

UNIVERSITAT DE BARCELONA

FACULTAT
FARMÀCIA

DEPARTAMENT
BIOQUÍMICA I BIOLOGIA MOLECULAR

LES PROPIETATS FÍSiques DE L'ADN EN ESCALA GENÒMICA

Josep Ramon Goñi Macià 2008

UNIVERSITAT DE BARCELONA

FACULTAT DE FARMÀCIA

DEPARTAMENT
BIOQUÍMICA I BIOLOGIA MOLECULAR

PROGRAMA DE DOCTORAT

BIENNI 2002-2004

TÍTOL DE LA TESI

LES PROPIETATS FÍSiques DE L'ADN EN ESCALA GENÒMICA

Memòria presentada per Josep Ramon Goñi Macià per optar al títol de doctor per la universitat de Barcelona

Modesto Orozco López

Josep Ramon Goñi Macià

Josep Ramon Goñi Macià 2008

Gràcies a la Maria Dolors, al Juan Miguel i especialment a la Meritxell pel vostre suport

RESUM

Es coneix com era post-genòmica els anys posteriors a la seqüenciació massiva dels genomes d'organismes i espècies superiors (entre elles la humana) i que ens permeten per primer cop, tenir una visió global dels mecanismes biològics que regulen la vida de cèl·lula. En aquests primers anys d'aquesta nova era, s'ha donat especial rellevància a la identificació i estudi funcional de motius de seqüència en el genoma, assumint la hipòtesi que la capa física (la molècula d'ADN) es un mer suport per a un codi que és interpretat per la cèl·lula.

El treball exposat en aquesta memòria posa en relleu la importància del nivell físic de l'ADN (ja sigui en la molècula de doble hèlix o en estructures d'ordre superior), en tots els aspectes funcionals de la molècula.

Els resultats exposats en les publicacions científiques que integren aquest treball es demostra la relació entre la estructura de l'ADN i les regions reguladores dels gens (els promotors). Es prova també la viabilitat de l'ús d'estructures no canòniques de l'ADN (com per exemple les triple hèlices) en teràpies anti-gèniques. Finalment es descriu la implementació d'una plataforma bioinformàtica que permet per primer cop, l'estudi dels factors fisicoquímics en escala genòmica.

LLISTA DE PUBLICACIONS

- I. Goñi JR, Fenollosa C, Pérez A, Torrents D, Orozco M. **DNAlive: A tool for the physical analysis of DNA at the genomic scale.** *Bioinformatics*. 2008; 24:1731-1732
- II. Goñi JR, Pérez A, Torrents D, Orozco M. **Determining promoter location based on DNA structure first-principles calculations.** *Genome Biol*. 2007;8:R263.
- III. Goñi JR, Vaquerizas JM, Dopazo J, Orozco M. **Exploring the reasons for the large density of triplex-forming oligonucleotide target sequences in the human regulatory regions.** *BMC Genomics*. 2006 7:63.
- IV. Goñi JR, de la Cruz X, Orozco M. **Triplex-forming oligonucleotide target sequences in the human genome.** *Nucleic Acids Res*. 2004;:354-60.

CONTINGUT

1	Introducció I	1
1.1	Objectius	2
2	Introducció II: Regulació genètica	5
2.1	El dogma central de la biologia molecular.....	5
2.1.1	Transferències de la informació biològica	6
2.2	Estructura dels gens en organismes eucariotes	12
2.2.1	Exons.....	14
2.2.2	Introns	15
2.2.3	Entroncament (<i>splicing</i>) alternatiu.....	16
2.2.4	Regions no traduïdes	17
2.2.5	Promotors de gens	18
2.2.6	Factors de transcripció.....	21
2.2.7	Potenciadors (<i>enhancers</i>) i silenciadors de la regulació.....	23
2.3	Genotips i fenotips.....	23
2.4	Regulació epigenètica.....	24
2.4.1	Metilació de l'ADN i la remodelació de la cromatina.....	25
2.4.2	Silenciament de gens basat en àcids nucleics.....	26
3	Introducció III: Estructura de l'ADN	33
3.1	Estructura primària de l'ADN	34
3.2	Estructura secundària de l'ADN.....	36
3.2.1	Ponts d'hidrogen.....	37
3.2.2	Interaccions d'apilament.....	39
3.2.3	L'efecte del solvent	41
3.3	Propietats fisicoquímiques de l'ADN	42
3.3.1	Propietats de fusió de l'ADN.....	42
3.3.2	Energies d'apilament	43
3.3.3	Geometria de la hèlix d'ADN.....	44
3.3.4	Propietats de flexibilitat de l'ADN	48
3.4	Formes no canòniques de l'ADN	49
3.4.1	Tríplex.....	50
3.4.2	Quàdruplex: ADN de quatre cadenes	53
3.5	Estructures d'ordre superior	55

3.5.1	Estructura del nucleosoma.....	56
3.5.2	Estructures superiors al nucleosoma.....	57
4	Mètodes.....	59
4.1	Eines bioinformàtiques.....	59
4.1.1	Mètodes de predicció de propietats físiques de l'ADN.....	60
4.1.2	Mètodes de predicció d'anotacions en genomes.....	67
4.1.3	Mètodes de predicció de la dinàmica de l'ADN.....	87
4.2	Bases de dades bioinformàtiques.....	94
4.2.1	Bases de dades d'ADN.....	96
4.2.2	Bases de dades de Proteïnes.....	100
5	Resultats.....	107
5.1	Estudi de les dianes de Tríplex en el genoma humà.....	107
5.2	Exploració de la super-població de TTS en promotors.....	107
5.3	Estudi de les propietats físiques de l'ADN en promotors.....	107
5.4	Desenvolupament de la plataforma DNALive.....	107
6	Discussió.....	109
6.1	Les seqüències formadores de tríplex en el genoma humà.....	109
6.2	El rol de les propietats físiques de l'ADN en el genoma humà.....	115
7	Conclusions.....	123
8	Referències bibliogràfiques.....	125
9	Publicacions.....	143

LLISTA D'ABREVIACIONS

3'UTR	Regió tres prima no traduïda (sigla en anglès; 3'Untranslated Region)
5'UTR	Regió cinc prima no traduïda (sigla en anglès; 5'Untranslated Region)
ADN	Àcid Desoxiribonucleic
AE	Promig d'errors en les prediccions; average mismatch of prediction
ARN	Àcid Ribonucleic
ARNi	ARN interferent
ARNm	ARN missatger
ARNt	ARN transferència
ASM	Promig de les puntuacions de les mesures ; averaged score measure
CAGE	Cap analysis of gene expression
dsRNA	ARN de doble cadena (en anglès double stranded RNA)
FN	Negatius Falsos (false negative)
FP	Positius Falsos (false positive)
GBD	Genome Browser Database
GBD	Genome Browser Database
GDIP	Distància generalitzada al predictor ideal
HMM	Models de Markov ocults (Hidden Markov Model)
INB	Institut Nacional de Bioinformàtica
MC	Simulació de Monte Carlo
MCMC	Monte Carlo amb Cadena de Markov
MD	Dinàmica moleculars
miARN	Micro-ARN
ORF	Pauta oberta de lectura (sigla en anglès; open reading frame)
PCR	Reacció en cadena per la polimerasa (sigla en anglès)
PPV	Percentatge de prediccions vàlides
PWM	Matriu de pesos, en l'anglès position weight matrix
Q	Coefficient associat de Yule
RBS	Lloc d'unió al ribosoma (sigla en anglès; ribosome binding site)
RMN	Ressonància magnètica nuclear
SENS	Sensibilitat
siARN	Petits ARN interferència (de l'anglès small interference RNA)
SNP	Nucleòtid d'un sol polimorfisme (sigla en anglès)
snRNA	Petits ARNs nuclears (de l'anglès small nuclear RNAs)
SPEC	Especificitat
TF	Factor de Transcripció (sigla en anglès)
TFBS	Llocs d'unió de factor de transcripció (sigla en anglès)
TFO	Seqüència formadora de triplex (en anglès triplex forming oligonucleotide)
TN	Negatius Verdaders (true negative)
TP	Positius Verdaders (true positive)
TSS	Inici de Transcripció (sigla en anglès)
TTS	Seqüència diana de triplex (en anglès triplex target sequence)
VIH	Virus de la Immunodeficiència Humana