

### ***3.1. Preparació de la proteïna***

A l'iniciar un experiment de cristal·lització de qualsevol proteïna s'han de complir una sèrie de condicions. La primera és una condició de puresa de la mostra. Ja que volem cristal·litzar una proteïna, o un complex concret, aquest ha de ser majoritari en la mostra, en una puresa aproximada del 95% com a mínim. La segona consisteix en que la proteïna ha d'estar en un tampó en la qual sigui estable, però que alhora contingui la mínima quantitat de reactius químics per tal de no interferir en les condicions de cristal·lització.

#### ***3.1.1. Expressió i purificació de la proteïna p4***

L'expressió i la purificació de la proteïna p4 va ser duta a terme en el laboratori de la Dra. Margarita Salas i està descrit en la bibliografia (Barthelemy *et al.*, 1987). L'últim pas de purificació consistia en una precipitació amb sulfat amònic al 70%, i així rebíem nosaltres la proteïna.

#### ***3.1.2. Solubilització de la proteïna***

Per tal d'eliminar el sulfat amònic, es va dissoldre el pellet salí en el que es trobava precipitada la proteïna amb el tampó que vàrem utilitzar al llarg de totes les proves de cristal·lització. Aquest tampó contenia: NaCl 200 mM, Tris 20 mM pH=7,5, EDTA 1 mM i DTT 1 mM.

Vàrem dissoldre la mostra en 10 ml de tampó, barrejant amb cura per tal que la mostra es redissolgués, però no que es desnaturalitzés. Un cop teníem la proteïna resolubilitzada, vàrem procedir a concentrar la mostra per ultrafiltració, utilitzant un centricon YM-10 (Amion-Millipore), seguint les recomanacions del fabricant en quant a força centrífuga aplicada i durada d'aquest.

Degut a que la capacitat màxima dels concentradors és de 2ml, es van fer varis cicles de reomplenat per tal de concentrar tota la mostra. Tan bon punt la mostra va estar concentrada a un volum final inferior de 0,5 ml , vàrem procedir a reomplir el centricon amb tampó amb molta cura, de manera que es formés una interfase entre la solució inicial i el tampó fresc addicionat.



Figura 3.1: Concentrador Centricon (Amion-Millipore) utilitzat durant el procés de concentració de la mostra de proteïna i la de complex de proteïna i DNA.

Vàrem reconcentrar i repetir el procés fins a un total de 5 vegades, amb el que consideràrem eliminat el sulfat amònic i teníem la mostra concentrada per poder fer els assajos de cristal·lització.

### ***3.1.3. Determinació de la concentració de proteïna***

La determinació de la concentració de proteïna total es va dur a terme pel mètode del coeficient d'extinció molar. Aquest mètode es fonamenta en l'absorció de llum de longitud d'ona de 280 nm per part dels grups aromàtics presents en els aminoàcids fenilalanina, tirosina, triptòfan i histidina; obtenint-se un coeficient

específic per cada proteïna. En aquest cas, varem utilitzar un valor d'extinció molar teòric, obtingut en la web de l'EXPASY (<http://ca.expasy.org/>). S'utilitza l'equació 3.1 per obtenir la concentració de proteïna.

$[\text{proteïna}] = (\text{Absorbància a } 280 \text{ nm} / \text{Coeficient d'extinció molar}) \times \text{Factor de dilució}$   
Equació 3.1.

### ***3.2. Preparació i purificació dels complexos***

Els complexos de proteïna i DNA es preparen a partir de la mostra de proteïna prèviament quantificada juntament amb DNA comercial. Aquest DNA passa per un procés d'*annealing* i posteriorment es barreja amb la proteïna en les condicions de molaritat que corresponguin en cada cas. Aquests complexos es purifiquen posteriorment per mètodes cromatogràfics per tal d'obtenir una mostra pura i homogènia, per realitzar els experiments de cristal·lització.

#### ***3.2.1. Preparació dels complexos***

El DNA que varem utilitzar va ser comercial, de la casa MWG. Les concentracions de cada cadena de DNA es van mesurar pel mètode del coeficient d'extinció molar, a partir del coeficient d'extinció molar teòric, obtingut utilitzant la pàgina web <http://www.pitt.edu/~rsup/OligoCalc.html>.

Per formar el DNA de doble cadena es van sotmetre la mescla de oligonucleòtids a un procés d'*annealing* de la següent manera. Barrejant de manera equimolar les dues cadenes de DNA i afegint una solució de KCl 4M fins a una concentració final de 20 mM. Es va preparar el bany a 85°C i es va escalfar la mostra durant 20 minuts. Un cop transcorregut el temps d'escalfat, es va apagar el bany i varem deixar que es refredés de manera molt lenta, equilibrant-se amb la temperatura ambient del laboratori. Aquest procés pot durar 4 hores o més, per lo qual és recomanable de deixar-ho tota la nit.

Per preparar el complex DNA proteïna, un cop refredat es va barrejar molt lentament amb la proteïna, en una proporció DNA:proteïna de 1:2. Es va deixar 15 minuts en gel, per que el complex es formés completament.

### **3.2.2. Purificació dels complexos p4-DNA**

La purificació dels complexos amb DNA es va dur a terme mitjançant una cromatografia d'exclusió molecular o gel filtració. Es va concentrar la mostra amb un *centricon* YM-10 (Amicon-Millipore), fins a obtenir un volum de 0.5 ml. La mostra es va carregar en una columna de gel filtració Superdex 75 HR 10/30 (Amersham Biosciences) muntada en un aparell de FPLC ÄKTA Basic (Pharmacia). La columna es va equilibrar prèviament amb un tampó Tris 20 mM pH=7.5, NaCl 200 mM, DTT 1 mM, EDTA 1 mM. Aquest tampó també es va utilitzar per eluir la columna, a un flux constant de 0.5 ml/min. L'eluit es va recollir en fraccions de 0.5 ml, que es conserven a 4°C fins dur a terme el seu anàlisi per SDS-PAGE al 15%. Es va ajuntar les fraccions més pures del pic corresponent al complex p4-DNA.

### **3.2.3. Detecció dels complexos per SDS-PAGE i tinció de plata**

Es va comprovar que els pics detectats en la cromatografia corresponien al complex de proteïna-DNA, mitjançant electroforesi en gel de poliacrilamida (SDS-PAGE) al 15% i una tinció de plata, que permet veure tant la proteïna com el DNA. El gel SDS-PAGE es va fer amb el kit de Mighty Small (Pharmacia) i a un percentatge de poliacrilamida del 15%, que permet separar proteïnes de pes molecular petit (entre 30-10 KDa).

La tinció de plata es va fer pel mètode alcalí de tinció positiva. El primer pas consisteix en la fixació del gel. Vàrem posar el gel en agitació en una solució de metanol al 50% durant un mínim de 1h.

Un cop transcorregut el temps de fixació, es va abocar la solució fixadora, sense

tocar el gel amb els dits (ja que interferiria amb la tinció) i hi vàrem afegir la solució de tinció. Per preparar la solució de tinció vàrem barrejar en un vas de precipitats ben net 40 ml d'aigua, 1.4 ml d'amoniac al 34%, i 190 µl de NaOH 10 M. En un altre vas de precipitats es van barrejar 0.8 g de nitrat de plata i 4 ml d'aigua. Aquesta barreja es va afegir amb molta cura utilitzant una pipeta pasteur al vas anterior, que contenia un agitador net.

Un cop teníem el gel amb la solució de tinció, ho vàrem mantenir 15 minuts en agitació. Es va abocar la solució de tinció i vàrem fer 2 rentats de 5 minuts amb aigua milliQ abundant, per disminuir el soroll. El següent pas és el de revelat, en el que l'ió plata es redueix a plata metàl·lica. Es van barrejar 75 µl de formaldehid al 36-38%, 35 µl d'àcid cítric 1M i vàrem afegir aigua fins a un volum final de 150 ml. Vàrem afegir una mica de la solució de revelat, ho vàrem remenar i ho vàrem llençar, per tal d'eliminar els restes de plata no units que puguin quedar. Es va afegir la resta i vàrem observar el revelat sota agitació. Quan es va arribar al grau de tinció desitjat, vàrem llençar la solució reveladora i vàrem afegir una solució metanol al 50% i àcid acètic al 12%, per aturar la reacció. Es va llençar la solució anterior i es va afegir aigua per hidratar el gel. Es deixa durant almenys 1 hora per aconseguir una bona rehidratació.

#### ***3.2.4. Determinació de la concentració de complex p4-DNA***

Degut a la gran absorbància del DNA a 280 nm, no ens va ser possible seguir utilitzant el mètode del coeficient d'extinció molar. Llavors, per quantificar la quantitat de complex del que disposàvem, vàrem utilitzar el mètode de Bradford (Bradford, 1976). Aquest es fonamenta en el desplaçament en medi àcid del màxim d'absorbància del compost blau de Coomassie 6-250 des de 465 nm fins a 595 nm quan es troba unit a proteïna. Es va fer una recta de calibració amb concentracions conegudes d'albumina per extrapolar, dins de la zona lineal de la recta, el valor d'absorbància de la mostra i trobar-ne la concentració.

### 3.3. Cristal·lització

Els cristalls es van créixer en plaques de cultiu Linbro® (6x4 pouets) (Figura 3.3) emprant el mètode de difusió de vapor en gota sentada (Figura 2.1). El volum de gota utilitzat va ser de 2  $\mu$ l, resultants de la mescla de 1  $\mu$ l de la solució que conté la proteïna p4 purificada a una concentració de 13 mg/ml amb 1  $\mu$ l de la solució de precipitant. En el cas del complex p4-DNA, es va utilitzar una concentració de 6 mg/ml. Un cop omplert el reservori i mesclada la gota, de tots els pouets, es segella la cavitat amb cinta adhesiva (Hampton). D'aquesta manera, s'aconsegueix l'hermeticitat necessària per l'equilibrat entre la condició del reservori i la gota (Figura 3.2).

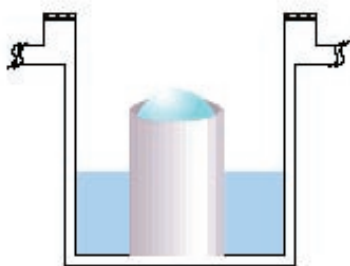


Figura 3.2: Esquema de una placa de cristal·lització, on es mostra la gota de proteïna i solució precipitant recolzada sobre el peu central, i la resta de la solució de cristal·lització al fons del pou.

Tots els experiments es van realitzar per duplicat, a 4°C i a 20°C. Les gotes es van deixar equilibrar a temperatura constant (20°C o 4°C), enfront d'un reservori que contenia 0.5 ml de la solució de cristal·lització (Figura 3.1). Entre la gota i el reservori s'estableix un equilibri per difusió de les espècies volàtils (aigua i dissolvents orgànics) fins que la pressió de vapor entre els dos punts s'iguali. La concentració de l'agent precipitant és major al reservori, doncs a la gota la solució precipitant ha estat mesclada amb la solució de proteïna. Així, l'intercanvi d'aigua es produeix des de la gota cap a la solució del pou, fet que provoca una reducció en el volum d'aquesta, de manera que augmenta progressivament la concentració dels seus constituents. Si s'han aconseguit trobar els percentatges de precipitant i proteïna adequats, la proteïna arriba a la sobresaturació i comença a cristal·litzar quan la gota i el reservori han assolit o estan a punt d'assolir l'equilibri .

Per tal de trobar les condicions a les quals la proteïna cristal·litza s'utilitzen diverses solucions de *screening* (Hampton), tenint-ne algunes d'específiques per complexos amb DNA i d'altres més generals.



Figura 3.3: Placa de tipus sitting drop de 24 pous utilitzada en els experiments de cristal·lització.

### ***3.3.1. Muntatge en crioloops, crioprotecció, i congelació dels cristalls***

Per tal de poder dur a terme els experiments de difracció, cal treure els cristalls de la solució de precipitant en que han crescut. Es treuen “pescant-los” de la solució mare amb uns bucles de niló, de manera que el cristall queda retingut a l’interior d’aquests gràcies a la tensió superficial que genera la capa fina de solució precipitant que l’envolta. Els cristalls de proteïna a temperatures properes a l’ambiental són sensibles als raigs X i sovint pateixen danys per radiació, especialment quan els experiments de difracció es duen a terme emprant les intenses línies de llum de sincrotró. Si aquests experiments es realitzen a temperatures criogèniques es redueix molt, o fins i tot s’arriba a eliminar, tant el dany per radiació com la dessecació del cristall i, per tant, s’obtenen dades de difracció de major qualitat i resolució, alhora que sovint permet la recollida de varis conjunts de dades d’un sol cristall, facilitant així la resolució del problema de les fases. Així doncs, la congelació dels cristalls és molt recomanable, alhora que facilita l’emmagatzematge i el transport dels cristalls.

Existeixen varies tècniques, des de la congelació directa del cristall, a la seva incubació prèvia en una solució estabilitzadora i/o l’addició d’algun agent crioprotector de manera directa o gradual. El mètode ideal és que el cristall creixi

ja en una solució precipitant que pugui actuar de crioprotector durant la recollida de dades.

### ***3.3.1.1. Harvesting: addició gradual de l'agent crioprotector***

Consisteix en canviar progressivament les condicions de la solució precipitant en que es troben els cristalls fins arribar a les de la solució crioprotectora final. L'objectiu és minimitzar el risc de que els cristalls es malmetin durant el procés de congelació, perdin el seu ordre intrínsec i, en conseqüència, poder de difracció.

Primer cal incubar els cristalls en una solució estabilitzant durant 5 o 15 minuts. Aquesta solució consisteix en la condició de cristal·lització multiplicant per 1,2 totes les concentracions. A continuació, i mantenint sempre els cristalls en aquesta solució estabilitzant, s'augmenta la substància que farà de crioprotector en percentatges del 5% fins arribar al valor desitjat. En cas de que els cristalls s'hagin crescut mitjançant la tècnica de la gota sentada es treu la cinta adhesiva de manera que la gota que conté el cristall quedi exposada. S'examina la gota sota la lupa binocular i s'afegeix, amb molta cura i al costat oposat a on hi ha el cristall, un volum igual al de la gota de la primera solució de *harvesting*. Es torna a col·locar la cinta adhesiva sobre el pou original. S'incuba uns 5 minuts a la temperatura en que s'han crescut els cristalls. És recomanable vigilar sota la lupa binocular que els cristalls no es deformin (per això va bé provar primer el protocol de *harvesting* amb cristalls no útils pels experiments de difracció). Es repeteixen els passos anteriors amb les següents solucions de *harvesting* fins arribar a la solució crioprotectora. Després del pas final, en què s'incuba durant 5' la gota que conté el cristall amb la solució crioprotectora, ja es pot procedir a la congelació dels cristalls, submergint ràpidament el *loop* (Figura 3.4) en que s'ha pescat el cristall en nitrogen líquid. Les solucions de *harvesting* provades en aquest treball van ser :

**5 – 10 – 15 % glicerol**, 12% PEG 6K, 2,4 M NaCl per la proteïna p4 sola

**5 – 10 – 15 – 20 % glicerol** 21,6% PEG 2K MME, 96 mM MgCl<sub>2</sub>, 60 mM Tris pH=8,5 pel complex p4-DNA.





Figura 3.4: Llaç o *loop*, utilitzat per pescar els cristalls de la gota de cristal·lització, muntat sobre una tija metàl·lica que té un peu imantat. Aquest imant s'utilitza per situar-lo en el goniòmetre de la línia de llum del sincrotró.

### 3.3.2. Derivatització de la proteïna amb àtoms pesants

El primer que cal fer és triar compostos que potencialment poden derivatitzar la proteïna i ens serveixin en la difracció. Per tant, cal assegurar-se que les diferències en les amplituds de la difracció provinents de la contribució de l'àtom pesant siguin majors que els error en la mesura de les dades. El tamany de l'àtom pesant (el numero atòmic) i el numero de llocs d'unió requerits per un fasejat exitós és proporcional al tamany (pes molecular) de la macromolècula. Els compostos més utilitzats són els de platí i els de mercuri. A més també s'utilitzen compostos que contenen pal·ladi, plata, reni, osmi, iridi, or, plom, lantànids o urani.

En el procés d'obtenir la incorporació d'un compost que conté un àtom pesant en el cristall de proteïna podem diferenciar dues estratègies ben diferenciades: la co-cristal·lització i el *soaking*. La co-cristal·lització consisteix en introduir la solució del compost que conté l'àtom pesant en la solució de cristal·lització en el moment de fer les gotes per cristal·litzar, per tal que mentre es vagi formant el cristall de proteïna es vagi incorporant el compost en la mateixa xarxa cristal·lina. La segona estratègia és el *soaking* que consisteix en introduir la solució del compost que conté l'àtom pesant en la solució de cristal·lització, un cop el cristall ja està format i que el compost difongui pels canals de solvent del cristall i s'incorpori de forma ordenada.

El procediment utilitzat per la derivatització és la incubació del cristall en una solució de harvesting (condició de cristallització x 1,2) i amb la solució de l'àtom pesant, durant diferents temps (de 6 hores a varis dies) i concentracions (entre 2mM a més de 50 mM). Un cop acomplert el temps d'incubació es procedeix amb la congelació del cristall.

Existeix una tercera estratègia, molt diferent a les anteriors i actualment molt més utilitzada. Aquesta consisteix en introduir aminoàcids que contenen l'àtom pesant dins la proteïna. L'aminoàcid utilitzat és la selenometionina, que té substituït el seu àtom de sofre per un de seleni. El procés d'incorporació consisteix en expressar la proteïna en una soca bacteriana auxòtrofa per metionina i aportar-li únicament la forma derivatitzada en el medi de cultiu. Així totes les proteïnes expressades en aquell moment tindran les metionines substituïdes per selenometionines.

### ***3.4. Recollida de les dades de difracció de raigs X***

En aquest treball hi ha tres recollides de dades diferents. La primera correspon a la difracció d'un cristall de la proteïna p4, al qual hi hem fet un *soaking* per derivatitzar-lo. La segona correspon a la difracció d'un cristall de proteïna p4 sol, i la tercera a un cristall del complex format per la proteïna p4 i un oligonucleòtid de 41 parells de bases.

Com a font de raigs X s'emprà la radiació de llum de sincrotró a diferents línies del ESRF a Grenoble (França). Per la proteïna derivatitzada vam utilitzar la línia BM-14, que permet variar la longitud d'ona dels raigs X i ens va permetre realitzar un experiment de tipus MAD. Per la proteïna p4 sola vam utilitzar la línia ID14-EH1, i pel complex p4-DNA vam utilitzar la línia microfocus ID13.

Per recollir un conjunt de dades complet, que permeti resoldre l'estructura, s'utilitza normalment la tècnica de rotació, on es va rotant el cristall mentre un feix de raigs X el bombardeja. El cristall gira al voltant d'un eix perpendicular

al feix dels raigs X. La rotació per cada imatge presa depèn de varis factors però és habitual que sigui propera a 1 grau. La rotació total necessària per recollir un conjunt de dades complert, depèn de la simetria del grup espacial al qual pertany el cristall. Grups espacials amb simetria més alta necessiten un menor número de imatges.

### ***3.4.1. Procediment de recollida de dades***

En un primer pas, es recullen unes poques imatges en diferents orientacions, normalment no seqüencials, separades per exemple 90°, per poder determinar la matriu d'orientació, la cel·la unitat i altres paràmetres del cristall com la mosaïcitat. També podem determinar la resolució màxima, el temps d'exposició, els angles que hem de recollir i quina rotació per imatge hem d'utilitzar. Un cop tenim tots els paràmetres clars, prosseguim amb la recollida de dades pròpiament dita. Generalment recollirem les imatges de manera seqüencial, però en altres casos ens pot convenir fer-ho d'una altra manera.

#### ***3.4.1.1. Experiment MAD: el cristall derivatitzat***

En l'experiment MAD (Multiwavelength Anomalous Dispersion) es recullen, típicament 3 ó més conjunts de dades de difracció, idèntics en quant al cristall i graus de rotació recollits, però diferents en quant a la longitud d'ona dels raigs X.

El primer pas a realitzar és recollir un espectre de fluorescència del cristall (XANES) a prop del canto d'absorció del dispersor anòmal. Això permet, a més de comprovar que l'àtom pesant està present en el cristall, calcular els valors de  $f'$  i  $f''$  en funció de la longitud d'ona. Aquestes funcions s'obtenen directament calculant la primera i la segona derivada de la funció de la fluorescència. A partir de les funcions derivades escollim les longituds d'ona més adequades per l'experiment MAD. Així recollim el primer conjunt de dades a l'energia on  $f''$  és màxima (pic), el segon on  $f'$  és mínima (inflexió) i la tercera a una energia llunyana d'aquestes

dues (remota).

En el nostre experiment vàrem recollir a unes longituds d'ona de: 1,0717 Å pel pic d'absorció; 1,0721 Å pel punt d'inflexió i 0,9762 Å per la remota (Taula 3.1).

Dades de l'experiment MAD	
Distància del detector (mm)	170
Angle de rotació (°)	1
Graus recollits (°)	210 / 156 / 170
Longitud d'ona (Å)	1.0717 / 1.0721 / 0.9762

Taula 3.1: Paràmetres de recollida de dades del cristall derivatitzat de proteïna p4

### **3.4.1.2. Cristall de la proteïna p4**

Per complementar les dades obtingudes pel cristall derivatitzat, es va recollir un sol conjunt de dades d'un cristall sense derivatitzar, intentant de millorar al màxim la resolució de la recollida. Si la difracció del cristall derivatitzat fos de molt bona qualitat i gran resolució seria un pas prescindible, però habitualment el cristall derivatitzat té una difracció molt pitjor que la del cristall natiu (Taula 3.2).

Dades de la recollida	
Distància del detector (mm)	179
Angle de rotació (°)	0.5
Graus recollits (°)	183
Longitud d'ona (Å)	0.934

Taula 3.2: Paràmetres de recollida de dades del cristall natiu de proteïna p4

### **3.4.1.3. Cristall del complex p4-DNA**

Es van recollir dades de difracció de raigs X del cristall del complex a la línia ID13 del ESRF. Aquesta línia té unes característiques especials en quant a la seva òptica permetent focalitzar molt els raigs X. S'anomena microfocus. La línia es molt brillant i el feix molt prim, permetent irradiar en una superfície molt petita del cristall. Això permet obtenir dades de difracció de cristalls molt petits que ens seria impossible d'obtenir en una línia convencional. Per contrapartida, el dany per

radiació que pateix el cristall és molt major, obligant a irradiar el cristall en diversos llocs per completar la recollida de dades. Es va recollir en 9 segments diferents solapant 2° entre segments per facilitar l'escalat de les dades (Taula 3.3).

	Dades de la recollida
Distància del detector (mm)	170
Angle de rotació (°)	0.5
Graus recollits (°)	200
Longitud d'ona (Å)	0.978

Taula 3.3: Paràmetres de recollida de dades del cristall del complex p4-DNA

### ***3.4.2. Dany per radiació***

L'energia alliberada per absorció dels raigs X en el cristall inevitablement el malmet. El procés de dany per radiació incrementa el desordre del cristall i redueix la intensitat de la dispersió.

No hi ha cap criteri definit per decidir quan un cristall està tan deteriorat per ser descartat. Però si les mesures han de ser de gran qualitat, qualsevol canvi observable ens farà plantejar aturar la recollida de dades. L'efecte més seriós ocorre en la part del patró de difracció de més alta resolució, on les intensitats observades de les reflexions de Bragg es veuen alterades més ràpidament.

El primer efecte observable del dany per radiació és normalment una reducció de les intensitats d'alt angle degut a un augment en el desordre. És usual aturar les mesures quan les intensitats de les mesures d'alta resolució es redueixen significativament.

Un cop a començat, el dany per radiació continua a una taxa fixa, encara que l'exposició s'aturi. Això és degut als efectes, relativament lents, iniciats per radicals lliures produïts pels raigs X en la mostra cristal·lina.

### ***3.5. Processament de les dades de difracció***

El conjunt de imatges obtingudes en la recollida de dades difracció de raigs X s'ha de processar per tal d'obtenir un fitxer amb les reflexions. Cada reflexió prové d'un pla de difracció i té una intensitat pròpia que depèn de la quantitat de densitat electrònica que conté el pla. Aquest fitxer de reflexions conté les intensitats corresponents a cada pla i és necessari per fer qualsevol càlcul cristal·logràfic.

#### ***3.5.1. Indexat, integració i escalat***

El processament de les dades el podem dividir en tres etapes: indexat, integració i escalat. El procés de indexat consisteix en determinar el grup espacial, l'orientació del cristall i la cel·la unitat a partir d'unes poques imatges, de la mateixa manera que s'ha descrit anteriorment a l'hora de determinar els paràmetres del cristall per definir l'estratègia de recollida de les dades. Això ens permetrà assignar a cada imatge una orientació, i a cada reflexió el seu índex de Miller hkl.

El següent pas és la integració, que consisteix en mesurar la intensitat de cada reflexió particular, tot ajustant la distribució de intensitat a una gaussiana. Cada reflexió s'envolta d'una caixa que s'utilitzarà per calcular el soroll de fons associat a cada reflexió i a cada mesura de intensitat. Això es fa degut a que el soroll de fons no és homogeni en tota la superfície de la imatge i que en determinades zones de l'imatge s'observen ombres o difracció difusa produïdes pel solvent desordenat del cristall. Tot el conjunt de mesures es recull en un fitxer que conté, entre d'altres, els índex de cada reflexió, amb la seva intensitat integrada i la seva desviació estàndard. Tots aquests paràmetres s'utilitzaran en el següent pas, que és l'escalat de les dades.

L'escalat tracta de promitjar totes les reflexions repetides recollides i de reconstruir les que estan disperses en varies imatges a causa de la mosaïcitat. Petites variacions en el temps d'exposició, la intensitat del feix o pel dany que pateix el cristall al llarg

de la recollida, fa que calgui aplicar un factor d'escala entre reflexions recollides en diferents imatges. Si estem recollint un experiment de tipus MAD, durant l'escalat s'han de tenir en compte els parells de Bijvoet, ja que en aquest cas no es complirà la llei de Friedel i les reflexions oposades no tindran la mateixa amplitud.

Les dades de la proteïna sola i la proteïna derivatitzada van ser indexades i integrades amb el programa MOSFLM (Leslie, 1991) i promitjades i escalades amb el programa SCALA (Evans, 1993). En el cas del complex p4-DNA, les dades es van indexar i integrar amb el programa XDS (Kabsch, 1993) i es van promitjar i escalar amb el programa XSCALE (Kabsch, 1993).

### ***3.6. Resolució del problema de la fase***

Les reflexions de cada cristall ens donen informació del cristall en l'espai recíproc. Per tal de transformar la informació en l'espai real, en el que vivim, s'ha d'utilitzar la funció de la densitat electrònica que representa la transformada de Fourier entre l'espai real i el recíproc. Per resoldre aquesta equació necessitem les amplituds dels factors d'estructura i les seves fases. Durant la recollida de les dades només som capaços de mesurar les intensitats de les ones, fàcilment convertibles a amplituds, però no la fase en la que es troben, que no és mesurable experimentalment.

Per poder estimar la fase de cada factor d'estructura existeixen diferents mètodes, que s'utilitzen depenent de la informació prèvia que es disposa, per exemple un model, o del tipus d'experiment que s'ha realitzat en la recollida de les dades. La resolució del problema de la fase suposa un pas, generalment crucial, en la resolució de l'estructura.

#### ***3.6.1. Resolució del problema de la fase per dispersió anòmala: el mètode MAD***

La dispersió anòmala és un fenomen que ocorre quan la freqüència dels raigs

X és propera a la freqüència d'un electró. En aquest moment l'electró absorbeix energia del raig X, experimenta una transició quàntica i reemet l'energia com a fotó amb la mateixa longitud d'ona, però amb la fase alterada.

Aquest fenomen és negligible per àtoms lleugers, incloent els de la majoria de macromolècules orgàniques, però els àtoms amb un número atòmic més gran, l'efecte és més significatiu, fins i tot molt evident, com en el cas d'àtoms realment pesants, com l'urani. Aquest efecte de dispersió anòmala té com a conseqüència que  $F_{hkl}$  i  $F_{-h-k-l}$  no tinguin el mateix valor d'amplitud, produint-se una violació de la llei de Friedel.

En un experiment de tipus MAD, com s'ha explicat anteriorment, recollim diferents conjunts de dades, normalment tres, cada un a una longitud d'ona diferent i les anomenem 'pic', 'inflexió' i 'remota'. Aquests valors de longitud d'ona provenen del càlcul de la primera ( $f'$ ) i segona ( $f''$ ) derivada de la funció de l'espectre de fluorescència (XANES) del cristall. A la longitud d'ona 'pic',  $f''$  té el valor més gran possible, a la longitud d'ona 'inflexió', la part normal del factor de dispersió atòmica  $f'$  és reduït al mínim valor possible. A la longitud d'ona 'remota', significativament més curta que les altres,  $f'$  és propera al seu valor *normal*, mentre que la component anòmala  $f''$  pot ser una mica significant.  $f'$  representa la contribució no anòmala (de la proteïna i dels àtoms pesants) i  $f''$  l'anòmala (només dels àtoms pesants).

Els canvis en les amplituds dels factors d'estructura sorgits de una dispersió anòmala són generalment petits i fer servir aquestes diferències requereix una acurada mesura de les intensitats. L'experiment MAD va sorgir com una alternativa dels experiments de tipus MIR, on es calculen les diferències isomorfes entre varis cristalls que contenen àtoms pesants. La tècnica MIR presenta diversos problemes: no-isomorfisme entre cristalls (canvis en la cel·la unitat, reorientacions de la proteïna, canvis conformacionals, canvis en els ions de sal i els de solvent), problemes en localitzar tots els àtoms pesants, problemes per refinar les posicions dels àtoms pesants, ocupàncies i paràmetres tèrmics i errors en la mesura de les intensitats. L'ús de la tècnica de MAD supera molts d'aquests problemes per la utilització d'un



sol cristall.

### **3.6.1.1. El programa SOLVE**

La cerca de les posicions de l'àtom pesant incorporat en la retícula cristal·lina es va realitzar amb el programa de fasejat automàtic SOLVE (Terwilliger, 1999). Aquest programa utilitza de manera iterativa la funció de Patterson, amb coeficients corresponents als quadrats de les diferències de les amplituds de Bijvoet ( $|F^+| - |F^-|$ )<sup>2</sup>, per trobar pics corresponents als parells d'àtoms pesants i calculant la diferència de Fourier obté un *Z-score*. La iteració consisteix en utilitzar la millor solució en el pic de Patterson per crear noves solucions a partir d'ella (*seed*), fins que es proven totes i s'escull el *Z-score* màxim (Terwilliger, 2002).

Durant el procés de cerca de posicions és important triar acuradament la resolució de treball, ja que a resolucions més altes hi ha més informació però menor qualitat en la difracció, fent menys significatives les diferències trobades en els parells de Bijvoet. Aquestes posicions d'àtom pesant són utilitzades per fer el càlcul de la fase.

### **3.6.1.2. Millora de les fases i modificació de la densitat electrònica**

Poques vegades les fases determinades experimentalment són suficientment acurades per donar-nos un mapa de densitat electrònica plenament interpretable. Les fases experimentals són sovint el punt de partida per la millora de les fases utilitzant diversos mètodes de modificació de la densitat, que estan basats en algun coneixement previ de l'estructura. El *solvent flattening*, *histogram matching* i el promitjat no cristal·logràfic són les tècniques majoritàries usades.

El *solvent flattening* és una tècnica molt potent que elimina la densitat electrònica negativa i dona a les zones de solvent el valor de densitat electrònica típic de 0.33 e Å<sup>-3</sup>, en lloc de la densitat electrònica típica de les proteïnes de 0.43 e Å<sup>-3</sup>. Els mètodes

automàtics s'utilitzen per definir la interfase proteïna-solvent (Wang, 1985; Leslie, 1988). El *histogram matching* altera valors puntuals de densitat electrònica per que estiguin en concordança amb uns valors de densitat electrònica que compleixen una distribució esperada per cristalls de proteïnes. El promitjat de simetria no cristal·logràfica imposa una equivalència en els valors de densitat electrònica quan més d'una molècula esta present en la unitat asimètrica.

La modificació de densitat sovint és un mètode cíclic, implicant transformacions del mapa de densitat electrònica modificat per obtenir fases modificades, recombinant aquestes fases amb les experimentals i calculant un nou mapa que és novament modificat, i així continua el cicle fins que convergeix. Aquests mètodes també poden ser utilitzats per estendre fases, assumint que tenim dades natives a major resolució. Les tècniques de modificació de la densitat no tornen bo un mapa dolent, però certament milloren mapes prometedors que mostren algunes característiques interpretables.

### ***3.6.1.3. Extensió de fases***

Quan utilitzem fasejat per la tècnica de MAD o altres mètodes, a vegades ens podem trobar que la resolució a la qual tenim bones dades per poder fer el càlcul inicial de fases no és la resolució màxima, i això suposa haver d'estendre les fases a reflexions de més alta resolució a partir de fases a menor resolució. Aquest és un procés iteratiu que utilitza les tècniques de modificació de densitat, ja sigui *histogram matching*, *solvent flattening*, simetria no cristal·logràfica, mitjançant l'ús de màscares i millora d'aquestes.

Durant el procés iteratiu s'utilitza la màscara per calcular les Fc i es comparen amb les experimentals en un rang de resolució lleugerament superior al de les fases obtingudes. Si la correlació és bona es calculen les fases en un rang de resolució una mica més alt.

#### **3.6.1.4. Simetria no cristal·logràfica**

Quan la proteïna té varies regions idèntiques, forma estructures quaternàries o es troba present en la unitat asimètrica en més d'una unitat, és possible obtenir una operació de simetria que les relacioni dins de l'unitat asimètrica. Aquesta operació de simetria no pertany al grup espacial, per això s'anomena simetria no cristal·logràfica. La major diferència de la simetria no cristal·logràfica és la seva naturalesa local. Un operador de simetria cristal·logràfica es pot aplicar a qualsevol punt de la cel·la unitat per trobar un punt equivalent. Un operador de simetria local només relaciona un volum específic amb un altre volum específic. Normalment aquest volum correspon a una cadena polipeptídica i l'operació de simetria local descriu com aquesta cadena està relacionada amb una altra.

Aquesta simetria ens pot ser útil per resoldre l'estructura en les seves diferents etapes. Si utilitzem la funció de autorotació (*self-rotation*) podem identificar pics no corresponents als eixos de simetria, que corresponen a les relacions entre les molècules dins de l'unitat asimètrica. En el procés de modificació de densitat, aquestes relacions ens seran útils per promitjar les densitats electròniques i disminuir el soroll de fons. Durant l'afinament de l'estructura, la simetria no cristal·logràfica ens permet introduir una restricció més, disminuint el nombre de graus de llibertat i de paràmetres a afinar, molt útil sobretot en els primers passos de l'afinat.

#### **3.6.1.5. El programa RESOLVE**

El programa RESOLVE (Terwilliger, 2002) identifica regions de simetria no cristal·logràfica i les regions de solvent per fer una modificació de la densitat electrònica i una posterior modificació de les fases. La modificació de la densitat es realitza per una actualització, en diferents cicles, de la distribució de probabilitats de fase per cada reflexió.

La distribució de probabilitat de fase per una reflexió és el producte de la

distribució obtinguda en l'experiment i la distribució de la funció de probabilitat del mapa. Aquesta funció de probabilitat del mapa és la suma de la certesa de densitat electrònica en cada punt de l'unitat asimètrica.

La simetria no cristal·logràfica és utilitzada per promitjar zones de densitat electrònica equivalents en la mateixa unitat asimètrica, reduint el soroll de fons i restringint les zones de solvent.

### ***3.6.2. Resolució del problema de la fase amb un model per homologia: el reemplaçament molecular i el programa AMoRe***

El reemplaçament molecular (Huber, 1965; Rossmann, 1962) és la tècnica més emprada actualment a l'hora de resoldre el problema de la fase. A mesura que el nombre d'estructures resoltes ha anat creixent, el nombre de nous plegaments o *foldings* nous redueix el seu ritme de creixement. El reemplaçament molecular és capaç de determinar l'orientació i posició d'una proteïna en el cristall a partir d'un model tridimensional d'una altra proteïna, de dominis homòlegs d'una altra proteïna, d'un model de ressonància magnètica nuclear, de models generats per homologia, o d'un model de microscòpia electrònica.

Consisteix en la cerca sistemàtica per correlació entre les intensitats difractades i les que produiria el model, tractat com un cos rígid, en successius empaquetaments en la cel·la cristal·logràfica. Es tracta d'un problema de 6 dimensions per molècula en la unitat asimètrica (tres per la rotació, els angles d'Euler i tres per la translació, les coordenades x,y,z). Aquest problema es pot simplificar separant la cerca de 6 dimensions en dues cerques de 3 dimensions, molt menys costoses computacionalment.

La primera cerca correspon a localitzar la orientació de cada molècula model situada a l'origen de la cel·la. De manera senzilla, es calculen les intensitats per cadascuna de les rotacions i es correlacionen amb les observades. Els valors de

millor correlació correspondran a l'orientació de la molècula en el cristall. Es treballa amb intensitats per que la transformada de Fourier d'aquestes dona lloc a un mapa que és l'autoconvolució de la cel·la unitat. En aquests mapes es mantenen les orientacions de la molècula, però totes les translacions es desplacen a l'origen.

De tots els valors de correlació, corresponents a orientacions de la molècula s'extreuen tantes orientacions com molècules en la unitat asimètrica. Generalment s'observen grups de solucions similars que corresponen a petites diferències en l'orientació, a causa del soroll. El nombre de molècules per unitat asimètrica es pot estimar fàcilment a partir del volum de Matthews ( $V_m$ ). És una llei empírica corresponent a cristalls amb una rang del 30% al 75% de solvent (Matthews, 1968).

Un cop resolt el problema de les orientacions, s'aplica un procediment similar per les translacions. Aquest és més ràpid, per que una translació només implica un canvi en la fase de les amplituds calculades a partir de les rotacions. Es trasllada, tridimensionalment, fins arribar a la millor correlació. Si hi ha més d'una molècula en l'unitat asimètrica, la funció de rotació ens donarà més d'una solució (per la segona, tercera, etc. molècules). Un cop trobada la solució de translació per la primera, podem fixar aquesta i fer una nova cerca de translació per la segona solució de rotació i així successivament.

El major problema de les cerques utilitzant models de baixa homologia és que, un cop s'obté una solució, les fases són una pobra estimació de la fase real, i hi ha una gran desviació cap a l'estructura utilitzada com a model, dificultant l'afinat de l'estructura correcta. En el nostre cas l'avantatge era clar, al disposar no d'un model d'homologia, sinó que posseïem un model de la mateixa proteïna, en un cas en un altre grup espacial i en l'altre, formant un complex amb el DNA. El programa utilitzat en la resolució del problema de la fase va ser el programa AMoRe (Navaza, 1994).

### 3.7. Construcció del model de proteïna

El càlcul de les fases permet resoldre l'equació de la densitat electrònica per obtenir un mapa tridimensional on es representa la densitat electrònica i en el que construirem el model de proteïna de manera que s'hi ajusti de la millor manera possible. Existeixen varis tipus de mapes que ens mostren diferent informació i també existeixen varis programes per construir el model de proteïna, de manera automàtica o manual.

#### 3.7.1. Càlcul del mapa de densitat electrònica

Un cop solucionat el problema de les fases, arriba el moment de visualitzar el mapa de densitat electrònica. El mapa de densitat electrònica és la transformada de Fourier dels factors d'estructura, amb les seves fases corresponents.

La transformada de Fourier (Equació 3.2) és calculada com un conjunt de valors de densitat electrònica a cada punt de la malla tridimensional.

$$\rho(xyz) = \frac{1}{V} \sum_h \sum_k \sum_l |F(hkl)| \cos 2\pi(hx + ky + lz - \phi(hkl))$$

Equació 3.2: Funció de la densitat electrònica definida en cada punt de la cel·la unitat, de coordenades  $x,y,z$ . Representa la transformada de Fourier entre l'espai real, definit per  $\rho(xyz)$ , i l'espai recíproc definit pels factors d'estructura  $F(hkl)$ .

Cal destacar el fet que la densitat electrònica en un punt ve determinada per la informació que contenen tots els factors d'estructura i, a la inversa, cada valor d'un factor d'estructura (mòdul i fase) ve determinat per les posicions de tots els àtoms que formen l'estructura.

La qualitat dels mapes de densitat electrònica depèn de l'exactitud de les dades de difracció i de la resolució. Quan més alta és la resolució, major és el detall al qual observem el mapa.

### ***3.7.2. El traçat de l'estructura***

Es considera que s'ha trobat la solució d'una estructura quan s'obté un mapa interpretable en termes de cadena d'aminoàcids (o dels àtoms que la formen). A partir d'aquest moment comença l'etapa de traçat i afinament.

El traçat consisteix a inspeccionar el mapa i anar situant la cadena polipeptídica al llarg d'ell. Depenent de la qualitat de les dades s'ha de fer gràficament, de manera manual, o es pot fer automàticament si la resolució és prou alta (aproximadament major de 2.5 Å). Generalment es comença introduint elements d'estructura secundària en els mapes inicials. Quan es té un traçat inicial, encara que només sigui aproximat o de certes parts de les cadenes, es calculen mapes de diferències  $2F_o-F_c$  i  $F_o-F_c$ . Aquests mapes s'obtenen fent la síntesi de Fourier amb les diferències entre les amplituds observades ( $F_o$ ) i les calculades amb el material traçat ( $F_c$ ), i com a fases s'utilitzen les disponibles, és a dir, en el cas del reemplaçament molecular les provinents del càlcul de  $F_c$  i, en el cas de l'experiment MAD o reemplaçament isomorf, de les posicions trobades dels àtoms pesants, normalment combinades amb les fases provinents del model parcial ja traçat.

Els mapes  $2F_o-F_c$  tenen un pes doble en l'amplitud observada respecte de la calculada. Els mapes  $F_o-F_c$  ajuden a la interpretació de parts conflictives dels mapes. Generalment són més intensos en les zones on hi ha errors en l'estructura, o aquesta no ha estat traçada encara. Les zones on falten àtoms tindran una densitat positiva ( $F_o > F_c$ ), i les zones on n'hi ha de mal traçats tindran densitat negativa ( $F_c > F_o$ ). En l'estructura final, aquests mapes haurien de presentar densitat propera a zero ja que s'igualen els dos components. Un cop feta la resta, com que s'està treballant amb coeficients propers a zero, els mapes  $F_o-F_c$  són molt més sorollosos que els  $2F_o-F_c$ , pel que és necessari inspeccionar-los a un nivell de sigma més elevat. Generalment

es treballa amb densitats de 1 sigma en els mapes 2Fo-Fc i de  $\pm 2$  ó 2.5 sigma en els mapes Fo-Fc.

Els programes automàtics utilitzen un altre tipus de mapes. Aquests programes treballen amb mapes 3Fo-2Fc, la suma dels dos mapes anteriors, tot i que en el treball manual és més pràctic separar els dos mapes. El treball de traçat i afinament és iteratiu. Entre cada cicle de traçat, manual o automàtic, s'intercalen cicles d'afinament fins que totes les cadenes de la proteïna resultin traçades el millor possible.

#### ***3.7.2.1. Construcció automàtica del model de proteïna amb el programa RESOLVE***

En un pas final, un cop s'han modificat les fases, el programa RESOLVE, i altres com ARP/wARP (Perrakis *et al.*, 1999), poden construir automàticament un model de proteïna que s'ajusti al mapa de densitat electrònica. El primer pas és la detecció de regions d'estructura secundària, tal com hèlix i fulles. El següent pas consisteix en l'ús de una llibreria de fragments de tripèptids per estendre les hèlix i les fulles en cada direcció i connectar els fragments. En el tercer, i últim pas, s'afegeixen les cadenes laterals i s'alinea la seqüència amb el model. L'opció de construcció automàtica del model depèn, en gran mesura, de la resolució màxima a la que es disposen les dades. Resolucions inferiors a 3 Å ens conduiran a construccions de models poc fiables i a assignacions incorrectes en la seqüència.

#### ***3.7.2.2. Construcció manual del model de proteïna***

On no arriba la construcció automàtica del model, hem de ser nosaltres mateixos, que a partir d'un mapa de densitat electrònica calculat a partir de les intensitats mesurades i les fases obtingudes a partir de qualsevol de les tècniques de fasejat disponibles, construir un model que s'hi ajusti. Aquesta construcció es realitza afegint residus un a un dins de la densitat electrònica de manera que quedin tots



els àtoms dins de la densitat electrònica i que el resultat sigui estereoquímicament raonable. També podem introduir fragments sencers d'elements d'estructura secundària, majoritàriament hèlix  $\alpha$ . El programa gràfic utilitzat en aquest treball s'anomena TURBO-FRODO (Rousell *et al.*, 1989), però existeixen alguns d'altres com O (Jones *et al.*, 1991), XtalView (McRee, 1999), COOT (Emsley *et al.*, 2004), etc. igualment vàlids.

### ***3.8. Afnament del model de proteïna***

La finalitat de l'afinament estructural és el d'ajustar el model de manera que obtenim la major correlació amb les observacions cristal·logràfiques. Es mesuren les discrepàncies entre les intensitats de les reflexions calculades a partir del model generat i les intensitats observades experimentalment. L'objectiu és minimitzar aquestes discrepàncies, fent que el model s'ajusti al màxim a les intensitats observades.

El model estructural que s'ha d'afinar està definit, en cada àtom, per les coordenades atòmiques (x,y,z) i un factor de temperatura (B), normalment isotròpic. Les coordenades expressen la posició de l'àtom en les tres dimensions de l'espai de la cel·la unitat. El factor de temperatura expressa la mobilitat tèrmica de l'àtom al voltant de la seva posició d'equilibri. En aquesta etapa s'ajusten els paràmetres estructurals, les coordenades i factor de temperatura, per arribar a un acord entre l'observat i el modelat.

En l'afinament d'estructures macromoleculares a resolució moderada, no hi ha prou observacions, en forma de reflexions, per poder realitzar un refinament de les posicions atòmiques independentment. Degut a la gran precisió en els coneixements estereoquímics dels diferents aminoàcids (distàncies d'enllaç, angles d'enllaç, angles dihedres, planaritat, etc.), es poden introduir restriccions estereoquímiques al model. Llavors l'estructura es defineix en termes de la seva estereoquímica, i aquesta s'utilitza en l'afinat, augmentant les dades i disminuint els graus de llibertat de les variables a afinar.

En els primers estadis d'afinament estructural es sol utilitzar el mètode del *simulated annealing*, basat en les tècniques de Monte Carlo i que equival al procés físic d'escalfar una proteïna fins a una temperatura que normalment està al voltant de 5000K i posteriorment refredar-la lentament. Utilitzant aquesta tècnica eliminem les rigideses imposades en la construcció o les diferències provinents de la solució de reemplaçament molecular, ja que s'assoleixen conformacions de la cadena polipeptídica que tenen una barrera energètica difícilment superable a baixes temperatures, o dit d'una altra manera, atrapades en mínims locals.

En el procés de minimització es tendeix a reduir l'energia del sistema, generalment utilitzant el sistema del gradient conjugat. L'exploració de l'espai conformacional és molt més reduïda que en el *simulated annealing*, i sovint tendeix a mínims locals, però ens permet fer un ajust fi quan les conformacions reals i les modelitzades són properes. També s'afinen altres paràmetres com el factor de temperatura i, de vegades, l'ocupància de cada àtom. Els valors més alts en el factor de temperatura els trobem en regions menys rígides, normalment exposades al solvent o zones actives de la proteïna, presentant diferents conformacions.

L'ocupació representa la proporció en la que un àtom individual es troba present en una posició determinada en totes les cel·les del cristall. Ens podem trobar en casos on hi hagi regions de la proteïna on sigui possible més d'una conformació estable, tant d'un residu com d'un grup de residus i llavors es tracen les conformacions alternatives, assignant una ocupància a cada conformació de manera proporcional. Un altre cas típic és en el cas dels àtoms pesants, ja que si són introduïts per *soaking*, la incorporació de l'àtom pesant no és del 100% i es veurà reflectit en l'ocupància.

En aquest treball s'ha utilitzat el programa CNS (Brünger *et al.*, 1998) per realitzar tot el refinament estructural. Les tècniques més utilitzades en el refinament amb CNS són la minimització d'estructures, el *annealing* o escalfament i l'afinament dels factors de temperatura o *B-factors*.

### **3.9. Avaluació del model de proteïna**

En el moment que considerem que l'afinament del model de proteïna ha finalitzat hem d'avaluar aquest model utilitzant diferents eines. La més utilitzada són els diagrames de Ramachandran (Ramachandran, *et al.*, 1968), però la informació que mostren pertany només a la cadena principal de la proteïna. El programa PROCHECK (Laskowski *et al.*, 1993) incorpora, a part del Ramachandran, altres mesures que avaluen la qualitat estereoquímica del model de proteïna d'una manera més extensa.

Existeix un altre grup d'eines que avaluen components no proteics del model com, en el nostre cas, una molècula d'àcid nucleic. Per avaluar els àcids nucleics disposem de diferents programes com el CURVES (Lavery *et al.*, 1988) i el 3DNA (Lu *et al.*, 2003).

#### **3.9.1. Estereoquímica de la cadena principal: els diagrames de Ramachandran**

Els diagrames de Ramachandran (Ramachandran *et al.*, 1968) són representacions en dos dimensions dels angles  $\phi$  (phi) i  $\psi$  (psi) de cada residu particular (Figura 3.5). En aquesta representació es mostren les conformacions de la cadena principal més favorables energèticament. Aquestes regions corresponen a les hèlix  $\alpha$  i a les fulles  $\beta$ . Aquest diagrama és molt general i mostra les regions favorables comunes a tots els aminoàcids, excepte la prolina, que al tenir la cadena lateral ciclada sobre la principal està totalment restringida conformacionalment.

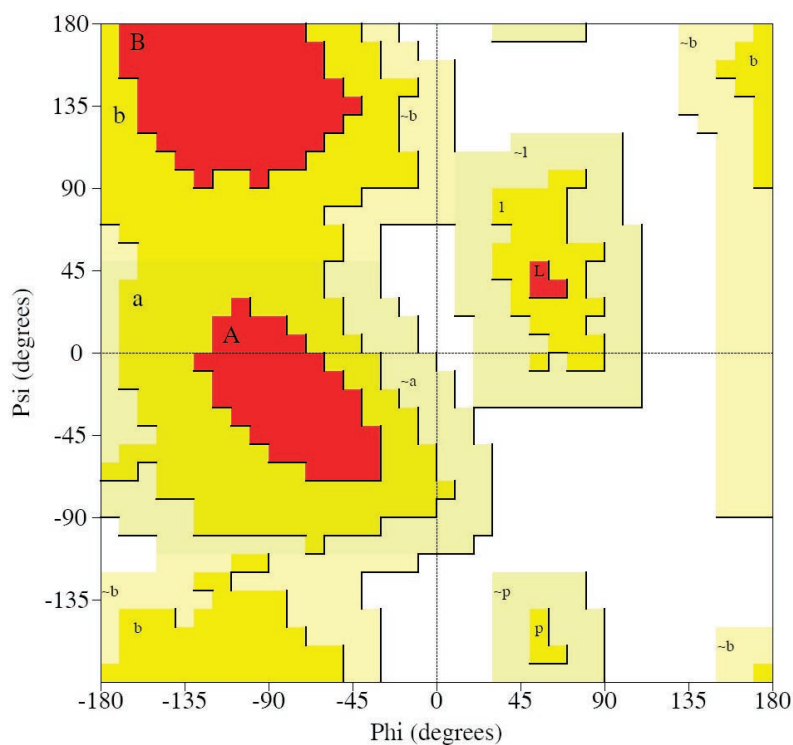


Figura 3.5: Representació del diagrama de Ramachandran on, en vermell hi ha les regions més favorables, en groc i groc pàl·lid les permeses i en blanc les no permeses.

Quan trobem residus en les regions ‘no permeses’ del Ramachandran, llavors els hem de justificar de manera que puguem explicar el seu estat d’alta energia.

### 3.9.1.1. Altres mesures estereoquímiques

Existeixen altres factors que permeten avaluar la qualitat de l’afinament del model de proteïna. Aquests, entre d’altres, són la planaritat de l’enllaç peptídic i la distorsió del tetràedre del carboni alfa. Això ens permet de veure, no si els residus estan en una conformació forçada o poc favorable energèticament, sinó una avaluació de la consistència del model de proteïna des d’una perspectiva química, preocupant-se essencialment per la idealitat geomètrica dels enllaços que formen els diferents àtoms del model proteic.

### 3.9.2. El *R factor*

El factor *R*, o *R factor* és un indicador de com de diferent és la “difracció” del nostre model respecte de la difracció real obtinguda experimentalment. S'utilitza com a mesura de l'afinament del model. Es tracta d'un índex de la desviació entre les amplituds observades ( $F_o$ ) i les calculades ( $F_c$ ). Els factors *R* s'utilitzen com a indicadors de qualitat durant tot el procés d'afinat. Un augment en el factor *R* ens indicarà que el que hem fet durant el cicle d'afinat ha empitjorat la correlació del model amb les dades. Aquest indicador necessita, un control intern, el *Rfree*, que és un subconjunt dins del *Rfactor* que no s'inclou dins del càlcul ni dins de l'afinat. El factor *Rfree* és el paràmetre que realment ens serveix per monitoritzar la qualitat de l'afinament.

Per tal de crear el conjunt de reflexions, que constituïran les utilitzades per calcular el *R free*, seleccionem aleatòriament entre un 5 i un 10% de les reflexions, que seran marcades en el fitxer de reflexions. Amb aquestes reflexions es farà un càlcul de *R factor* a l'igual que l'anterior. Així tenim, per una banda, un *R factor* de reflexions que s'ha utilitzat en l'afinat i un *R factor* de reflexions no utilitzades en l'afinat. Per aquesta raó el *Rfree* té, normalment, un valor lleugerament superior al *R factor*.

El *R factor* pateix una gran variació depenent de la resolució i altres factors. La modelització només aproximada del solvent també té una influència notable en el valor més aviat alt del *Rfactor* en les estructures de proteïnes.

### 3.9.3. El model de DNA

Per avaluar el model de DNA s'han d'utilitzar uns paràmetres i unes eines totalment diferents a les utilitzades pel model de proteïna. Els paràmetres que hem utilitzat en aquest treball són el grau de curvatura global del dúplex de DNA i l'amplada dels dos solcs de la molècula de DNA, el major i el menor. El programa utilitzat és el

programa CURVES (Lavery *et al.*, 1988), però n' existeixen d'altres, com el 3DNA (Lu *et al.*, 2003) i el FREEHELIX (Dickerson, 1998), els dos igualment vàlids.

### **3.10. El fitxer PDB**

Les coordenades dels àtoms de cada model molecular s'emmagatzemen en un fitxer amb extensió .pdb. Aquest fitxer té un format estàndard que permet el seu ús en tots els programes de cristal·lografia i els de representació tridimensional. El fitxer es troba en un format ASCII, fàcilment editable, i que assigna una fila a cada àtom, especificant el tipus d'àtom i el residu al que pertany, les coordenades x,y,z i el seu factor de temperatura. També conté altres entrades, no corresponents amb les coordenades del model, que especifiquen el grup espacial, les dimensions de la cel·la, les condicions de cristal·lització i la publicació derivada del model molecular, entre d'altres.

Existeix un banc de dades mundial (<http://www.rcsb.org/pdb/>) que emmagatzema totes les coordenades dels models de proteïna determinats per diferents tècniques, tant experimentals com teòriques.