org/>), que utiliza os dados disponíveis de várias bases de dados proteômicos, a fim de correlacionar os espectros obtidos dos peptídeos resultantes da digestão das proteínas com outros já identificados anteriormente.

3.10. Imunodeteção para confirmação da proteína Pur- α :

Para confirmar a presença da proteína Pur- α no conjunto de proteínas purificadas pelo o sistema de afinidade com esferas metálicas/DNA alvo, realizou-se ensaio de imunodeteção (*dot-blot*) utilizando o anticorpo policlonal de camundongo anti-Pur- α (Abnova, EUA). Um total de 20 μ l da solução de proteínas nucleares totais (50 μ g/ μ l) e da solução de proteínas purificadas (1,5 μ g/ μ l) utilizando-se o sistema esferas/DNA alvo foram depositadas sobre membranas de nitrocelulose (Biorad, EUA) equilibradas com tampão TBS (10 mM Tris, 150 mM NaCl, pH 7,5) contendo 20% de metanol. Após secagem natural, as membranas foram lavadas com tampão TBS e bloqueadas com leite desnatado 5% diluído em TBS a 0,05%, por 1 hora sob agitação. Na seqüência, a membrana foi lavada três vezes com tampão TBS-T (TBS + 0,05% Tween-20) por 5 minutos e incubadas com o anticorpo de camundongo anti-Pur- α por pelo menos 12 horas, a 4°C. Após incubação a membrana foi lavada várias vezes com tampão TBS-T e incubada com o anticorpo secundário (anti-IgG de camundongo) conjugado com peroxidase, por 1 hora em temperatura ambiente. Para revelação, a membrana foi tratada com reagentes ECL (Amershan Biosciences, EUA) e exposta ao filme fotográfico (Kodak, EUA).

3.11. Análise de Fluxos Metabólicos

O modelo matemático que descreve a atuação da FMRP nos neurônios foi simulado e analisado segundo a técnica de Análise de Fluxos Metabólicos descrita por Stephanopoulos e colaboradores (1998). Todos os procedimentos computacionais foram implementados usando-se o pacote de programação computacional MATLAB® (The MathWorks Inc., MA, EUA).

A Análise de Fluxos Metabólicos é baseada em um modelo estequiométrico que representa as interações moleculares e bioquímicas da célula, a serem estudadas. As reações podem ser obtidas na literatura ou a partir de dados experimentais. O modelo estequiométrico é baseado em balanços de massa dos componentes considerados. Para os complexos internos é usada a hipótese de estado pseudo-estacionário, ou seja, o volume não muda e as concentrações não variam em um curto período de tempo.

As equações de balanço de massa dos componentes ou grupo de componentes (complexos protéicos) de cada etapa do modelo biológico proposto são determinadas conforme a equação (1). Para o estado estacionário,

$$\sum_j v r_j = 0 \quad (1)$$

onde, a soma de todos os fluxos relativos à formação ou consumo (utilização) de um dado componente ou complexo é igual a zero, sendo v o coeficiente estequiométrico e r_j a velocidade de consumo ou formação do componente ou complexo. Nesta equação, os fluxos que contribuem para a formação de um complexo são caracterizados com sinal positivo e os fluxos de contribuem para o consumo do complexo são de sinal negativo.

A partir destas equações é determinada a matriz estequiométrica do modelo, chamada de matriz G , onde cada coluna da matriz corresponderá ao fluxo de um componente ou complexo do modelo, e as linhas corresponderão às reações que envolvem esses componentes. Todos os fluxos não medidos experimentalmente serão obtidos, em primeira análise, utilizando a Eq. (2) na forma abaixo:

$$v_c = -(G'_c)^{\#} \cdot G'_m \cdot v_m \quad (2)$$

onde, v_c é o valor do fluxo calculado, G'_c representa a matriz transposta de G em que foram selecionadas as colunas referentes aos fluxos que se deseja calcular, isto é, todos os fluxos não medidos. G'_m é a matriz transposta de G referente aos fluxos medidos, v_m é o vetor dos fluxos medidos, e o símbolo $\#$ representa a pseudo-inversa da matriz.

A partir do uso de rotinas de álgebra linear são identificadas as reações dependentes e independentes e calculada a matriz de redundância (R), conforme a equação (3),

$$R = G_m - G_c \cdot G_c^{\#} \cdot G_m \quad (3)$$

onde, G_c é a matriz de coeficientes estequiométricos dos fluxos calculados, e $G_c^\#$ é a matriz pseudo-inversa da matriz G_c .

A sensibilidade da matriz utilizada pode ser medida através do número condicional, C , definido na equação 4, cuja magnitude fornece informações importantes sobre os requisitos de precisão nos fluxos medidos:

$$C = \|G_c\| \|(G_c)^\#\| \quad (4)$$

Para uma matriz estequiométrica ser considerada bem condicionada, este número deve estar entre 1 e 100, conforme descrito por Stephanopoulos e colaboradores (1998).#

Outro modo de avaliar o modelo é através da equação 5, que fornece informação sobre a sensibilidade dos cálculos no que diz respeito às variações dos fluxos calculados às pequenas perturbações nas medidas (Stephanopoulos et al., 1998).

$$\frac{dv_c}{dv_m} = -G_c^\# \cdot G_m \quad (5)$$

4. RESULTADOS E DISCUSSÃO

4.1. Determinação de possíveis sítios para fatores de regulação da transcrição do gene *FMR1*

O alinhamento e a comparação entre seqüências genômicas ou seqüências de proteínas homólogas vêm sendo muito utilizados como um primeiro passo na caracterização de dados ainda desconhecidos de seqüências de DNA ou aminoácidos. Visando o estudo da regulação da expressão do gene *FMR1*, a partir do uso de técnicas de genômica comparativa, inicialmente verificou-se o grau de homologia entre as proteínas FMRP de algumas espécies de mamíferos, que apresentavam o gene *FMR1*.

A Figura 4.1 mostra o alinhamento entre as seqüências das proteínas FMRP humana e suas homólogas de outras quatro espécies de mamíferos.

O alinhamento entre as seqüências mostra que os principais domínios da proteína FMRP, relacionadas à ligação ao RNA, são conservadas, o que indica provável conservação de sua função. Visto que a proteína FMRP possui os mesmos domínios e provavelmente a mesma função nestas espécies, pode-se sugerir que a regulação da expressão deste gene seja controlada por elementos comuns. Esta evidência estimulou a comparação entre as seqüências à montante deste gene nestes organismos, em busca de regiões conservadas ao longo da evolução, que pudessem estar relacionadas com a regulação da expressão do gene.

Os estudos de genômica comparativa, entre a seqüência de DNA humano e as seqüências homólogas das quatro espécies comparadas, que incluíram o gene *FMR1* e até 111.000 pares de base à montante do seu códon de início, num total de **150.000 pb**, foram realizados utilizando-se o programa Vista. Os resultados obtidos a partir da análise comparativa das seqüências mostraram várias regiões conservadas entre as seqüências homólogas e a seqüência humana de interesse, incluindo a região atualmente reconhecida como região promotora do gene, que localiza-se entre 0 e 350 pares de bases à montante do início do gene, aproximadamente (Drouin *et al.*, 1997; Kumari e Usdin, 2001).

```

M.musculus -----MEELVVEVRGSN-----GAFYK----- 17
R.norvegicus -----MEELVVEVRGSN-----GAFYK----- 17
C.familiaris -----MEELVVEVRGSN-----GAFYK----- 17
H.sapiens -----MEELVVEVRGSN-----GAFYK----- 17
P.troglodytes MAYSPFPKHLPLPAGSGPVVWVVEERGA LQPGRLLRQEGFRSPLASLFRSSVGTSESST 60
: :*** **: .:::

M.musculus -----AFVKDVHEDSITVAFENNWPQERQIPFHDFRFPVPPVGYNKDINESDEVEV 67
R.norvegicus -----AFVKDVHEDSITVAFENNWPQERQIPFHDFRFPVPPVGYNKDINESDEVEV 67
C.familiaris -----AFVKDVHEDSITVAFENNWPQERQIPFHDFRFPVPPVGYNKDINESDEVEV 67
H.sapiens -----AFVKDVHEDSITVAFENNWPQERQIPFHDFRFPVPPVGYNKDINESDEVEV 67
P.troglodytes VLGEGCEERVAFVKDVHEDSITVAFENNWPDRQIPFHDFRFPVPPVGYNKDINESDEVEV 120
*****:*****:*****

M.musculus YSRANEKEPCCWLLAKVRMIKGEFYVIEAACDATYNEIVTIERLRSVNPKNPATKDTFH 127
R.norvegicus YSRANEKEPCCWLLAKVRMIKGEFYVIEAACDATYNEIVTIERLRSVNPKNPATKDTFH 127
C.familiaris YSRANEKEPCCWLLAKVRMIKGEFYVIEAACDATYNEIVTIERLRSVNPKNPATKDTFH 127
H.sapiens YSRANEKEPCCWLLAKVRMIKGEFYVIEAACDATYNEIVTIERLRSVNPKNPATKDTFH 127
P.troglodytes YSRANEKEPCCWLLAKVRMIKGEFYVIEAACDATYNEIVTIERLRSVNPKNPATKDTFH 180
*****

M.musculus KIKLEVPEDLRQMC AKESAHKDFKKA VGAFSVTYDPENYQLVILSINEVTSKRAHMLIDM 187
R.norvegicus KIKLEVPEDLRQMC AKESAHKDFKKA VGAFSVTYDPENYQLVILSINEVTSKRAHMLIDM 187
C.familiaris KIKLDVPE DLRQMC AKESAHKDFKKA VGAFSVTYDPENYQLVILSINEVTSKRAHMLIDM 187
H.sapiens KIKLDVPE DLRQMC AKESAHKDFKKA VGAFSVTYDPENYQLVILSINEVTSKRAHMLIDM 187
P.troglodytes KIKLDVPE DLRQMC AKESAHKDFKKA VGAFSVTYDPENYQLVILSINEVTSKRAHMLIDM 240
*****:*****:***** **:*:*****:**:*:*****:****

M.musculus HFRSLR TKLSLILRN EEA SKQLESSRQLASRFHEQFIVREDLMGLAIGTHGANIQOARKV 247
R.norvegicus HFRSLR TKLSLILRN EEA SKQLESSRQLASRFHEQFIVREDLMGLAIGTHGANIQOARKV 247
C.familiaris HFRSLR TKLSLILRN EEA SKQLESSRQLASRFHEQFIVREDLMGLAIGTHGANIQOARKV 247
H.sapiens HFRSLR TKLSLILRN EEA SKQLESSRQLASRFHEQFIVREDLMGLAIGTHGANIQOARKV 247
P.troglodytes HFRSLR TKLSLILRN EEA SKQLESSRQLASRFHEQFIVREDLMGLAIGTHGANIQOARKV 300
*****:*****:*****:*****:*****:*****

KH
M.musculus PGVTAIDLDED TCTFHIYGEDQDAVKKARSFLEFAEDVIQVPRNLVGVKVIKNGKLIQEI 307
R.norvegicus PGVTAIDLDED TCTFHIYGEDQDAVKKARSFLEFAEDVIQVPRNLVGVKVIKNGKLIQEI 307
C.familiaris PGVTAIDLDED TCTFHIYGEDQDAVKKARSFLEFAEDVIQVPRNLVGVKVIKNGKLIQEI 307
H.sapiens PGVTAIDLDED TCTFHIYGEDQDAVKKARSFLEFAEDVIQVPRNLVGVKVIKNGKLIQEI 307
P.troglodytes PGVTAIDLDED TCTFHIYGEDQDAVKKARSFLEFAEDVIQVPRNLVGVKVIKNGKLIQEI 360
*****:*****:*****:*****:*****:*****

KH-1
M.musculus VDKSGVVRVRIEAENEKSVPOEEEIMPPSSLPSNNSRVGPNSSSEK KHLDTKEN-THFSQ 366
R.norvegicus VDKSGVVRVRIEAENEKSVPOEEEIMPPSSLPSNNSRVGPNSSSEK KHLDTKEN-THFSQ 366
C.familiaris VDKSGVVRVRIEAENEKSVPOEEEIMPPSSLPSNNSRVGPTPSEK KHLDIKENSTHFSQ 367
H.sapiens VDKSGVVRVRIEAENEKSVPOEEEIMPPSSLPSNNSRVGPNAP EKKHLDIKENSTHFSQ 367
P.troglodytes VDKSGVVRVRIEAENEKSVPOEEEIMPPSSLPSNNSRVGPNAP EKKHLDIKENSTHFSQ 420
*****:*****:*****:*****:*****:*****

M.musculus PNSTKVQRVLVSSIVAGGPQKPEPKAWQGMVFFVFGTKDSIANATVLLDYLHNLNLYKEV 426
R.norvegicus PNSTKVQR-----GMVFFVFGTKDSIANATVLLDYLHNLNLYKEV 405
C.familiaris PNSTKVQRVLVSSIVAGESQKPEL KAWQGMVFFVFGTKDSIANATVLLDYLHNLNLYKEV 427
H.sapiens PNSTKVQRVLVASSVAGESQKPEL KAWQGMVFFVFGTKDSIANATVLLDYLHNLNLYKEV 427
P.troglodytes PNSTKVQRVLVASSVAGESQKPE-----LKA WQEV 451
*:**

M.musculus DQLRLERLQIDEQLRQIGASSRPPPNRTDKEKGYVTDGQGMGRGSRPYRNRGHGRRGPG 486
R.norvegicus DQLRLERLQIDEQLRQIGASSRPPPNRTDKEKGYVTDGQGMGRGSRPYRNRGHGRRGPG 465
C.familiaris DQLRLERLQIDEQLRQIGASSRPPPNRTDKEKGYVTDGQGMGRGSRPYRNRGHGRRGPG 487
H.sapiens DQLRLERLQIDEQLRQIGASSRPPPNRTDKEKGYVTDGQGMGRGSRPYRNRGHGRRGPG 487
P.troglodytes DQLRLERLQIDEQLRQIGASSRPPPNRTDKEKGYVTDGQGMGRGSRPYRNRGHGRRGPG 511
*****:*****:*****:*****:*****:*****

M.musculus YTSGTNSEASNASETESDHRDELSDWSLAPTEEERESFLRRGDGRRRGGGGRRGQGGRRGRG 546
R.norvegicus YTS-----ESDHRDELSDWSLAPTEEERESFLRRGDGRRRGGGGRRGQGGRRGRG 513
C.familiaris YTS-----APTEEERENFLRRGDGRRRGGGGRRGQGGRRGRG 522
H.sapiens YTSGTNSEASNASETESDHRDELSDWSLAPTEEERESFLRRGDGRRRGGGGRRGQGGRRGRG 547
P.troglodytes YTSGTNSEASNASETESDHRDELSDWSLAPTEEERESFLRRGDGRRRGGGGRRGQGGRRGRG 571
*:* * :*****:*****:*****:*****

RGG box
M.musculus GGFKGNDDHSRTDNRPRNPREAKGRTADG-----SLQ SASSEGSRLRT 589
R.norvegicus GGFKGNDDHSRTDNRPRNPREAKGRTADG-----SLQ TSSEGSRLRT 586
C.familiaris GGFKGNDDHSRTDNRPRNPREAKGRTADGSLQIRVDCNNERSVHTKTLQNTSSEGNRLRT 582
H.sapiens GGFKGNDDHSRTDNRPRNPREAKGRTADGSLQIRVDCNNERSVHTKTLQNTSSEGSRLRT 607
P.troglodytes GGFKGNDDHSRTDNRPRNPREAKGRTADGSLQIRVDCNNERSVHTKTLQNTSSEGSRLRT 631
* :*****:*****:*****:*****:*****:*****

M.musculus GKDRNQKKEKPDSDVGLQPLVNGVP 614
R.norvegicus GKDRNQKKEKPDSDVGLQPLVNGVP 581
C.familiaris GKDRNQKKEKPDSDVGLQPLVNGVP 607
H.sapiens GKDRNQKKEKPDSDVGLQPLVNGVP 632
P.troglodytes GKDRNQKKEKPDSDVGLQPLVNGVP 656
**:*:*****:*****:*****

```

Figura 4.1: Alinhamento global entre as proteínas homólogas da FMRP encontradas em diferentes organismos obtido a partir do programa ClustalW. As seqüências grifadas correspondem a alguns domínios também conservados.

Os resultados desta comparação são apresentados na Figura 4.2, onde os picos representam o grau de conservação destas regiões nas diferentes comparações: (A) seqüência humana alinhada com a seqüência do chimpanzé, (B)

seqüência humana alinhada com a seqüência do cachorro, (C) seqüência humana alinhada com a seqüência do rato e (D) seqüência humana alinhada com a seqüência do camundongo. Nota-se nestas comparações que os éxons do gene, representados na cor azul escura, são muito conservados entre as espécies, assim como as suas regiões 5' UTR e 3' UTR, marcadas em azul claro.

Do total de regiões conservadas entre as espécies estudadas, buscaram-se aquelas que são conservadas em todas as cinco espécies e que apresentam um grau de conservação maior que 75%. Como parâmetro de comparação, utilizaram-se, também, os resultados obtidos com o uso do programa rVista, que combina o alinhamento obtido pelo mVista com uma pesquisa no principal banco de dados de sítios de ligação para fatores de transcrição, TRANSFAC Professional, para localizar dentro destas regiões conservadas os possíveis sítios de ligação para fatores de transcrição conhecidos.

Este método *in silico*, chamado de *phylogenetic footprinting*, foi importante para auxiliar na escolha das regiões do DNA que foram estudadas na segunda etapa da pesquisa. Um resultado típico do rVista é apresentado na Figura 4.3, que mostra o alinhamento entre uma seqüência de DNA humana e sua homóloga no rato, localizada entre +900 pb e -19.000 pb, tendo como referência a primeira base do códon de início do gene. Os resultados do rVista indicaram várias seqüências de consenso reconhecíveis por proteínas de transcrição presentes nestas regiões conservadas. Destas seqüências, algumas se destacam por estarem localizadas em regiões com alto nível de conservação entre todas as espécies estudadas. Todos os possíveis sítios de transcrição encontrados nesta seqüência, que podem ser reconhecidos por proteínas cadastradas no TRANSFAC, estão identificados através dos traços verticais. O nome das proteínas está listado ao lado esquerdo da figura.

Os sítios de interação encontrados nas regiões conservadas apresentados como resultado do programa rVista são um indicador significativo da relevância destas seqüências de DNA que se encontram conservadas ao longo da evolução. Pode-se considerar que qualquer uma das proteínas listadas à esquerda no gráfico é potencialmente um fator de transcrição para o gene *FMR1*.

Porém, muitas proteínas que atuam como fatores de transcrição não foram ainda identificadas e listadas na base de dados TRANSFAC e, portanto, seus sítios de interação não são apresentados aqui, o que indica que o resultado do programa rVista é um indicativo conservador, se não consideramos os possíveis resultados falso-positivos.

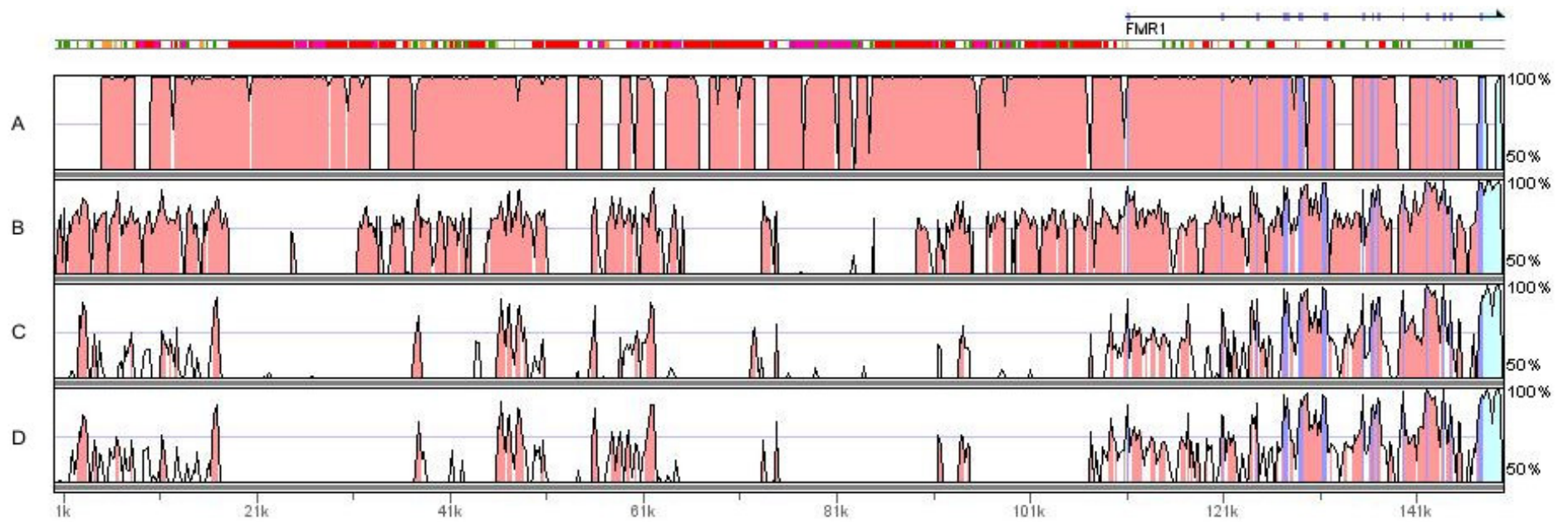


Figura 4.2: Resultado de alinhamento global obtido com o programa mVista. As quatro linhas representam o alinhamento global entre a seqüência à montante e codificante do gene *FMR1* humano alinhada com a seqüência homóloga: (A) *Pan troglodytes*, (B) *Canis familiaris*, (C) *Rattus norvegicus* e (D) *Mus musculus*. Os picos indicam os níveis de conservação ao longo das seqüências. As regiões marcadas de azul correspondem aos éxons do gene *FMR1* humano e dos genes homólogos. Cada linha apresenta a comparação da seqüência genômica de uma espécie com a seqüência humana de interesse.

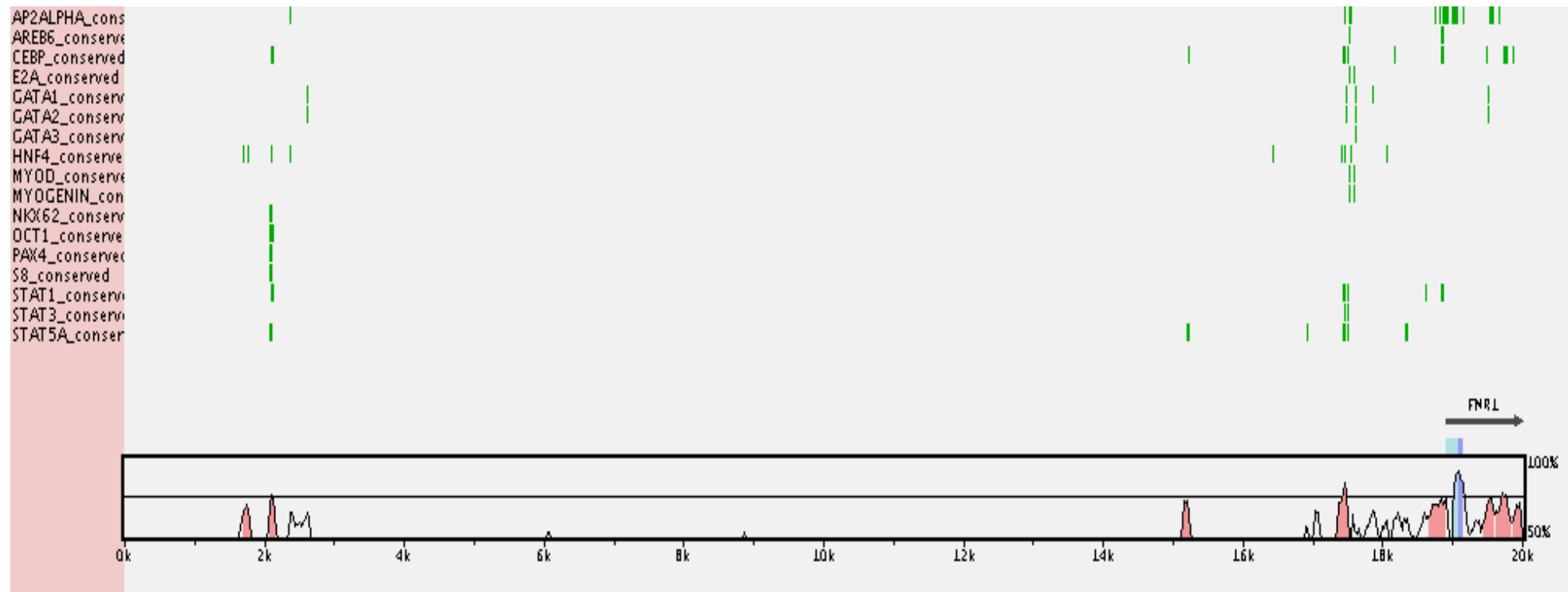


Figura 4.3: Resultados de *Phylogenetic footprinting* obtidos com o programa rVista no estudo da região imediatamente à montante do gene (20 kpb). Os TFBS mais conservados estão marcados na posição relativa em cada pico e o nome das proteínas que reconhecem esses sítios de interação estão apresentados ao lado.

Utilizando-se desta metodologia verificou-se toda a seqüência de 111.000 pb à montante do gene, a fim de identificar as regiões conservadas que mais apresentavam TFBS, segundo os critérios do rVista e com grau de conservação elevado (+75%), dado o alinhamento obtido do mVista. A junção destes parâmetros levou à identificação de cinco potenciais regiões conservadas (S1, S2, S3a/b, S4 e S5), que estão apresentadas na Figura 4.4. A figura apresenta o alinhamento global entre a seqüência humana e a seqüência do rato, incluindo o gene completo e a região de 111 kpb à montante do gene.

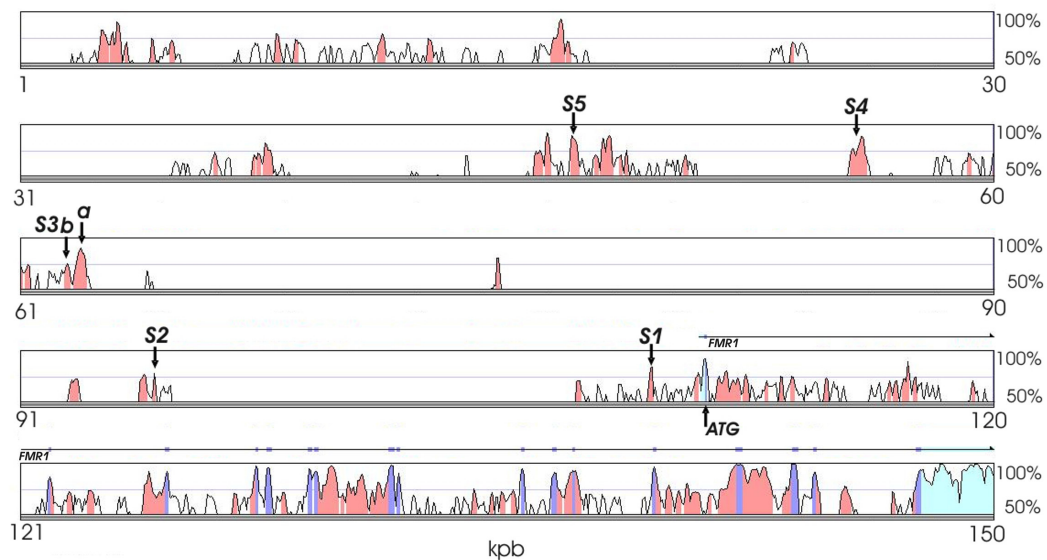


Figura 4.4: Alinhamento global entre a seqüência humana e a seqüência do rato (*Rattus norvegicus*), incluindo o gene completo e a região de 111 kpb à montante do gene *FMR1*, perfazendo um total de 150 kpb. Os picos indicam os níveis de conservação ao longo das seqüências, variando de 50 a 100%. As regiões marcadas de azul escuro correspondem aos éxons do gene *FMR1* humano e do gene homólogo. As regiões marcadas em azul claro correspondem às regiões 5' UTR e 3' UTR. As setas indicam as cinco regiões conservadas que foram objeto dos estudos subseqüentes (S1, S2, S3 a/b, S4 e S5). O códon de início (ATG) do gene também está indicado com uma seta.

Para a comprovação experimental das evidências obtidas *in silico*, foram escolhidas as comparações entre as seqüências humana e do *Rattus norvegicus*, que foi o animal modelo do qual foram extraídos os componentes necessários para os experimentos.

Os alinhamentos globais obtidos através do uso do programa mVista para as cinco regiões escolhidas são apresentados nas Figuras 4.5 a 4.9.

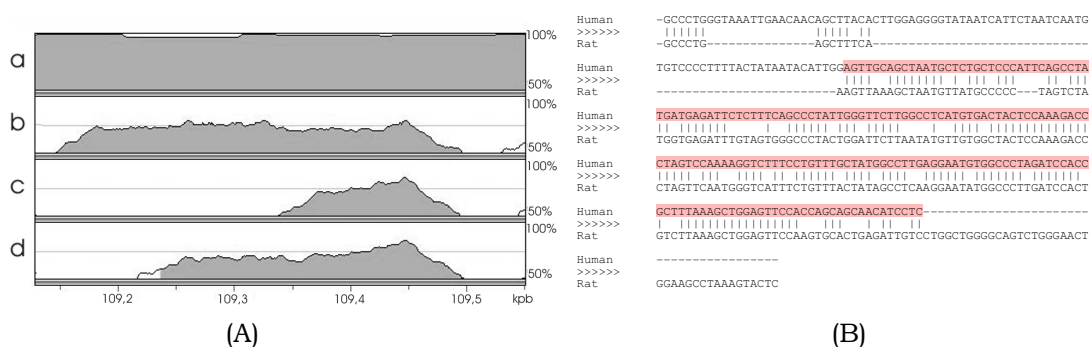


Figura 4.5: Resultados de genômica comparativa obtidos através do programa mVista, para a região S1. (A) Representação gráfica dos alinhamentos globais obtidos através do uso do programa mVista da região S1 do genoma humano em função da sua região homóloga em (a) *Pan troglodites*, (b) *Canis familiaris*, (c) *Rattus norvegicus* e (d) *Mus musculus*. (B) Alinhamento da seqüência S1 humana e do rato. Os nucleotídeos em destaque fazem parte da região também indicada em A.

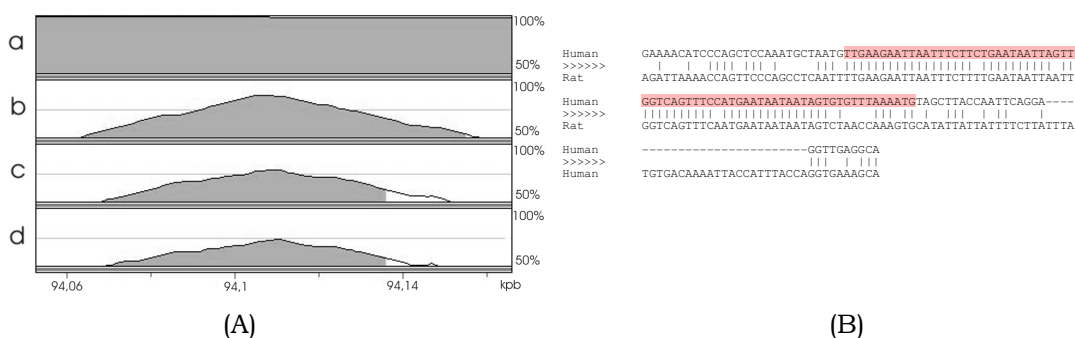


Figura 4.6: Resultados de genômica comparativa obtidos através do programa Vista, para a região S2. (A) Representação gráfica dos alinhamentos globais obtidos através do uso do programa mVista da região S2 do genoma humano em função da sua região homóloga em (a) *Pan troglodites*, (b) *Canis familiaris*, (c) *Rattus norvegicus* e (d) *Mus musculus*. (B) Alinhamento da seqüência S2 humana e do rato. Os nucleotídeos em destaque fazem parte da região também indicada em A.

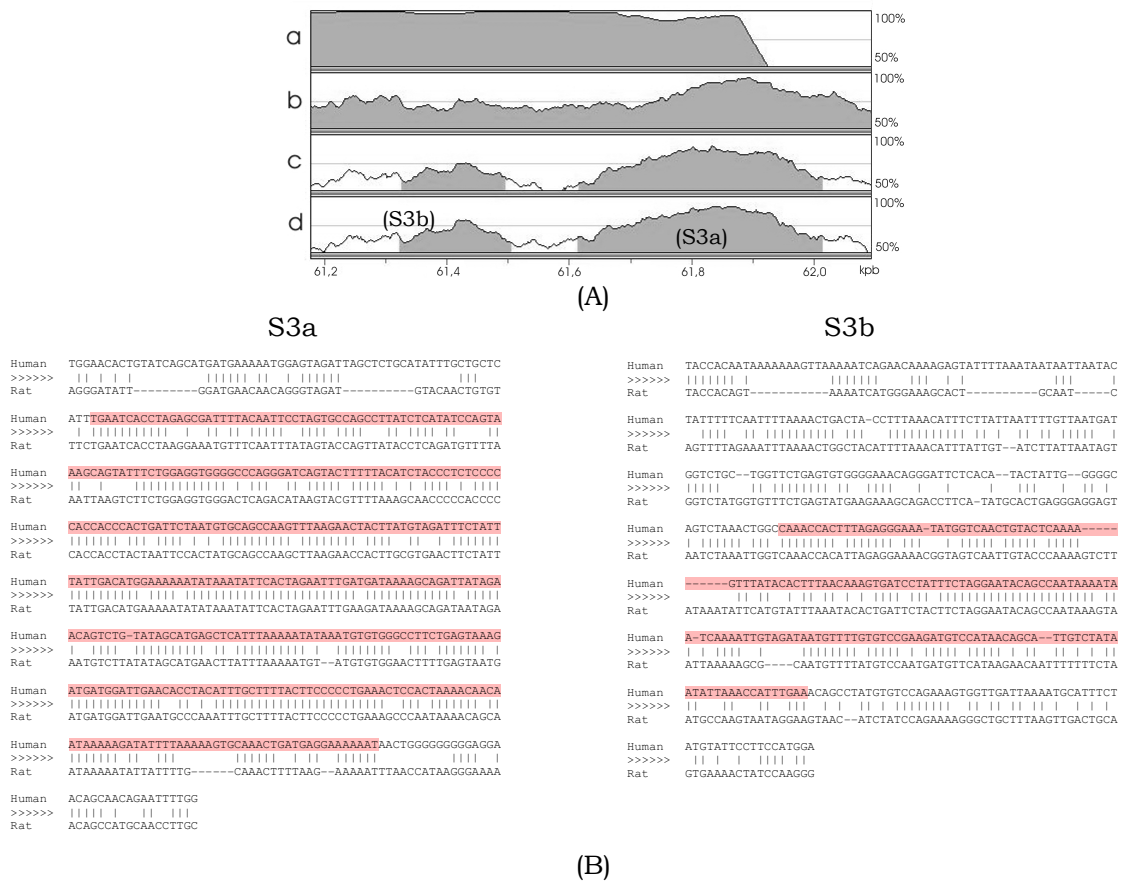


Figura 4.7: Resultados de genômica comparativa obtidos através do programa Vista, para a região S3. (A) Representação gráfica dos alinhamentos globais obtidos através do uso do programa mVista da região S3 do genoma humano em função da sua região homóloga em (a) *Pan troglodites*, (b) *Canis familiaris*, (c) *Rattus norvegicus* e (d) *Mus musculus*. (B) Alinhamento da seqüência S3 humana e do rato. Os nucleotídeos em destaque fazem parte da região também indicada em A.

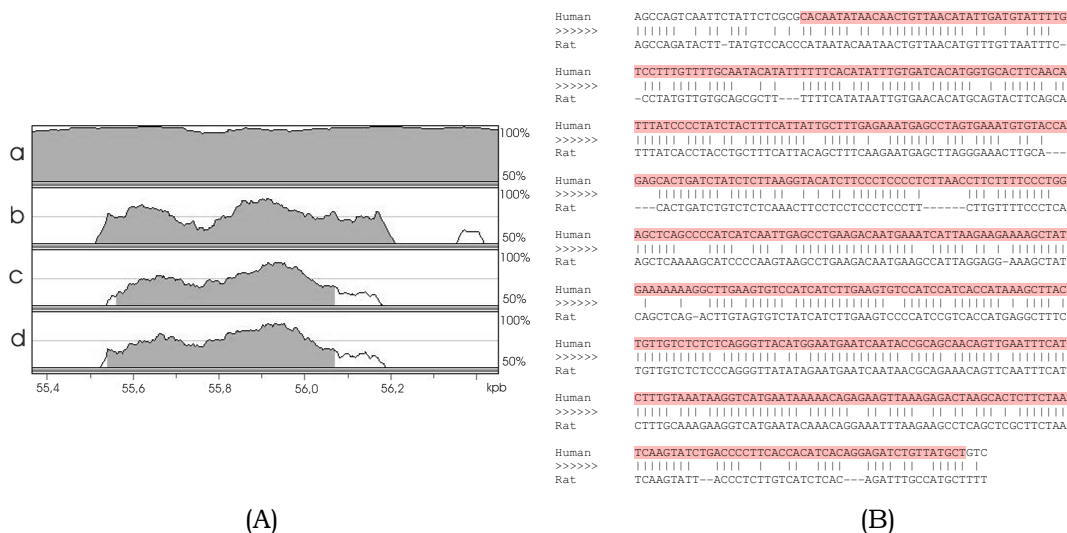


Figura 4.8: Resultados de genômica comparativa obtidos através do programa Vista, para a região S4. (A) Representação gráfica dos alinhamentos globais obtidos através do uso do programa mVista da região S4 do genoma humano em função da sua região homóloga em (a) *Pan troglodites*, (b) *Canis familiaris*, (c) *Rattus norvegicus* e (d) *Mus musculus*. (B) Alinhamento da seqüência S4 humana e do rato. Os nucleotídeos em destaque fazem parte da região também indicada em A.

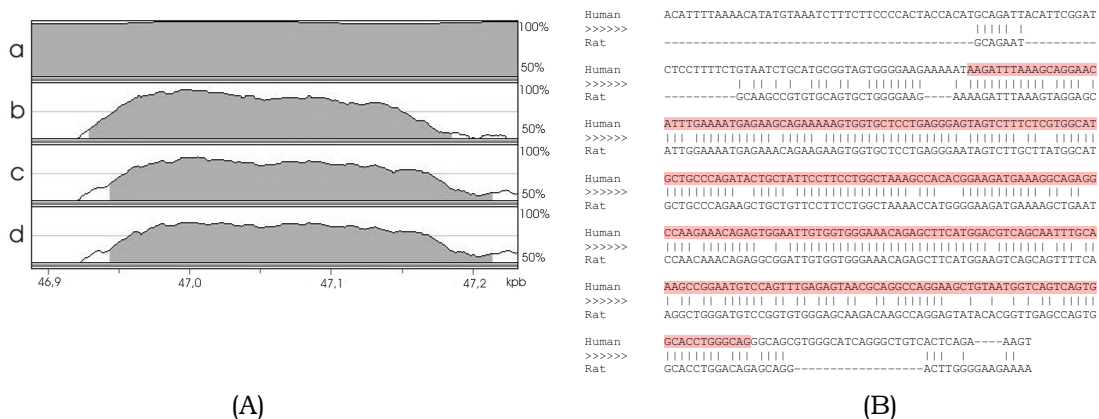


Figura 4.9: Resultados de genômica comparativa obtidos através do programa Vista, para a região S5. (A) Representação gráfica dos alinhamentos globais obtidos através do uso do programa mVista da região S5 do genoma humano em função da sua região homóloga em (a) *Pan troglodites*, (b) *Canis familiaris*, (c) *Rattus norvegicus* e (d) *Mus musculus*. (B) Alinhamento da seqüência S5 humana e do rato. Os nucleotídeos em destaque fazem parte da região também indicada em A.

Esses resultados mostram que uma análise da região reguladora de genes, apenas utilizando genômica comparativa e programas de reconhecimento de padrões de seqüências como o rVista pode ser uma ferramenta poderosa para o direcionamento dos estudos genômicos. No entanto, muitos dos resultados encontrados, utilizando-se técnicas puramente computacionais e comparativas, são falsos positivos, e a única maneira de descartá-los é a associação destes resultados com testes laboratoriais *in vitro* e/ou *in vivo*, que possam comprovar essas evidências.

4.2. Interação DNA–proteína nas seqüências conservadas

A análise experimental para determinação dos sítios de interação com proteínas de transcrição nas cinco regiões indicadas pelo estudo *in silico*, foi realizada em sua maior parte no Department of Biochemistry and Molecular Medicine, University of California, School of Medicine, Davis, Califórnia, Estados Unidos.

As cinco seqüências do DNA correspondentes aos alinhamentos apresentados nas Figuras 4.5 a 4.9 foram amplificadas pela técnica de Polymerase Chain Reaction – PCR (Reação em Cadeia da Polimerase). Esta técnica permite criar múltiplas cópias da seqüência de DNA (oligo) a ser estudada, utilizando a enzima DNA polimerase sem o uso de um organismo vivo. Para isso foram desenhados pares de *primers* (iniciadores) específicos para cada seqüência a ser amplificada. No caso da seqüência 3 (S3), houve a necessidade de dividi-la em duas partes (S3a e S3b), devido ao seu tamanho (872 pb) que ultrapassou os limites experimentais

relacionados à atuação da enzima. Sendo assim, foram criados 6 pares de *primers*, que anelam perfeitamente com o início e com o final das seqüências a serem amplificadas:

Após otimização do processo de amplificação das seqüências alvo foram obtidos os produtos finais dos PCR. A Figura 4.10 apresenta os produtos dos PCR em gel de agarose 1% corados com brometo de etídio.

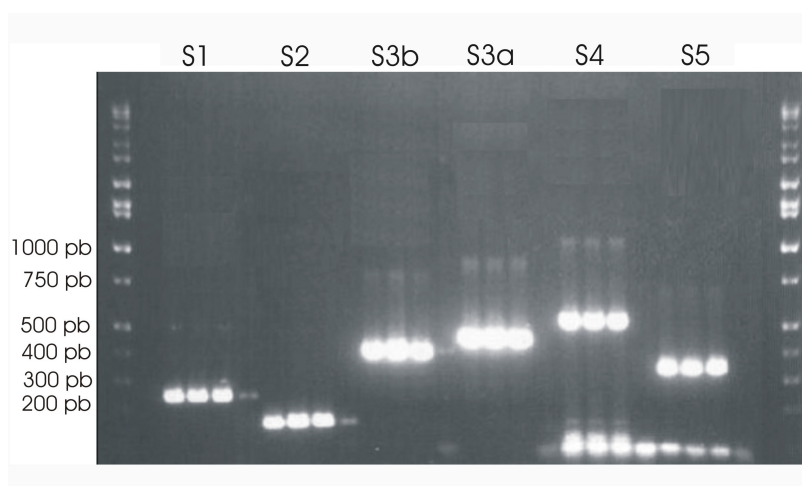


Figura 4.10: Produtos dos PCR realizados para amplificação dos oligonucleotídeos alvo, em gel de agarose 1% corado com brometo de etídio. As bandas apresentadas em triplicata correspondem as seqüências 1 a 5 conforme indicação na parte superior do gel.

Os produtos da amplificação, S1, S2, S3a, S3b, S4 e S5, resultantes dos PCR, correspondem aos tamanhos de 243, 153, 469, 403, 553 e 351 pares de bases, respectivamente.

As seqüências de DNA resultantes dos PCR foram então marcadas com digoxigenina (DIG), uma lactona esteróide cristalina ($C_{23}HO_5$), obtida pela hidrólise da digoxina. A digoxigenina vem sendo utilizada como marcador para oligonucleotídeos em substituição ao fósforo marcado radioativamente. As seqüências de nucleotídeos marcadas foram utilizados para a realização de experimentos de *gel shift* para identificação da interação DNA/proteína. Promoveu-se a reação de interação, entre proteínas nucleares extraídas de cérebro de ratos Wistar e as seqüências alvo, com subsequente separação dos complexos formados, por eletroforese nativa em gel de acrilamida 5%. Os resultados destes experimentos estão apresentados na Figura 4.11.

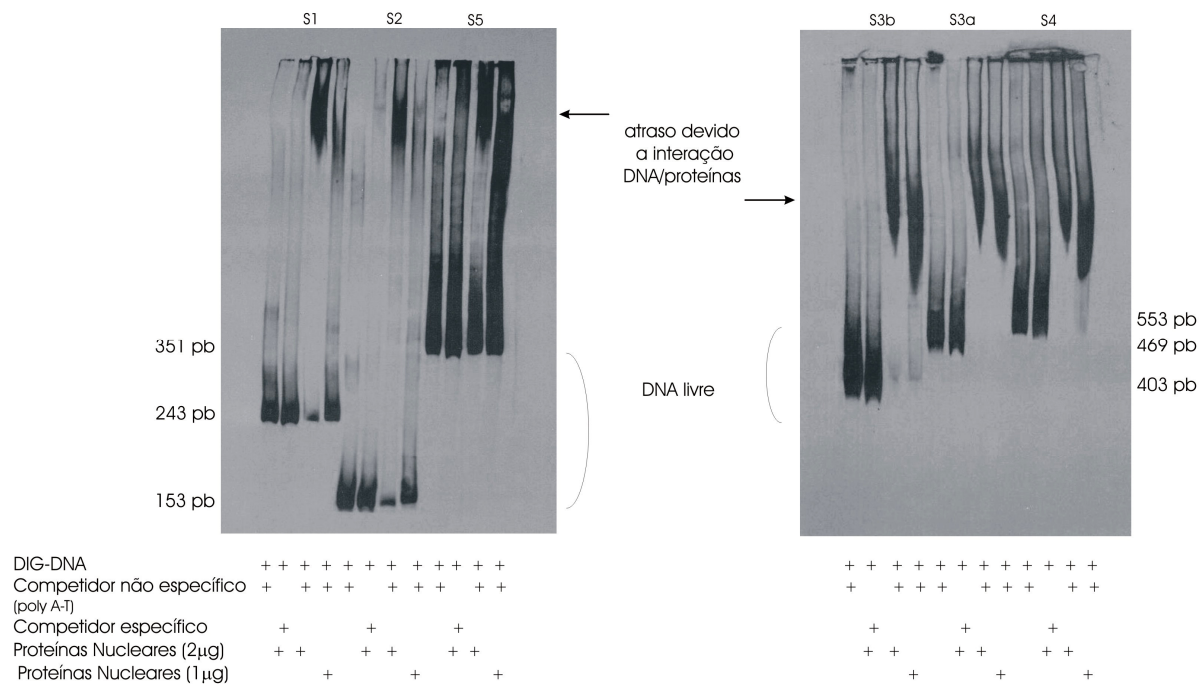


Figura 4.11: Resultado dos experimentos de *gel shift* para identificação da formação dos complexos DNA/proteína entre as proteínas nucleares e os oligonucleotídeos alvo (S1,S2, S3a, S3b, S4 e S5). Os tamanhos dos oligonucleotídeos livres estão indicados ao lado de cada gel. As setas indicam o atraso da migração no gel de acrilamida devido à interação dos oligonucleotídeos com proteínas nucleares. O esquema apresentado abaixo dos géis indica a composição de cada uma das reações de ligação. A detecção foi feita por quimioluminescência por anticorpo anti-DIG marcado com fosfatase alcalina.

O atraso na migração das seqüências de nucleotídeos ocorreu devido à sua interação com as proteínas nucleares. As reações de interação foram compostas pelos produtos dos PCR marcados com digoxigenina (oligos alvo), proteínas nucleares em duas concentrações diferentes (1 e 2 µg/reacção) e oligonucleotídeos competidores específicos ou não específicos, conforme o esquema apresentado na Figura 4.11. Como competidores específicos foram utilizados os próprios produtos das amplificações (S1, S2, S3a, S3b, S4 e S5, respectivamente, em concentração 100× maior do que as dos produtos marcados, porém sem marcação com DIG. Como competidor não específico foi adicionado à reacção de ligação um oligonucleotídeo de poli-(AT)₂₅, também em concentração 100× superior às seqüências. A adição de competidor não específico tem o objetivo de eliminar as interações não específicas durante a ligação DNA/proteína).

Todas as seqüências estudadas mostraram afinidade por proteínas nucleares, uma vez que todas as reações contendo oligonucleotídeo marcado, proteínas nucleares, e competidor não específico, apresentaram um atraso da migração no gel. As reações contendo competidor específico mostraram que a interação DNA/proteína é específica para aquela seqüência, uma vez que este competidor, em maior concentração, competiu com o oligonucleotídeo alvo pela

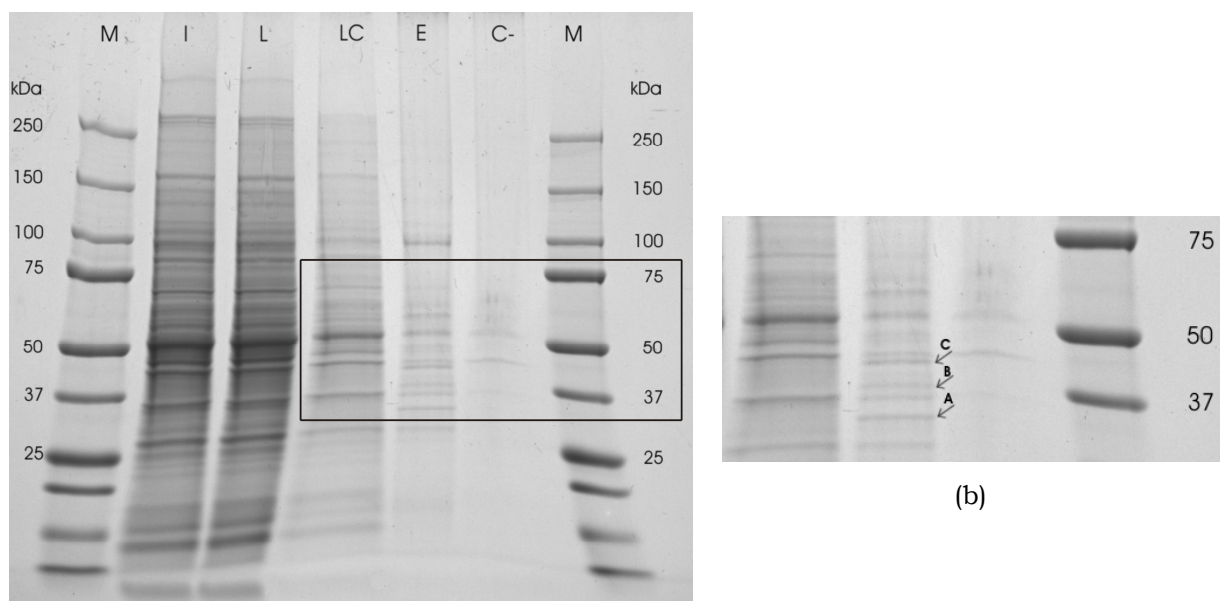
interação com as proteínas e o deslocou, permitindo que o oligonucleotídeo marcado migrasse livre no gel.

Não foi possível identificar bandas nítidas na migração dos complexos DNA/proteínas, o que pode indicar que mais de um tipo de proteína interagiu com o DNA, formando um complexo multiprotéico. No entanto, pode-se afirmar que todas as cinco seqüências têm potencial para abrigar sítios de interação para proteínas, sejam elas fatores de transcrição ou não.

4.3. Purificação e identificação das proteínas nucleares que se ligam especificamente aos segmentos de DNA conservados

Conforme descrito anteriormente, para identificar as proteínas que interagiram com os oligonucleotídeos alvo foram realizados experimentos de purificação através da precipitação magnética por afinidade. Nestes experimentos, as seqüências de DNA em estudo foram amplificadas utilizando-se um dos *primers* ligados à molécula de biotina (*primers* biotilados). Os produtos biotilados resultantes destas ampliações foram imobilizados em esferas metálicas contendo streptavidina covalentemente ligada. As esferas contendo o DNA alvo foram utilizadas como matriz para a separação das proteínas nucleares que apresentam afinidade pelo oligonucleotídeo especificamente ligado. Após as reações de ligação entre proteínas e DNA alvo imobilizados nas esferas metálicas, os complexos formados foram magneticamente precipitados e lavados, e as proteínas que se ligaram especificamente ao oligonucleotídeo alvo foram eluídas com tampão contendo altas concentrações de KCl. As proteínas eluídas foram dessalinizadas e separadas em gel gradiente de acrilamida 4 a 15% em eletroforese SDS-PAGE. As Figuras 4.12 e 4.13 apresentam as eletroforeses das purificações das proteínas nucleares ligadas aos oligonucleotídeos das seqüências 1 e 2 (S1 e S2), respectivamente.

Depois da separação por eletroforese SDS-PAGE, algumas proteínas eluídas com tampão contendo alta concentração de KCl e que não foram identificadas no controle negativo (C-), foram cortadas do gel de acrilamida e sujeitas à digestão *in gel* por tripsina. Para esta etapa foram escolhidas as proteínas marcadas com as setas (A, B, C e X, Y e Z nas Figuras 4.12 e 4.13, respectivamente). Os peptídeos resultantes da digestão foram submetidos à análise de espectrometria de massa para identificação das proteínas.



(a)

Figura 4.12: (a) Eletroforese SDS-PAGE, em gel de poliacrilamida e corado com "Comassie Blue", representativa da purificação das proteínas nucleares que se ligam especificamente à seqüência S1 estudada. Faixas: (I) proteínas nucleares totais, (L) proteínas nucleares que saíram na lavagem com tampão de ligação, (LC) proteínas nucleares que saíram na lavagem com tampão contendo competidor não-específico (oligo poli-(AT)₂₅), (E) proteínas eluídas pelo tampão de eluição com alta concentração de KCl, (C-) controle negativo: proteínas que adsorveram nas esferas sem DNA imobilizado e foram eluídas com o tampão de eluição, (M) marcadores de peso molecular. (b) Aproximação da eletroforese correspondente ao retângulo marcado em (a). As setas A, B e C indicam proteínas que adsorveram especificamente nas esferas contendo oligonucleotídeo alvo e não adsorveram no controle negativo ou saíram na lavagem com competidor não-específico. Estas proteínas foram cortadas do gel e submetidas à espectrometria de massa para identificação.

Os resultados obtidos através da espectrometria de massa foram analisados com o auxílio do programa X! Tandem, que combina os espectros de massa obtidos, com os espectros de seqüências de peptídeos já conhecidos e disponíveis em bancos de dados proteômicos. Com isto foi possível identificar as proteínas, candidatas a atuarem como fatores de transcrição, a partir de sua interação com os oligonucleotídeos alvo.

As principais proteínas identificadas nas bandas A, B e C, que aparecem na Figura 4.12, estão listadas nas Tabelas 4.1, 4.2 e 4.3, respectivamente. Estas proteínas foram identificadas com a utilização do servidor de dados proteômicos The Global Proteome Machine (<http://gpmdb.thegpm.org/>), que utiliza várias bases de dados proteômicos, a fim de correlacionar os espectros obtidos dos peptídeos resultantes da digestão das proteínas em estudo, com outros já identificados anteriormente. Todas as proteínas listadas nestas tabelas tiveram mais de um peptídeo único identificado e log(e) menor que -10. O Log(e) indica a probabilidade de a proteína identificada ser ou não um resultado randômico. Quanto mais negativo for este número, maior será a probabilidade de o resultado ser verdadeiro.

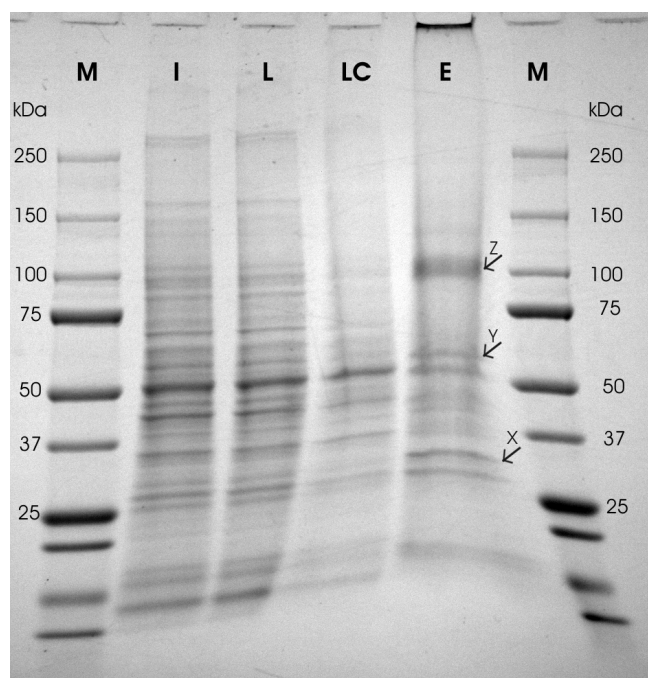


Figura 4.13: (a) Eletroforese SDS-PAGE, em gel de poliacrilamida e corado com "Comassie Blue", representativa da purificação das proteínas nucleares que se ligam especificamente à seqüência S2 estudada. Faixas: (I) proteínas nucleares totais, (L) proteínas nucleares que saíram na lavagem com tampão de ligação, (LC) proteínas nucleares que saíram na lavagem com tampão contendo competidor não-específico (oligo poli-(AT)₂₅), (E) proteínas eluídas pelo tampão de eluição com alta concentração de KCl, (M) marcadores de peso molecular. As setas X, Y e Z indicam as proteínas que adsorveram especificamente nas esferas contendo oligonucleotídeo alvo e não adsorveram no controle negativo ou saíram na lavagem com competidor não-específico. Estas proteínas foram cortadas do gel e submetidas à espectrometria de massa para identificação.

A Tabela 4.1 mostra as cinco proteínas mais promissoras encontradas pela análise de espectrometria de massa na banda A. Esta banda corresponde às proteínas de peso molecular entre 30 e 35 kDa, o que é consistente com o peso molecular de todas as proteínas listadas. O número de peptídeos únicos encontrados para cada proteína identificada e o percentual de cobertura são dados importantes que indicam a relevância dessas proteínas na amostra.

Tabela 4.1: Cinco principais proteínas identificadas por espectrometria de massa a partir da Banda A indicada na Figura 4.12.

Proteínas	Massa Molec. (kDa)	% Cobertura *	Peptídeos Únicos **
Sub-unidade 4 do complexo THO (THO4)	27.0	18	3
Tubulina alfa 2	50.1	16	5
Fator eucariótico de alongação de tradução 1 beta 2	24.7	15	3
Proteína ativadora da transcrição Pur-beta	34.9	15	2
Tubulina beta 2C	49.8	12	4

* percentual de resíduos nos peptídeos identificados

** número de peptídeos únicos identificados

É comum a existência de contaminantes e outras proteínas ou pedaços de proteínas presentes na mesma banda de uma eletroforese, principalmente sendo esta uma eletroforese de apenas uma dimensão. No entanto, o número de peptídeos únicos, o percentual de cobertura e o log(e), contribuem para a confirmação dos resultados apresentados.

A banda B corresponde às proteínas com peso molecular aproximado de 38 a 40 kDa e os resultados mais relevantes obtidos pela análise de espectrometria de massa desta amostra estão listados na Tabela 4.2.

Tabela 4.2: Seis principais proteínas identificadas por espectrometria de massa a partir da Banda B indicada na Figura 4.12.

Proteínas	Massa Molec. (kDa)	% Cobertura*	Peptídeos Únicos**
Proteína ativadora da transcrição Pur-beta	33.5	44	6
Sub-unidade beta-2 do canal de potássio	41.0	36	9
Proteína ativadora da transcrição Pur-alfa	34.9	26	6
Cadeia beta da tubulina	49.9	22	8
Ribonucleoproteínas heterogêneas nucleares A2/B1	37.3	19	6
Cadeia leve LC2 da proteína associada a microtúbulos 1A	112.3	9,1	8

* percentual de resíduos nos peptídeos identificados

** número de peptídeos únicos identificados

Todas as proteínas listadas têm seu peso molecular consistente com a faixa na qual se encontravam no gel, com exceção da cadeia leve LC2 da proteína MAP1A, que foi identificada e que tem um peso molecular de 112 kDa. Acredita-se que neste caso o peptídeo identificado é resultante de uma forma truncada da proteína, proveniente de degradação ou tradução inacabada do peptídeo completo. Os ativadores de transcrição, proteínas Pur- α e Pur- β , foram identificados nesta amostra. Ambas apresentaram alto percentual de cobertura e seis peptídeos únicos cada uma. Os resultados de espectrometria de massa referentes aos peptídeos listados na Tabela 4.2 são apresentados na Figura 4.14. Alguns dos peptídeos das proteínas Pur- α , Pur- β e um da proteína hnRNP A2/B1 foram identificados.

Os picos foram normalizados para facilitar a comparação entre eles. A intensidade relativa (IR) dos picos das proteínas Pur- α e β e da proteína hnRNP A2/B1 confirmam que estas proteínas eram abundantes nas amostras testadas.

A Tabela 4.3 mostra as proteínas identificadas a partir da banda C, do gel de eletroforese.

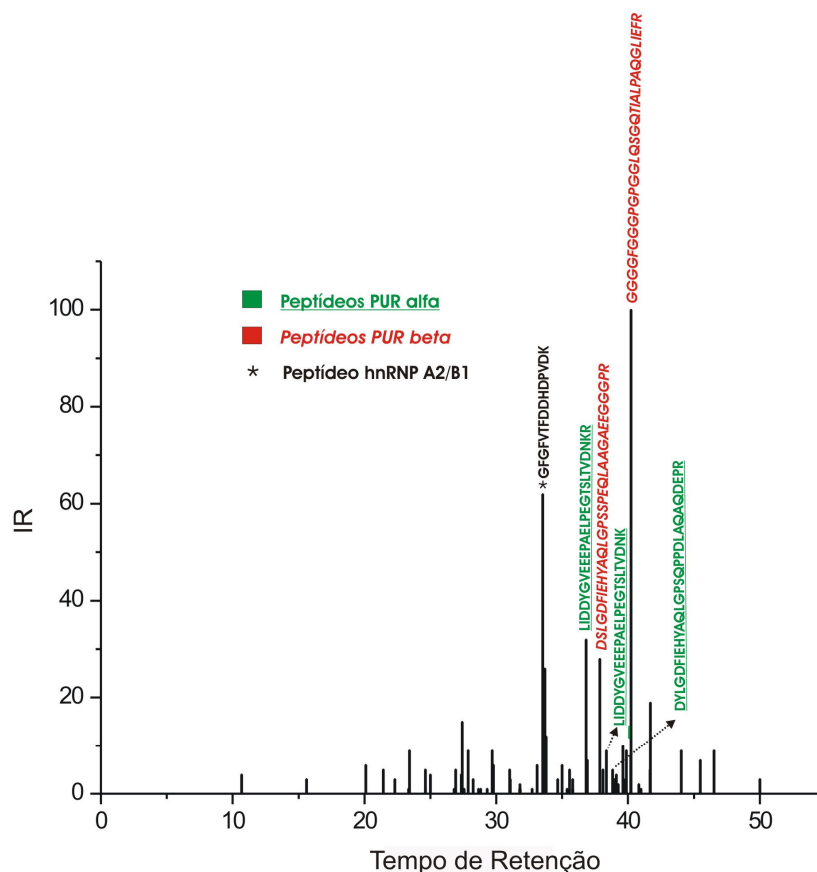


Figura 4.14: Análise de espectrometria de massa das proteínas da banda B. Alguns peptídeos das proteínas Pur- α and Pur- β estão identificados. (*) Pico relativo à proteína hnRNP A2/B1. IR é a intensidade relativa.

A banda C corresponde às proteínas com peso molecular ao redor de 45 kDa. Nesta banda também foram identificados peptídeos pertencentes à proteína Pur- α , sendo inclusive a proteína mais abundante identificada. A proteína associada a microtúbulo (MAP1A) também foi identificada, porém seu peso molecular não corresponde à faixa de peso esperada para a banda C. Novamente explica-se este caso como sendo peptídeos resultantes de uma proteína truncada.

Tabela 4.3: Cinco principais proteínas identificadas por espectrometria de massa a partir da Banda C indicada na Figura 4.12.

Proteínas	Massa Molec. (kDa)	% Cobertura *	Peptídeos Únicos **
Proteína ativadora da transcrição Pur-alfa	34.9	36	12
Actina gama 1	41.8	34	12
Actina alfa 1	42.0	15	3
Tubulina beta 3	50.4	11	4
Proteína associada a microtúbulo 1 A	324.3	3,3	7

* percentual de resíduos nos peptídeos identificados

** número de peptídeos únicos identificados

três proteínas em especial chamaram a atenção: as proteínas **Pur- α** e **Pur- β** e a proteína **hnRNP**.

Tabela 4.4: Principais proteínas das bandas X, Y e Z identificadas por espectrometria de massa.

	Proteínas	Massa Molec. (kDa)	% Cobertura *	Peptídeos Únicos **
Banda X	Tubulina beta 2	49.8	58	21
	Proteína ribossomal 40S S3	26.7	40	8
	ATPase, transportadora de H ⁺	28.3	40	8
	Similar a tubulina alfa-2	50.1	36	12
	Gliceraldeído-3-fosfate desidrogenase	35.8	33	9
Banda Y	Tubulina beta 2	49.8	70	28
	Tubulina beta Neurônio-específica classe III	49.9	24	6
	Proteína associada a microtúbulo 1A (MAP 1A)	299.5	23	55
	Proteína associada a microtúbulo 1B (MAP 1B)	269.5	20	32
Banda Z	Tubulina beta-5	49.6	20	6
	Dinamina-1	97.5	36	31
	Proteína associada a microtúbulo 1A (MAP 1A)	299.5	25	63
	Proteína associada a microtúbulo 1B (MAP 1B)	269.5	23	67
	Dinamina-3	98.0	4,8	3

* percentual de resíduos nos peptídeos identificados

** número de peptídeos únicos identificados

Para confirmar a presença da proteína Pur- α , como uma das proteínas que se ligam à seqüência S1 estudada, realizou-se uma análise de imunodeteção, contra o anticorpo anti-Pur- α das amostras obtidas durante o processo de purificação. Conforme apresentado na Figura 4.16, o resultado desta análise confirmou a presença da referida proteína na solução protéica obtida após a purificação, o que valida ainda mais o resultado obtido pela espectrometria de massa.

A proteína Pur- α é uma proteína que se liga tanto às fitas simples como às fitas duplas de DNA de forma específica. Ela tem preferência por regiões ricas em purinas, numa seqüência, que foi definida por alguns autores, como repetições do trinucleotídeo GGN (Gallia *et al.*, 2000; Wortman *et al.*, 2005), onde N pode ser qualquer nucleotídeo.

Segundo os estudos de Khalili e colaboradores (2003), realizado com camundongos *knockout* para o gene da proteína Pur- α , esta proteína é essencial para o desenvolvimento pós-natal do cérebro e também está relacionada à proliferação celular. Os camundongos modificados aparentam normalidade no momento do nascimento, porém após duas semanas de idade, começam a manifestar severos problemas de tremor e crises espontâneas. Estes indivíduos,

com expressão nula da proteína Pur- α , também possuem baixíssimo número de células neuronais na região do hipocampo e cerebelo. Estes fenótipos vêm de encontro com aqueles típicos apresentados pelos pacientes afetados pela Síndrome do X-Frágil, permitindo-se estabelecer uma relação entre Pur- α e a síndrome.

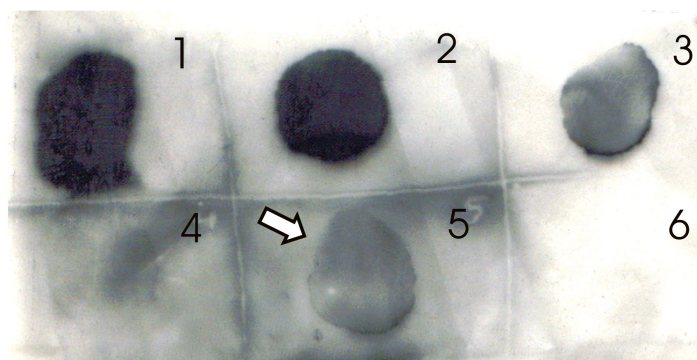


Figura 4.16: Análise de imunodeteção da proteína Pur- α . Amostras (dots): 1 - Solução de proteínas nucleares extraídas das células de ratos Wistar; 2 - Proteínas que não aderiram à seqüência de DNA S1; 3 - Proteínas que saíram na lavagem (com tampão de ligação) durante a purificação das proteínas que se ligam à seqüência S1; 4 - Proteínas que saíram na lavagem com tampão contendo alta concentração de inibidor inespecífico (oligo poli-AT); 5 - Proteínas que eluíram com tampão de eluição (proteínas purificadas) após ligaram-se especificamente à seqüência S1; 6 - Controle negativo, proteínas que se ligaram às esferas metálicas sem DNA.

Kobayashi e colaboradores (2000) sugerem, através de seus estudos, que a proteína Pur- α associa-se com o RNA BC1 e está envolvida com a regulação da transcrição do gene deste RNA. O RNA BC1 é preferencialmente expresso nas células neuronais e forma partículas de ribonucleoproteínas (RNP) nos dendritos dos neurônios. Da mesma forma, Zalfa e colaboradores (2003) identificaram que a proteína FMRP também associa-se ao RNA BC1 e regula a transcrição de RNA mensageiros específicos nas sinapses.

Em outro estudo, Jin e colaboradores (2007) identificaram que as proteínas Pur- α e nhRNP A2/B1 ligam-se de forma específica às repetições CGG do RNA do gene *FMR1* e aparentemente trafegam pelo citoplasma ligadas a ele fazendo parte do complexo de ribonucleoproteínas (Ohashi *et al.*, 2002).

Aliando-se as conclusões destes autores com os resultados obtidos neste trabalho, pode-se propor um modelo da regulação da transcrição do gene *FMR1* pela proteína Pur- α ou Pur- β , conforme apresentado na Figura 4.17. Neste modelo as proteínas Pur- α ou Pur- β , ou ainda um dímero das duas proteínas, se ligam nos dois pontos do DNA, região promotora com repetições CGG e seqüência S1, e atuam como um amplificador ou inibidor do acoplamento da RNA polimerase, através da

formação de um *looping* (alça) na fita de DNA. Esta estrutura tridimensional do DNA é típica para formação de complexos de iniciação da transcrição.

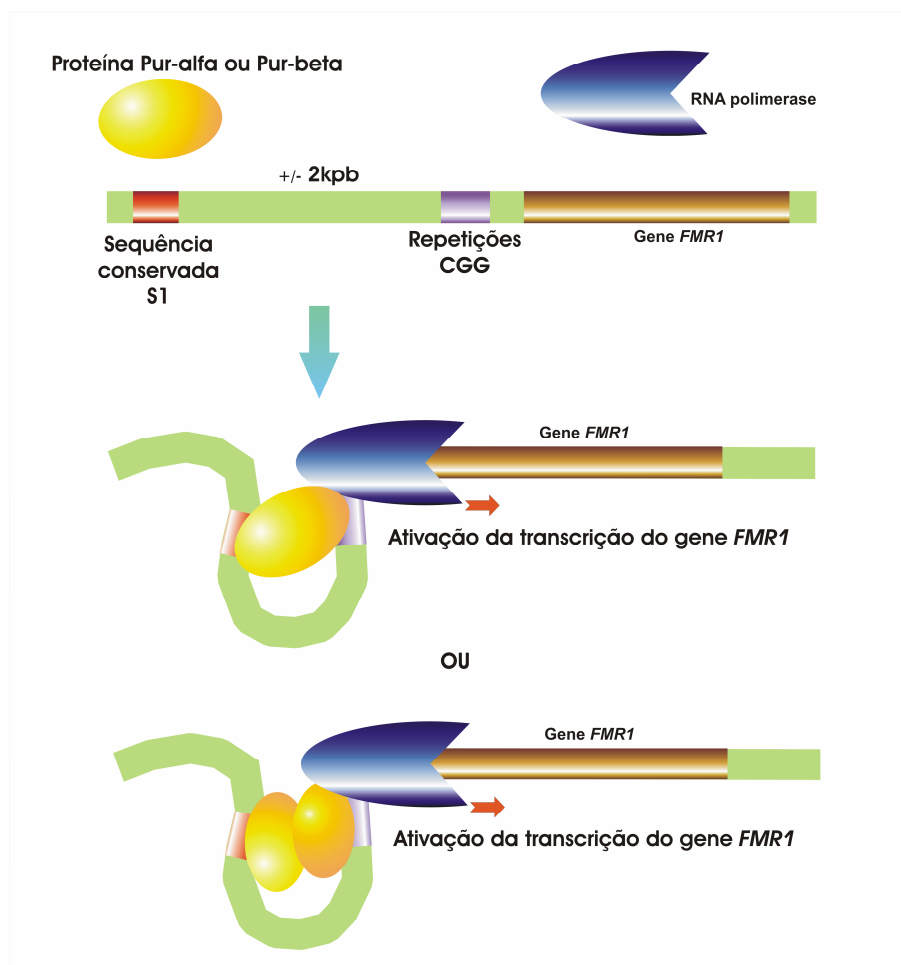


Figura 4.17: Modelo proposto da regulação da transcrição do gene *FMR1* pelas proteínas Pur- α e Pur- β . A ligação das proteínas Pur- α e Pur- β nas regiões reguladoras levam à formação de uma alça no DNA.

As evidências que associam as proteínas Pur- α e Pur- β com a regulação da transcrição de alguns genes e possivelmente com a regulação do gene *FMR1* parecem promissoras. Embora estas proteínas ainda não sejam consideradas fatores de transcrição e desta forma não estejam relacionadas nos bancos de dados de fatores de transcrição, não se pode negligenciar a sua contribuição quanto ao reconhecimento de seqüências específicas do DNA e conseqüente ligação a estes sítios, resultando na regulação da expressão do gene relacionado. Com base nestas evidências, é possível traçar um paralelo entre a atuação da proteína FMRP e os sintomas típicos dos pacientes afetados pela Síndrome do X-Frágil, com a proteína Pur- α , cujo papel precisa ser melhor explorado.

4.4. Análise de fluxos metabólicos no estudo do modelo funcional teórico para a proteína FMRP

Como forma de incrementar o conhecimento relacionado com a atuação da proteína FMRP e como a sua expressão afeta a expressão de outras proteínas estudou-se a distribuição dos fluxos dos complexos protéicos, contendo a proteína, relacionados com o transporte e regulação da tradução de mRNA, utilizando valores teóricos para alguns fluxos. A análise dos demais fluxos foi feita utilizando-se a técnica conhecida como Análise de Fluxos Metabólicos (AFM) (Stephanopoulos *et al.*, 1998; Chatziioannou *et al.*, 2003; Ozkan *et al.*, 2005).

Com a AFM pode-se estimar valores para as velocidades de transformação e transporte de moléculas ou complexos que participam de um dado processo. Através de balanços de massa e com alguns dados experimentais, constrói-se a matriz estequiométrica que representa os balanços de todos os componentes, como base para a AFM.

A análise dos fluxos do modelo proposto foi baseada em valores adimensionais, obtidos na literatura, conforme balanço de massa e relações estequiométricas do modelo biológico aqui proposto para a atuação da proteína FMRP, na regulação da tradução de RNA mensageiros específicos nos dendritos.

A Figura 4.18 apresenta de forma esquemática uma proposta de modelo biológico da via de atuação da proteína FMRP em células neuronais. Este modelo foi construído a partir das evidências experimentais encontradas na literatura, obtidas com células de pacientes afetados e com a utilização de camundongos *knockout* (Jin e Warren, 2003; Antar *et al.*, 2004; Bear *et al.*, 2004; Weiler *et al.*, 2004), e de tal modo que seja possível formular um modelo de AFM com os dados conhecidos até agora.

A FMRP atua na localização e exportação de RNA mensageiro do núcleo, bem como participa da regulação da tradução destes mensageiros no corpo e nos dendritos dos neurônios.

O procedimento ideal para a obtenção dos fluxos medidos, o que permite a estimativa dos demais fluxos do modelo, seria a sua determinação experimental a partir de um sistema *in vivo*, por exemplo de cultivo de células. Para tal foi utilizado o sistema de cultivo de células progenitoras neuronais obtido a partir de tecido pós-morte de um paciente com Síndrome do X-Frágil (Schwartz *et al.*, 2005) (Trabalho realizado em colaboração com o Grupo de Pesquisa do Department of Biochemistry and Molecular Medicine, University of California em Davis, Estados Unidos). No

entanto, problemas relacionados com o acondicionamento das células no *freezer* não permitiram a aquisição de dados confiáveis para a adequada alimentação ao modelo. Desta forma, alguns fluxos tiveram que ser estimados a partir de dados da literatura. A Figura 4.18 resume o modelo proposto para a análise de fluxos metabólicos (AFM da proteína FMRP).

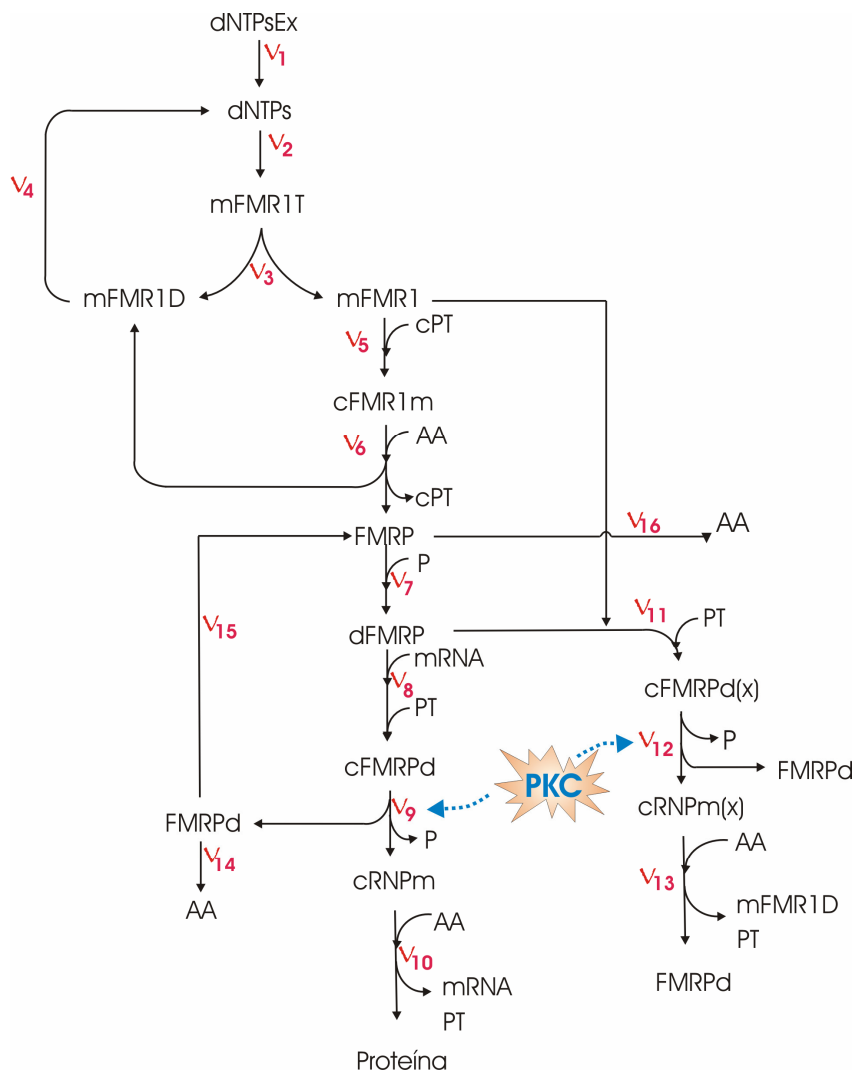


Figura 4.18: Modelo biológico proposto, neste trabalho, para o transporte e regulação da expressão de mRNA pela FMRP em células neuronais. Os números indicam os fluxos (v) dos complexos.

De acordo com o diagrama apresentado na Figura 4.18, e segundo a Equação 1, pode-se determinar um conjunto de 17 equações derivadas dos balanços de massa de cada um dos complexos envolvidos no regime estacionário. Assim, para os seguintes complexos, temos:

$$dNTPs: \quad v_1 - v_2 + v_4 = 0 \quad (4)$$

$$FMR1mT: \quad v_2 - v_3 = 0 \quad (5)$$

$$\begin{aligned}
FMR1m: & \quad 0,9 \cdot v_3 - v_5 - v_{11} = 0 & (6) \\
FMR1mD: & \quad 0,1 \cdot v_3 - v_4 + v_6 + v_{13} = 0 & (7) \\
cPT: & \quad -v_5 + v_6 = 0 & (8) \\
cFMR1m: & \quad v_5 - v_6 = 0 & (9) \\
AA: & \quad -v_6 - v_{10} - v_{13} + v_{14} + v_{16} = 0 & (10) \\
FMRP: & \quad v_6 - 2v_7 + v_{15} - v_{16} = 0 & (11) \\
P: & \quad -2v_7 + 2v_9 + 2v_{12} = 0 & (12) \\
dFMRP: & \quad v_7 - v_8 - v_{11} = 0 & (13) \\
mRNA: & \quad -v_8 + v_{10} = 0 & (14) \\
PT: & \quad -v_8 + v_{10} - v_{11} + v_{13} = 0 & (15) \\
cFMRPd: & \quad v_8 - v_9 = 0 & (16) \\
FMRPd: & \quad 2v_9 + 2v_{12} + v_{13} - v_{14} - v_{15} = 0 & (17) \\
cRNPm: & \quad v_9 - v_{10} = 0 & (18) \\
cFMRPd(x): & \quad v_{11} - v_{12} = 0 & (19) \\
cRNPm(x): & \quad v_{12} - v_{13} = 0 & (20)
\end{aligned}$$

Os fluxos apresentados nestas equações são baseados nas seguintes velocidades: entrada de nucleotídeos ($dNTPs$) no núcleo (v_1); transcrição do gene *FMR1* no seu mensageiro (*FMR1T*) (v_2); diferenciação entre mRNA que será traduzido (*FMR1m*) e a parte que será degradada (*FMR1mD*) (v_3) (Primerano *et al.*, 2002); degradação de mRNA (v_4); formação do complexo riboprotéico (*cFMR1m*) contendo o mRNA da FMRP e as ribonucleoproteínas (*cPT*) (v_5); tradução da FMRP (v_6); fosforilação e dimerização da FMRP (*dFMRP*) (v_7); formação do complexo protéico contendo FMRP, outras proteínas (*PT*) e o mRNA (*cFMRPd*) que terá sua tradução controlada negativamente pela FMRP (v_8); desfosforilação do complexo, liberando a proteína FMRP no dendrito (*FMRPd*) e o complexo ribonucleoprotéico (*cRNPm*) para a tradução (v_9); tradução de uma proteína reguladora da internalização da GLUR1, devido à sinalização neuronal (v_{10}); formação do complexo protéico contendo FMRP, outras proteínas (*PT*) e o seu próprio mRNA (*cFMRPd(x)*) que terá sua tradução controlada no neurônio, devido ao estímulo sináptico, conforme postulado por Weiler e colaboradores (1997) (v_{11}); desfosforilação do complexo contendo FMRP e seu mRNA, liberando a proteína FMRP no dendrito (*FMRPd*) e o complexo ribonucleoprotéico (*cRNPm(x)*) para a tradução (v_{12}); tradução da FMRP no dendrito regulada pela sinalização neuronal

(v_{13}); decomposição da FMRP no dendrito (v_{14}); transporte para o retorno da FMRP para o corpo do neurônio (v_{15}); decomposição da FMRP no corpo do neurônio (v_{16}).

Os fluxos v_9 e v_{12} são ativados devido à cascata de ativação da PKC com a desfosforilação da FMRP e conseqüente liberação do complexo ribonucleoprotéico para a tradução, como sugerido por alguns autores (Siomi *et al.*, 2002; Todd e Malter, 2002) mas que, no entanto, ainda não possui comprovação experimental.

A partir destas equações determinou-se a matriz estequiométrica, onde as colunas da matriz são relativas aos fluxos ou reações e as linhas correspondem a cada um dos complexos que participam do processo.

Dentre os fluxos podemos distinguir os medidos que, neste caso, foram estimados a partir de dados da literatura ($v_2, v_4, v_6, v_{13}, v_{14}, v_{16}$), e os calculados, assim determinados a partir do modelo e da AFM. Da literatura, utilizou-se uma velocidade média de transcrição de 45 dNTPs por segundo e uma velocidade média de tradução de 15 aminoácidos por segundo (Gardioli *et al.*, 1999), para a estimativa de alguns fluxos (chamados medidos no modelo). No caso da transcrição e degradação de *mFMR1* (v_2 e v_4) considerou-se 4362 dNTPs e para a FMRP 632 aminoácidos (v_6, v_{13}, v_{14} e v_{16}), com base nas seqüências obtidas no GenBank (<http://www.ncbi.nih.gov>).

Para estudar a variação nos fluxos em resposta a um estímulo nervoso, alteraram-se os fluxos de transcrição e tradução do gene *FMR1*, uma vez que, segundo vários autores, os níveis de mRNA e da FMRP aumentam em torno de 30 a 120%, dependendo do tecido e da metodologia empregados nos experimentos (Li *et al.*, 2002; Antar *et al.*, 2004; Weiler *et al.*, 2004). Optou-se por um aumento em torno de 50% para a velocidade de transcrição do gene *FMR1* e um aumento de 10% a 100% para a velocidade do fluxo de tradução da proteína FMRP nos dendritos. Assim, os fluxos medidos foram deduzidos e estão apresentados na Tabela 4.5. Os demais fluxos foram calculados através da Equação 2, com o uso do modelo.

Tabela 4.5: Fluxos medidos a partir de dados experimentais encontrados na literatura.

Fluxos Medidos	Sem estímulo (mmol/s)	Com estímulo (mmol/s)
v_2	0,0103	0,0150
v_4	0,0103	0,0103
v_6	0,0237	0,0237
v_{13}	0,0237	0,0260 a 0,0474
v_{14}	0,0237	0,0237
v_{16}	0,0237	0,0237

Para o modelo proposto, calculou-se ainda a matriz de redundância (R) e o *rank* desta matriz, que é igual a 4. O *rank* de R corresponde aos graus de liberdade do sistema.

Como o sistema é formado por 17 equações e apenas 16 variáveis (v), temos um sistema sobre-determinado, o que permite uma melhor estimativa dos fluxos a serem calculados e um melhoramento dos fluxos medidos, através da reconciliação estatística das medidas, conforme descrito por Stephanopoulos e colaboradores (1998). Os novos valores medidos podem então ser estimados a partir da Equação 21.

$$v_{m,new} = \left(I - R_r' (R_r R_r')^{-1} R_r \right) v_m \quad (21)$$

onde $R_r = K \cdot R$ é a matriz de redundância reduzida e K é um vetor de dimensões ($1 \times j$), sendo j o número de colunas da matriz G , com 1 nas posições referentes às linhas linearmente dependentes, obtidas após a análise do *rank* da matriz de redundância, e zeros nas demais posições. Do mesmo modo, os novos valores calculados podem ser obtidos a partir da Equação 2, substituindo-se os novos valores medidos. A variação percentual dos valores dos fluxos, após a reconciliação estatística, e já considerando as variações percentuais devido ao estímulo neuronal, são apresentados na Figura 4.19.

Nesta figura, nota-se que com o aumento da velocidade de tradução da FMRP nos dendritos (v_{13}) e com o aumento de aproximadamente 50% da velocidade de transcrição do gene *FMR1* (v_2), nos dados de entrada do modelo, houve uma diminuição dos fluxos 8 a 10, que correspondem aos fluxos relacionados com a tradução da proteína reguladora da internalização do GLUR1, regulada pela FMRP (v_{10}). A diminuição destes fluxos é coerente com o modelo apresentado por Huber e colaboradores (2002), apresentado na Figura 2.6, visto que no modelo os autores sugerem que a FMRP regula negativamente a internalização do receptor glutamatérgico GLUR1.

Nota-se também uma diminuição acentuada dos fluxos relacionados com a tradução da FMRP no corpo do neurônio (v_5 e v_6). Essa diminuição pode estar relacionada com o aumento da demanda de mRNA transcritos do gene *FMR1* nos dendritos para que ocorra a tradução localizada da FMRP, embora também deva haver um aumento da velocidade de transcrição do gene *FMR1* (v_2), como sugere a reconciliação estatística do modelo. A velocidade de transcrição de um gene pode

ser afetada por vários fatores, principalmente relacionados à presença ou não de determinados fatores de transcrição. No caso do gene *FMRI*, além dos fatores já conhecidamente relacionados com a sua regulação, este trabalho propôs um modelo de regulação através da ligação das proteínas Pur- α e Pur- β numa região localizada a aproximadamente 2 kpb à montante do gene.

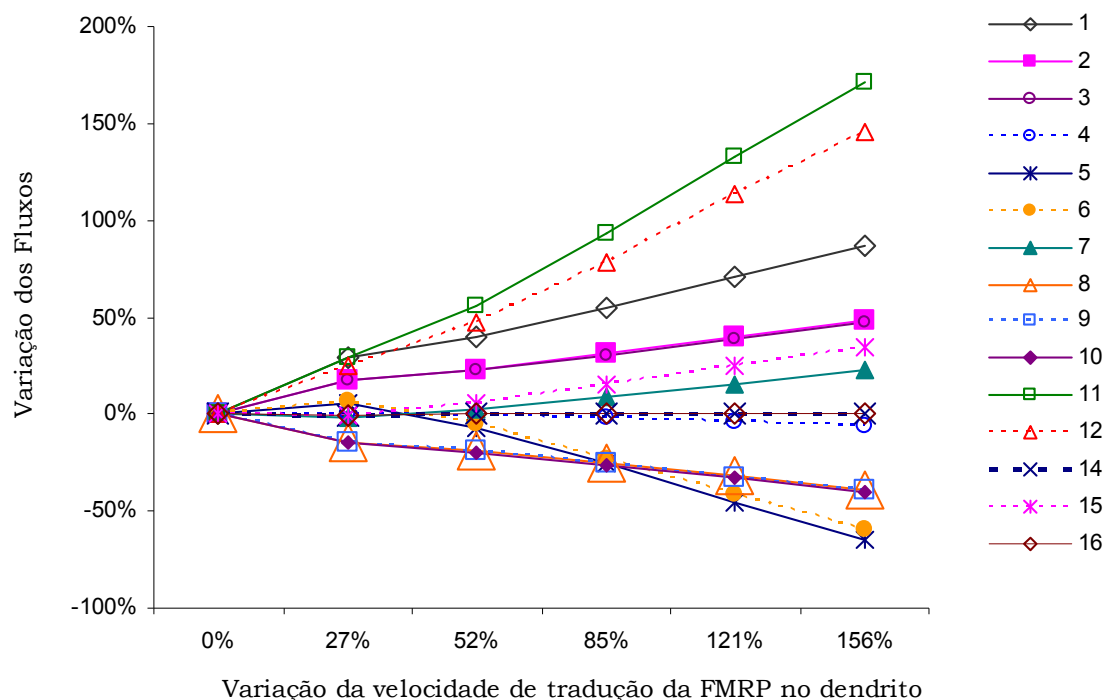


Figura 4.19: Variação dos fluxos medidos e calculados segundo o modelo proposto e com a variação de aproximadamente 50% da velocidade de transcrição do gene *FMRI* (v_2) e de 10 a 100% da velocidade de tradução da FMRP nos dendritos (v_{13}). Os valores apresentados no eixo x correspondem às variações de v_{13} já considerando a reconciliação estatística. A legenda ao lado identifica os fluxos. Fluxos relacionados: (1) entrada de nucleotídeos (dNTPs) no núcleo; (2) transcrição do gene *FMRI* no seu mensageiro; (3) diferenciação entre mRNA que será traduzido e a parte que será degradada; (4) degradação de mRNA; (5) formação do complexo riboprotéico contendo o mRNA da FMRP e as ribonucleoproteínas; (6) tradução da FMRP; (7) fosforilação e dimerização da FMRP; (8) formação do complexo protéico contendo FMRP, outras proteínas e o mRNA que terá sua tradução controlada negativamente pela FMRP; (9) desfosforilação do complexo, liberando a proteína FMRP no dendrito e o complexo ribonucleoprotéico para a tradução; (10) tradução de uma proteína reguladora da internalização da *GLUR1*, devido à sinalização neuronal; (11) formação do complexo protéico contendo FMRP, outras proteínas e o seu próprio mRNA que terá sua tradução controlada no neurônio; (12) desfosforilação do complexo contendo FMRP e seu mRNA, liberando a proteína FMRP no dendrito e o complexo ribonucleoprotéico para a tradução; (13) tradução da FMRP no dendrito regulada pela sinalização neuronal; (14) decomposição da FMRP no dendrito; (15) transporte para o retorno da FMRP para o corpo do neurônio; (16) decomposição da FMRP no corpo do neurônio.

Os fluxos v_{11} e v_{12} , relacionados com a tradução da FMRP nos dendritos também aumentam, conforme seria previsto.

O modelo biológico da função da proteína FMRP, como proteína que atua na repressão da tradução de alguns mRNA ao nível dos dendritos, tem sido estudado por vários grupos de pesquisa (Jin e Warren, 2003; Antar *et al.*, 2004; Bear *et al.*, 2004; Weiler *et al.*, 2004). No entanto, os dados quantitativos ainda são muito rudimentares e em condições muito restritas. Por isso, optou-se por utilizar para este estudo os fluxos de transcrição e tradução do gene *FMR1*, como dados de entrada.

Estes valores iniciais nos permitem uma análise do modelo. Todavia é importante lembrar que os valores absolutos podem não refletir a realidade. Por isso, os resultados são dados como variações percentuais. Para um modelo do qual pouco se sabe quantitativamente, estes resultados proporcionam uma possibilidade de análise da rede reguladora como um todo.

A relação entre os fluxos pode ser comparada através da análise de sensibilidade. A Tabela 4.6 apresenta as sensibilidades calculadas entre os fluxos medidos e os fluxos calculados. Através desta tabela, pode-se estabelecer uma comparação entre os impactos causados aos fluxos calculados, devido às alterações dos fluxos medidos.

As variações da velocidade de tradução da FMRP no corpo do neurônio (v_6) e da velocidade de degradação da proteína FMRP no corpo do neurônio (v_{16}) são as que mais causam impacto nos fluxos calculados, o que é bastante coerente, uma vez que são essas velocidades que vão determinar a quantidade de proteína FMRP que estará disponível para atuar como reguladora da tradução. Nota-se ainda pelas sensibilidades específicas para cada fluxo medido que o fluxo de FMRP do dendrito para o corpo do neurônio (v_{15}) é muito sensível às variações de v_6 e v_{16} .

Tabela 4.6: Sensibilidades calculadas entre os fluxos medidos e calculados:

Calculados	Medidos					
	v_2	v_4	v_6	v_{13}	v_{14}	v_{16}
v_1	1,0000	-1,0000	0,0000	0,0000	0,0000	0,0000
v_3	0,7404	0,0740	0,2117	0,2117	-0,0054	-0,0054
v_5	0,1401	0,0140	0,8257	-0,1743	0,0030	0,0030
v_7	0,1085	0,0108	-1,0382	-0,0382	0,9056	0,9056
v_8	-0,1049	-0,0105	-0,8631	-0,8631	0,9913	0,9913
v_9	-0,0362	-0,0036	-0,9873	-0,9873	1,0315	1,0315
v_{10}	0,0000	0,0000	-1,0000	-1,0000	1,0000	1,0000
v_{11}	0,2459	0,0246	-0,2865	0,7135	-0,0140	-0,0140
v_{12}	0,1392	0,0139	-0,1990	0,8010	0,0288	0,0288
v_{15}	0,2116	0,0212	-2,7245	0,2755	1,4659	2,4659
acumulada	2,7267	1,1727	8,1359	5,0645	5,4455	6,4455

Pode-se avaliar a qualidade do modelo proposto calculando-se o número condicional. Segundo Stephanopoulos e colaboradores (1998), este deve estar entre 1 e 100 para caracterizar um sistema bem definido e com baixa propagação de erro. No caso do modelo proposto o número condicional é igual a 20, o que nos permite afirmar que o modelo está bem condicionado.

As equações de conservação permitiram modelar a atuação da proteína da Síndrome do X-Frágil (FMRP) no sistema de regulação da tradução de mRNA em neurônios e juntamente com dados reais obtidos a partir de sistemas *in vivo*, podem auxiliar na compreensão da dinâmica associada à atuação da proteína FMRP e aos sistemas de regulação associados à Síndrome do X-Frágil.

5. CONCLUSÕES E SUGESTÕES PARA TRABALHOS FUTUROS

Ferramentas de engenharia genômica foram utilizadas a fim de compreender a regulação da expressão do gene *FMR1* e a atuação da proteína FMRP. A metodologia utilizada incluiu o uso de ferramentas computacionais de bioinformática de alinhamento e comparação genômica; no plano experimental, foram utilizadas ferramentas de biologia molecular e proteômica e estudos *in vitro* para determinação de interação DNA-proteína. A caracterização da função de proteínas chave para o processo de transcrição e tradução foi auxiliada por técnicas de proteômica e modelagem matemática de sistemas biológicos.

Os estudos sobre a regulação da transcrição do gene *FMR1* revelaram a existência de várias regiões do DNA genômico à montante do gene *FMR1*, que se mantêm conservadas entre mamíferos, e que têm potencial para abrigarem sítios de interação para proteínas de transcrição ou outros elementos relacionados com o controle da transcrição. Dentre estas regiões destaca-se a região localizada a aproximadamente dois mil pares de bases do códon de início do gene *FMR1*, que apresentou interação com proteínas típicas de controle de transcrição, como as proteínas Pur- α e Pur- β e ainda a proteína hnRNP A2/B1. O papel da Pur α e Pur β como elemento na regulação da transcrição do gene *FMR1* não tinha sido até então demonstrado, sendo esse o primeiro trabalho que propõe esta relação.

Outras conclusões relevantes a partir dos dados obtidos com este estudo foram:

- a. A escolha das seqüências genômicas e dos organismos utilizados para as técnicas de genômica comparativa foram adequados, uma vez que com o uso destas técnicas foi possível a identificação de elementos de regulação conservados;
- b. Pela primeira vez a região intergênica completa à montante do gene *FMR1* foi estudada e alguns elementos conservados foram particularmente caracterizados.
- c. Todos os oligonucleotídeos das seqüências estudadas interagiram, especificamente, por afinidade com as proteínas nucleares, uma vez que todas as reações, contendo o oligonucleotídeo marcado apresentaram um atraso na migração no gel, e o oligonucleotídeo não marcado foi capaz de deslocar as proteínas.

- d. Sugere-se que há a formação de um complexo multiprotéico, já que não foi possível identificar bandas nítidas na migração dos complexos DNA/proteínas.
- e. A identificação da proteína Pur- α foi evidenciada pela análise de espectrometria de massa e pela análise de imunodeteção.

Foi apresentado um modelo funcional composto pela distribuição dos fluxos dos complexos protéicos, relacionados com o transporte e regulação da tradução de mRNA, contendo a proteína FMRP. O modelo foi elaborado segundo o método de Análise de Fluxos Metabólicos, aliado aos dados encontrados na literatura, que descrevem este processo. Dentro deste contexto, o presente trabalho buscou acrescentar uma visão sistemática sobre a Síndrome do X-Frágil e possíveis soluções relacionadas às interações moleculares a fim de contribuir para a compreensão do processo biológico relacionado com esta doença. O conhecimento ou previsão destas interações permitirão, no futuro, identificar possíveis agentes terapêuticos e prever a sua ação antes da intervenção em pacientes.

Outras conclusões importantes resultantes deste estudo:

- a. Com o aumento da velocidade de tradução da FMRP nos dendritos e da velocidade de transcrição do gene *FMR1* houve diminuição dos fluxos relacionados com a tradução da proteína reguladora da internalização do GLUR1. Esta diminuição é coerente com o modelo apresentado por Huber e colaboradores (2002), que sugere que a FMRP regula negativamente a internalização do receptor glutamatérgico GLUR1.
- b. As variações da velocidade de tradução da FMRP no corpo do neurônio e da velocidade de degradação da proteína FMRP no corpo do neurônio são as que mais causam impacto nos fluxos calculados.

5.1. Sugestões para trabalhos futuros

- Verificar outras regiões à montante do gene, que também são conservadas e que possam abrigar sítios de interação para proteínas de transcrição;
- Purificar e identificar as proteínas que se ligam às seqüências S3, S4 e S5;

- Verificar a contribuição funcional das proteínas identificadas na expressão do gene *FMR1*.
- Determinar experimentalmente, a partir de cultivo de células, as velocidades de tradução da proteína FMRP e da proteína que tem sua tradução regulada por ela e também a velocidade de transcrição do gene *FMR1*, para alimentar o modelo proposto com dados reais.
- Acrescentar a regulação da transcrição do gene *FMR1* realizada pelas proteínas Pur- α e Pur- β no modelo matemático proposto e utilizar os dados experimentais relacionados com a contribuição funcional destas proteínas na expressão do gene para alimentar o modelo.

6. REFERÊNCIAS BIBLIOGRÁFICAS

Adinolfi, S., A. Ramos, S. R. Martin, F. Dal Piaz, P. Pucci, B. Bardoni, J. L. Mandel e A. Pastore. The N-terminus of the fragile X mental retardation protein contains a novel domain involved in dimerization and RNA binding. *Biochemistry*, 42, 10437-10444, 2003.

Ahituv, N., E. M. Rubin e M. A. Nobrega. Exploiting human-fish genome comparisons for deciphering gene regulation. *Human Molecular Genetics*, 13, R261-R266, 2004.

Antar, L. N., R. Afroz, J. B. Dichtenberg, R. C. Carroll e G. J. Bassell. Metabotropic glutamate receptor activation regulates fragile x mental retardation protein and FMR1 mRNA localization differentially in dendrites and at synapses. *J Neurosci*, 24, 2648-55, 2004.

Aschrafi, A., B. A. Cunningham, G. M. Edelman e P. W. Vanderklish. The fragile X mental retardation protein and group I metabotropic glutamate receptors regulate levels of mRNA granules in brain. *Proceedings of the National Academy of Sciences of the United States of America*, 102, 2180-2185, 2005.

Bakker, C. E., C. Verheij, R. Willemsen, R. Vanderhelm, F. Oerlemans, M. Vermey, A. Bygrave, A. T. Hoogeveen, B. A. Oostra, E. Reyniers, K. Deboulle, R. Dhooge, P. Cras, D. Vanvelzen, G. Nagels, J. J. Martin, P. P. Dedeyn, J. K. Darby e P. J. Willems. Fmr1 Knockout Mice - a Model to Study Fragile-X Mental-Retardation. *Cell*, 78, 23-33, 1994.

Baldwin, M. A. Protein identification by mass spectrometry - Issues to be considered. *Molecular & Cellular Proteomics*, 3, 1-9, 2004.

Bardoni, B., A. Schenck e J. L. Mandel. The Fragile X mental retardation protein. *Brain Research Bulletin*, 56, 375-382, 2001.

Bear, M. F., K. M. Huber e S. T. Warren. The mGluR theory of fragile X mental retardation. *Trends in Neurosciences*, 27, 370-377, 2004.

Beilina, A., F. Tassone, P. H. Schwartz, P. Sahota e P. J. Hagerman. Redistribution of transcription start sites within the FMR1 promoter region with expansion of the downstream CGG-repeat element. *Human Molecular Genetics*, 13, 543-549, 2004.

Berman, B. P., Y. Nibu, B. D. Pfeiffer, P. Tomancak, S. E. Celniker, M. Levine, G. M. Rubin e M. B. Eisen. Exploiting transcription factor binding site clustering to identify cis-regulatory modules involved in pattern formation in the Drosophila genome. *Proceedings of the National Academy of Sciences of the United States of America*, 99, 757-762, 2002.

Berry-Kravis, E., G. A. W., L. S. Crnic e W. T. Greenough. Understanding fragile X syndrome. *current Paediatrics*, 12, 316-324, 2002.

Blanchette, M. e M. Tompa. Discovery of regulatory elements by a computational method for phylogenetic footprinting. *Genome Research*, 12, 739-748, 2002.

- Brown, V., P. Jin, S. Ceman, J. C. Darnell, W. T. O'donnell, S. A. Tenenbaum, X. Jin, Y. Feng, K. D. Wilkinson, J. D. Keene, R. B. Darnell e S. T. Warren. Microarray identification of FMRP-associated brain mRNAs and altered mRNA translational profiles in fragile X syndrome. *Cell*, 107, 477-87, 2001.
- Brown, V., K. Small, L. Lakkis, Y. Feng, C. Gunter, K. D. Wilkinson e S. T. Warren. Purified recombinant Fmrp exhibits selective RNA binding as an intrinsic property of the fragile X mental retardation protein. *Journal of Biological Chemistry*, 273, 15521-15527, 1998.
- Carey, J. Gel Retardation. *Methods in Enzymology*, 208, 103-117, 1991.
- Castellvi-Bel, S. e M. Mila. Genes responsible for nonspecific mental retardation. *Molecular Genetics and Metabolism*, 72, 104-108, 2001.
- Ceman, S., V. Brown e S. T. Warren. Isolation of an FMRP-associated messenger ribonucleoprotein particle and identification of nucleolin and the fragile X-related proteins as components of the complex. *Molecular and Cellular Biology*, 19, 7925-7932, 1999.
- Ceman, S., W. T. O'donnell, M. Reed, S. Patton, J. Pohl e S. T. Warren. Phosphorylation influences the translation state of FMRP-associated polyribosomes. *Hum Mol Genet*, 12, 3295-305, 2003.
- Centonze, D., S. Rossi, V. Mercaldo, I. Napoli, M. T. Ciotti, V. De Chiara, A. Musella, C. Prosperetti, P. Calabresi, G. Bernardi e C. Bagni. Abnormal striatal GABA transmission in the mouse model for the fragile X syndrome. *Biological Psychiatry*, 63, 963-973, 2008.
- Chatziioannou, A., G. Palaiologos e F. N. Kolisis. Metabolic flux analysis as a tool for the elucidation of the metabolism of neuro transmitter glutamate. *Metabolic Engineering*, 5, 201-210, 2003.
- Chen, L., S. W. Yun, J. Seto, W. Liu e M. Toth. The fragile X mental retardation protein binds and regulates a novel class of mRNAs containing U rich target sequences. *Neuroscience*, 120, 1005-1017, 2003.
- Couronne, O., A. Poliakov, N. Bray, T. Ishkhanov, D. Ryaboy, E. Rubin, L. Pachter e I. Dubchak. Strategies and tools for whole-genome alignments. *Genome Research*, 13, 73-80, 2003.
- D'agata, V., S. T. Warren, W. Q. Zhao, E. R. Torre, D. L. Alkon e S. Cavallaro. Gene expression profiles in a transgenic animal model of fragile X syndrome. *Neurobiology of Disease*, 10, 211-218, 2002.
- Deissler, H., A. Behnkrappa e W. Doerfler. Purification of nuclear proteins from human HeLa cells that bind specifically to the unstable tandem repeat (CGG), in the human FMR1 gene. *Journal of Biological Chemistry*, 271, 4327-4334, 1996.
- Dermitzakis, E. T. e A. G. Clark. Evolution of transcription factor binding sites in mammalian gene regulatory regions: Conservation and turnover. *Molecular Biology and Evolution*, 19, 1114-1121, 2002.
- Drouin, R., M. Angers, N. Dallaire, T. M. Rose, E. W. Khandjian e F. Rousseau. Structural and functional characterization of the human FMR1 promoter reveals

- similarities with the hnRNP-A2 promoter region. *Human Molecular Genetics*, 6, 2051-2060, 1997.
- Dubchak, I. e K. Frazer. Multi-species sequence comparison: the next frontier in genome annotation. *Genome Biology*, 4, 2003.
- Eberhart, D. E., H. E. Malter, Y. Feng e S. T. Warren. The fragile X mental retardation protein is a ribonucleoprotein containing both nuclear localization and nuclear export signals. *Human Molecular Genetics*, 5, 1083-1091, 1996.
- Fogel, G. B., D. G. Weekes, G. Varga, E. R. Dow, A. M. Craven, H. B. Harlow, E. W. Su, J. E. Onyia e C. Su. A statistical analysis of the TRANSFAC database. *Biosystems*, 81, 137-54, 2005.
- Frazer, K. A., L. Elnitski, D. M. Church, I. Dubchak e R. C. Hardison. Cross-species sequence comparisons: A review of methods and available resources. *Genome Research*, 13, 1-12, 2003.
- Frazer, K. A., L. Pachter, A. Poliakov, E. M. Rubin e I. Dubchak. VISTA: computational tools for comparative genomics. *Nucleic Acids Research*, 32, W273-W279, 2004.
- Gabrielsen, O. S., E. Hornes, L. Korsnes, A. Ruet e T. B. Oyen. Magnetic DNA Affinity Purification of Yeast Transcription Factor-Tau - a New Purification Principle for the Ultrarapid Isolation of near Homogeneous Factor. *Nucleic Acids Research*, 17, 6253-6267, 1989.
- Gallagher, S. M., C. A. Daly, M. F. Bear e K. M. Huber. Extracellular signal-regulated protein kinase activation is required for metabotropic glutamate receptor-dependent long-term depression in hippocampal area CA1. *Journal of Neuroscience*, 24, 4859-4864, 2004.
- Gallia, G. L., E. M. Johnson e K. Khalili. Pur alpha: a multifunctional single-stranded DNA- and RNA-binding protein. *Nucleic Acids Research*, 28, 3197-3205, 2000.
- Ganapathi, M., G. P. Singh, K. S. Sandhu, S. K. Brahmachari e V. Brahmachari. A whole genome analysis of 5' regulatory regions of human genes for putative cis-acting modulators of nucleosome positioning. *Gene*, 391, 242-251, 2007.
- Garber, K., K. T. Smith, D. Reines e S. T. Warren. Transcription, translation and fragile X syndrome. *Current Opinion in Genetics & Development*, 16, 270-275, 2006.
- Gardiol, A., C. Racca e A. Triller. Dendritic and Postsynaptic Protein Synthetic Machinery. *Journal of Neuroscience*, 19, 168-179, 1999.
- Gheldof, N., T. M. Tabuchi e J. Dekker. The active FMR1 promoter is associated with a large domain of altered chromatin conformation with embedded local histone modifications. *Proceedings of the National Academy of Sciences of the United States of America*, 103, 12463-12468, 2006.
- Greco, C. M., R. J. Hagerman, F. Tassone, A. E. Chudley, M. R. Del Bigio, S. Jacquemont, M. Leehey e P. J. Hagerman. Neuronal and astrocytic intranuclear inclusions in Fragile X carriers with cerebellar tremor/ataxia and cognitive changes. *Journal of Neuropathology and Experimental Neurology*, 61, 450-450, 2002.

- Greenough, W. T., A. Y. Klintsova, S. A. Irwin, R. Galvez, K. E. Bates e I. J. Weiler. Synaptic regulation of protein synthesis and the fragile X protein. *PNAS*, 98, 7101-7106, 2001.
- Grehan, S., E. Tse e J. M. Taylor. Two distal downstream enhancers direct expression of the human apolipoprotein E gene to astrocytes in the brain. *Journal of Neuroscience*, 21, 812-822, 2001.
- Hagerman, R. J. e P. J. Hagerman. Fragile X syndrome: A model of gene-brain-behavior relationships. *Molecular Genetics and Metabolism*, 74, 89-97, 2001.
- Hagerman, R. J. e P. J. Hagerman. The fragile X premutation: into the phenotypic fold. *Current Opinion in Genetics & Development*, 12, 278-283, 2002a.
- Hagerman, R. J. e P. J. Hagerman. Fragile X Syndrome: Diagnosis, Treatment, and Research: Johns Hopkins University Press. 2002b. 552 p p.
- Hanson, J. E. e D. V. Madison. Presynaptic Fmr1 genotype influences the degree of synaptic connectivity in a mosaic mouse model of fragile X syndrome. *Journal of Neuroscience*, 27, 4014-4018, 2007.
- Hayashi, M. L., B. S. S. Rao, J. S. Seo, H. S. Choi, B. M. Dolan, S. Y. Choi, S. Chattarji e S. Tonegawa. Inhibition of p21-activated kinase rescues symptoms of fragile X syndrome in mice. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 11489-11494, 2007.
- Hellman, L. M. e M. G. Fried. Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. *Nature Protocols*, 2, 1849-1861, 2007.
- Huang, T., L. Y. Li, Y. Shen, X. B. Qin, Z. L. Pang e G. Y. Wu. Alternative splicing of the FMR1 gene in human fetal brain neurons. *American Journal of Medical Genetics*, 64, 252-255, 1996.
- Huber, K. M., S. M. Gallagher, S. T. Warren e M. F. Bear. Altered synaptic plasticity in a mouse model of fragile X mental retardation. *PNAS*, 99, 7746-7750, 2002.
- Huber, K. M., M. S. Kayser e M. F. Bear. Role for rapid dendritic protein synthesis in hippocampal mGluR-dependent long-term depression. *Science*, 288, 1254-1256, 2000.
- Ishizuka, A., M. C. Siomi e H. Siomi. A Drosophila fragile X protein interacts with components of RNAi and ribosomal proteins. *Genes & Development*, 16, 2497-2508, 2002.
- Jin, P., R. H. Duan, A. Qurashi, Y. L. Qin, D. H. Tian, T. C. Rosser, H. J. Liu, Y. Feng e S. T. Warren. Pur alpha binds to rCGG repeats and modulates repeat-mediated neurodegeneration in a Drosophila model of fragile X tremor/ataxia syndrome. *Neuron*, 55, 556-564, 2007.
- Jin, P. e S. T. Warren. New insights into fragile X syndrome: from molecules to neurobehaviors. *Trends in Biochemical Sciences*, 28, 152-158, 2003.
- Kel-Margoulis, O. V., A. E. Kel, I. Reuter, I. V. Deineko e E. Wingender. TRANSCmpel: a database on composite regulatory elements in eukaryotic genes. *Nucleic Acids Res*, 30, 332-4, 2002.

- Khalili, K., L. Del Valle, V. Muralidharan, W. J. Gault, N. Darbinian, J. Otte, E. Meier, E. M. Johnson, D. C. Daniel, Y. Kinoshita, S. Amini e J. Gordon. Pur alpha is essential for postnatal brain development and developmentally coupled cellular proliferation as revealed by genetic inactivation in the mouse. *Molecular and Cellular Biology*, 23, 6857-6875, 2003.
- Kirkpatrick, L. L., K. A. Mcilwain e D. L. Nelson. Comparative genomic sequence analysis of the FXR gene family: FMR1, FXR1, and FXR2. *Genomics*, 78, 169-177, 2001.
- Kobayashi, S., K. Agui, S. Kamo, Y. M. Li e K. Anzai. Neural BC1 RNA associates with Pur alpha, a single-stranded DNA and RNA binding protein, which is involved in the transcription of the BC1 RNA gene. *Biochemical and Biophysical Research Communications*, 277, 341-347, 2000.
- Kolchanov, N. A., O. A. Podkolodnaya, E. A. Ananko, E. V. Ignatieva, I. L. Stepanenko, O. V. Kel-Margoulis, A. E. Kel, T. I. Merkulova, T. N. Goryachkovskaya, T. V. Busygina, F. A. Kolpakov, N. L. Podkolodny, A. N. Naumochkin, I. M. Korostishevskaya, A. G. Romashchenko e G. C. Overton. Transcription regulatory regions database (TRRD): its status in 2000. *Nucleic Acids Res*, 28, 298-301, 2000.
- Kumari, D. e K. Usdin. Interaction of the transcription factors USF1, USF2, and alpha-Pal/Nrf-1 with the FMR1 promoter - Implications for Fragile X mental retardation syndrome. *Journal of Biological Chemistry*, 276, 4357-4364, 2001.
- Laggerbauer, B., D. Ostareck, E. M. Keidel, A. Ostareck-Lederer e U. Fischer. Evidence that fragile X mental retardation protein is a negative regulator of translation. *Human Molecular Genetics*, 10, 329-338, 2001.
- Lee, J. S., J. W. Grisham e S. S. Thorgeirsson. Comparative functional genomics for identifying models of human cancer. *Carcinogenesis*, 26, 1013-1020, 2005.
- Lenhard, B., A. Sandelin, L. Mendoza, P. Engstrom, N. Jareborg e W. W. Wasserman. Identification of conserved regulatory elements by comparative genome analysis. *J Biol*, 2, 13, 2003.
- Levy, S. e S. Hannenhalli. Identification of transcription factor binding sites in the human genome sequence. *Mammalian Genome*, 13, 510-514, 2002.
- Li, J. X., M. R. Pelletier, J. L. P. Velazquez e P. L. Carlen. Reduced cortical synaptic plasticity and GluR1 expression associated with fragile X mental retardation protein deficiency. *Molecular and Cellular Neuroscience*, 19, 138-151, 2002.
- Li, Z. Z., Y. Y. Zhang, L. Ku, K. D. Wilkinson, S. T. Warren e Y. Feng. The fragile X mental retardation protein inhibits translation via interacting with mRNA. *Nucleic Acids Research*, 29, 2276-2283, 2001.
- Lim, J. H., A. B. Booker e J. R. Fallon. Regulating fragile X gene transcription in the brain and beyond. *Journal of Cellular Physiology*, 205, 170-175, 2005.
- Lim, J. H., A. B. Booker, T. Luo, T. Williams, Y. Furuta, O. Lagutin, G. Oliver, T. D. Sargent e J. R. Fallon. AP-2 alpha selectively regulates fragile X mental retardation-1 gene transcription during embryonic development. *Human Molecular Genetics*, 14, 2027-2034, 2005.

- Lin, D., D. L. Tabb e J. R. Yates. Large-scale protein identification using mass spectrometry. *Biochimica Et Biophysica Acta-Proteins and Proteomics*, 1646, 1-10, 2003.
- Loesch, D. Z., Q. M. Bui, R. M. Huggins, R. J. Mitchell, R. J. Hagerman e F. Tassone. Transcript levels of the intermediate size or grey zone fragile X mental retardation 1 alleles are raised, and correlate with the number of CGG repeats. *Journal of Medical Genetics*, 44, 200-204, 2007.
- Loots, G. G. e I. Ovcharenko. rVISTA 2.0: evolutionary analysis of transcription factor binding sites. *Nucleic Acids Research*, 32, W217-W221, 2004.
- Loots, G. G., I. Ovcharenko, L. Pachter, I. Dubchak e E. M. Rubin. rVista for comparative sequence-based discovery of functional transcription factor binding sites. *Genome Research*, 12, 832-839, 2002.
- Lowry, O. H., N. J. Rosebrough, A. L. Farr e R. J. Randall. Protein Measurement with the Folin Phenol Reagent. *Journal of Biological Chemistry*, 193, 265-275, 1951.
- Lu, R., H. P. Wang, Z. Liang, L. Ku, W. T. O'donnell, W. Li, S. T. Warren e Y. Feng. The fragile X protein controls microtubule-associated protein 1B translation and microtubule stability in brain neuron development. *PNAS*, 101, 15201-15206, 2004.
- Mayor, C., M. Brudno, J. R. Schwartz, A. Poliakov, E. M. Rubin, K. A. Frazer, L. S. Pachter e I. Dubchak. VISTA : visualizing global DNA sequence alignments of arbitrary length. *Bioinformatics*, 16, 1046-1047, 2000.
- Muller-Hartmann, H., H. Deissler, F. Naumann, B. Schmitz, J. Schroer e W. Doerfler. The human 20-kDa 5'-(CGG)(n)-3'-binding protein is targeted to the nucleus and affects the activity of the FMR1 promoter. *Journal of Biological Chemistry*, 275, 6447-6452, 2000.
- Nandi, T., C. B-Rao e S. Ramachandran. Comparative genomics using data mining tools. *Journal of Biosciences*, 27, 15-25, 2002.
- Nobrega, M. A. e L. A. Pennacchio. Comparative genomic analysis as a tool for biological discovery. *Journal of Physiology-London*, 554, 31-39, 2004.
- Nolin, S. L., F. A. Lewis, L. L. Ye, G. E. Houck, A. E. Glicksman, P. Limprasert, S. Y. Li, N. Zhong, A. E. Ashley, E. Feingold, S. L. Sherman e W. T. Brown. Familial transmission of the FMR1 CGG repeat. *American Journal of Human Genetics*, 59, 1252-1261, 1996.
- Nordhoff, E., A. M. Krogsdam, H. F. Jorgensen, B. H. Kallipolitis, B. F. C. Clark, P. Roepstorff e K. Kristiansen. Rapid identification of DNA-binding proteins by mass spectrometry. *Nature Biotechnology*, 17, 884-888, 1999.
- Ohashi, S., K. Koike, A. Omori, S. Ichinose, S. Ohara, S. Kobayashi, T. A. Sato e K. Anzai. Identification of mRNA/protein (mRNP) complexes containing Pur alpha, mStaufen, Fragile X Protein, and myosin Va and their association with rough endoplasmic reticulum equipped with a kinesin motor. *Journal of Biological Chemistry*, 277, 37804-37810, 2002.
- Oostra, B. A. Functions of the fragile X protein. *Trends in Molecular Medicine*, 8, 102-103, 2002.

- Orlov, S. V., K. B. Kuteykin-Teplyakov, I. A. Ignatovich, E. B. Dizhe, O. A. Mirgorodskaya, A. V. Grishin, O. B. Guzhova, E. B. Prokhortchouk, P. V. Guliy e A. P. Perevozchikov. Novel repressor of the human FMR1 gene - identification of p56 human (GCC)(n)-binding protein as a Kruppel-like transcription factor ZF5. *Febs Journal*, 274, 4848-4862, 2007.
- Otero, Y. D., L. A. Severijnen, G. Van Cappellen, M. Schrier, B. Oostra e R. Willemsen. Transport of fragile X mental retardation protein via granules in neurites of PC12 cells. *Molecular and Cellular Biology*, 22, 8332-8341, 2002.
- Ozkan, P., B. Sariyar, F. O. Utkur, U. Akman e A. Hortacsu. Metabolic flux analysis of recombinant protein overproduction in Escherichia coli. *Biochemical Engineering Journal*, 22, 167-195, 2005.
- Pan, L. Y., Y. Q. Zhang, E. Woodruff e K. Broadie. The Drosophila fragile X gene negatively regulates neuronal elaboration and synaptic differentiation. *Current Biology*, 14, 1863-1870, 2004.
- Paradee, W., H. E. Melikian, D. L. Rasmussen, A. Kenneson, P. J. Conn e S. T. Warren. Fragile X mouse: Strain effects of knockout phenotype and evidence suggesting deficient amygdala function. *Neuroscience*, 94, 185-192, 1999.
- Pennacchio, L. A. Insights from human/mouse genome comparisons. *Mammalian Genome*, 14, 429-436, 2003.
- Prasad, S. e K. Singh. Age- and sex-dependent differential interaction of nuclear trans-acting factors with Fmr-1 promoter in mice brain. *Neurochemical Research*, 33, 1028-1035, 2008.
- Price, T. J., M. H. Rashid, M. Millecamps, R. Sanoja, J. M. Entrena e F. Cervero. Decreased nociceptive sensitization in mice lacking the fragile X mental retardation protein: Role of mGluR1/5 and mTOR. *Journal of Neuroscience*, 27, 13958-13967, 2007.
- Primerano, B., F. Tassone, R. J. Hagerman, P. Hagerman, F. Amaldi e C. Bagni. Reduced FMR1 mRNA translation efficiency in Fragile X patients with premutations. *Rna-a Publication of the Rna Society*, 8, 1482-1488, 2002.
- Sawtell, N. B., K. M. Huber, J. C. Roder e M. F. Bear. Induction of NMDA receptor-dependent long-term depression in visual cortex does not require metabotropic glutamate receptors. *Journal of Neurophysiology*, 82, 3594-3597, 1999.
- Schaeffer, C., B. Bardoni, J. L. Mandel, B. Ehresmann, C. Ehresmann e H. Moine. The fragile X mental retardation protein binds specifically to its mRNA via a purine quartet motif. *Embo Journal*, 20, 4803-4813, 2001.
- Schaeffer, C., M. Beaulande, C. Ehresmann, B. Ehresmann e H. Moine. The RNA binding protein FMRP: new connections and missing links. *Biology of the Cell*, 95, 221-228, 2003.
- Schwartz, P. H., F. Tassone, C. M. Greco, H. E. Nethercott, B. Ziaecian, R. J. Hagerman e P. J. Hagerman. Neural progenitor cells from an adult patient with fragile X syndrome. *Bmc Medical Genetics*, 6, -, 2005.

- Siomi, M. C., K. Higashijima, A. Ishizuka e H. Siomi. Casein kinase II phosphorylates the fragile X mental retardation protein and modulates its biological properties. *Molecular and Cellular Biology*, 22, 8438-8447, 2002.
- Siomi, M. C., H. Siomi, W. H. Sauer, S. Srinivasan, R. L. Nussbaum e G. Dreyfuss. Fxr1, an Autosomal Homolog of the Fragile-X Mental-Retardation Gene. *Embo Journal*, 14, 2401-2408, 1995.
- Slegtenhorst-Eegdeman, K. E., D. G. De Rooij, M. Verhoef-Post, H. J. G. Van De Kant, C. E. Bakker, B. A. Oostra, J. A. Grootegoed e A. P. N. Themmen. Macroorchidism in FMR1 knockout mice is caused by increased Sertoli cell proliferation during testicular development. *Endocrinology*, 139, 156-162, 1998.
- Smith, K. T., B. Coffee e D. Reines. Occupancy and synergistic activation of the FMR1 promoter by Nrf-1 and Sp1 in vivo. *Human Molecular Genetics*, 13, 1611-1621, 2004.
- Smith, K. T., R. D. Nicholls e D. Reines. The gene encoding the fragile X RNA-binding protein is controlled by nuclear respiratory factor 2 and the CREB family of transcription factors. *Nucleic Acids Research*, 34, 1205-1215, 2006.
- Stephanopoulos, G. N., A. A. Aristidou e J. Nielsen. Metabolic engineering - Principles and Methodology. San Diego: Academic Press. 1998. 724 p.
- Tamanini, F., R. Willemsen, L. Vanunen, C. Bontekoe, H. Galjaard, B. A. Oostra e A. T. Hoogeveen. Differential expression of FMR1, FXR1 and FXR2 proteins in human brain and testis. *Human Molecular Genetics*, 6, 1315-1322, 1997.
- Tassone, F., R. J. Hagerman, W. D. Chamberlain e P. J. Hagerman. Transcription of the FMR1 gene in individuals with fragile X syndrome. *American Journal of Medical Genetics*, 97, 195-203, 2000.
- Tassone, F., R. J. Hagerman, D. N. Ikle, P. N. Dyer, M. Lampe, R. Willemsen, B. A. Oostra e A. K. Taylor. FMRP expression as a potential prognostic indicator in fragile X syndrome. *American Journal of Medical Genetics*, 84, 250-261, 1999.
- Thomas, J. W., J. W. Touchman, R. W. Blakesley, G. G. Bouffard, S. M. Beckstrom-Sternberg, E. H. Margulies, M. Blanchette, A. C. Siepel, P. J. Thomas, J. C. McDowell, B. Maskeri, N. F. Hansen, M. S. Schwartz, R. J. Weber, W. J. Kent, D. Karolchik, T. C. Bruen, R. Bevan, D. J. Cutler, S. Schwartz, L. Elnitski, J. R. Idol, A. B. Prasad, S. Q. Lee-Lin, V. V. B. Maduro, T. J. Summers, M. E. Portnoy, N. L. Dietrich, N. Akhter, K. Ayele, B. Benjamin, K. Cariaga, C. P. Brinkley, S. Y. Brooks, S. Granite, X. Guan, J. Gupta, P. Haghghi, S. L. Ho, M. C. Huang, E. Karlins, P. L. Laric, R. Legaspi, M. J. Lim, Q. L. Maduro, C. A. Masiello, S. D. Mastrian, J. C. McCloskey, R. Pearson, S. Stantripop, E. E. Tiongson, J. T. Tran, C. Tsurgeon, J. L. Vogt, M. A. Walker, K. D. Wetherby, L. S. Wiggins, A. C. Young, L. H. Zhang, K. Osoegawa, B. Zhu, B. Zhao, C. L. Shu, P. J. De Jong, C. E. Lawrence, A. F. Smit, A. Chakravarti, D. Haussler, P. Green, W. Miller e E. D. Green. Comparative analyses of multi-species sequences from targeted genomic regions. *Nature*, 424, 788-793, 2003.
- Thompson, J., D. Higgins e T. Gibson. Clustal W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, 22, 4673-4680, 1994.

- Todd, P. K. e J. S. Malter. Fragile X mental retardation protein in plasticity and disease. *Journal of Neuroscience Research*, 70, 623-630, 2002.
- Van 'T Padje, S., B. Engels, L. Blonden, L. A. Severijnen, F. Verheijen, B. Oostra e R. Willemsen. Characterisation of Fmrp in zebrafish: evolutionary dynamics of the *fmr1* gene. *Development Genes and Evolution*, 215, 198-206, 2005.
- Weiler, I. J. e W. T. Greenough. Metabotropic Glutamate Receptors Trigger Postsynaptic Protein-Synthesis. *PNAS*, 90, 7168-7171, 1993.
- Weiler, I. J. e W. T. Greenough. Synaptic synthesis of the Fragile X protein: Possible involvement in synapse maturation and elimination. *American Journal of Medical Genetics*, 83, 248-252, 1999.
- Weiler, I. J., S. A. Irwin, A. Y. Klintsova, C. M. Spencer, A. D. Brazelton, K. Miyashiro, T. A. Comery, B. Patel, J. Eberwine e W. T. Greenough. Fragile X mental retardation protein is translated near synapses in response to neurotransmitter activation. *PNAS*, 94, 5395-5400, 1997.
- Weiler, I. J., C. C. Spangler, A. Y. Klintsova, A. W. Grossman, S. H. Kim, V. Bertaina-Anglade, H. Khaliq, F. E. De Vries, F. A. E. Lambers, F. Hatia, C. K. Base e W. T. Greenough. Fragile X mental retardation protein is necessary for neurotransmitter-activated protein translation at synapses. *PNAS*, 101, 17504-17509, 2004.
- Weisbrod, S. e H. Weintraub. Isolation of a Subclass of Nuclear Proteins Responsible for Conferring a Dnase I-Sensitive Structure on Globin Chromatin. *Proceedings of the National Academy of Sciences of the United States of America*, 76, 630-634, 1979.
- Willemsen, R., A. Smits, S. Mohkamsing, H. Vanbeerendonk, A. Dehaan, B. Devries, A. Vandenouweland, E. Sijtermans, H. Galjaard e B. A. Oostra. Rapid antibody test for diagnosing fragile X syndrome: A validation of the technique. *Human Genetics*, 99, 308-311, 1997.
- Woo, A. J., J. S. Dods, E. Susanto, D. Ulgiati e L. J. Abraham. A proteomics approach for the identification of DNA binding activities observed in the electrophoretic mobility shift assay. *Molecular & Cellular Proteomics*, 1, 472-478, 2002.
- Wortman, M. J., E. M. Johnson e A. D. Bergemann. Mechanism of DNA binding and localized strand separation by Pur alpha and comparison with Pur family member, Pur beta. *Biochimica Et Biophysica Acta-Molecular Cell Research*, 1743, 64-78, 2005.
- Yates, J. R. Mass spectrometry and the age of the proteome. *Journal of Mass Spectrometry*, 33, 1-19, 1998.
- Zalfa, F., M. Giorgi, B. Primerano, A. Moro, A. Di Penta, S. Reis, B. Oostra e C. Bagni. The fragile X syndrome protein FMRP associates with BC1 RNA and regulates the translation of specific mRNAs at Synapses. *Cell*, 112, 317-327, 2003.
- Zhang, Y., J. P. Oconnor, M. C. Siomi, S. Srinivasan, A. Dutra, R. L. Nussbaum e G. Dreyfuss. The Fragile-X Mental-Retardation Syndrome Protein Interacts with Novel Homologs Fxr1 and Fxr2. *Embo Journal*, 14, 5358-5366, 1995.