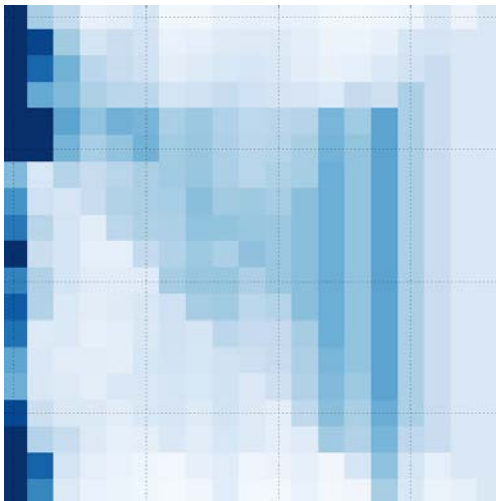


ANÀLISI COMPUTACIONAL DE PROTEÏNES

Efecte de la tendència a la formació d'hèlix- α per part de la proteïna ACTR (IDP) en el procés d'acoblament a la proteïna CREB.

Hipòtesi de la Longitud d'hèlix- α

Treball de recerca de Batxillerat



Magalí Luna

Tutor: **David Cosials**

Institut Poeta Maragall

Setembre 2015- Gener 2017

Índex

I. Motivació

Motivació de la recerca (7)

II. Introducció

1. Introducció (11)

- 1.1. Les IDP i el seu comportament biològic
- 1.2. Les proteïnes ACTR i CREB, el domini NCBD i l'hèlix- α
- 1.3. L'anàlisi computacional del comportament de plegament de les proteïnes: el mètode Monte Carlo

2. Hipòtesi i objectius del treball de recerca (17)

- 2.1. Objectiu inicial
- 2.2. Hipòtesi de la Longitud d'hèlix- α
- 2.3. Objectiu específic

3. Material i mètodes (19)

- 3.1. Software
- 3.2. Hardware
- 3.3. Centre de recerca i guiatge acadèmic
- 3.4. Mètodes

III. Recerca

4. Les proteïnes implicades: ACTR i CREB (25)

- 4.1. ACTR
- 4.2. La proteïna CREB
- 4.3. El lloc d'unió NCBD de la proteïna CREB
- 4.4. Les hipòtesis d'acoblament d'ACTR i CREB
 - 4.4.1. Hipòtesi de l'acoblament induït
 - 4.4.2. Hipòtesi de la selecció conformacional
- 4.5. Resultats experimentals previs
- 4.6. Hipòtesi de Longitud d'hèlix- α

5. Les restriccions a l'hèlix- α d'ACTR: importància per a l'anàlisi computacional (33)

- 5.1. L'estructura de les proteïnes
- 5.2. L'hèlix- α
- 5.3. Els angles de rotació de l'enllaç peptídic:
factor de càlcul probabilístic en l'anàlisi computacional

6. El mètode Monte Carlo i el programari utilitzat (43)

- 6.1. El mètode de Monte Carlo
- 6.2. Els cicles de càlcul de Monte Carlo
- 6.3. Les Protein Data base (PDB)
- 6.4. El software d'anàlisi de proteïnes: Profasi i Phaistos
 - 6.4.1. Profasi
 - 6.4.2. Phaistos
 - 6.4.3. VMD

7. Els mètodes computacionals utilitzats (49)

- 7.1. Simulacions amb Profasi
- 7.2. Simulacions amb Phaistos
- 7.3. iPython Notebook
- 7.4. Visualització dels resultats

IV. Resultats i conclusions

8. Resultats i discussió (55)

- 8.1. Resultats amb Profasi
- 8.2. Resultats amb Phaistos

9. Conclusions (63)

10. Recerca futura (65)

Agraïments (67)

Índex de figures (69)

Referències i bibliografia consultada (71)

V. Annex

Computational Analysis of the Effect of Helical Propensity in Intrinsically Disordered Proteins to the Binding Process (77)

Article presentat al Programa Joves i Ciència per optar a participar en el programa RSI 2016 al Massachusetts Institute of Technology

I. Motivació

Motivació del treball

El 1958, Frederick Sanger va presentar a la comunitat científica la primera seqüenciació d'una proteïna: la insulina. Aquest descobriment li va fer merèixer aquell mateix any el Premi Nobel de Química.

L'any següent, el 1959, Max Perutz i Sir John Cowdery Kendrew van determinar les primeres estructures proteiques, la de l'hemoglobina i la mioglobina respectivament. Per aconseguir-ho, van utilitzar una tècnica anomenada de difracció de raigs X. Aquesta aportació al coneixement els va permetre compartir el Premi Nobel de Química del 1962.

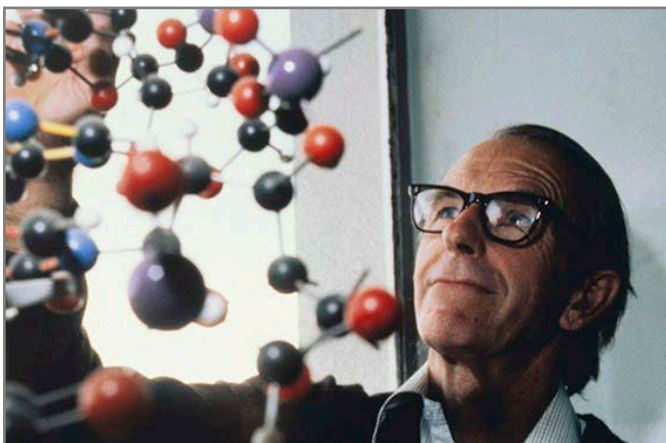


Figura 1. El bioquímic britànic Frederick Sanger, guardonat en dues ocasions amb el premi Nobel de Química (1958 i 1980), estudiant un dels seus models estructurals.

Font: Alchetron.com

Després d'aquests descobriments, després d'aquests grans avenços, es podria pensar que ja s'havien aclarit tots els misteris al voltant d'aquestes macromolècules. Però només en van ser el principi.

Durant les dècades següents es va observar que hi havia proteïnes l'estructura de les quals no podia ser descrita per la tècnica de difracció de raigs X. Eren proteïnes tant inestables que no podien ser convertides en cristalls. Les van anomenar proteïnes intrínsecament desordenades, *Intrinsically Disordered Proteins* (IDP).

L'any 2000, estudis bioinformàtics van predir que aquestes IDP eren molt més comunes del que es podria pensar inicialment, i que eren especialment abundants si, en concret, es parlava de proteïnes relacionades amb malalties.

Aquesta afirmació va fer que la ciència tornés a enfocar la seva curiositat cap a aquestes macromolècules.

Aquest projecte està enfocat a investigar el comportament d'aquestes proteïnes intrínsecament desordenades (IDP); en concret, s'ha centrat en la proteïna ACTR i en la seva unió amb el domini NCDB de la proteïna CREB, involucrada en la Síndrome de Rubinstein-Taybi i en la Leucèmia Mieloide.

La part pràctica d'aquest projecte va ser realitzada al Institut de Química Avançada de Catalunya (iQAC), inscrit en el Consejo Superior de Investigaciones Científicas (CSIC), sota la direcció del Doctor Ramon Crehuet.

Vaig estar treballant en el laboratori de bioquímica computacional de l'iQAC durant cinc mesos per arribar als resultats que presentaré a continuació, i actualment hi segueixo treballant per tal de preparar aquest projecte per a ser publicat en una revista científica.

Aquesta col·laboració va ser possible gràcies al programa Joves i Ciència, una iniciativa de la Fundació Catalunya - La Pedrera que proporciona a estudiants interessats en la ciència l'oportunitat de dur a terme recerca d'avantguarda en laboratoris d'investigació tant nacionals com internacionals.

El programa Joves i Ciència patrocina als seus estudiants durant tres estius. L'estiu del 2015, vaig passar dues setmanes fent recerca biomèdica a l'edifici singular Món Natura Pirineus situat a Les Planes de Són, al Pallars Sobirà.

L'estiu del 2016, gràcies al treball realitzat a l'iQAC que presentaré a continuació, vaig tenir el plaer de poder participar en el prestigiós programa del *Research Science Institute* que té a lloc al *Massachusetts Institute of Technology* (MIT) a Cambridge, Boston.

Durant sis setmanes vaig estar col·laborant en el *MIT Kavli Institute of Astrophysics and Space Research* sota la direcció del Doctor Rongmon Bordoloi, on hi vaig poder fer una investigació en l'àmbit de la cosmologia que vaig presentar en un article que porta el títol «*Measurement of the Circumgalactic Gas Around $1.8 < z < 3.0$ Galaxies Using Foreground -Background Galaxy Pairs*». Hi segueixo treballant a distància en l'article que també té prevista la publicació en una revista científica durant els propers mesos.



Figura 2. A dalt, participants al programa RSI 2016 al Massachussets Institute of Technology (MIT). A l'esquerra, amb el biòleg Howard Robert Horvitz, Premi Nobel de Medicina l'any 2002.

Font: Pròpia

Totes aquestes experiències tenen un gran valor per a mi, i m'han ajudat a definir-me en l'àmbit acadèmic i en el personal. Però tot aquest viatge va començar amb la motivació més bàsica de totes, aquella que fa girar el món: preguntar-se el per què de les coses.

II. Introducció

1. Introducció

1.1. Les IDP i el seu comportament biològic

Les **proteïnes** són macromolècules complexes que a compleixen funcions biològiques essencials. En general, presenten una estructura de plegament estable que s'organitza en quatre nivells: estructura primària, secundària, terciària i quaternària, nivells que segueixen una jerarquia de complexitat tridimensional creixent.

En canvi, les anomenades **proteïnes intrínsecament desordenades (IDP**, de l'anglès *Intrinsically Disordered Proteins*) són una família de proteïnes que, aparentment, tenen una funció biològica sense una estructura estable ben definida (més enllà de la seqüència d'aminoàcids) en condicions fisiològiques normals [1].

En general, les IDP aconseguixen una estructura de plegament estable **a mesura que s'uneixen** a altres macromolècules per accomplir una funció biològica determinada.

No obstant això, s'ha observat que, en molts casos, aquestes proteïnes tenen una estructura secundària **transitòria** abans del procés d'acoblament. Per ara no es coneix completament la manera en què aquesta estructura secundària transitòria afecta l'eficàcia de l'acoblament.

1.2. Les proteïnes ACTR i CREB, el domini NCBD i l'hèlix - α

La proteïna **ACTR** és una IDP que interactua amb el domini, o lloc d'unió, **NCBD** de la proteïna **CREB**.

CREB és una proteïna codificada en humans pel gen CREBBP. Aquesta proteïna compleix un paper biològic essencial, ja que és reguladora de l'activitat de molts altres gens. Està relacionada amb l'activació del procés de transcripció de l'ADN, el desenvolupament embrionari, el control de creixement i l'homeòstasi. Determinades disfuncions en aquest gen provocades per mutacions poden causar, per exemple, la síndrome de Rubinstein-Taybi o la leucèmia mieloide aguda.

ACTR és precisament una de les proteïnes coactivadores de CREB. Quan ACTR s'acobla al domini NCBD de la proteïna CREB, aquesta pot desenvolupar la seva funció reguladora de la transcripció d'altres gens.

Com ja s'ha comentat, ACTR és una IDP, i és precisament quan s'uneix al domini NCBD de la proteïna CREB quan **ACTR es plega** en una de les estructures proteíniques secundàries més comunes, l'**hèlix- α** o **hèlix alfa**, i pot començar a desenvolupar la seva funció coactivadora.

Però, com es veurà durant l'exposició d'aquest treball, encara estan en discussió **dues hipòtesis**: si l'hèlix- α es forma justament durant el **procés** d'acoblament o si, contràriament, ja existeix una estructura secundària transitòria prèvia. Alguns resultats experimentals obtinguts en laboratori, mostren que la segona hipòtesi, anomenada "Hipòtesi de la selecció conformacional", podria ser la correcta [2].

Per dur a terme aquests estudis de laboratori esmentats anteriorment, s'han creat experimentalment algunes **mutacions de ACTR** que afavoreixen, o bé pertorben, la formació d'una hèlix- α en ACTR abans de l'acoblament. Algunes d'aquestes mutacions seran utilitzades en el nostre treball.

1.3. L'anàlisi computacional del plegament de les proteïnes: el mètode Monte Carlo

El plegament de les cadenes proteiques en estructures tridimensionals és un fenomen molt complex en el qual intervenen factors diversos, entre els quals cal destacar, principalment, la seqüència d'aminoàcids de la cadena i les característiques químiques de cada un dels grups funcionals: carboxil, amino i lateral. Els aminoàcids interaccionen entre ells i amb el seu medi cel·lular per produir una forma tridimensional ben definida.

L'estudi de l'estructura tridimensional de les proteïnes ha ocupat els bioquímics des dels inicis del segle XX, i s'ha portat a terme amb les tècniques bioquímiques tradicionals com l'anàlisi per difracció de raigs X, l'electroforesi o la ressonància magnètica nuclear (RMN). El procés és **lent, laboriós i costós**.

En els darrers temps, l'anàlisi amb utilització d'ordinadors, o **anàlisi computacional**, és utilitzat en moltes branques de la ciència en que la quantitat de factors és elevada i de difícil control per part de l'investigador; com ara la meteorologia, l'ecologia o, com en el nostre cas, l'estudi de l'estructura de plegament de les proteïnes. L'anàlisi computacional permet la **realització de simulacions**, que a més de la informació i els avantatges que aporten al procés d'investigació, **rebaixen notablement el temps invertit i les despeses**.

Existeix una gran quantitat de software o programari dissenyat per a la investigació computacional de l'estructura de les proteïnes. **En el nostre cas hem utilitzat dos d'aquests programes: Profasi i Phaistos**.

Aquests dos programes, com gran part del software dissenyat per dur a terme un anàlisi computacional de plegament de proteïnes utilitzen el mètode de **Monte Carlo**. Monte Carlo és un sistema d'**anàlisi estadístic** basat en l'estudi de la probabilitat que es doni un fenomen concret. En una simulació Monte Carlo, els elements d'estudi es mouen a l'atzar dins de les opcions que els són possibles. Per exemple, el mètode considerarà que un aminoàcid formarà, o no, un enllaç químic o un altre amb l'aminoàcid contigu. Aquest atzar es basarà en la probabilitat d'una opció o de l'altra segons les lleis fisicoquímiques. A partir d'aquí seguirà amb el següent aminoàcid i així successivament.

Amb aquest mètode és possible determinar, per exemple, la probabilitat que una determinada cadena d'aminoàcids adopti, o no, una estructura específica: en el nostre cas una hèlix- α per part d'ACTR.

2. Hipòtesi i objectius del treball de recerca

2.1 Objectiu inicial

L'objectiu inicial d'aquest treball de recerca és l'ús de **simulacions computacionals**, mitjançant software d'anàlisi de proteïnes basat en el mètode **Monte Carlo**, per determinar la probabilitat que la proteïna **ACTR** formi una estructura secundària, concretament una **hèlix- α** , **abans de l'acoblament** al lloc d'unió **NCBD** de la proteïna **CREB**.

En particular, **ens proposem aportar proves** en favor de la hipòtesi de la **selecció conformacional**, que afirma que ja existeix una estructura secundària **transitòria prèvia**, enfront de la hipòtesi alternativa que diu que l'hèlix- α es forma justament **durant el procés** d'acoblament, o **hipòtesi de l'acoblament induït**. Com ja ha estat comentat, diversos resultats experimentals, avancen que la hipòtesi de la selecció conformacional podria ser la correcta [2].

2.2 Hipòtesi de la longitud d'hèlix- α

L'aportació innovadora d'aquest treball de recerca és la **Hipòtesi de la longitud d'hèlix- α** .

A diferència del criteri "tradicional" utilitzat en les anàlisis computacionals clàssiques, que dóna compte de la probabilitat de formació d'una hèlix- α pel mètode de sumar la probabilitat que ho faci cada aminoàcid individualment, la nostra recerca va proposar una variant. Aquesta alternativa consisteix en utilitzar l'anàlisi computacional per fer una simulació que ens informés de la **longitud total d'estructura d'hèlix- α** de la qual forma part l'aminoàcid estructurat. En aquest cas es té en compte no només el percentatge de formacions individuals d'hèlix- α de cada aminoàcid, sinó la llargada de les estructures en forma d'hèlix- α de les quals aquests formen part.

En conseqüència, **la hipòtesi d'aquest treball de recerca** és comprovar **si la longitud o llargada total d'estructura d'hèlix- α és una variable significativa** a l'hora de dur a terme el procés d'unió de la proteïna **ACTR** al lloc d'unió **NCBD** de la proteïna **CREB**.

La comprovació d'aquesta hipòtesi aportaria també un valor útil per a una millor discriminació entre les dues hipòtesis de partida, a favor de la **Hipòtesi de la selecció conformacional**.

2.3 Objectiu específic

Finalment, i també molt important en un sentit pràctic, un **objectiu específic d'aquest treball de recerca** és comprovar la correspondència entre els resultats experimentals de laboratori i els mètodes computacionals; és a dir, verificar en quina mesura els resultats experimentals de laboratori es poden predir mitjançant els mètodes computacionals. Això pot ser útil per reduir el temps, el treball humà i també les despeses econòmiques de les investigacions.

En els següents capítols es presenten en detall els conceptes i els mecanismes implicats en la recerca, així com les metodologies utilitzades per al treball.

3. Material i mètodes

Com acabem de comentar, l'objectiu d'aquest treball de recerca ha estat l'ús de **simulacions computacionals** per determinar tendències de formació d'hèlix- α en **ACTR**, mitjançant ordinadors i software especialitzat en l'anàlisi d'estructura de proteïnes basat en el mètode **Monte Carlo**.

Així doncs, per desenvolupar-lo, a banda del material bibliogràfic i de recerca que es ressenya en l'apartat de bibliografia, s'han fet servir el material i els mètodes següents:

3.1. Software

- Programa d'anàlisi computacional de proteïnes **Profasi**
- Programa d'anàlisi computacional de proteïnes **Phaistos**
- Llenguatge de programació **Python – iPython Notebook**
- Programa de generació de gràfics **Gnuplot i Matplotlib**
- Programa visor de proteïnes **VMD** (Molecular Visual Dynamics)

3.2. Hardware

- **Clúster computacional** del l'Institut de Química Avançada de Catalunya (iQAC).
- **Ordinador personal** (sistema operatiu Linux Mint)

3.3. Centre de treball i guiatge acadèmic

Tota la recerca es va portar a terme a les instal·lacions de l'Institut de Química Avançada de Catalunya (iQAC), sota el guiatge del Doctor Ramon Crehuet.

3.4. Mètodes

Per fer les simulacions computacionals de la seqüència d'aminoàcids original i de totes les mutacions que s'han proposat, s'ha fet servir el mètode de Monte Carlo.

Es tracta d'un mètode numèric que explora les diverses possibilitats de plegament de proteïnes mitjançant opcions a l'atzar utilitzant la tècnica matemàtica del *Importance sampling*. Els fonaments d'aquest mètode s'expliquen en el capítol corresponent al mètode de Monte Carlo.

Per determinar la probabilitat que la proteïna **ACTR** formi una estructura secundària, concretament una **hèlix- α** , **abans de l'acoblament** (estat "no acoblat") al lloc d'unió **NCBD** de la proteïna **CREB**, s'ha analitzat la proteïna natural, és a dir, l'estat anomenat "**salvatge**", així com diversos casos de **mutació de la seqüència d'aminoàcids** dissenyats experimentalment en laboratoris.

Així, s'han fet un total de **9 anàlisis computacionals**, d'acord amb la distribució següent (s'indica la denominació dels tipus mutants):

- Anàlisi de la seqüència natural o "salvatge"
 - ✓ WT Wild Type (tipus "salvatge" en anglès)
- Anàlisi de 4 seqüències experimentals amb 1 mutació
 - ✓ T1054Q
 - ✓ S1043M
 - ✓ A1047G
 - ✓ D1050E
- Anàlisi de 3 seqüències experimentals amb 2 mutacions
 - ✓ ST_doble
 - ✓ SD_doble
 - ✓ DT_doble
- Anàlisi de 1 seqüència experimental amb 3 mutacions
 - ✓ Mutant Triple (conté les mutacions T1054Q, S1043M i D1050E)

La recerca de la probabilitat o “**tendència d'hèlix- α** ” es fa amb dos criteris diferents:

Criteri 1 o tradicional: probabilitat acumulativa, o “aminoàcid a aminoàcid”

És el criteri “tradicional”, utilitzat habitualment en els estudis bioquímics. S'obté calculant, primer, la probabilitat que cada un dels aminoàcids, individualment, estigui formant part d'una hèlix- α , independentment de la presència o manca d'estructura de la resta d'aminoàcids de la cadena; i, posteriorment, calculant un “**acumulatiu**” o sumatori de la probabilitat d'ordenació de tots els aminoàcids, la qual cosa ens dona una probabilitat o **tendència general d'hèlix- α** de tota la cadena de l'ACTR. És a dir, el mètode dona una mesura de la probabilitat concreta de que cada un dels aminoàcids estigui formant una hèlix- α , i en base a aquesta informació s'elabora un índex de probabilitat, o tendència a hèlix- α , de tota la cadena.

Criteri 2, Hipòtesi innovadora del nostre treball de recerca”: probabilitat simultània, o de “longitud total de l'hèlix- α ”

Aquest ha estat el criteri en el qual es basa la hipòtesi del nostre Treball de recerca. A diferència del criteri “tradicional”, “aminoàcid a aminoàcid”, utilitzat en les anàlisis computacionals clàssiques, la nostra recerca va proposar la variant consistent a utilitzar l'anàlisi computacional per fer una simulació que ens informés de la **longitud total d'estructura d'hèlix- α** que es produïa en un mateix moment, és a dir, que donés compte del nombre d'aminoàcids que estaven formant part d'un tram d'hèlix- α simultàniament, en un mateix moment.

La introducció en el nostre treball de recerca d'aquesta hipòtesi, de “longitud total de l'hèlix- α ”, combinat amb l'anàlisi de les diverses mutacions estudiades, ens pot ajudar a confirmar, o rebutjar, una o altra de les dues hipòtesis anteriorment descrites: si l'hèlix- α es forma durant el procés d'acoblament o si, en canvi, existeix una estructura secundària transitòria prèvia com indica la hipòtesi de la “selecció conformacional” que sembla ser certa segons els estudis experimentals previs.

Adicionalment i com a novetat en la recerca, la nostra **Hipòtesi de la longitud d'hèlix- α** , ens ha permès introduir un nou paràmetre no tradicionalment contemplat que estudia si, a més a més de la probabilitat individual de formació d'hèlix- α , **la llargada de l'hèlix- α de la qual forma part**

l'aminoàcid estructurat és un factor important a l'hora de dur a terme el procés d'acoblament de la proteïna ACTR el domini NCBD de la proteïna CREB.

Com es veurà en l'apartat dedicat a les **"Conclusions"**, la **Hipòtesi de la longitud d'hèlix- α sembla ser correcta**, d'acord amb els resultats de les simulacions computacionals; la qual cosa dóna suport no només a la nostra mateixa hipòtesi sinó també a la hipòtesi de la selecció conformacional, tal com apuntaven els experiments previs de laboratori [2] i la nostra pròpia recerca.

III. Recerca

4. Les proteïnes implicades: ACTR i CREB

4.1. ACTR

ACTR (*Activator for Thyroid hormone and Retinoid receptors*), la proteïna que és **objecte d'estudi** d'aquest treball de recerca, és un **coactivador** d'un receptor nuclear: la proteïna CREB. ACTR és també una **proteïna intrínsecament desordenada** o IDP. Quan s'uneix al domini NCBD de la proteïna CREB, es plega i comença a fer la seva funció coactivadora. La proteïna va ser descrita per primera vegada el 1997 a la revista *Cell*. [3].

Un **coactivador** (en el nostre cas, ACTR) és una proteïna que incrementa la transcripció de l'ADN, o expressió gènica, quan s'uneix a un receptor nuclear d'una altra proteïna anomenada també activador o factor de transcripció (en el nostre cas, CREB). El receptor conté diversos dominis o llocs d'unió: almenys un lloc d'unió amb el coactivador i també un lloc d'unió a l'ADN que és el que permet realitzar la funció de transcripció o expressió gènica: el procés pel qual a partir d'aquest gen en concret se sintetitza el producte gènic corresponent, la seva proteïna.



Figura 3.: Article original de la revista *Cell* en què es descriu per primera vegada la proteïna **ACTR** i la seves funcions biològiques. (1997)

4.2. La proteïna CREB

La proteïna **CREB** (*Cyclic-AMP-Response-Element-Binding protein*) té diverses funcions biològiques relacionades amb l'activació del procés de transcripció de l'ADN, i per tant té un paper molt important en aspectes biològics fonamentals com el desenvolupament embrionari, el control de creixement i l'homeòstasi o regulació de l'estabilitat interna en els éssers vius. [4].

De fet, CREB és una proteïna relacionada amb els coneguts gens HOX que regulen l'anomenada transcripció "corrent avall" (*downstream*) de tot un conjunt de gens importants en el desenvolupament embrionari de les espècies. CREB està sintetitzada pel gen CREBBP i va ser descrita per primer cop el 1987 a la revista *Nature*.

The image shows a screenshot of the original article page from the journal Nature. The page features the journal's logo and title at the top, followed by an access notice. The article title is 'Binding of a nuclear protein to the cyclic-AMP response element of the somatostatin gene', authored by Marc R. Montminy and Louise M. Bilezikjian. The article text begins with 'Many hormones act on neuroendocrine cells by activating second messenger pathways. Two of these, the phosphoinositol and cAMP-dependent pathways, cause changes in cellular activity through specific protein kinases. By phosphorylating cytoplasmic and nuclear proteins, these kinases apparently coordinate cellular processes, including the biosynthesis and release of neuropeptides. Somatostatin biosynthesis and release, for example, are both positively regulated by the second messenger cAMP in hypothalamic cells¹, and cAMP also induces somatostatin gene transcription 8-10-fold in transfected PC12 pheochromocytoma cells². Transcriptional induction requires a 30-nucleotide cAMP response element (CRE) which is conserved in other cAMP-responsive genes²⁻⁴. This element also confers cAMP responsiveness when placed upstream of the heterologous simian virus 40 (SV40) promoter. The somatostatin gene does not, however, respond to cAMP in mutant PC 12 cells which lack cAMP-dependent protein kinase type II activity². Activation of somatostatin gene transcription may consequently require the phosphorylation of a nuclear protein which binds to the CRE. Using a DNase I protection assay, we have characterized a nuclear protein in PC 12 cells which binds selectively to the CRE in the somatostatin gene. We have purified this protein which is of relative molecular mass 43,000 (M_r 43K) by sequence-specific DNA affinity chromatography. This 43K CRE binding protein (CREB) is phosphorylated *in vitro* when it is incubated with the catalytic subunit of cAMP-dependent protein kinase. Stimulating PC 12 cells with forskolin, an activator of adenylyl cyclase, causes a 3-4-fold increase in the phosphorylation of this protein. We conclude that the cAMP-dependent pathway may regulate gene transcription in response to hormonal stimulation by phosphorylating this CREB protein.'

Figura 4. Article original de la revista *Nature* en què es descriu per primera vegada la proteïna CREB i la seves funcions biològiques. (1987)

Es tracta d'una proteïna molt important, ja que a més de les funcions que realitza, les mutacions en el seu gen causen diverses **malalties** com la síndrome de Rubinstein-Taybi (RTS) [5] i la leucèmia mieloide [6].

La **síndrome de Rubinstein-Taybi** és una malaltia genètica caracteritzada per trets facials característics en els pacients, baixa estatura, malformacions als polzes i primers dits dels peus (dits amples, en porra o maça) i graus variables de retard mental.

L'origen de la síndrome es troba en una mutació del gen CREBBP, en el cromosoma 16, mutació que impedeix la transcripció adequada de la **proteïna CREB**. Aquesta síndrome afecta una de cada 125.000 persones. El primer cas va ser observat en 1957 pels doctors grecs Matsoukas i Theodoru.



Figura 5. Imatge de Mollie, una nena amb la síndrome de Rubinstein-Taybi. S'hi poden observar els ulls lleument separats, les parpelles caigudes, l'expressió dolça i els polzes gruixuts, algunes de les característiques d'aquesta síndrome.

Foto: Frederikke Brostrup

La **leucèmia mieloide aguda** és un tipus de càncer causat per la reproducció ràpida de cèl·lules blàstiques o mieloides (les precursors dels leucòcits) que s'acumulen a la medul·la òssia i dificulten la producció de glòbuls vermells normals. Sovint està precedida per una fase sense símptomes, la leucèmia mieloide crònica, que afecta només la base genètica.

Diversos investigadors han descobert que la proteïna **CREB** es sobreexpressa en les cèl·lules blàstiques de pacients amb leucèmies agudes. [7]

És la leucèmia més habitual en els adults, el seu risc augmenta amb l'envelliment i és responsable del 1% de les morts per càncer. És el tipus de leucèmia, per exemple, a què es dedica la Fundació Josep Carreras contra la leucèmia (<http://www.fcarreras.org/>).



Figura 6. Imatge de la presentació de la Fundació Josep Carreras contra la leucèmia en la seva web (esquerra), i logotip de la Fundació (a dalt).

Font: www.fcarreras.org

4.3. El lloc d'unió NCBD de la proteïna CREB

La proteïna CREB també és coneguda pels seus **dominis actius**, o llocs d'unió, com ara NCBD (*Nuclear-receptor Co-activator Binding Domain*), que interactua amb **ACTR**, el **coactivador que ens interessa**.

Dit d'una altra manera, el **domini NCBD de la proteïna CREB** és el lloc on **ACTR s'uneix a CREB**. A partir del moment de l'acoblament, la proteïna CREB pot desenvolupar la seva funció de transcripció de l'ADN i els gens corresponents es poden expressar, és a dir, poden sintetitzar les seves proteïnes. Com ja hem vist, aquestes tenen una gran importància en el desenvolupament embrionari i en altres funcions biològiques bàsiques, així com un paper destacat en disfuncions com la síndrome de Rubinstein-Taybi i la leucèmia mieloide.

El **domini actiu d'una proteïna** és una part d'aquesta proteïna amb estructura terciària que pot funcionar independentment de la resta de la proteïna. Cada domini és una estructura tridimensional estable per si mateixa. Moltes proteïnes es componen de diversos dominis estructurals. Al mateix temps, un domini que codifica per una certa macromolècula concreta pot ser trobat en diverses proteïnes diferents.

Això vol dir que la bioquímica cel·lular utilitza aquests dominis com a **blocs de construcció** o mòduls que estan presents en diferents llocs per crear proteïnes amb diferents funcions. Per aquesta mateixa raó, i gràcies a les tecnologies de l'**enginyeria genètica**, els dominis poden ser intercanviats entre una proteïna i una altra per produir noves proteïnes diferents, també anomenades proteïnes quimèriques [8].

4.4. Les hipòtesis d'acoblament d'ACTR i CREB

Com ja ha estat comentat, l'objectiu del treball de recerca és l'ús de simulacions computacionals per determinar la probabilitat de que la proteïna **ACTR** formi una **hèlix- α** , **abans de l'acoblament** al domini **NCBD** de la proteïna receptora **CREB**.

En definitiva, es tracta de determinar si les simulacions donen suport a la hipòtesi que l'hèlix- α es forma justament durant el **procés d'acoblament**: **hipòtesi de l'acoblament induït**; o si, per contra, donen suport a l'altra hipòtesi, és a dir, que ja existeix una estructura secundària transitòria prèvia: **hipòtesi de la selecció conformacional**.

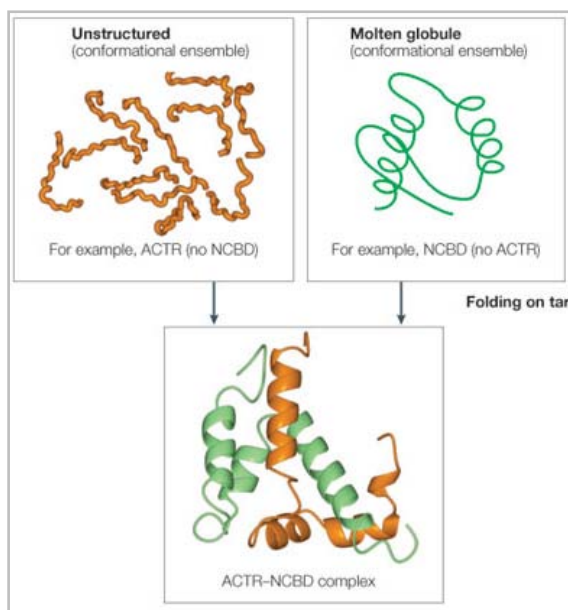


Figura 7. Acoblament d'ACTR al punt d'unió NCBD de CREB.

Font: Nature Reviews. Molecular Cell Biology. www.nature.com

En els dos apartats següents presentem aquestes dues hipòtesis i **una de nova que ha aportat el nostre treball de recerca i que ens permet reforçar les nostres conclusions a favor de la hipòtesi de la selecció conformacional**, hipòtesi que també rep el suport d'alguns resultats experimentals, així com **introduir un nou paràmetre** en l'estudi de l'acoblament de les proteïnes.

4.4.1. Hipòtesi de l'acoblament induït

La hipòtesi de l'acoblament induït defensa que les IDPs no tenen una estructura secundària fins que s'acoblen a una altra proteïna (en el nostre cas, CREB). Segons aquesta hipòtesi, la proteïna desordenada adquireix l'estructura secundària, com per exemple una hèlix- α , després del procés d'acoblament amb l'altra proteïna.

4.4.2. Hipòtesi de la selecció conformacional

La hipòtesi de la selecció conformacional defensa que les IDP tenen algun tipus d'estructura secundària transitòria abans que s'uneixin a una altra proteïna. Aquesta estructura secundària transitòria **atrau l'altra proteïna i afavoreix el procés d'acoblament**.

4.5. Resultats experimentals previs

Com ja hem posat de manifest anteriorment en aquest treball, alguns resultats experimentals obtinguts en laboratori mostren que la hipòtesi de la selecció conformacional podria ser la correcta [2].

Aquesta suposició es basa diversos experiments que mostren que **l'increment de la formació prèvia d'una estructura d'hèlix- α transitòria abans de la unió fa que el procés d'acoblament sigui més ràpid**. Aquests experiments utilitzen formes de la proteïna ACTR amb la **seqüència original de aminoàcids alterada**, generades per una mutació de la seqüència que la codifica. D'ara en endavant anomenarem a aquestes seqüències alterades "mutacions".

Nosaltres hem treballat amb aquestes mutacions, ja que són les més utilitzades en els experiments de laboratori que acabem d'esmentar. Els autors de les proves experimentals han triat aquestes mutacions perquè aquestes incrementen la tendència d'ACTR a formar hèlix- α abans

d'interactuar amb el lloc d'unió. En concret, totes elles faciliten la formació espontània d'hèlix excepte una: la mutació anomenada "A1047G", que la dificulta i s'utilitza com a referència o control.

Utilitzant les mateixes mutacions per a les nostres simulacions, serem capaços de comparar els nostres resultats amb els resultats experimentals i podrem així confirmar o descartar la hipòtesi.

4.6. Hipòtesi de la longitud d'hèlix- α

Com ja ha estat presentat en l'apartat d'objectius, la **Hipòtesi de la longitud d'hèlix- α** és la nostra hipòtesi apostada al treball de recerca. Per estudiar aquesta hipòtesi cal idear i utilitzar una metodologia concreta.

En el camp de l'anàlisi computacional de proteïnes, la metodologia tradicional calcula el grau de formació d'estructura d'hèlix- α per part d'una proteïna donada utilitzant el sistema de sumar la probabilitat de que cada un dels aminoàcids de la cadena estigui estructurat en forma d'hèlix- α .

Tot i ser una mesura àmpliament utilitzada en els estudis bioquímics per la seva utilitat, les característiques de la nostra recerca, en la qual la mesura de la formació d'una hèlix alfa és decisiva, ens va portar a considerar la possibilitat de calcular no només la probabilitat individual de cada aminoàcid de formar una hèlix- α sinó també la **longitud total d'aquesta estructura d'hèlix- α** . És a dir, el contingut (la quantitat) d'aminoàcids que formen hèlix- α **simultàniament**.

En conseqüència, **el nostre treball de recerca** va posar a punt i va fer servir aquesta metodologia **dissenyada per la Hipòtesi de la longitud d'hèlix- α** per a una millor discriminació entre les dues hipòtesis de partida, la de l'acoblament induït i la de la selecció conformacional; a més d'aportar un **nou paràmetre a tenir en compte a l'hora d'estudiar el procés d'acoblament entre proteïnes**.

5. Les restriccions de l'hèlix- α d'ACTR: importància per a l'anàlisi computacional

Fins aquí hem presentat diversos antecedents necessaris per enquadrar l'**objectiu del nostre treball de recerca**: l'anàlisi computacional de la tendència de la proteïna ACTR a formar hèlix- α prèviament al el procés d'unió al domini NCBD de la proteïna receptora CREB.

En l'anàlisi computacional que farem cal tenir molt en compte aspectes com ara el **mecanisme com es genera l'hèlix- α** que, en definitiva, serà la **informació que tindrà en compte l'ordinador per calcular les probabilitats** que es conformi l'hèlix.

Com ja ha estat comentat anteriorment, aquesta probabilitat serà comptada per a cada aminoàcid individualment i per a tota la cadena, **metodologia per a la hipòtesi del nostre Treball de recerca, o Hipòtesi de la longitud d'hèlix- α** , utilitzant percentatges en cada cas.

Aquest capítol està dedicat a aprofundir en aquests mecanismes.

5.1 El marc conceptual: l'estructura tridimensional de les proteïnes

Com que l'**hèlix- α** implicada en la unió d'ACTR al lloc d'unió NCBD de la proteïna CREB és una estructura secundària, i com que el domini o lloc d'unió de CREB depèn de l'estructures terciària i quaternària d'aquesta proteïna, val la pena fer un repàs de les principals característiques del plegament de les molècules proteïques.

Com és ben conegut, les proteïnes són macromolècules biològiques formades per una **seqüència d'aminoàcids**. Els aminoàcids són els monòmers o peces que formaran el polímer o cadena.

Cada aminoàcid està format pel carboni alfa (C α) central i, a més a més, per d'un grup funcional amino (—NH₂), un altre de carboxil (—COOH) i finalment una cadena lateral (—R) que varia entre els diferents aminoàcids.

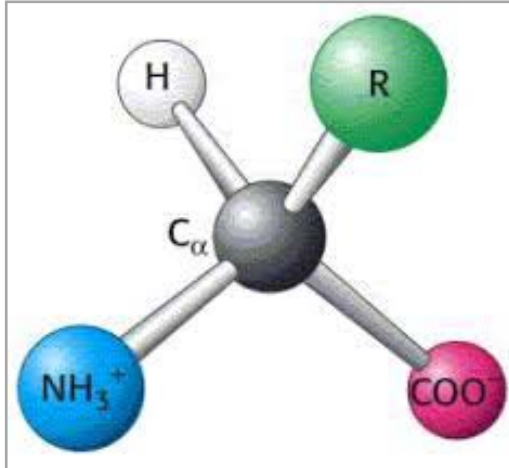


Figura 8. Aminoàcid en la seva forma ionitzada o dipolar, amb càrrega negativa al grup carboxil (en vermell) i positiva al grup amino (en blau). La cadena lateral “—R” es mostra en color verd.

Font: Berg, Tymoczko & Stryer. *Bioquímica*.

La cadena de la proteïna es forma durant el procés de traducció, quan l'ARN missatger és descodificat i els diferents aminoàcids resultants de la descodificació s'uneixen entre sí per formar una cadena peptídica o polipèptid, mitjançant l'enllaç peptídic.

L'enllaç peptídic resulta de la unió entre dos aminoàcids quan el grup amino d'un d'ells reacciona amb el grup carboxil de l'altre, amb la generació d'una molècula d'aigua. L'enllaç es repeteix amb els aminoàcins següents fins a completar el polipèptid.

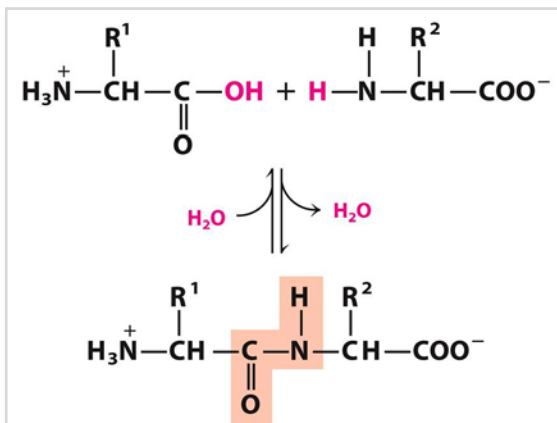


Figura 9. Enllaç peptídic

Font: Lehninger. *Bioquímica*.

Tot i que aquesta seqüència determina la composició fonamental de la molècula, **la funció de la proteïna es porta a terme gràcies a la seva forma o estructura tridimensional.**

Aquesta estructura tridimensional és fruit del replegament del polipèptid gràcies als enllaços que es formen entre els diversos aminoàcids entre si, normalment **ponts d'hidrogen**, a causa de les característiques fisicoquímiques de cada aminoàcid present en la seqüència.

El resultat final és una mena de cabdell irregular, no gaire diferent als **cabdells de corda o filferro** que podem trobar als nostres calaixos de les eines, però que a diferència d'aquests no es formen a l'atzar sinó que respon a lleis físiques i químiques molt determinades i precises.



Figura 10. Al·legoria d'una proteïna plegada amb un cabdell de corda embolicat.

L'estructura tridimensional és fonamental perquè les proteïnes puguin fer la seva **funció**. Si a causa d'una mutació, un aminoàcid no és el "correcte" i la proteïna no es plega en la forma adequada, les proteïnes no poden desenvolupar la seva funció correctament. Una gran part de les **malalties d'origen genètic** tenen el seu origen en proteïnes que no s'han transcrit i plegat correctament.

Com és ben conegut, l'estructura de plegament de les proteïnes s'organitza en quatre nivells que segueixen una jerarquia de complexitat tridimensional creixent, que repassem tot seguit:

- **Estructura primària**, marcada per la seqüència d'aminoàcids del polipèptid. L'estructura primària té un caràcter lineal, o més aviat "planar", encara que químicament no ho és ben bé del tot.
- **Estructura secundària**. La cadena d'aminoàcids té tendència a plegar-se i a adoptar una estructura regular característica, que guanya estabilitat gràcies als enllaços de pont d'hidrogen. El plegament es produeix a causa de diversos factors com les característiques fisicoquímiques de cada aminoàcid o les condicions del medi en què es mouen. Majoritàriament, les proteïnes adopten dos tipus d'estructura

secundària: l'hèlix- α , com la que adopta la nostra proteïna ACTR, i el full plegat o full- β ; i, menys habitualment, altres estructures com els girs- β o la superhèlix- α de les fibres de col·lagen (pell, ossos) i de la queratina (ungles, pèl).

- **Estructura terciària.** L'estructura secundària es replega sobre sí mateixa i adopta formacions tridimensionals esfèriques o globulars més o menys irregulars, en una forma que és específica de cada proteïna i que es manté estable pels enllaços de pont d'hidrogen, els de Van der Waals i els iònics. A més a més, aquesta forma característica crea espais puntuals amb una microestructura, una polaritat i unes càrregues específiques capaces de formar les unions amb altres molècules: són els receptors, els **llocs d'unió o acoblament també anomenats dominis, com és el cas del domini NCBD de la nostra proteïna CREB.**
- **Estructura quaternària.** Moltes proteïnes, com per exemple l'hemoglobina o la mateixa CREB, estan formades per la unió de més d'una cadena d'aminoàcids. Així, aquestes molècules consten de diverses formacions o subunitats.

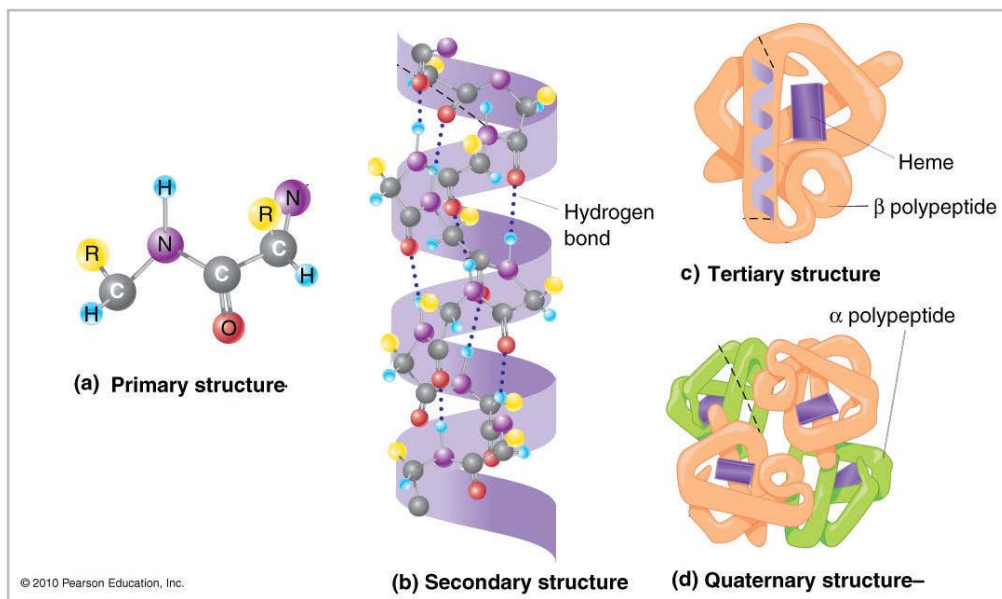


Figura 11. Estructura tridimensional de les proteïnes. a) Estructura primària en un dipèptid; b) Estructura secundària, en aquest cas una hèlix- α ; c) Estructura terciària: l'hèlix- α i/o els fulls- β s'han recargolat sobre si mateixos; d) Estructura quaternària: 4 subunitats amb plegament terciari s'han acoblat per formar una proteïna de 4 cadenes, en aquest cas l'hemoglobina.

Font: Memorial University of Newfoundland. (mun.ca). © 2010 PJ Russell, iGenetics 3 ed.

5.2 L'hèlix- α

L'hèlix- α és l'estructura secundària de plegament més habitual de les cadenes peptídiques.

És l'estructura de plegament que **adopta la nostra proteïna d'estudi ACTR en la seva transició entre una proteïna intrínscament desordenada (IDP) i una proteïna estructurada durant el seu procés d'acoblament** al domini NCBD de la proteïna receptora CREB. Per això és interessant fer un repàs de les seves característiques.

L'hèlix- α va ser descoberta el 1951 pel químic nordamericà Linus Pauling a partir d'estudis de cristal·lografia. Tres anys més tard (1954), Pauling va rebre el Premi Nobel de química. Pauling va rebre anys després el Premi Nobel de la Pau (1962).

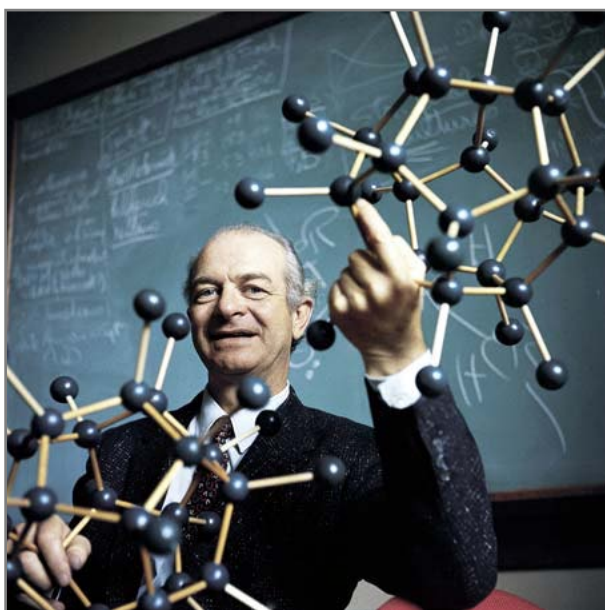


Figura 12. Linus Pauling, descobridor de l'hèlix- α , és una de les quatre persones que han rebut dos premis Nobel, juntament amb Frederick Sanger (també implicat en l'estudi de les proteïnes), Marie Curie i el físic nordamericà John Bardeen.

Font: Cienciaonline.com.

L'hèlix- α (o hèlix alfa) és una estructura espiral, o helicoïdal (com el seu nom indica). En aquesta estructura helicoïdal, cada aminoàcid es relaciona amb el següent de la cadena **mitjançant un gir de 100° sexagesimals**, cosa que fa que hi hagi 3,6 aminoàcids per volta d'hèlix.

En aquesta estructura espiral, cada grup amino (-NH) de l'aminoàcid situat en la posició concreta forma un pont d'hidrogen amb el grup carbonil (CO) de l'aminoàcid situat quatre llocs més endavant en la seqüència peptídica.

L'aparença final de l'estructura és una forma cilíndrica a l'interior de les quals hi ha els enllaços peptídics i a l'exterior els grups laterals (-R).

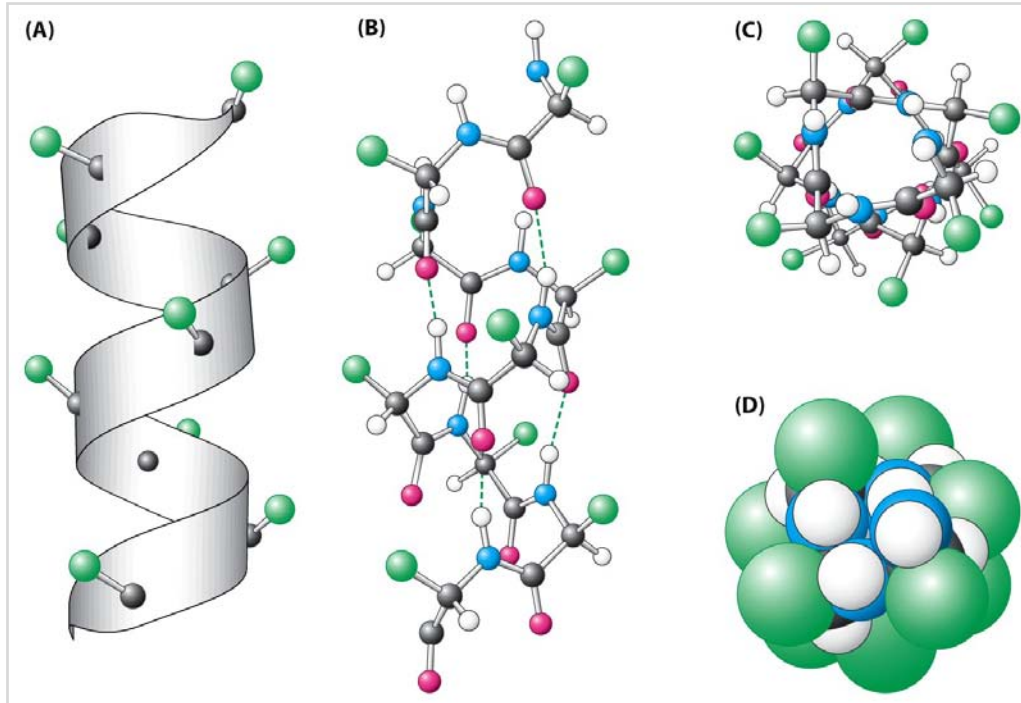


Figura 13. Estructura tridimensional de l'hèlix- α . A) Representació del cilindre o cinta, amb els $C\alpha$ en el centre (en negre) i els radicals laterals (-R) cap a fora (en verd); B) Vista lateral amb un model de boles i pals, que mostre els ponts d'hidrogen entre els grups amino i carboni (línies discontinües); C) Vista vertical, amb la part interior de l'hèlix i els radicals laterals (-R) cap a fora (en verd); D) Vista vertical d'una estructura de boles, amb la part interior de l'hèlix i els radicals laterals (-R) cap a fora (en verd).

Font: Berg, Tymoczko & Stryer. *Bioquímica*.

Les raons per les quals es dona, o no, cada gir i el posterior enllaç de pont d'hidrogen amb el següent (en definitiva si es forma hèlix- α o no entre dos aminoàcids) s'analitzen al capítol següent i són la base de treball de l'estudi probabilístic computacional.

5.3. Els angles de rotació de l'enllaç peptídic: factor de càlcul probabilístic en l'anàlisi computacional

L'enllaç peptídic es troba bàsicament en un mateix pla. Per a cada parell d'aminoàcids units en un enllaç peptídic, trobem 6 àtoms en aquest mateix pla: l'àtom del carboni- α i el grup CO del primer aminoàcid, i el grup NH i l'àtom del carboni- α del segon aminoàcid.

La naturalesa de l'enllaç químic en un pèptid explica el caràcter pla de l'enllaç, ja que té un considerable caràcter de **doble enllaç**, la qual cosa impedeix la rotació al seu voltant i dóna **rigidesa a l'enllaç**.

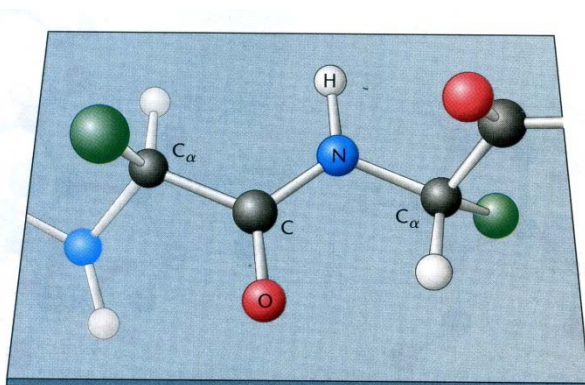


Figura 14. L'enllaç peptídic forma un pla en el qual hi ha 6 àtoms (C α , C, O, N, H, C α). Les cadenes laterals -R es mostren en color verd.

Font: Berg, Tymoczko, Stryer.
Bioquímica.

En canvi, i contràriament al que passa amb l'enllaç peptídic estricte, els enllaços entre el grup amino i l'àtom de carboni- α , i entre l'àtom del carboni- α i el grup carbonil són **enllaços senzills purs**. Les dues unitats peptídiques rígides adjacents poden rotar al voltant d'aquests enllaços, en diverses orientacions. **Aquesta llibertat de rotació al voltant de dos enllaços de cada aminoàcid permet que les proteïnes es pleguin de maneres molt diferents**, i permet la formació d'estructures secundàries ja esmentades, com els fulls- β o l'hèlix- α , com en el cas de la nostra proteïna ACTR.

La **llibertat de rotació** d'aquests dos enllaços es concreten en els **angles de torsió** que es poden veure a les figures següents. L'angle de rotació al voltant de l'enllaç entre el nitrogen i el carboni- α s'anomena Φ (ϕ), mentre que l'angle de rotació al voltant de l'enllaç entre l'àtom de carboni- α i el del carboni carbonil s'anomena Ψ (ψ).

Aquests marges de rotació són els que permeten que la cadena polipeptídica adopti o no una determinada estructura secundària, **per exemple una hèlix- α** .

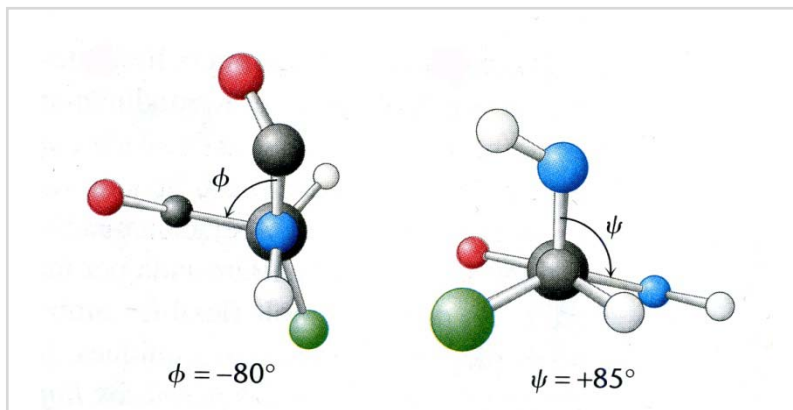
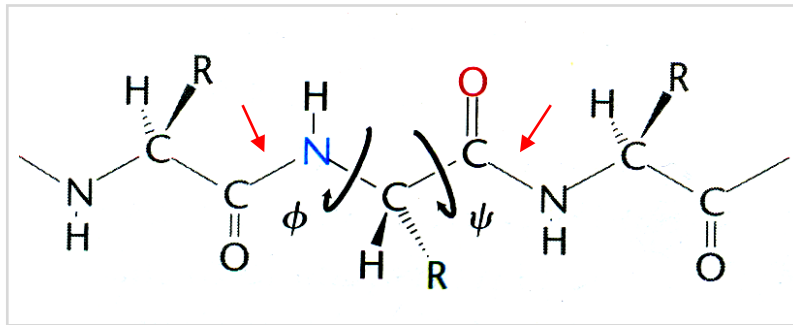


Figura 15. Rotació al voltant dels enllaços senzills del polipèptid (fletxes negres). A la figura superior, ϕ (ϕ) és l'angle de rotació al voltant de l'enllaç entre el N i el C- α , mentre que ψ (ψ) és l'angle de rotació al voltant de l'enllaç entre el C- α i el CO; les fletxes vermelles assenyalen l'enllaç peptídic rígid. A sota, la representació tridimensional dels mateixos angles de rotació ϕ (ϕ) i ψ (ψ).

Font: Berg, Tymoczko, Stryer. *Bioquímica*.

Cal tenir en compte que no totes les combinacions de ϕ (ϕ) i ψ (ψ) són possibles. El físic hindú Gopalasamudram Ramachandran es va adonar que moltes combinacions estan prohibides a causa dels impactes que es produirien entre àtoms durant els plegaments, anomenats col·lisions estèriques, és a dir, que dos àtoms no poden estar alhora al mateix lloc.

Els valors permesos es poden visualitzar en una representació gràfica anomenada **diagrama de Ramachandran**: tres quartes parts de les combinacions de possibles de ϕ (ϕ) i ψ (ψ) són impossibles a causa de col·lisions estèriques.

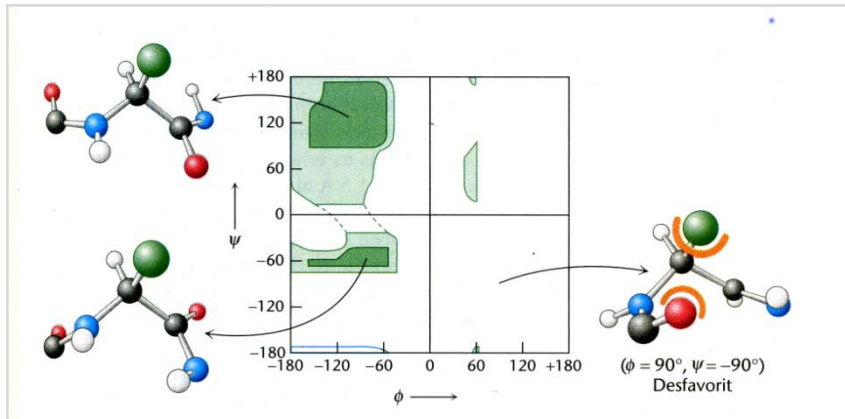


Figura 16. Gràfica de Ramachandran, que mostra que no són possibles tots el valors dels angles de Fi (ϕ) i Psi (ψ) a causa de les col·lisions estèriques entre els àtom dels aminoàcids. Les regions més favorables es mostren en verd fosc i les intermedies en verd clar. La resta estan desfavorides a causa de les restriccions estèriques.

Font: Berg, Tymoczko, Stryer. *Bioquímica*.

En principi, d'acord amb les lleis de la termodinàmica, un polímer qualsevol es pot plegar a l'atzar en una estructura qualsevol, cosa que genera una varietat de conformacions molt alta. En canvi, la rigidesa relativa de l'enllaç peptídic permet una gamma restringida d'angles Fi (ϕ) i Psi (ψ), i això limita el plegament de les proteïnes i el nombre d'estructures possibles.

Són precisament aquestes possibilitats i aquestes restriccions dels angles de rotació, el factor que utilitza el nostre anàlisi computacional per determinar la probabilitat de formació d'una estructura d'hèlix- α per part de la nostra proteïna d'estudi, ACTR.

6. El mètode Monte Carlo i el programari utilitzat

6.1. El mètode de Monte Carlo

Per fer simulacions computacionals de la seqüència d'aminoàcids original i de totes les mutacions objecte d'estudi, s'ha fet servir un mètode de **Monte Carlo**. Es tracta d'un mètode numèric que explora les diverses possibilitats de plegament de proteïnes mitjançant opcions a l'atzar utilitzant la tècnica estadística matemàtica de l'*Importance sampling*. A grans trets, la idea és que alguns valors de la variable aleatòria apareixen amb més freqüència, cosa que ens permet "descartar" les combinacions més inusuals.

Els mètodes de Monte Carlo es van començar a desenvolupar l'any 1944 i s'anomenen així, com el seu nom indica, pel famós casino de Montecarlo, a Mònaco. Per entendre la noció del mecanisme dels mètodes de Monte Carlo es recorre sovint a l'exemple del **joc dels vaixells**.



Figura 17. Joc d'enfonsar vaixells.

Font: Todocoleccion.net

A partir de les primeres jugades, el jugador va anotant les caselles on ha fet “aigua”, per exemple “H-4”, “H-5”, “H-6”, “H-7”, etc. ; aquestes anotacions li serveixen per no haver de “*tornar a recórrer el camí recorregut*”, i en una jugada posterior seguir endavant sabent que el tram H4/H7 és “aigua”, sense haver de jugar caselles a l’atzar. De la mateixa manera, si “H-8” resulta ser un vaixell, aquesta informació també serveix al jugador per “començar a explorar” si el vaixell és d’1, 2, 3 o 4 caselles, i si està en posició horitzontal (H-8/H11) o vertical (I-7/G-7/F-7, etc.).

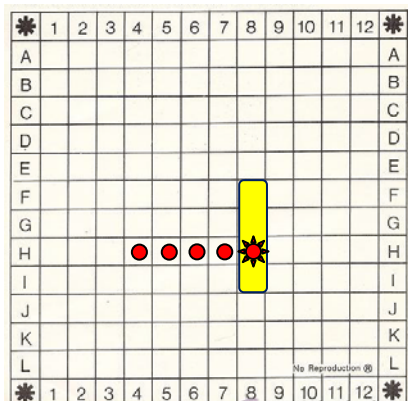


Figura 18. Un camí d’exploració a l’atzar del mètode Monte Carlo es pot comparar a l’atzar al joc d’enfonsar vaixells

Font: Elaboració pròpia

D’una manera similar, i utilitzant els mecanismes de plegament de les cadenes polipeptídiques que hem exposat en el capítol anterior, el mètode de Monte Carlo, amb el software de simulació de proteïnes que ara presentarem (Profasi i Phaistos), és capaç de calcular la probabilitat que un determinat aminoàcid de la cadena formi part o no una hèlix- α (criteri 1 o tradicional: “**aminoàcid a aminoàcid**”); i fins i tot pot determinar també quina longitud total pot adoptar aquesta hèlix- α , per confirmar o descartar la nostra **Hipòtesi de la Longitud de l’estructura d’hèlix- α** .

Òbviament, en el nostre cas les opcions no vénen determinades com en el joc d’enfonsar vaixells pels valors “aigua” i “tocat” sinó pel **valor de rotació dels angles Φ i Ψ dels amidaïds** de la cadena de la nostra proteïna ACTR, valors que estan ben establerts en els diagrames de Ramachandran presentats també en el capítol anterior.

6.2. Els cicles de càlcul de Monte Carlo

Les simulacions de Monte Carlo que utilitzarem per estudiar el comportament de la nostra proteïna ACTR treballen en els anomenats “**cicles de Monte Carlo**”.

Es defineix un cicle de Monte Carlo com cada una de les simulacions completes del comportament de plegament de la seqüència d'aminoàcids. En el nostre cas, treballem amb un fragment de la proteïna **ACTR de 19 aminoàcids**. Així doncs, un cicle de Monte Carlo simulant la conformació d'aquest fragment equival a una possible conformació tridimensional dels dinou aminoàcids. Monte Carlo és un mètode que es basa en decisions atzaroses segons les propietats fisicoquímiques dels aminoàcids de la proteïna que analitza, però aquestes decisions, a partir del segon cicle de Monte Carlo, estaran també influenciades per la conformació resultant del **cicle anterior** (igual com el jugador del joc d'enfonsar vaixells “recorda” les caselles que ja ha “utilitzat”).

Per naturalesa, una proteïna busca adoptar la conformació més estable, i aquesta estabilitat ve determinada (entre d'altres variables) per la energia interna de la proteïna.

Si en la simulació, en el cicle anterior s'ha aconseguit adoptar una conformació poc energètica, les simulacions resultants dels següents cicles buscaran aquesta estabilitat efectuant només canvis subtils en la conformació. No obstant, encara que s'hagi arribat a una **estabilitat aparent**, és molt probable que aquesta conformació no sigui la menys energètica possible de la proteïna estudiada. Per a aconseguir una estabilitat més forta, primer s'ha de augmentar momentàniament l'energia interna de la proteïna, desestabilitzar-la, per després assolir un **estat encara menys energètic**, més estable que la suposada estabilitat anterior.

El fet de que es requereixi aquesta desestabilització temporal per a arribar a una conformació més estable que la anterior és el que ocasiona que les simulacions «s'estanquin» en una primera estabilitat i no presentin un decreixement d'energia lineal.

Podríem comparar aquest procés amb un animal que neix al cim d'una muntanya. Si un dia baixa fins a la base d'aquest turó, podria pensar que aquell terreny és el menys elevat de tot. Podria pensar que ja no es pot arribar més avall. Però, si un dia creua la petita carena que s'aixeca davant de la base la primera muntanya inicial i que condueix a un poble costaner, comprovarà que existia un pla encara menys elevat que el que ell havia considerat com la

base absoluta. Per posar un exemple més concret, seria el cas d'un animal que viu al Montseny i baixa al Vallès. Quan pensa que el Vallès és la plana més baixa, un dia puja a Collserola i darrera troba el mar, encara més baix que el Vallès.

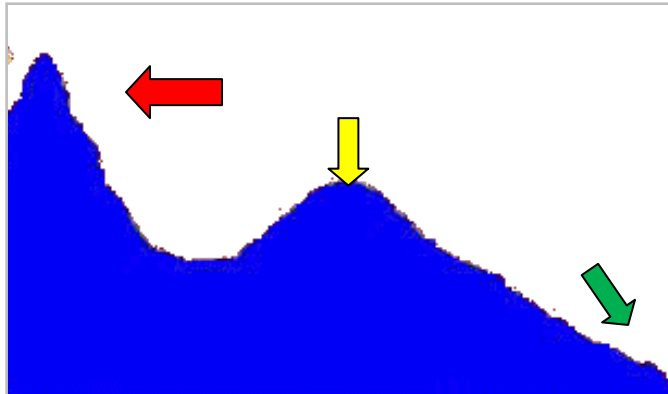


Figura 19. Nivells energètics del plegament d'una proteïna comparats amb un perfil topogràfic. Per aconseguir el nivell energètic més baix (verd) des del cim (vermell), cal primer superar el nivell mitjà (groc).

Font: Elaboració pròpia

Aplicat a les simulacions de proteïnes, aquestes muntanyes són els estats més energètics, i els plans són les estabilitats temporals. El que fa que un cicle de Monte Carlo es decideixi a «creuar la carena», a desestabilitzar-se, és simplement el nombre de cicles que porta «aturat» en una mateixa conformació.

Aquest fenomen físic ens dicta que, com més cicles de Monte Carlo es calculin, més realista serà la simulació. Ja que en la natura la proteïna adoptarà sempre la conformació menys energètica de totes. I per a assolir-la s'haurien de calcular infinits cicles.

Com que no es disposa del temps necessari per a computar aquest nombre de cicles de Monte Carlo, es decreta un **número de cicles prou gran** per assolir una conformació molt estable, **però prou petit** perquè els càlculs es puguin fer en una quantitat de temps raonable.

6.3. Les Protein Data base (PDB)

Un cop realitzades les simulacions, les dades s'extreuen mitjançant un arxiu de **base de dades de proteïnes o PDB**.

Els arxius PDB o **Protein Data Bank** són bases de dades que contenen en forma de text tota la informació necessària per que un software especialitzat sigui capaç de llegir-los en forma de simulacions tridimensionals d'una proteïna. Aquesta informació inclou la seqüència d'aminoàcids de la

macromolècula, les característiques fisicoquímiques intrínseques de cada aminoàcid (tals com la càrrega elèctrica) i les característiques fisicoquímiques dels aminoàcids en l'espai (la hidrofobicitat, els angles de gir, etc.).

Els PDB són l'arxiu base utilitzat internacionalment per la comunitat científica en l'àmbit de la biologia computacional. Existeix un suport web d'accés lliure anomenat **RCSB Protein Data Bank** on estan publicats tots els arxius PDB descrits fins a l'actualitat. Aquest banc de dades facilita en gran mesura el progrés de la recerca en el camp de l'estudi de les proteïnes.

6.4. El software d'anàlisi de proteïnes: Profasi i Phaistos

Per dur a terme les simulacions de Monte Carlo hem utilitzat dos softwares professionals d'anàlisi de proteïnes, utilitzats habitualment pel CTCG (Computational and Theoretical Chemistry Group), el grup de recerca dirigit per Dr. Ramon Crehuet, de l'Institut de Química Avançada de Catalunya (iQAC): Profasi i Phaistos.

6.4.1. Profasi

Profasi (PROtein Folding and Aggregation Simulator - Simulador de plegament i agregació de proteïnes) és un programa de programari lliure en llenguatge C++ específic per a les simulacions de Monte Carlo de proteïnes [9]. Proporciona un conjunt d'eines per a la realització de simulacions i anàlisi de les dades generades (Bases de dades de proteïnes).

El model utilitza una descripció de tots els àtom de proteïnes i dels graus de llibertat dels angles de torsió dels aminoàcids, creant models de proteïnes amb les longituds i els angles dels enllaços.

Profasi és programari lliure i està disponible lliurement sota GNU General Public Licence. La versió utilitzada és la vigent v1.5 Profasi.

Malgrat la seva solvència, Profasi és un programa més dissenyat per treballar amb proteïnes ja plegades que amb processos de simulació. **És per això que finalment no ens va resultar del tot adequat.**

6.4.2. Phaistos

Phaistos es també un programari que utilitza el mètode Monte Carlo de mostreig per a la simulació, la predicció i la deducció de l'estructura de les proteïnes. Utilitza els últims avenços en la metodologia de Monte Carlo, i les darreres actualitzacions dels models probabilístics de l'estructura de les proteïnes. Una configuració flexible d'interfície permet als usuaris establir simulacions amb la configuració desitjada de forma ràpida.

Phaistos és també un software de programari lliure disponible i es distribueix sota la versió 3.0 Llicència pública general de GNU a Internet. Com Profasi, Phaistos utilitza el llenguatge C++ i s'ha provat en les plataformes Linux i OSX.

A diferència de Profasi, Phaistos sí que és un programa més dissenyat per treballar amb simulacions de plegament de proteïnes. **I per això al final va ser més adequat que Profasi.**

6.4.3. VMD

VMD (Visual Molecular Dynamics) ens permet visualitzar d'una manera ràpida i en pantalla els resultats de les anàlisis de configuració fets amb els altres programes, Profasi i Phaistos. Això ens ha permès seleccionar els cicles de Monte Carlo més útils i descartar aquells que no aportaven informació d'interès (sobretot els inicials, quan les simulacions de plegament d'ACTR no havien adquirit encara consistència).

VMD és un programa dissenyat per a la modelització, la visualització i l'anàlisi dels sistemes biològics, com ara proteïnes o àcids nucleics. El programa pot llegir els arxius de les dades de bases de proteïnes i presentar en pantalla la seva estructura. També ofereix una àmplia varietat de mètodes per a la representació i la coloració d'una molècula: punts i línies simples, esferes i cilindres, tubs i cintes mitjançant animacions. VMD es pot utilitzar per animar i analitzar la trajectòria d'una simulació de plegament de proteïnes, que **és l'ús que li hem donat nosaltres.**

7. Els mètodes computacionals utilitzats

7.1. Simulacions amb Profasi

Amb el programa Profasi, s'han calculat les simulacions de cicles de 100, 100.000 i 100.000.000 cicles de Monte Carlo. Després dels càlculs numèrics hem extret els arxius de bases de dades de proteïnes PDB (cada 1, 3 i 800 cicles respectivament) que contenen tota la informació necessària per a representar les proteïnes tridimensionals.

Finalment es va treballar amb les simulacions 100.000.000 de cicles de Monte Carlo, i amb els arxius PDB que n'hem extret.

Vam fer aquesta reducció perquè, tot i tenir tots els càlculs que fan les representacions gràfiques de cada cicle de Monte Carlo, extreure tots els fitxers PDB requeriria molt de temps. A més, no ens donaria una diferència estadística per justificar aquesta despesa de temps i de treball computacional.

Amb aquestes dades hem representat alguns gràfics de contingut d'hèlix- α aminoàcid a aminoàcid amb el programa de disseny de gràfics Gnuplot.

Amb aquest sistema hem generat "línies de temps" (*time-lines*), és a dir, l'evolució de contingut d' hèlix- α per a cada simulació, que hem pogut visualitzar amb el programa de visualització de proteïnes VMD, per tal d'avaluar l'evolució del contingut d' hèlix- α per a cada simulació.

Hem comprovat que després dels primers 1000 cicles de Monte Carlo, totes les simulacions tenen gairebé el mateix contingut d'hèlix- α estable. Per tant, hem utilitzat les etapes de 1000 a 10.000.000 per fer la comparació entre una mutació i una altre.

Els passos anteriors al nombre de 1.000 han estat descartats per raons estadístiques, ja que resulten més irregulars i absurds que la proteïna real. Hem acabat el rang en 10.000.000 causa de que les següents simulacions tenien gairebé el mateix contingut hèlix- α estable.

És per això que hem triat aquest rang per fer els càlculs següents i per tal de poder crear els gràfics d'una manera més ràpida i fàcil.

Aquesta reducció específica s'ha aplicat amb cada una de les 8 mutacions i amb el tipus salvatge (WT), és a dir, la seqüència d'aminoàcids original sense cap mutació. Amb aquestes dades, hem representat un gràfic general de contingut d'hèlix- α "aminoàcid a aminoàcid" per poder veure totes les variacions de contingut d'hèlix- α "aminoàcid a aminoàcid" entre cada una de les mutacions.

7.2. Simulacions amb Phaistos

Per generar estructures de proteïnes amb el programa Phaistos es va utilitzar un mètode de dinàmica de Xarxa Bayesiana, un tipus de model probabilístic [10].

Phaistos és un programa dissenyat per a simular tot tipus de proteïnes [11]. També pot analitzar molècules utilitzant arxius PDB, o pot fer servir un arxiu PDB inicial fer els càlculs matemàtics, sense necessitat de transcriure tota aquesta informació numèrica en arxius PDB. Phaistos té la capacitat per diferenciar una IDP d'una proteïna normal. En l'estudi de les IDP, això li permet obtenir **resultats més fiables i realistes** que amb programes dissenyats per simular el plegament de proteïnes.

Amb el programa Phaistos, i amb l'ajuda de l'eina iPython Notebook, hem creat alguns gràfics que comparen diferents factors del nostre interès. Per exemple, vam poder calcular hèlix- α per aminoàcid individual (aminoàcid a aminoàcid) i també la llargada d'hèlix- α (mitjançant gràfics de longitud total).

Amb aquesta informació volem saber si, a més de l'efecte que pugui tenir sobre la formació d'hèlix- α dels aminoàcids individuals, la longitud de l'estructura d'hèlix- α podria tenir un paper. Això s'ha comprovat amb la comparació dels pics de les nostres gràfiques de llargada d'hèlix- α amb els resultats experimentals de laboratori referits al temps transcorregut en el procés d'unió d'ACTR el domini NCBD de la proteïna CREB [2].

7.3. iPython Notebook

iPython notebook és un entorn computacional interactiu que utilitza el llenguatge de programació Python [12]. S'utilitza principalment per programar arxius de codi, però es pot usar també per combinar codi de programació amb

notes de text o explicacions, i fins i tot els gràfics i símbols matemàtics. És útil per reunir càlculs i anàlisi de dades en el mateix arxiu.

Hem utilitzat iPython Notebook que fer plots i per analitzar les dades.

7.4. Visualització dels resultats

Hèlix- α per aminoàcid individual

La tendència a la formació d'hèlix- α aminoàcid a aminoàcid mostra la freqüència d'estructura d'hèlix- α (normalitzada de 0 a 1) per a cada aminoàcid, per separat i de forma independent dels seus veïns més propers o qualsevol altre.

Hèlix- α per longitud

La tendència a la formació d'hèlix- α llarga mostra la freqüència de estructura d'hèlix- α de cada aminoàcid i quina llargada correspon a aquesta hèlix- α en concret.

Hem fet dos tipus de gràfics de longitud d'hèlix- α : els primers que mostren la llargada mitjana de l'hèlix- α que conté cada aminoàcid (en una freqüència normalitzada de 0 a 2,5). Els altres mostren la probabilitat de la formació i estabilització d'aquesta estructura secundària, en un gradient de color de blanc a blau, en què el blanc és un aminoàcid no estructurat i el blau fosc és un aminoàcid ben estructurat en hèlix- α [13].

Vam fer també dues gràfiques que comparen el contingut d' hèlix- α amb el temps empleat en el procés d'acoblament. Una d'elles mostra la relació entre el contingut d'hèlix- α de cada mutació i els resultats experimentals de laboratori del temps utilitzat en el procés d'acoblament.

L'altre mostra una regressió lineal entre el contingut d'hèlix- α de la cadena més llarga d'hèlix- α (19 aminoàcids) i el temps usat en l'acoblament de les proteïnes.

IV. Resultats i conclusions

8. Resultats i discussió

8.1. Resultats amb Profasi

En els gràfics de **longitud d'hèlix- α** (Figures 20, 21 i 22) que hem obtingut amb els càlculs amb Profasi, podem veure que entre ells no hi ha una notable diferència. Tots dos mostren una tendència a hèlix- α similar i estable. Aquests resultats no són realistes tenint en compte que estem treballant amb una IDP, que té una estructura secundària no estable. Això es pot comprovar mitjançant la comparació de la gràfica mutació A1047G amb les altres. La mutació A1047G ha estat dissenyada per interrompre les hèlix- α , i les altres mutacions per augmentar-la; i entre els seus gràfics gairebé no hi ha diferència significativa.

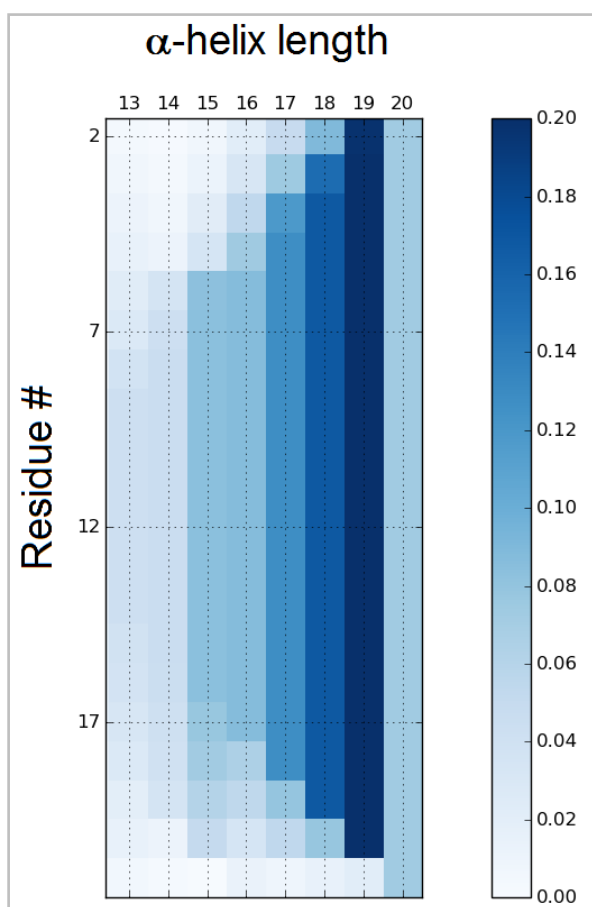


Figura 20. Gràfica de la longitud d'hèlix- α de la mutació A1047G usant els càlculs de Profasi.

Font: Elaboració pròpia

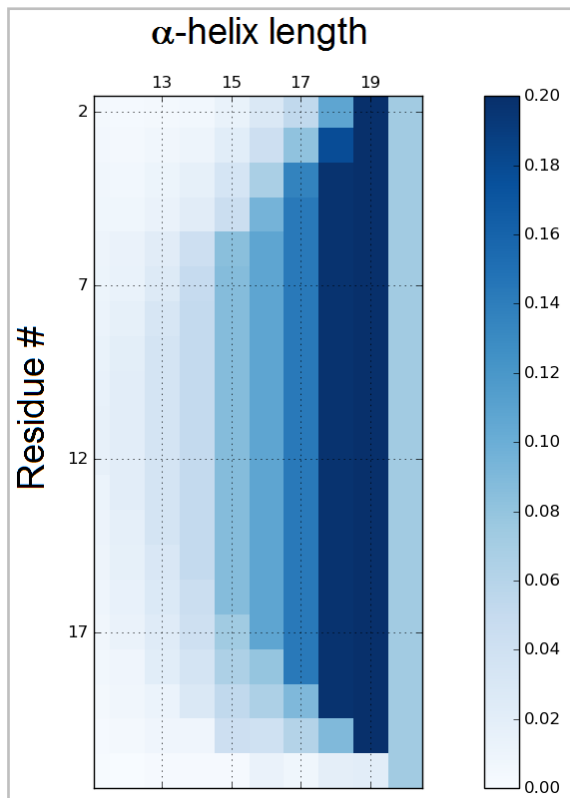


Figura 21. Gràfica de llargada d'hèlix- α de la mutació D1050E usant els càlculs de Profasi.

Font: Elaboració pròpia

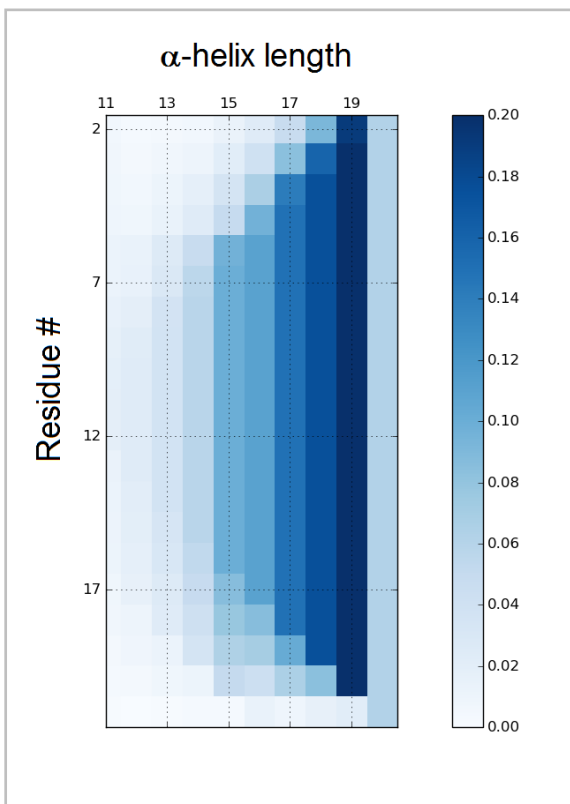


Figura 22. Gràfica de llargada d'hèlix- α de la mutació S1043M usant els càlculs de Profasi.

Font: Elaboració pròpia

A la Figura 23 podem veure gairebé el mateix resultat per a totes les mutacions i el tipus salvatge, a excepció de la mutació A1047G, que té un menor contingut d'hèlix- α al principi, però que també s'estabilitza després de l'11è aminoàcid. Com en el cas anterior, aquesta gràfica mostra resultats poc realistes, ja que ACTR presenta una estructura hèlix- α estabilitzada al llarg de tota la seqüència d'aminoàcids. Aquests resultats contradiuen els experiments de laboratori que mostren que ACTR no fa gaire estructura hèlix- α estabilitzada, i que entre les mutacions hi ha diferències significatives (especialment entre la mutació A1047G i les altres).

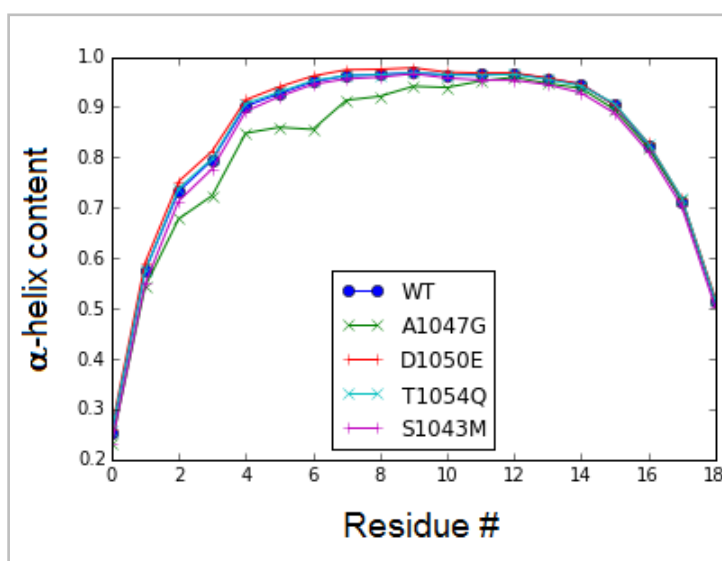


Figura 23. Gràfica d'hèlix- α "aminoàcid a aminoàcid" comparant algunes mutacions individuals i el tipus salvatge utilitzant els càlculs de Profasi.

Font: Elaboració pròpia

Això ens mostra que un programa dissenyat per simular proteïnes clàssiques amb estructura secundària estable, com Profasi, no és una bona opció per treballar amb les IDP, ja que no pot diferenciar entre una IDP i una proteïna normal ben estructurada. Per tant, els seus càlculs tracten l'IDP com una proteïna estructurada amb un alt contingut d'estructura secundària llarga i estable.

8.2. Resultats amb Phaistos

A diferència dels gràfics de longitud d'hèlix- α creats amb els càlculs de Profasi, els que van ser generats utilitzant els càlculs de Phaistos mostren uns resultats més plausibles.

Si comparem els gràfics de longitud d'hèlix- α de Phaistos de les mutacions S1043M i D1050E (Figures 25 i 26) amb el gràfic de longitud d'hèlix- α de la mutació A1047G (Figura 24) hi podem apreciar una diferència notable.

La figura 24 té un contingut molt menor de d'hèlix- α , la qual cosa coincideix amb el disseny original d'aquesta mutació, A1047G, per interrompre l'estructura hèlix- α , i de les mutacions S1043M i D1050E per augmentar-la.

En el mateix sentit, la gràfica de longitud d'hèlix- α per al Mutant Triple (Figura 27) mostra més estructura d'hèlix- α , més llarga i estabilitzada. Aquest és un resultat més realista i que concorda millor amb els resultats experimentals obtinguts al laboratori.

Aquesta gràfic (Figura 27) mostra també un nivell més creïble d'estructura hèlix- α en una IDP, encara que té una mutació per afavorir-la.

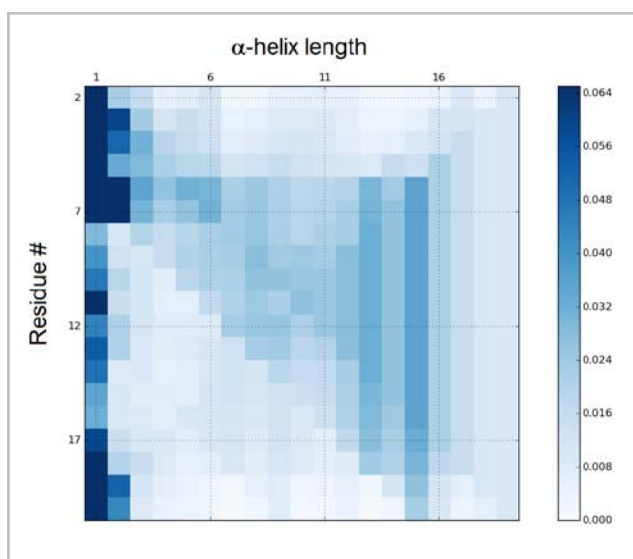


Figura 24: Gràfica de llargada d'hèlix- α de la mutació A1047G utilitzant els càlculs de Phaistos.

Font: Elaboració pròpia

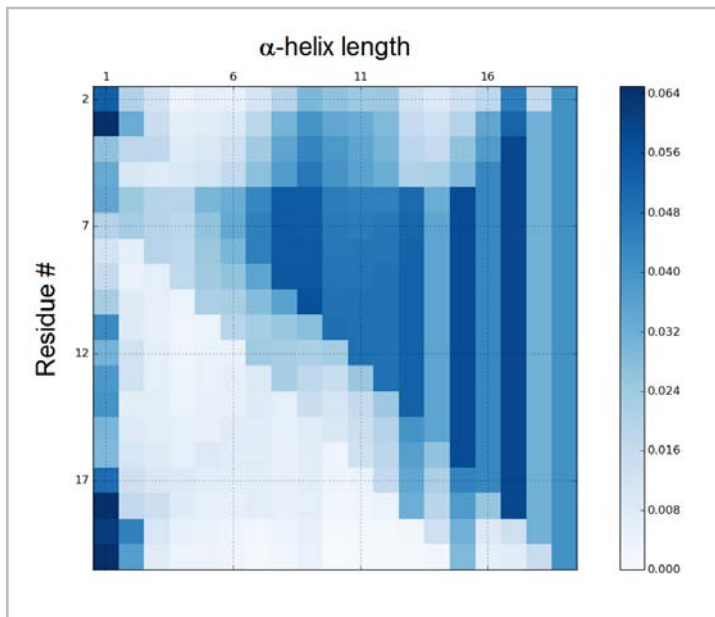


Figura 25: Gràfica de llargada d'hèlix- α de la mutació S1043M utilitzant els càlculs de Phaistos.

Font: Elaboració pròpia

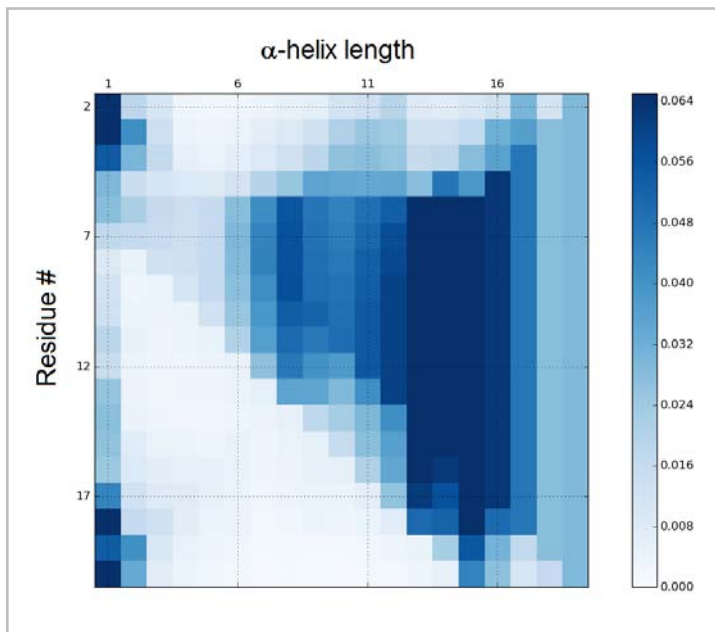


Figura 26: Gràfica de llargada d'hèlix- α de la mutació D1050E utilitzant els càlculs de Phaistos.

Font: Elaboració pròpia

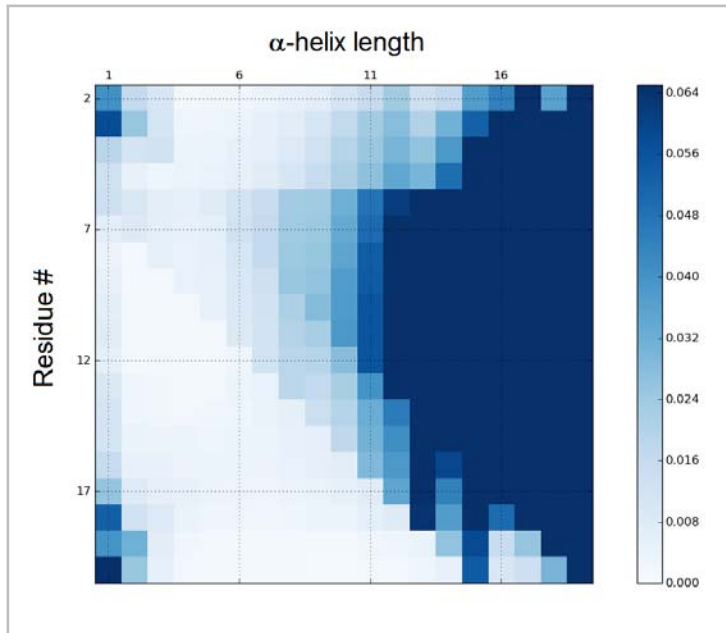


Figura 27: Gràfica de llargada d'hèlix- α del Mutant Triple (S1043M, D1050E i T1054Q) utilitzant els càlculs de Phaistos.
Font: Elaboració pròpia

La figura 28 mostra que A1047G té, com s'esperava, un baix contingut de hèlix- α "aminoàcid a aminoàcid". Les altres mutacions individuals tenen un major contingut d'estructura secundària que el tipus salvatge (WT, Wild Tipus). Els mutants dobles tenen un gran contingut en hèlix- α , resultat ja previst degut a la combinació de més d'una mutació per afavorir l'helicitat de la proteïna. Es mostren ordenats, de més estructurat a menys, DT_doble, ST_doble i SD_doble. El Mutant Triple té un contingut molt més alt de hèlix- α que qualsevol de les altres mutacions o el WT. Aquest també és un resultat esperat a la vista dels experiments de laboratori anteriors, que mostren que el Mutant Triple té un major contingut de hèlix- α , així com també una major velocitat d'acoblament.

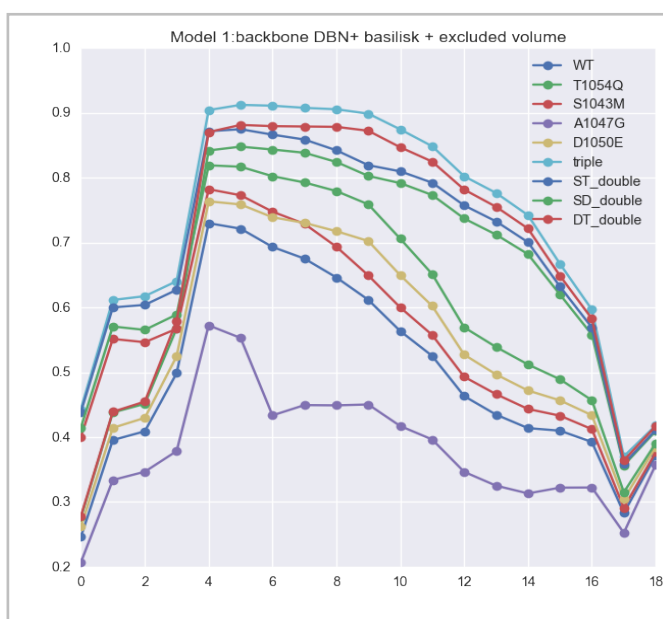


Figura 28: Gràfic d'hèlix- α "aminoàcid a aminoàcid" comparant totes les mutacions i el WT utilitzant els càlculs de Phaistos.
Font: Elaboració pròpia

La gràfica de llargada d'hèlix- α (linies phaist llargada) mostra que A1047G té un previst baix contingut d'hèlix- α i les altres mutacions individuals tenen un major contingut d'estructura secundària que el WT.

Els mutants més amb la major llargada d' hèlix- α són en aquest cas el Mutant Triple i el mutant DT_doble, que com el seu nom indica incorpora les dues mutacions individuals D1050E i T1054Q.

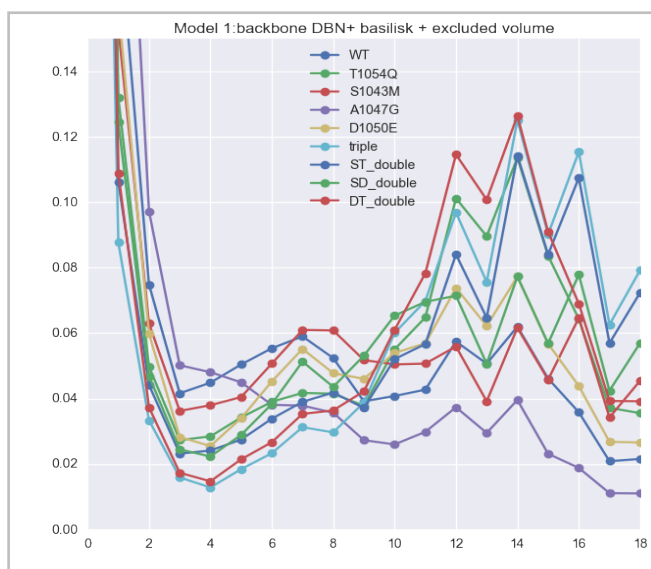


Figura 29: Gràfic de llargada d'hèlix- α comparant totes les mutacions i el WT utilitzant els càlculs de Phaistos.

Font: Elaboració pròpia

Si fem una comparació amb els resultats de laboratori del temps en què cada mutació gasta en el procés d'unió, podem veure una correlació significativa entre aquest i el contingut d'hèlix- α "aminoàcid a aminoàcid" (Figura 30).

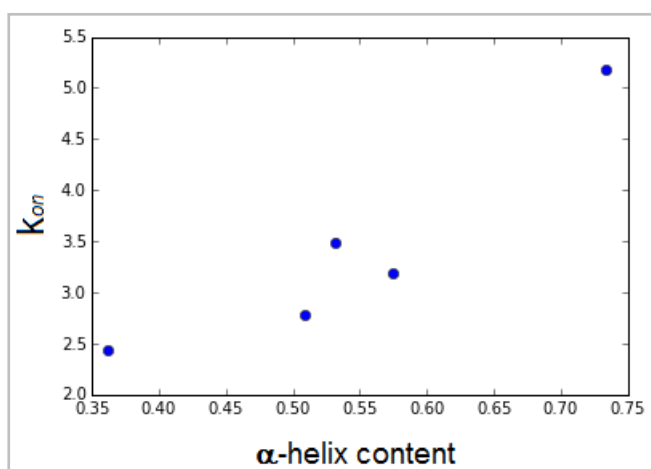


Figura 30: Relació entre el contingut d'hèlix- α "aminoàcid a aminoàcid" i el temps utilitzat en el procés d'acoblament de les dues proteïnes.

Font: Elaboració pròpia

La figura 31 mostra una regressió lineal entre el contingut d'hèlix- α "aminoàcid a aminoàcid" de la hèlix- α més llarga de totes (la que és resultat de la simulació del Mutant Triple) i el temps empleat en el pocés d'acoblament d'ACTR amb el domini NCBD de la proteïna CREB. Aquest gràfic mostra una relació significativa que ens permet donar suport a la hipòtesi de selecció conformacional explicada anteriorment en el punt 4.4 de la introducció.

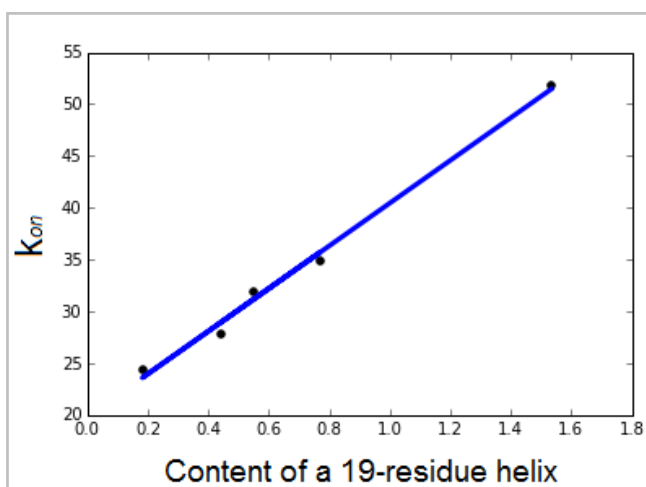


Figura 31: Regressió lineal entre el contingut d'hèlix- α "aminoàcid a aminoàcid" de la hèlix- α més llarga (la del Mutant Triple) i la reducció del temps utilitzat en el procés d'unió de les dues proteïnes.

Font: Elaboració pròpia

9. Conclusions

Aquest treball té la seva base en estudis bioquímics experimentals realitzats en laboratori. No obstant, el que hem presentat no és una mera variant computacional dels experiments previs, sinó que **hi hem afegit nous paràmetres i hipòtesis pròpies innovadores, de les quals no existeix en l'actualitat cap publicació descrita.**

En concret, la nostra més gran aportació ha sigut la **Hipòtesi de la longitud d'hèlix- α .**

Després d'analitzar les gràfiques i les diferents dades obtingudes, podem concloure que **la Hipòtesi de la longitud d'hèlix- α , presentada l'inici del treball, sembla ser certa.** Basem aquesta afirmació en la relació entre els resultats de la nostres gràfiques d'hèlix- α "aminoàcid a aminoàcid" amb les dades experimentals de laboratori obtingudes en l'estudi previ sobre el temps d'acoblament entre ACTR i el domini NCBD de la proteïna CREB. Podem observar que a mesura que augmenta la llargada d'hèlix- α augmenta també la reducció del temps empleat en el procés d'acoblament.

Podem veure aquesta relació en la figura 30.

Aquesta relació mostra que la mutació que més bons resultats ha donat a la investigació de laboratori quant a minimitzar el temps empleat en l'acoblament, **el Mutant Triple, és també la mutació que adopta la conformació amb l' hèlix- α més llarga** d'entre totes les diferents mutacions i l'estat «salvatge».

Podem comprovar aquesta teoria en la figura 31.

Aquests resultats semblen mostrar que, **d'acord amb la nostra hipòtesi inicial, la llargada de la hèlix- α és un factor molt important a l'hora de reduir el temps utilitzat en el procés d'acoblament entre les dues proteïnes.**

Ja que tant la llargada de la hèlix- α com el temps d'acoblament entre proteïnes són factors regits únicament per condicions fisicoquímiques comunes en totes les proteïnes en general, es pot estendre l'afirmació de que **la llargada de la hèlix- α és un factor determinant a l'hora de reduir el temps utilitzat en el procés d'acoblament a tot el conjunt de proteïnes conegudes;**

especialment en el cas de les proteïnes intrínsecament desordenades que poden variar en gran mesura el seu plegament.

Hem pogut comprovar també altres resultats que ja esperàvem a la vista dels resultats de la recerca experimental prèvia feta en laboratori. Hem pogut corroborar, analitzant les gràfiques d' hèlix- α per aminoàcid obtingudes, que **les mutacions individuals S, T, la mutació ST_ doble i especialment el Mutant Triple són les variacions que més bons resultats donen** a l'hora de calcular l'increment d'hèlix- α per aminoàcid d'ACTR, en comparació amb l'estat «salvatge».

També confirmem que **la mutació A1047G**, que va ser dissenyada per desafavorir al màxim el contingut d' hèlix- α obté un resultat de **contingut d' hèlix- α per aminoàcid menor que qualsevol altra mutació** i fins i tot que el de l'estat «salvatge».

Adicionalment, com a conclusions de caire més pràctic i enfocades a la recerca científica en general, vam voler analitzar si la biologia computacional, més segura i barata que l'experimentació biològica de laboratori, podria considerar-se una possible alternativa amb el mateix nivell de fiabilitat.

Hem comprovat que es poden predir resultats d'experimentació de laboratori utilitzant mètodes computacionals. No obstant això, només hem sigut capaços d'arribar a uns resultats comuns fent servir el programa Phaistos per calcular les simulacions. Les simulacions obtingudes a partir dels càlculs fets amb el programa Profasi, degut a la seva excessiva predisposició a forçar el plegament de les proteïnes, no eren realistes i van haver de ser descartades. **Això indica que els mètodes computacionals estan encara lluny de ser prou fiables i precisos com per ser capaços de presentar una alternativa excel·lent** als experiments de laboratori.

Però al mateix temps, **els bons resultats obtinguts amb el programa Phaistos mostren que en un futur on es disposi de millor software potser podria arribar a ser un mètode tan vàlid** com un d'experimental.

10. Recerca futura

Com sempre passa en la recerca científica, aquest projecte no està acabat i té moltes possibles vies per les quals es podria seguir investigant.

El següent pas d'aquest treball podria ser utilitzar altres opcions de software per corroborar encara més les dades obtingudes. Actualment seguim treballant en una ampliació d'aquesta recerca i estem utilitzant **nous programes de simulació de proteïnes com el Campari**.

Una altra implementació interessant seria buscar **noves mutacions** que afavorissin la formació de hèlix- α llargues com a factor principal, a més a més d'incrementar la quantitat individual d'hèlix- α per aminoàcid.

Adicionalment, es podria **simular el procés d'acoblament** complet utilitzant mètodes computacionals i comparar les dades obtingudes amb les descrites pel investigadors experimentals.

En l'apartat de conclusions, hem detallat que tot sembla indicar que la nostra hipòtesi, la **Hipòtesi de la longitud d'hèlix- α , sembla ser certa**. No obstant, es podria provar de confirmar aquest resultat amb més fiabilitat mirant si es mirés d'incrementar la relació entre la llargada de l'hèlix- α i el decreixement del temps utilitzant en el procés d'acoblament al introduir en la equació informació sobre la llargada d' hèlix- α que encara no s'ha estudiat experimentalment en un laboratori.

En relació a les conclusions orientades al treball científic pràctic, es podrien **estudiar, i desenvolupar, nous mètodes computacionals** més acurats fins a trobar o crear un software que permeti predir els resultats dels experiments de laboratori amb prou precisió com per permetre reduir-los amb nous mètodes de càlcul per computació.

Agraïments

Voldria mostrar el meu agraïment especial al meu tutor extern, el Doctor Ramon Crehuet, per tot el temps, el suport i la instrucció que m'ha proporcionat i encara em proporciona en aquesta recerca. Aquest projecte no s'hauria pogut dur a terme sense ell.

Voldria agrair a l'iQAC i al CSIC la cessió de les seves instal·lacions i temps de computació dels seus clústers per poder desenvolupar els meus càlculs.

Vull agrair al meu tutor del Treball de Recerca, el meu professor de física José David Cosials, el seu temps i les seves recomanacions sobre el treball, així com els seus consells per a presentar-lo adequadament.

Vull agrair al meu tutor de l'article desenvolupat per al programa Joves i Ciència, Tiago Cordeiro, la gestió de la meua col·laboració amb l'iQAC, i la seva revisió constant del meu progrés.

Vull agrair al meu pare la seva ajuda i paciència incansable en el procés de redacció del Treball de Recerca. Si no fos per ell encara seguiria rumiant davant d'un paper blanc.

Voldria agrair a les coordinadores del programa Joves i Ciència, l'Eva Calvés i la Carla Conejo, la seva atenció durant aquesta meravellosa iniciativa.

I finalment, vull mostrar el meu agraïment al programa Joves i Ciència i a la Fundació Catalunya-La Pedrera per donar-me aquesta oportunitat única de treballar, aprendre i entendre la ciència molt més del que mai hauria pogut imaginar.

A tots ells, els vull donar les gràcies.

Referències i bibliografia consultada

Referències

- [1] I. C. Felli et al., *Intrinsically Disordered Proteins, in NMR of Biomolecules: Towards Mechanistic Systems Biology*, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany (2012).
- [2] V. Iesmantavicius, M. Kjaergaard et al. *Helical Propensity in an Intrinsically Disordered Protein Accelerates Ligand Binding*, *Angew. Chem. Int. Ed.*, 53, 1548-1551, (2014).
- [3] H. Chen et al., *Nuclear Receptor Coactivator ACTR Is a Novel Histone Acetyltransferase and Forms a Multimeric Activation Complex with P/CAF and CBP/p300*, *Cell*, 90, 569–580, (1997).
- [4] Marc R. Montminy & Louise M. Bilezikjian, *Binding of a nuclear protein to the cyclic-AMP response element of the somatostatin gene*. *Nature* 328, 175-178 (9 July 1987)
- [5] F. Petrij et al., *Rubinstein-Taybi syndrome caused by mutations in the transcriptional co-activator CBP*, *Nature*, 376, 348–51, (1995).
- [6] D.B. Shankar, K.M. Sakamoto. *The role of cyclic-AMP binding protein (CREB) in leukemia cell proliferation and acute leukemias*. *Leuk Lymphoma*. 2004 Feb; 45(2):265-70.
- [7] JL. Vizmanos et al., *t(10;16)(q22;p13) and MORF-CREBBP fusion is a recurrent event in acute myeloid leukemia*, *Genes Chromosomes Cancer*, 36, 402–5, (2003).
- [8] L.L. Looger, M.A. Dwyer, J.J. Smith, H.W. Hellinga, *Computational design of receptor and sensor proteins with novel functions*, *Nature* 423, 185-190, (May, 2003)
- [9] A. Irbäck, S. Mohanty, *PROFASI: A Monte Carlo simulation package for protein folding and aggregation*, *J. Comput. Chem.*, 27, 1548–1555, (2006).
- [10] W. Boomsma, K. V. Mardia, C. C. Taylor, J. Ferkinghoff-Borg, A. Krogh, T. Hamelryck, *A Generative Probabilistic Model of Local Protein Structure*, *Proc. Natl. Acad. Sci.*, 105, 8932–8937, (2008).

- [11] W. Boomsma et al., *PHAISTOS: a framework for Markov chain Monte Carlo simulation and inference of protein structure*, *J Comput Chem.*, 34, 1697–1705, (2013).
- [12] F. Pérez et al., *IPython: A System for Interactive Scientific Computing*, *{it Computing in Science and Engineering* , 9, 21-29, (2007).
- [13] J. Iglesias, M. Sanchez-Martínez, R. Crehuet, *Visualizing Cooperative Secondary Structure Elements in Protein Ensembles SS-Map*, *Intrinsically Disord. Proteins* , 1, e25323, (2013).

Bibliografía consultada

Berg J.M., Timoczko J.L. & Stryer I.; *Bioquímica*. 2007, Barcelona, Editorial Reverté.

Lehninger, A., Nelson, D. L., Cox, M.M.; *Principios de bioquímica*. 3a ed. 2001, Barcelona, Omega, cop.

Índex de figures

Figura 1. El bioquímic britànic Frederick Sanger, guardonat en dues ocasions amb el premi Nobel de Química (1958 i 1980), estudiant un dels seus models estructurals. **Font:** Alchetron.com

Figura 2. A dalt, participants al programa RSI 2016 al Massachusetts Institute of Technology (MIT). A l'esquerra, amb el biòleg Howard Robert Horvitz, Premi Nobel de Medicina l'any 2002. **Font:** Pròpia

Figura 3.: Article original de la revista *Cell* en què es descriu per primera vegada la proteïna **ACTR** i la seves funcions biològiques. (1997)

Figura 4. Article original de la revista *Nature* en què es descriu per primera vegada la proteïna **CREB** i la seves funcions biològiques. (1987)

Figura 5. Imatge de Mollie, una nena amb la síndrome de Rubinstein-Taybi. S'hi poden observar els ulls lleument separats, les parpelles caigudes, l'expressió dolça i els polzes gruixuts, algunes de les característiques d'aquesta síndrome. Foto: Frederikke Brostrup

Figura 6. Imatge de la presentació de la Fundació Josep Carreras contra la leucèmia en la seva web (esquerra), i logotip de la Fundació (a dalt). **Font:** www.fcarreras.org

Figura 7. Acoblament d'ACTR al punt d'unió NCBD de CREB.
Font: Nature Reviews. Molecular Cell Biology. www.nature.com

Figura 8. Aminoàcid en la seva forma ionitzada o dipolar, amb càrrega negativa al grup carboxil (en vermell) i positiva al grup amino (en blau). La cadena lateral “-R” es mostra en color verd. **Font:** Berg, Tymoczko & Stryer. *Bioquímica*.

Figura 9. Enllaç peptídic. **Font:** Lehninger. *Bioquímica*.

Figura 10. Al·legoria d'una proteïna plegada amb un cabdell de corda embolicat.

Figura 11. Estructura tridimensional de les proteïnes. a) Estructura primària en un dipèptid; b) Estructura secundària, en aquest cas una hèlix- α ; c) Estructura terciària: l'hèlix- α i/o els fulls- β s'han recargolat sobre si mateixos; d) Estructura quaternària: 4 subunitats amb plegament terciari s'han acoblat per formar una proteïna de 4 cadenes, en aquest cas l'hemoglobina. **Font:** Memorial University of Newfoundland. (mun.ca). © 2010 PJ Russell, iGenetics 3 ed.

Figura 12. Linus Pauling, descobridor de l'hèlix- α , és una de les quatre persones que han rebut dos premis Nobel, juntament amb Frederick Sanger (també implicat en l'estudi de les proteïnes), Marie Curie i el físic nordamericà John Bardeen. **Font:** Cienciaonline.com.

Figura 13. Estructura tridimensional de l'hèlix- α . A) Representació del cilindre o cinta, amb els C α en el centre (en negre) i els radicals laterals (-R) cap a fora (en verd); B) Vista lateral amb un model de boles i pals, que mostre els ponts d'hidrogen entre els grups amino i carboni (línies discontinües); C) Vista vertical, amb la part interior de l'hèlix i els radicals laterals (-R) cap a fora (en verd); D) Vista vertical d'una estructura de boles, amb la part interior de l'hèlix i els radicals laterals (-R) cap a fora (en verd). **Font:** Berg, Tymoczko & Stryer. *Bioquímica*.

Figura 14. L'enllaç peptídic forma un pla en el qual hi ha 6 àtoms (C, O, N, H, C α). Les cadenes laterals -R es mostren en color verd. **Font:** Berg, Tymoczko, Stryer. *Bioquímica*.

Figura 15. Rotació al voltant dels enllaços senzills del polipèptid (fletxes negres). A la figura superior, Φ (phi) és l'angle de rotació al voltant de l'enllaç entre el N i el C- α , mentre que Ψ (psi) és l'angle de rotació al voltant de l'enllaç entre el C- α i el CO; les fletxes vermelles assenyalen l'enllaç peptídic rígid. A sota, la representació tridimensional dels mateixos angles de rotació Φ (phi) i Ψ (psi). **Font:** Berg, Tymoczko, Stryer. *Bioquímica*.

Figura 16. Gràfica de Ramachandran, que mostra que no són possibles tots els valors dels angles de Φ (phi) i Ψ (psi) a causa de les col·lisions estèriques entre els àtoms dels aminoàcids. Les regions més favorables es mostren en verd fosc i les intermedies en verd clar. La resta estan desfavorides a causa de les restriccions estèriques. **Font:** Berg, Tymoczko, Stryer. *Bioquímica*.

Figura 17. Joc d'enfonsar vaixells. **Font:** Todocoleccion.net

Figura 18. Un camí d'exploració a l'atzar del mètode Monte Carlo es pot comparar a un camí a l'atzar al joc d'enfonsar vaixells **Font:** Elaboració pròpia

Figura 19. Nivells energètics del plegament d'una proteïna comparats amb un perfil topogràfic. Per aconseguir el nivell energètic més baix (verd) des del cim (vermell), cal primer superar el nivell mitjà (groc). **Font:** Elaboració pròpia

Figura 20. Gràfica de llargada d'hèlix- α de la mutació A1047G usant els càlculs de Profasi. **Font:** Elaboració pròpia

Figura 21. Gràfica de llargada d'hèlix- α de la mutació D1050E usant els càlculs de Profasi. **Font:** Elaboració pròpia

Figura 22. Gràfica de llargada d'hèlix- α de la mutació S1043M usant els càlculs de Profasi. **Font:** Elaboració pròpia

Figura 23. Gràfica d'hèlix- α "aminàcid a aminoàcid" comparant algunes mutacions individuals i el tipus salvatge utilitzant els càlculs de Profasi. **Font:** Elaboració pròpia

Figura 24: Gràfica de llargada d'hèlix- α de la mutació A1047G utilitzant els càlculs de Phaistos. **Font:** Elaboració pròpia

Figura 25: Gràfica de llargada d'hèlix- α de la mutació S1043M utilitzant els càlculs de Phaistos. **Font:** Elaboració pròpia

Figura 26: Gràfica de llargada d'hèlix- α de la mutació D1050E utilitzant els càlculs de Phaistos. **Font:** Elaboració pròpia

Figura 27: Gràfica de llargada d'hèlix- α del Mutant Triple (S1043M, D1050E i T1054Q) utilitzant els càlculs de Phaistos. **Font:** Elaboració pròpia

Figura 28: Gràfic d'hèlix- α "aminoàcid a aminoàcid" comparant totes les mutacions i el WT utilitzant els càlculs de Phaistos. **Font:** Elaboració pròpia

Figura 29: Gràfic de llargada d'hèlix- α comparant totes les mutacions i el WT utilitzant els càlculs de Phaistos. **Font:** Elaboració pròpia

Figura 30: Relació entre el contingut d'hèlix- α "aminoàcid a aminoàcid" i el temps utilitzat en el procés d'acoblament de les dues proteïnes. **Font:** Elaboració pròpia

Figura 31: Regressió lineal entre el contingut d'hèlix- α "aminoàcid a aminoàcid" de la hèlix- α més llarga (la del Mutant Triple) i la reducció del temps utilitzat en el procés d'unió de les dues proteïnes. **Font:** Elaboració pròpia

V. Annex

Computational Analysis of the Effect of Helical Propensity in Intrinsically Disordered Proteins to the Binding Process

Article presentat al Programa Joves i Ciència per optar a participar en el programa Research Science Institute (RSI) 2016 al Massachusetts Institute of Technology

Computational Analysis of the Effect of Helical Propensity in Intrinsically Disordered Proteins to the Binding Process

Author: Magalí Luna i Perelló
Advisor: Ramon Crehuet

January 18, 2016

Abstract

Abstract: Intrinsically Disordered Proteins (IDPs) do not have a stable ordered fold. Usually, they achieve the stable fold structure as they bind to other macromolecules. However, in many cases these proteins have some transient secondary structure before the binding process. For now, it is not completely known how this transient secondary structure affects the binding efficiency. The ACTR protein is an IDP which interacts with the NCBD domain of the CREB protein. Some mutations have been created of ACTR that favor or disrupt the formation of α -helix before binding. The goal of this research is to use computational Monte Carlo (MC) simulations in order to determine the effect of the mutation of the residue sequence on the formation of secondary structure in unbound state, looking for the presence of α -helix in each residue and for the length of the α -helix structures to which these residues belong.

1 Introduction

1.1 IDPs

Intrinsically Disordered Proteins (IDPs) are a family of proteins which apparently have a function without a well-defined structure in physiological conditions [1]. The discovery of the IDPs challenges the theory that defends that the proteins must have a well-defined structure to have a function, because the IDPs have a useful function by themselves. But they

also have the ability to start doing a new function after binding another protein. IDPs are common in nature, and abundantly found to be involved in the pathogenesis of various diseases. They have numerous biological functions that complement functional repertoires of ordered proteins although they lack structure. They also participate in one-to-many and many-to-one interactions at the vital cellular processes.

1.2 Binding Process

Induced-fit Hypothesis The Induced-fit Hypothesis defends that IDPs have no secondary structure until they bind to another protein. According to the Induced-fit Hypothesis, it achieves that secondary structure, such as an α -helix, after the binding process.

Conformational Selection Hypothesis The Conformational Selection Hypothesis defends that IDPs have some transient secondary structure before they bind to another protein. This transient secondary structure attracts the other protein and favors the binding process.

1.2.1 Previous Experimental Results

Some previous laboratory experimental results show that the second hypothesis, the Conformational Selection Hypothesis could be the realistic one [2]. This is based on experiments that show that increasing the previous formation of a transient α -helix structure in

the free state, by doing some mutations in the amino acid sequence, makes the binding process faster.

We have worked with these mutations because they are the ones used in the laboratory experiments. They choose these mutations because these ones change the helicity without interacting with the binding site. Using the same mutations for the simulations we are able to compare our results to confirm or discard our hypothesis.

1.2.2 Length Hypothesis

One of the hypothesis of this work is the Length Hypothesis. It defends that the reduction of the binding time is not only regulated by the α -helix content of each residue individually and independently, but it is also determined by the length of this α -helix structure.

1.3 ACTR and NCBD-CREB

ACTR is a nuclear receptor co-activator that is a potent histone acetyltransferase and also an IDP [3]. When it binds to the NCBD domain of the CREB protein, it folds and starts having another useful function.

NCBD is a domain of the CREB protein. The CREB protein has a function related to the activation of the transcription process, the embryonic development, the growth control, and the homeostasis [4]. Mutations in this gene cause Rubinstein-Taybi syndrome (RTS) [5]. It has also been associated with myeloid leukemia [6]. But it is also known because of its active domains, such as NCBD, who interacts with ACTR, the co-activator we are interested in.

One of the targets of this project is to check if the laboratory experimental results can be predicted by computational methods. That can be useful to reduce time, human work and also economical expenses.

2 Computational Methods

To do computational simulations of the original residue sequence and all the mutations that have been proposed, a Monte Carlo method has been used. It is

a numerical method that explores the protein folding possibilities through random choices using importance sampling.

We have used Profasi to perform Monte Carlo simulations. Profasi is a C++ program designed to calculate simulations of protein folding and aggregation using the Monte Carlo method [7]. It provides an all-atom description of the protein chains, creating protein models with fixed bond lengths and bond angles.

With the Profasi program, simulations of 100, 100.000 and 100.000.000 Monte Carlo (MC) cycles have been calculated. After the numeric calculations we have extracted pdb files (every 1, 3 and 800 cycles respectively) that contain all the information needed to represent the three-dimensional proteins. We finally worked with the 100.000.000 MC cycles simulations, and with the pdb files that we extracted from them.

We made this reduction because, even having all the calculations to make the graphical representations of each MC cycle, extracting all pdb files would require a lot of time. Additionally, it would not make such a statistical difference to justify that expense of time and computational work.

With this data we have plotted some preview graphs of α -helix content by residue with the Gnuplot program. We have as well generated time-lines of alpha helix content for each simulation with the VMD program, in order to evaluate the evolution of the α -helix content for every result of the simulation.

We have checked that after the first 1000 MC cycles, all the simulations have almost the same stable α -helix content. Therefore, we used the steps from 1000 to 10.000.000 to do the comparison between one mutation and another.

The steps before number 1000 have been discarded for statistical reasons because they are more irregular than the real protein would be. We finished the range at 10.000.000 because the following simulations had almost the same stable α -helix content. For this reason we chose this range to make the next calculations and creations of graphs faster and easier.

This specific reduction has been applied to each mutation and to the wild type (WT), i.e., the original amino acid sequence without any mutation. With

this data we have represented a global graph of α -helix content by residue to compare all the variations of our residue sequence.

We also used a Dynamics Bayesian Network method [8] to generate protein structures as implemented in Phaistos. Phaistos is a program designed to simulate all types of proteins [9]. It can also analyze them using pdb files or using one single initial pdb file and calculating the trajectory directly from mathematical calculations, without need to transcript all these numeric information into pdb files. It has the capacity to differentiate an IDP from a normal protein. In the study of the IDPs, this allows you to obtain more reliable and realistic results than with programs designed to simulate protein folding.

With Phaistos program, and with the help of the iPython Notebook, we created some graphs comparing different factors of our interest. We made α -helix by residue plots and also α -helix by length graphs.

With this information we want to know if, apart from the effect on the α -helix formation of the individual residues, the length of the α -helix structure could play a role. This has been checked comparing the peaks of our α -helix by length plots of each mutation with the laboratory experimental results. These lab results measured the elapsed time in the binding process of ACTR to the NCBP domain of the CREB protein in each case [2].

iPython notebook is an interactive computational environment that uses the python programming language [10]. It can be used to write text files, and also to combine programming code with text notes or explanations, and even graphs and mathematical symbols. It is useful to gather computations and data analysis in the same file. We have used iPython notebook to do plots and to analyze data.

2.1 Visualization of the results

The α -helix by residue plots show the frequency of α -helix structure (normalized from 0 to 1) for each residue, individually and independently of their nearest neighbors or any other one.

The α -helix by length plots show the frequency of α -helix structure of each residue and how long is the corresponding α -helix. We have made two types of

α -helix by length graphs: the first ones show the average length of the α -helix structures containing each residue (in a frequency normalized from 0 to 2.5). The other ones show as well the probability of the formation and stabilization of this secondary structure, in a color gradient from white to blue, in which white is a non-structured residue and dark blue is a well-structured residue [11].

We made also two plots comparing the α -helix content with the association rate constant. One shows the relation between the α -helix content of each mutation and the laboratory experimental results of the association rate constant (k_{on})[2]. The other one shows a linear regression between the α -helix content of the longest α -helix (19 residues) and k_{on} .

3 Results and Discussions

3.1 Profasi

In the α -helix by length graphs (Figure 1 and Figure 2) that we have obtained with the Profasi calculations, we can see that between them there is not a remarkable difference. They both show a similar and stable α -helix propensity. These results are not realistic considering that we are working with an IDP, which has non-stable secondary structure. This can be checked comparing the A1047G mutation graph with the other one. The A1047G mutation has been designed to disrupt the α -helix the most, and the other mutation to increase it; and between their graphs there is almost no difference.

In Figure 3 we can see almost the same result for all the mutations and the WT except for the A1047G mutation, who has a bit lower α -helix content at the beginning, but it is also stabilized after the 11th residue. As in the previous case, this plot shows unrealistic results in which ACTR has a stabilized α -helix structure along all the residue sequence. This results are challenged by the laboratory experiments which show that ACTR has not that much stabilized α -helix structure [2], and that between the mutations there are remarkable differences (especially between A1047G mutation and the other ones).

That shows us that a program designed to simulate

protein foldings, like Profasi, is not a good choice for working with IDPs because it cannot differentiate between an IDP and a normal well-structured protein. Its calculations therefore treat the IDP like a structured protein with a high content of long and stable secondary structure.

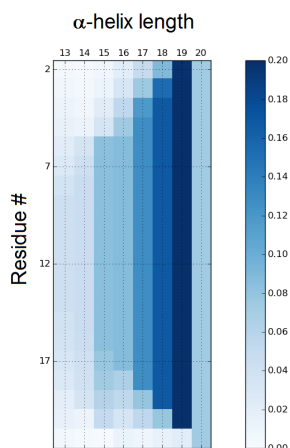


Figure 1: α -helix by length graph of the A1047G mutation using the Profasi calculations.

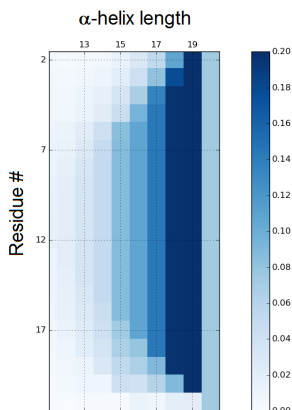


Figure 2: α -helix by length graph of the D1050E mutation using the Profasi calculations.

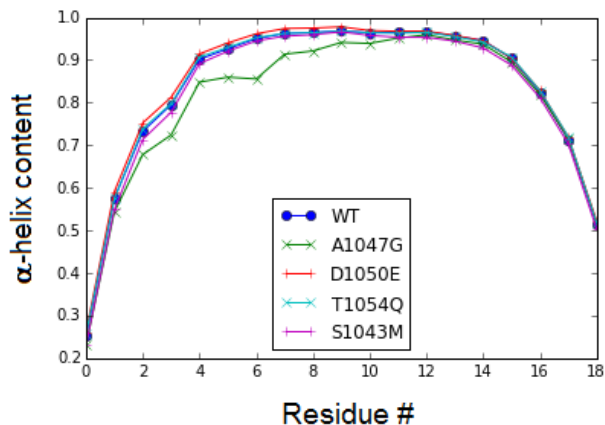


Figure 3: α -helix by residue plot comparing all the individual mutations and the WT using the Profasi calculations.

3.2 Phaistos

Unlike the α -helix by length graphs created using the Profasi calculations, the ones that were created using the Phaistos calculations show a more plausible results. Comparing the S1043 and D1050E α -helix by length Phaistos graphs (Figure 4, Figure 6) with the A1047G α -helix by length Phaistos graph (Figure 5) we can see a remarkable difference. Figure 5 has a much lower α -helix content and length, so it agrees with the original design of the mutations (S1043M and D1050E to increase the α -helix structure and A1047G to disrupt it). The α -helix by length graph (Figure 7) of the Triple Mutant shows the most stabilized α -helix structure, and the longest one. This is a realistic result according to the previous hypothesis and the laboratory experimental results [2]. This graphs show also a more believable level of α -helix structure in a IDP, even if it has a mutation to favor it.

The α -helix by residue Phaistos plot (Figure 8) shows that A1047G has an expected low content of α -helix. The other individual mutations have a higher content of secondary structure than the WT, even though the S1043M mutation is very similar to it. The Triple

Mutant has a much higher content of α -helix structure than any of the other mutations or the WT. This is also an expected result because of the previous laboratory experiments, that show that the Triple Mutant has a higher content of α -helix structure and also a higher velocity during the binding process [2]. The α -helix by length Phaistos plot (Figure 9) shows that A1047G has an expected low content of α -helix and the other individual mutations have a higher content of secondary structure than the WT too. However, in this case the D1050E mutation is the one that has a similitude with the WT (they both have their length peak in the 14th residue).

On the other hand, the S1043M mutation and the Triple Mutant both have their length peak at the 16th residue.

If we compare this to the laboratory results of the time that each mutation spends in the binding process [2], we can see a significant correlation between the α -helix content and the association rate constant (k_{on}) (Figure 10).

In Figure 11 we can also see a linear regression between the α -helix content of the longest α -helix and the association rate constant (k_{on}). This plot shows a significant correlation that allows us to support the Length Hypothesis explained above.

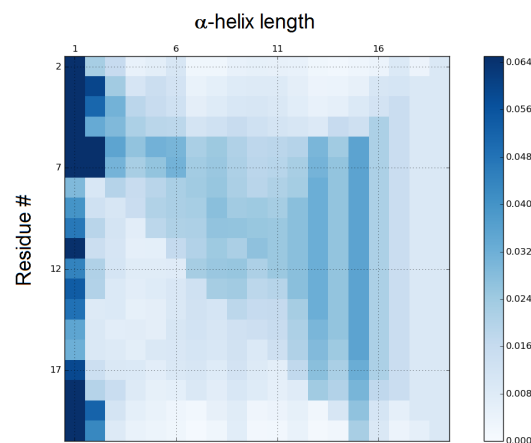


Figure 5: α -helix by length graph of the A1047G mutation using the Phaistos calculations.

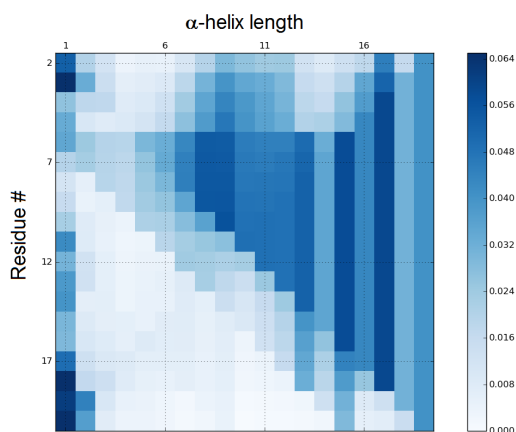


Figure 4: α -helix by length graph of the S1043M mutation using the Phaistos calculations.

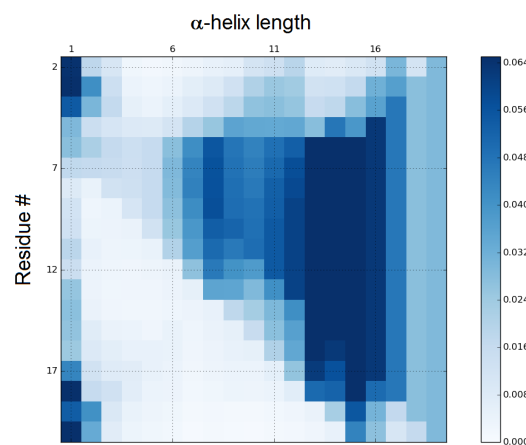


Figure 6: α -helix by length graph of the D1050E mutation using the Phaistos calculations.

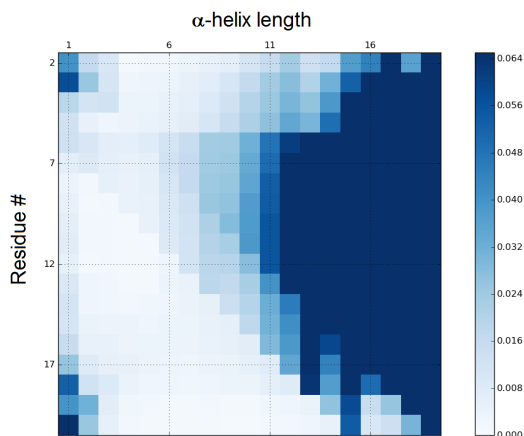


Figure 7: α -helix by length graph of the Triple Mutant combination (S1043M, D1050E and T1054Q) using the Phaistos calculations.

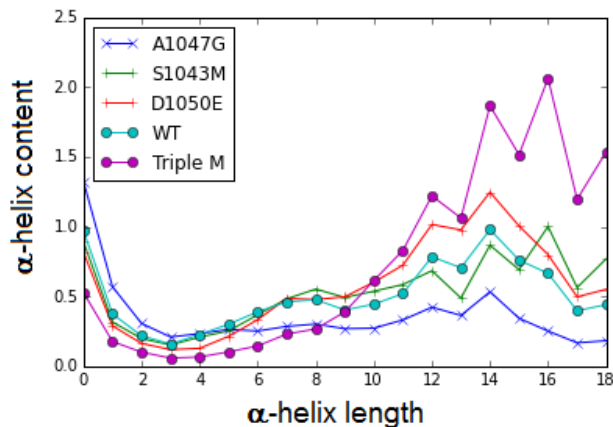


Figure 9: α -helix by length plot comparing S1043M, A1047G and D1050E individual mutations, the Triple Mutant and the WT using the Phaistos calculations.

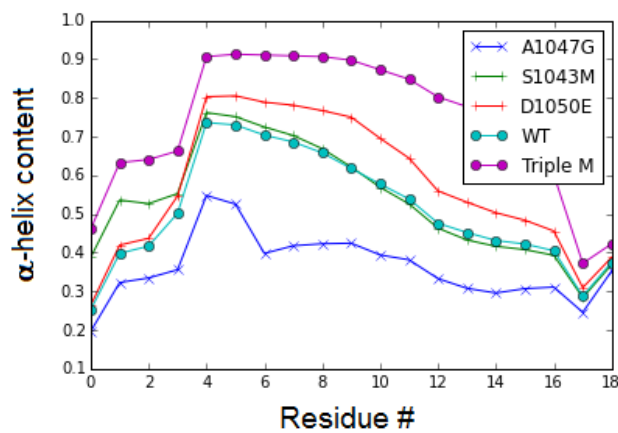


Figure 8: α -helix by residue plot comparing S1043M, A1047G and D1050E individual mutations, the Triple Mutant and the WT using the Phaistos calculations.

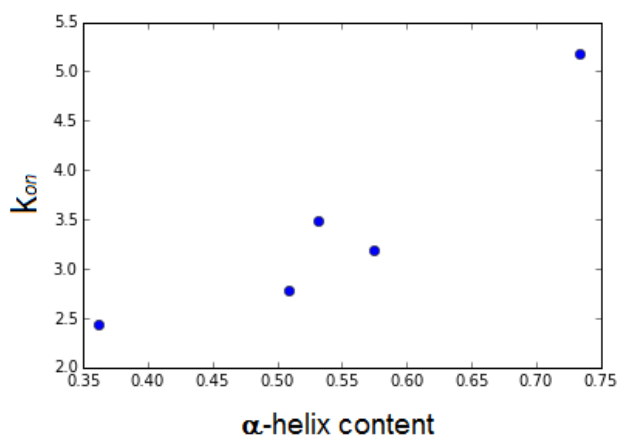


Figure 10: Relation between the α -helix content and the laboratory experimental results of the association rate constant (k_{on}).

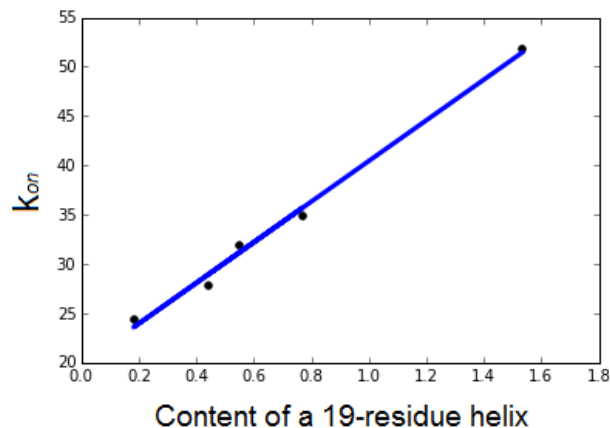


Figure 11: Linear regression between the α -helix content of the longest α -helix and the laboratory experimental results of the association rate constant (k_{on}).

4 Conclusions

We have been able to predict some laboratory experimental results in a faster, cheaper and safer way. But it is also true that we only have been able to do this using the Phaistos calculations, not the Profasi ones. That means that the computational methods are far from being enough accurate and reliable for all cases to be an alternative to the laboratory experiments. However they are excellent tools to check this experimental results and to design new prototypes that can be used in the laboratory for next experiments.

As we expected, in the α -helix by residue plots the mutations that have been designed to increase the α -helix, the S1043M and D1050E individual mutations and especially the Triple Mutant, have a higher α -helix content than the WT. The mutation that has been designed to disrupt the α -helix the most, the A1047G mutation, has a lower α -helix content than the WT or any other mutation.

Observing the α -helix by length graphs we can conclude that the Length Hypothesis we have presented at the beginning of our research seems to be true. This conclusion is based on the comparison of our results in the individual α -helix by length graphs and

the global one with the laboratory experimental data, taking into account the elapsed time during the binding process [2]. We can see this relation in Figure 10. This comparison shows that the mutation that has the best results in the laboratory experiments, the Triple Mutant, has also the longest α -helix structure of all the mutations. This correlation is shown in 11. This seems to mean that the length of the α -helix structure is also an important factor in the reduction of the binding time.

The next step of our research could be to simulate the binding process using computational methods, and also to think some other mutations to increase the α -helix structure considering the α -helix length as a main factor at the same time as the α -helix individual content.

We can also explore another computational programs or tools to find the realistic ones, to go forward with the prediction work and to look for a computational method to predict the laboratory experimental results with enough precision to reduce the laboratory experiments doing the calculations in a novel way.

One of the hypothesis that we could be able to confirm following this work's trajectory is the Length Hypothesis. As a next step, we could find out if we will be able to increase the relation between the α -helix content and the laboratory experimental results of the association rate constant (Figure 10) introducing information about the α -helix length that has not been seen yet experimentally.

5 Acknowledgements

I want to thank specially my advisor Ramon Crehuet for all the time, support and information that he has given to me during all the time I have been working in this research. This project could not have been accomplished without him. I thank the iQAC Institute as well for letting me work in their computational labs.

I want to thank Tiago Cordeiro for managing my stage at the iQAC Institute, for worrying about the progress of my research and for always resolving all the doubts I have had during this project. I want to thank also my other three Joves i Ciència teach-

ers for orientating and encouraging me to begin this research.

I want to thank Eva Calvés for her work coordinating all our projects in the Joves i Ciència Program, and also for keeping us informed and for resolving all our doubts.

And finally I really want to thank the Joves i Ciència Program and also the Fundació Catalunya-La Pedrera for giving me this unique opportunity to work, learn and understand so much more science than I could have ever imagined.

To all of them, I want to say thank you.

References

- [1] I. C. FELLI ET AL., Intrinsically Disordered Proteins, in NMR of Biomolecules: Towards Mechanistic Systems Biology, *Wiley-VCH Verlag GmbH & Co. KGaA*, Weinheim, Germany (2012).
- [2] V. IEŠMANTAVIČIUS, M. KJAERGAARD ET AL., Helical Propensity in an Intrinsically Disordered Protein Accelerates Ligand Binding, *Angew. Chem. Int. Ed.*, **53**, 1548-1551, (2014).
- [3] H. CHEN ET AL., Nuclear Receptor Coactivator ACTR Is a Novel Histone Acetyltransferase and Forms a Multimeric Activation Complex with P/CAF and CBP/p300, *Cell*, **90**, 569-580, (1997).
- [4] H. SIDDIQUE, VN. RAO, ES. REDDY, CBP-mediated post-translational N-glycosylation of BRCA2, *Int J Oncol.*, **35**, 16387-91, (2009).
- [5] F. PETRIJ ET AL., Rubinstein-Taybi syndrome caused by mutations in the transcriptional coactivator CBP, *Nature*, **376**, 348-51, (1995).
- [6] JL. VIZMANOS ET AL., t(10;16)(q22;p13) and MORF-CREBBP fusion is a recurrent event in acute myeloid leukemia, *Genes Chromosomes Cancer*, **36**, 402-5, (2003).
- [7] A. IRBÄCK, S. MOHANTY, PROFASI: A Monte Carlo simulation package for protein folding and aggregation, *J. Comput. Chem.*, **27**, 1548-1555, (2006).
- [8] W. BOOMSMA, K. V. MARDIA, C. C. TAYLR, J. FERKINGHOFF-BORG, A. KROGH, T. HAMELRYCK A Generative Probabilistic Model of Local Protein Structure, *Proc. Natl. Acad. Sci.*, **105**, 8932-8937, (2008).
- [9] W. BOOMSMA ET AL., PHAISTOS: a framework for Markov chain Monte Carlo simulation and inference of protein structure, *J Comput Chem.*, **34**, 1697-1705, (2013).
- [10] F. PÉREZ ET AL., IPython: A System for Interactive Scientific Computing, *Computing in Science and Engineering*, **9**, 21-29, (2007).
- [11] J. IGLESIAS, M. SANCHEZ-MARTÍNEZ, R. CREHUET, Visualizing Cooperative Secondary Structure Elements in Protein Ensembles SS-Map, *Intrinsically Disord. Proteins*, **1**, e25323, (2013).