

La costruzione del modello molecolare

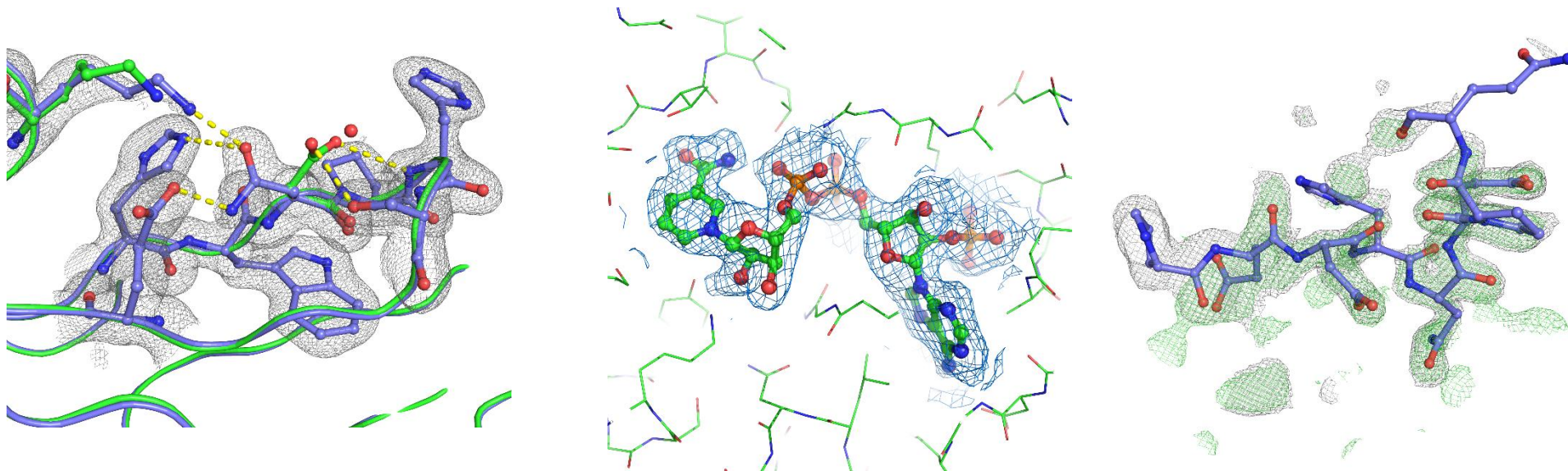
Laurea Magistrale in Biotecnologie Mediche
Curriculum Nanobiotecnologie

A.A. 2020-21

Costruzione di un modello molecolare

Una volta ottenute le **fasi iniziali** dei fattori di struttura o per via sperimentale (derivato isomorfo o MAD/SAD) o per mezzo del Molecular Replacement siamo in grado di calcolare la **prima mappa di densità elettronica** $\rho(r)$.

Lo step successivo è quello di costruire un **modello molecolare coerente con la mappa di densità elettronica**, ovvero con il dato sperimentale.



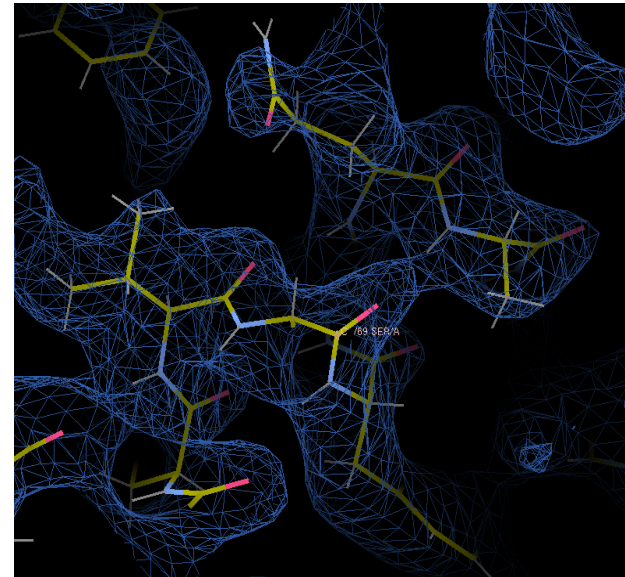
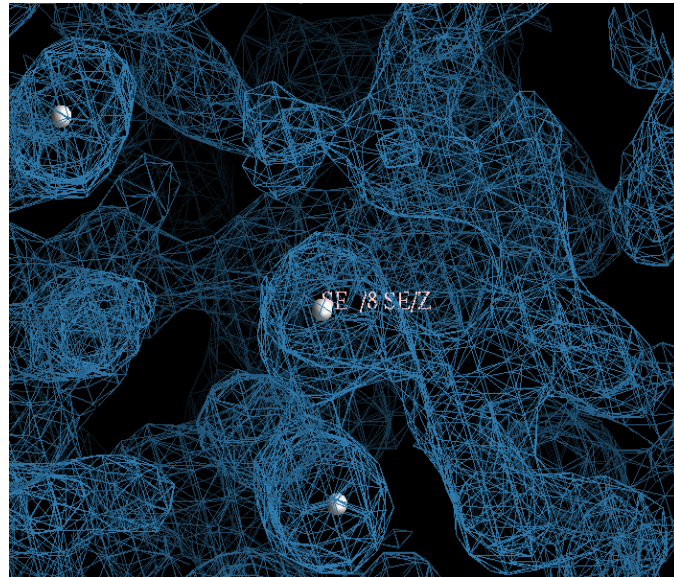
La mappa iniziale

La mappa di densità elettronica iniziale può essere più o meno facile da interpretare, ovvero può essere più o meno complicato 'costruire' un modello molecolare chimicamente sensato a partire dalle mappe.

Le mappe di densità ottenute da MIR o MAD/SAD possono essere complesse e di difficile interpretazione

Le mappe ottenute da Molecular Replacement possono invece essere **piuttosto semplici**, soprattutto se il modello usato ha **omologia elevata** con la proteina target

Fasi SAD
3.2 Å



Fasi MR
2.2 Å

Il modello molecolare

Costruire un modello molecolare significa:

- **Determinare le coordinate x, y, z degli atomi che costituiscono l'unità asimmetrica**
- **Determinare i parametri termici associati agli atomi**

Oltre agli atomi costituenti la macromolecola possono essere presenti

- **Atomi di acqua che coordinano con la proteina (*shell* di idratazione)**
- **Costituenti del buffer, della soluzione precipitante o del crio-protettore**
- **Ligandi introdotti volontariamente (per esempio inibitori di un enzima)**

La possibilità di ottenere un modello molecolare quanto più dettagliato dipende dalla qualità dei dati e in ultima analisi dalla risoluzione della struttura.

Quanto più è elevata la risoluzione tanto più sarà elevato il dettaglio del modello molecolare che possiamo costruire!

Miglioramento delle mappe di densità 'sperimentali'

Le mappe ottenute con metodi di fase sperimentale (MIR, MAD/SAD) sono affette da errori sperimentali più o meno grandi quali:

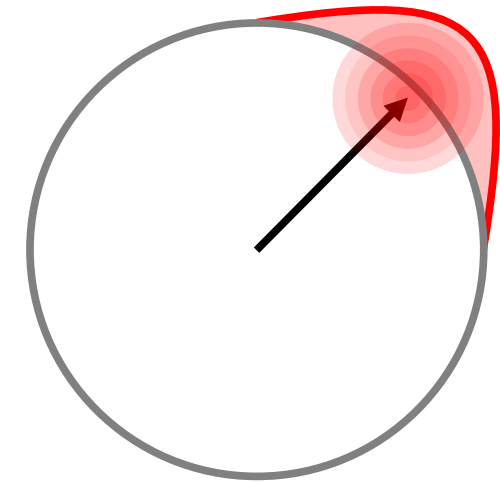
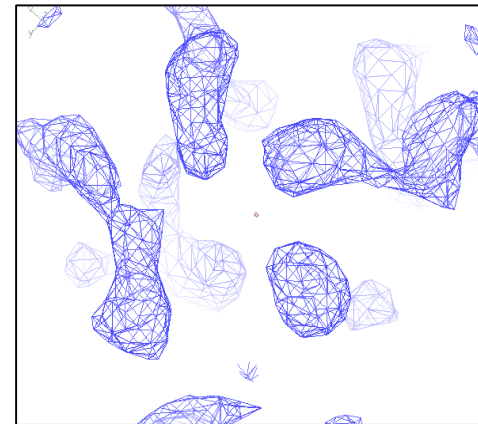
- Non isomorfismo
- Danno da radiazione

Questi errori influenzano i valori delle **fasi** determinate e in ultima analisi la qualità delle mappe di densità elettronica.

Le mappe ottenute da Molecular Replacement possono soffrire di '**bias**' ovvero di un **effetto memoria** della molecola usata come modello

Per migliorare la qualità delle mappe, bisogna migliorare le fasi (**Phase Improvement**).

Il modo più efficace di migliorare le fasi, e quindi le mappe, va sotto il nome generico di **Density Modification**



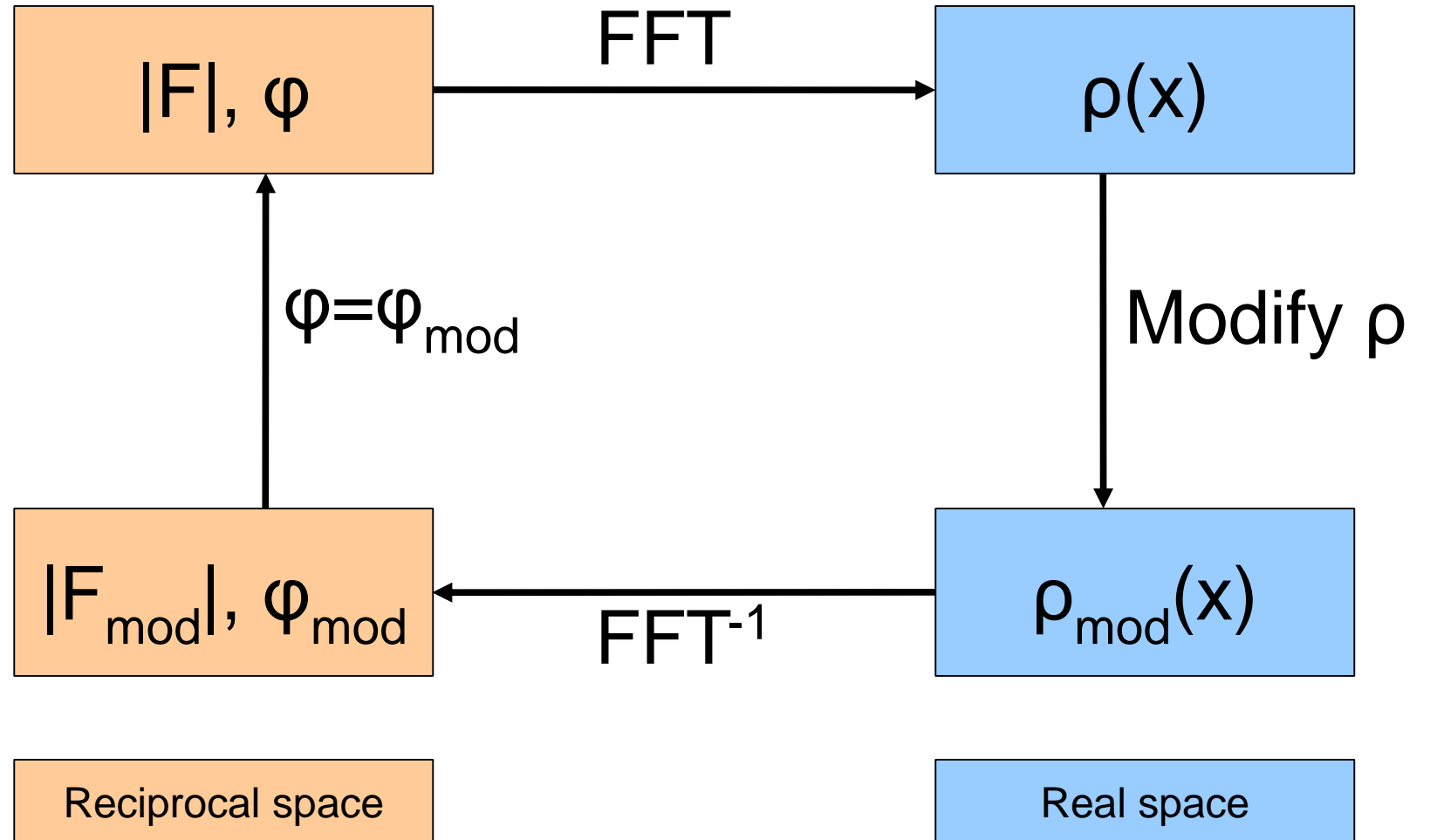
In virtù degli errori sperimentali le fasi sono espresse con una loro probabilità (di essere corrette)

Density Modification

Density Modification

L'idea di fondo della density modification consiste nel modificare la mappa di densità, **sulla base di proprietà note delle mappe medesime**, quindi contro-trasformare la mappa modificata in **nuove fasi e fattori di struttura** modificati.

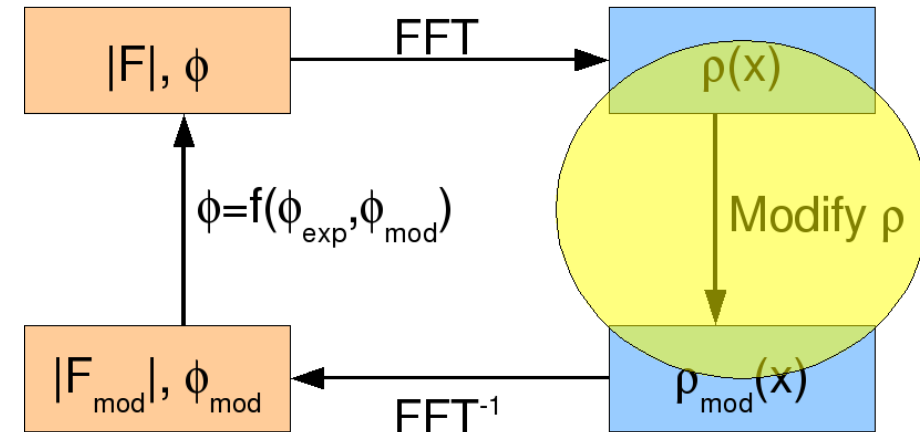
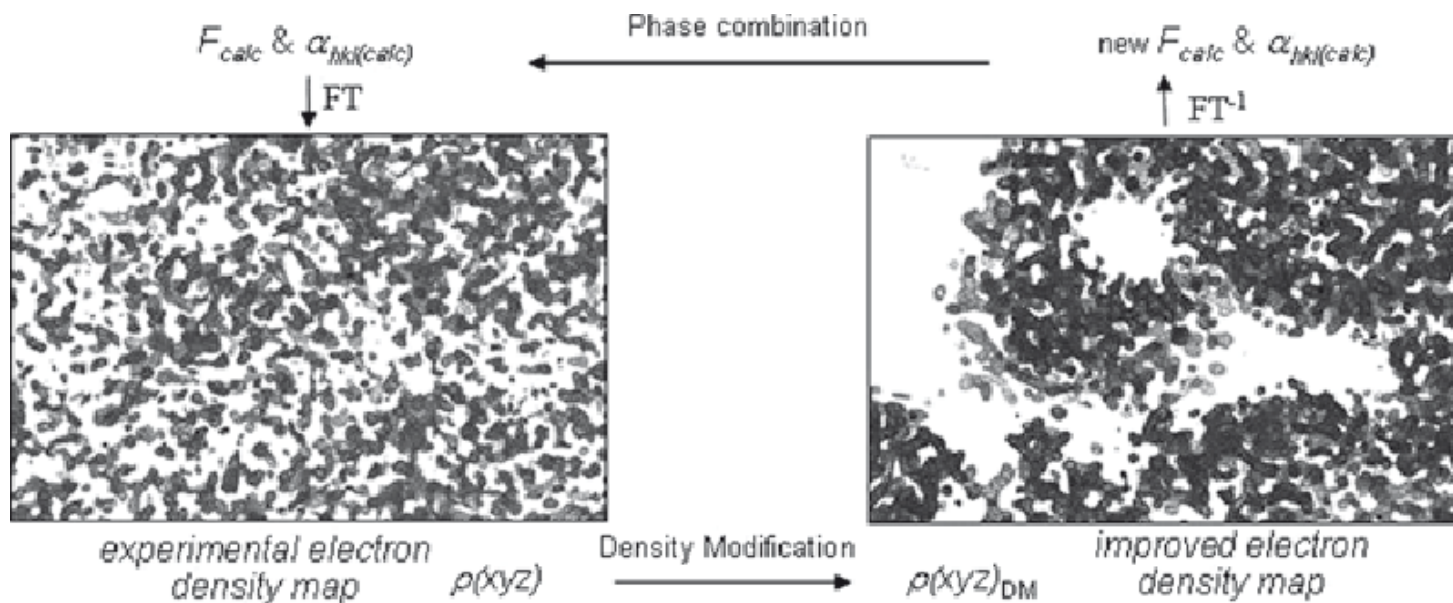
Il processo è ripetuto in modo iterativo fino a convergenza.



Density Modification

La Density Modification è una **metodologia ibrida** in cui si combinano sia **informazioni presenti nello spazio reale** (**vincoli sulla densità elettronica**) che **quelle nello spazio reciproco** (F_{hkl} e ϕ_{hkl}).

Le informazioni nello spazio reale sono utilizzate per modificare le fasi e i fattori di struttura nello spazio reciproco.



Density Modification

Con il termine **Density Modification** si intende una varietà di metodi che introducendo dei vincoli (constraints) sulla densità elettronica nello spazio reale, e che quindi determinano una modifica di F_{hkl} e φ_{hkl} nello spazio reciproco.

Sono utilizzati principalmente tre metodi

- **Solvent Flattening**
- **Histogram Matching**
- **Non-Crystallographic Symmetry (Averaging)**

Il Solvent Flattening e l'applicazione della Non-Crystallographic Symmetry sono particolarmente efficaci

Solvent Flattening

Solvent Flattening

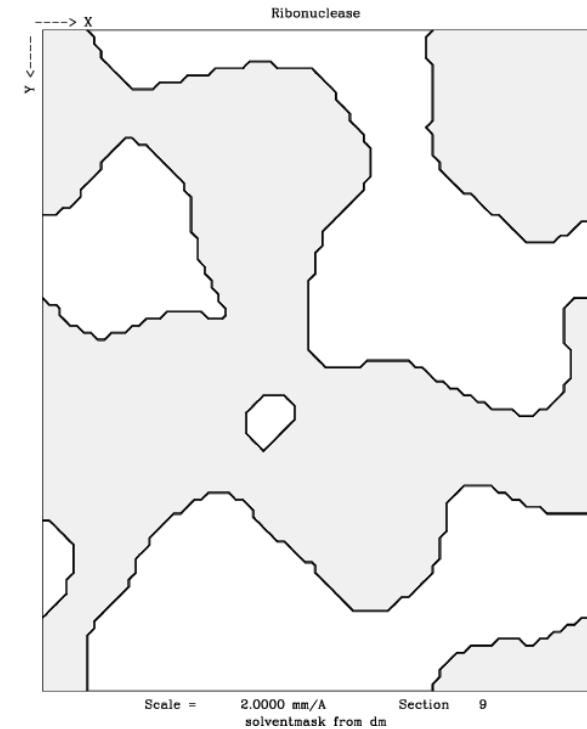
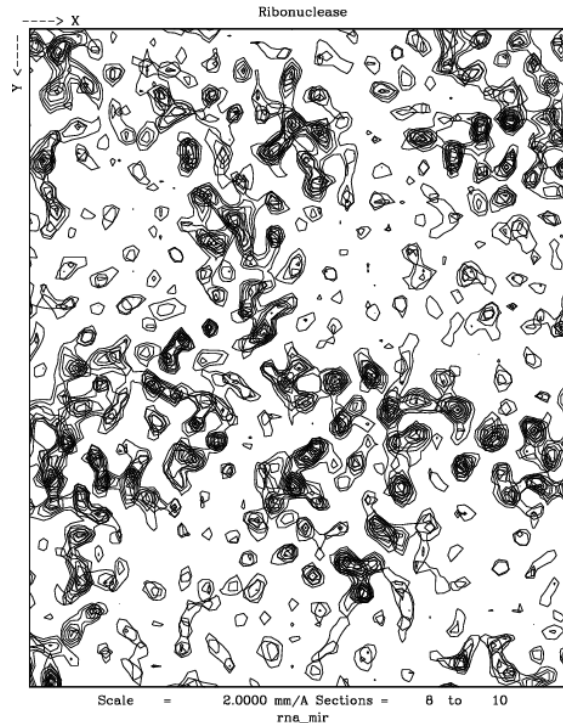
L'idea di base è molto semplice: in un cristallo di proteine una larga parte del volume del cristallo è riempito dal solvente.

Come prima approssimazione si può assumere che lo spazio occupato dal solvente abbia una densità elettronica costante: $\langle \rho_{\text{solv}}(r) \rangle = 0.33 \text{ e}^-/\text{\AA}^3$

Nel solvent Flattening, a tutto il volume attribuito al solvente viene imposto un valore di densità pari a $\langle \rho_{\text{solv}}(r) \rangle$.

La densità così modificata viene contro-trasformata per dare dei F_{hkl} e φ_{hkl} modificati.

La mappa di densità risultante dovrebbe risultare più 'pulita' se le zone di solvente sono state correttamente assegnate.



Solvent Flattening

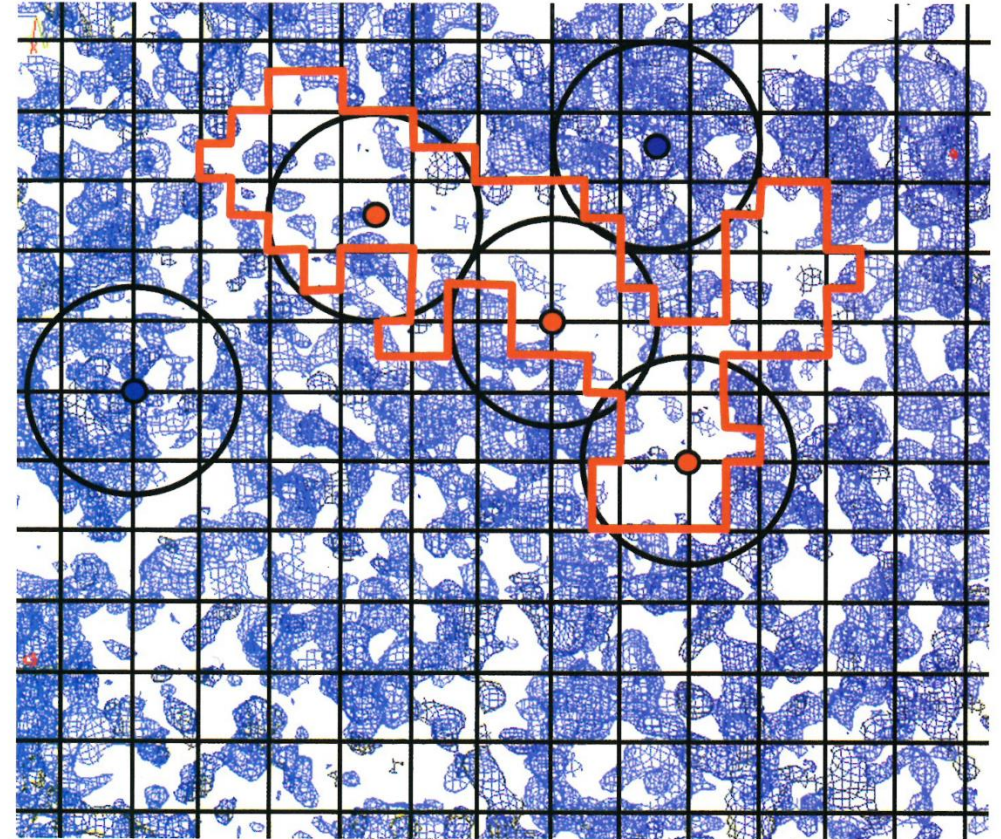
Partendo da una mappa sperimentale può non essere facile discriminare tra zone occupate dalla proteina e zone occupate dal solvente.

La cella viene divisa in una griglia tridimensionale e per ogni elemento della griglia (voxel) viene calcolato il contenuto medio di densità.

Sulla base del contenuto medio di densità nell'elemento di griglia e di un valore di soglia pre-definito, **ogni elemento della griglia è assegnato o alla proteina o al solvente.**

Viene quindi costruita una maschera (*solvent mask*), dove elementi contigui vanno a costruire alternativamente le regioni di solvente e le regioni della proteina.

Figure 10-24 Principle of mask generation. A 3-dimensional grid somewhat coarser (exaggerated in this drawing) than the map grid is placed over the initial protein map. At each grid point, the average electron density within a sphere of a given radius (shown as circles) is computed. If the density is above a certain cutoff value (circles with a blue center), the grid point is deemed to be in protein and assigned the corresponding average density ρ_{ave} . If the value is below the cutoff, the grid point is assigned to the solvent region (red-centered circles). If the map is of redeemable quality, contiguous solvent and protein regions will be found.

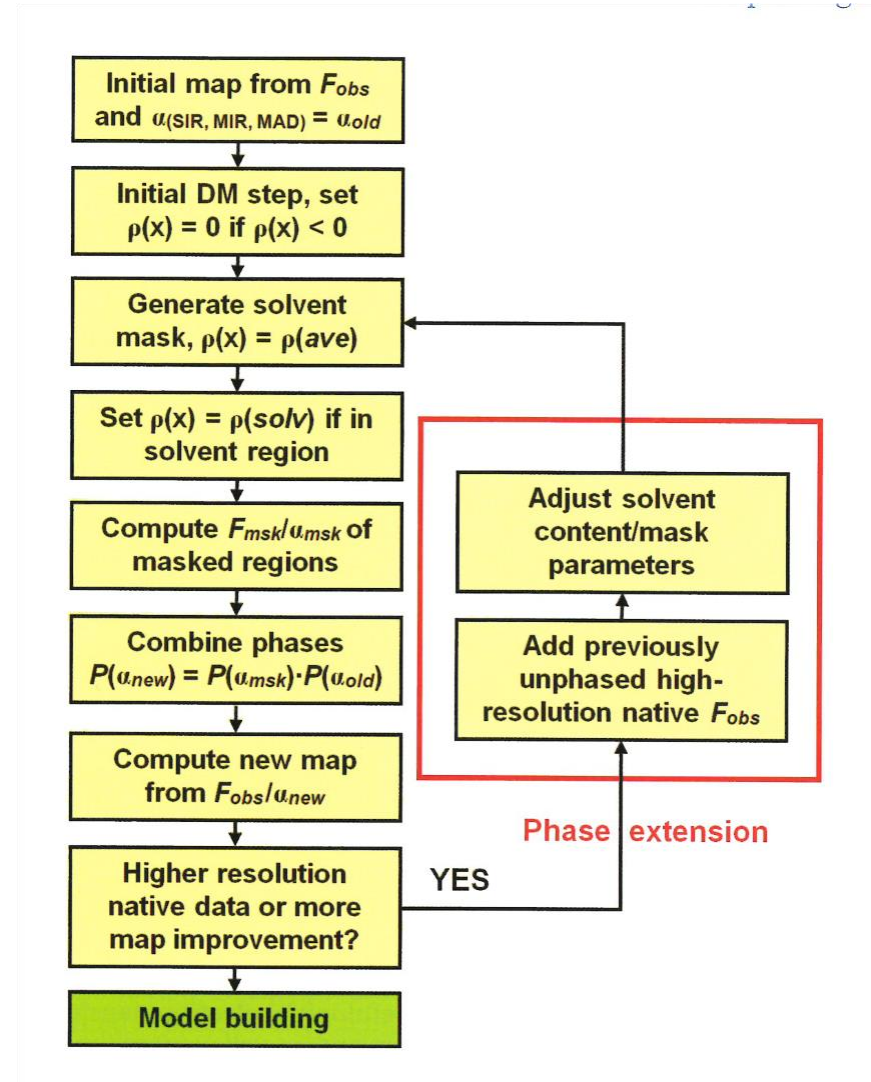


Solvent Flattening

Poiché il Solvent Flattening può contribuire a rendere la $\rho(r)$ più 'pulita' e a migliorare le fasi, anche la determinazione della solvent mask può diventare più chiara.

Il procedimento è ripetuto iterativamente per un certo numero di cicli.

In molti casi la procedura è effettuata per intervalli di risoluzione crescente (le fasi a risoluzione più alta sono maggiormente affette dagli errori sperimentali).



Histogram Matching

Histogram Matching

L'Histogram matching deriva dai metodi di trattamento delle immagini.

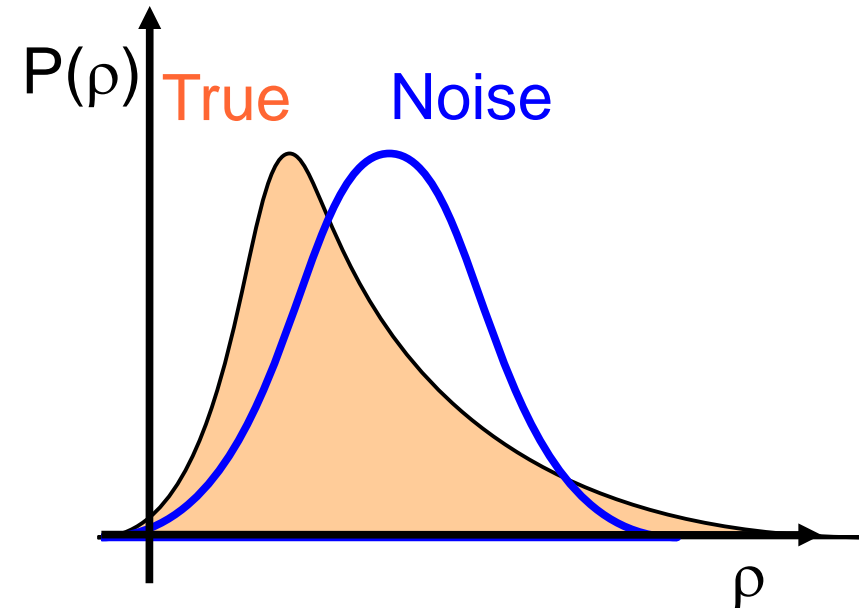
Se dividiamo la nostra cella con una griglia tridimensionale, ogni elemento di volume (voxel) avrà una certa 'intensità' che è il valore della $\rho(r)$ per l'elemento di volume.

Si può creare un **istogramma (distribuzione) dei valori $\rho(r)$** .

Nel caso di una mappa costituita da rumore l'istogramma sarà sempre descritto da una gaussiana.

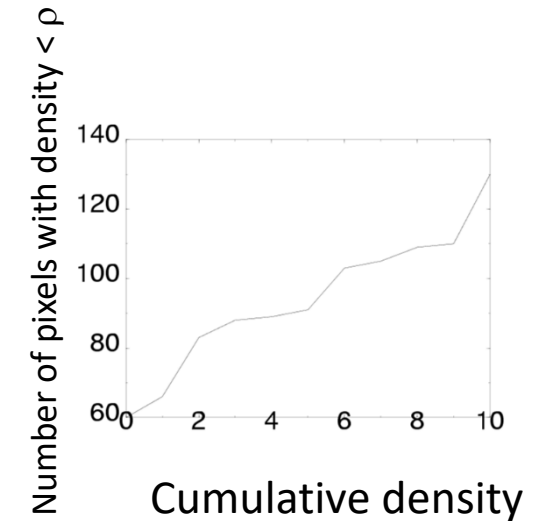
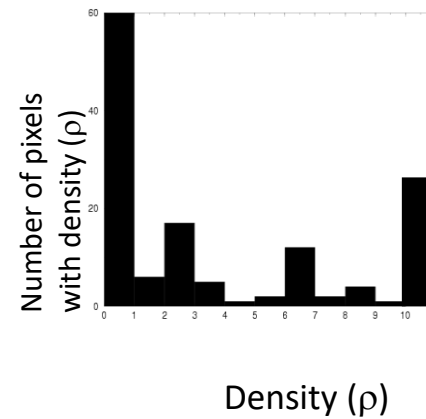
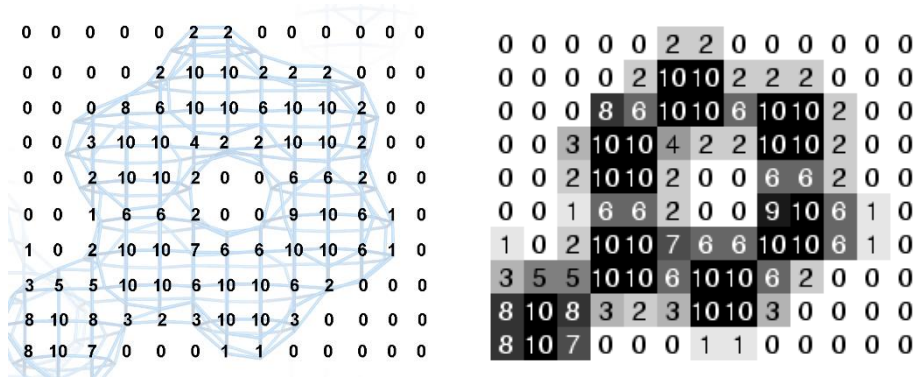
Nel caso di una mappa costruita a partire da fasi corrette, la distribuzione sarà asimmetrica (skewed) con una forma ben precisa.

La forma dell'istogramma è sostanzialmente uguale per qualsiasi mappa con fasi corrette (alla stessa risoluzione).



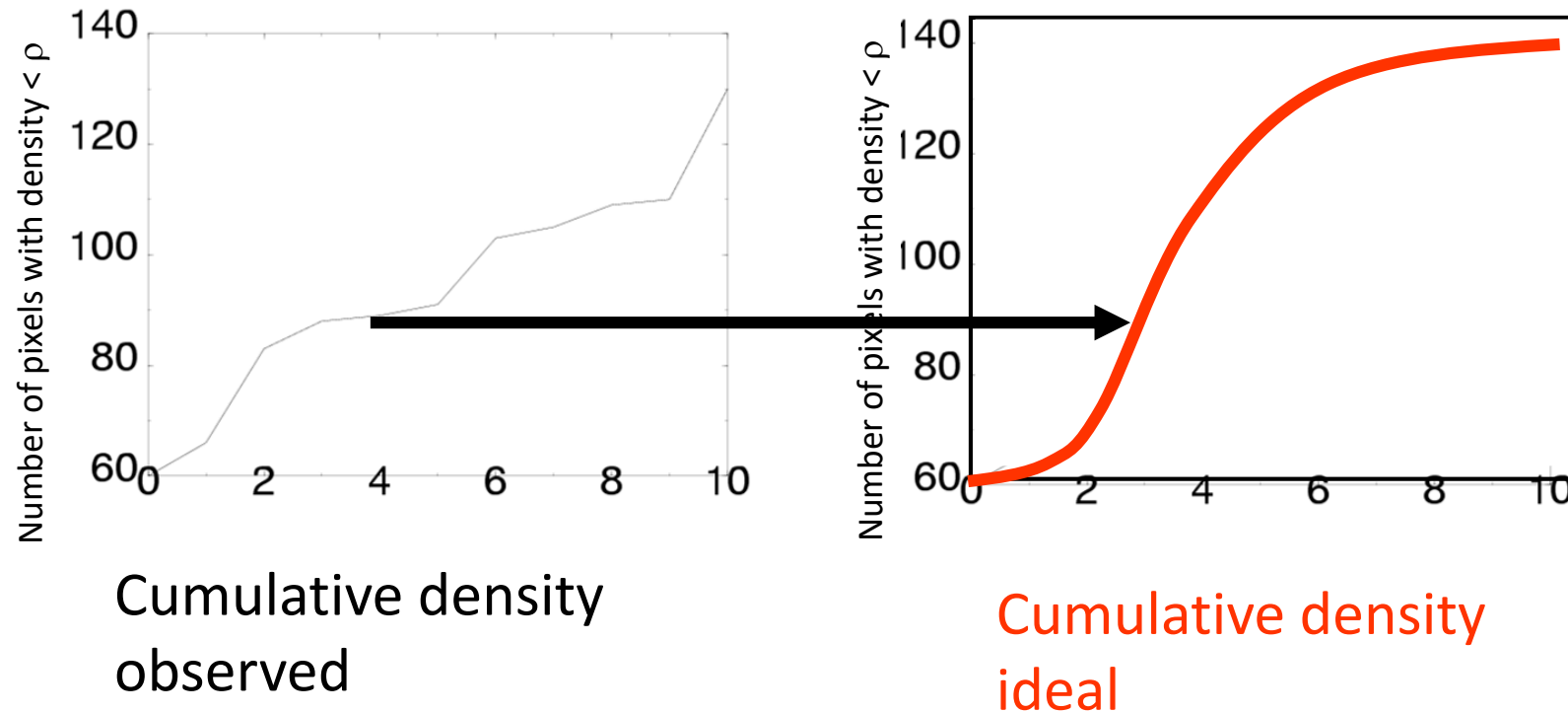
Histogram Matching

In una struttura con fasi ancora approssimate si interviene sul valore della $\rho(r)$ in modo che l'istogramma della densità abbia la medesima forma osservata per strutture cristallografiche ben determinate e con fase corretta



Histogram Matching

Pixels con $\rho=4$ nella mappa originale saranno modificati in $\rho= 2.5$ nella mappa modificata per essere in accordo con l'istogramma previsto



Simmetria Non Cristallografica

Simmetria Non-Cristallografica

La simmetria cristallografica stabilisce una relazione esatta tra molecole 'simmetriche'. Nel cristallo, le coordinate delle molecole collegate da relazioni di simmetria cristallografica obbediscono a relazioni matematiche esatte e ben determinate (dal gruppo spaziale).

Due molecole collegate da relazione di simmetria cristallografica sono esattamente uguali.

Esiste però un secondo tipo di simmetria, detta **non-cristallografica** (Non-Crystallographic Symmetry: **NCS**), in cui le molecole simmetriche **non sono perfettamente uguali**, ovvero **non sono collegate da una relazione matematica esatta**, ma solo approssimata.

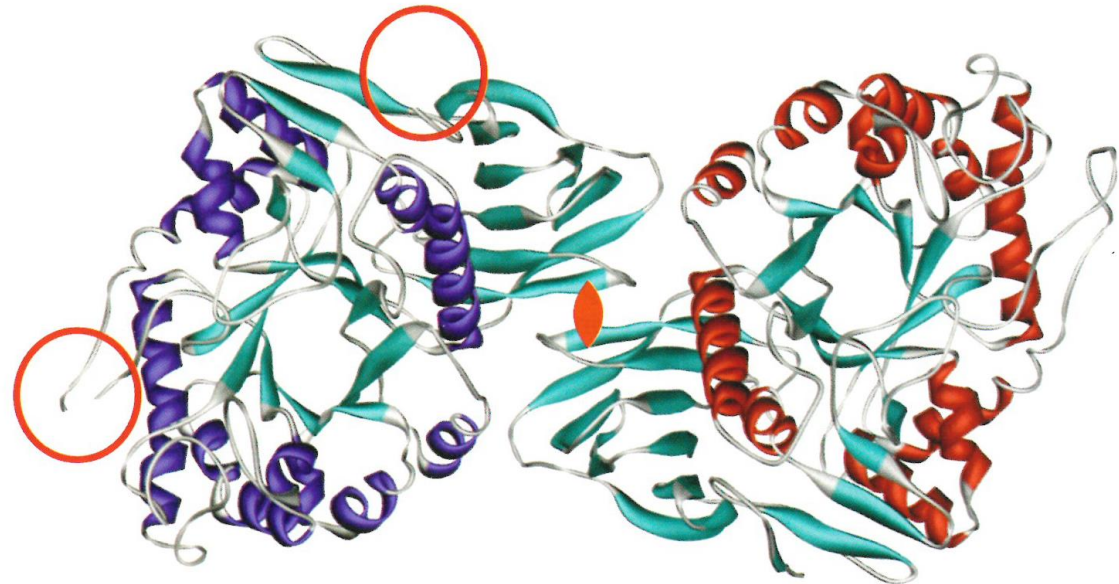


Figure 5-19 A non-crystallographic homo-dimer. The two copies of a molecule related by a 2-fold NCS axis generally are not quite identical: The molecule to the left has lost a part of its N-terminus and the extended loop is not traced (red circles), perhaps because of increased flexibility and an absence of electron density. In general, there are many small differences between the two molecules of NCS-related dimers. To emphasize the differences between the molecules clearly, the color of the helices has been changed. The NCS pseudo-dyad is indicated by the red dyad symbol.

Simmetria non-cristallografica

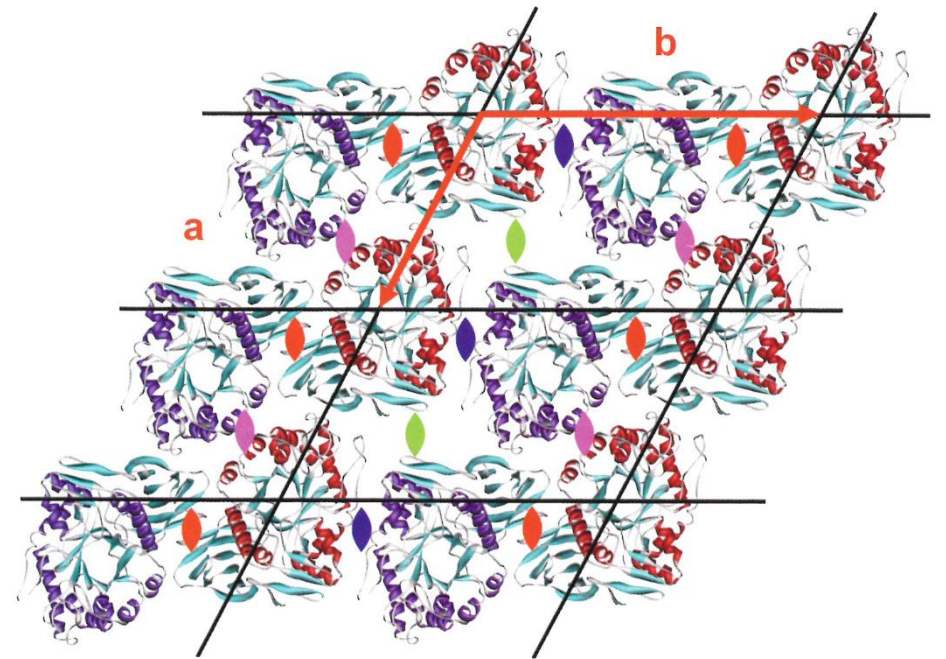
Gli elementi di simmetria cristallografici possono essere presenti solo in determinate posizioni della cella unitaria, in accordo con il gruppo spaziale (le loro posizioni sono riportate per ogni gruppo spaziale, nelle Tabelle Internazionali).

Gli elementi di simmetria non-cristallografici possono invece essere presenti in qualsiasi punto della cella unitaria.

La simmetria non-cristallografica (NCS) non stabilisce relazioni esatte tra molecole, ma solo approssimate. **Sono** elementi di simmetria locali

La deviazione quadratica media (rmsd) tra le posizioni dei C α di due molecole sovrapposte a seguito dell'applicazione della NCS, fornisce un'idea di quanto sia rispettata la simmetria locale

Figure 5-20 Non-crystallographic dimeric symmetry in a p1 structure. In analogy to the p2 case depicted in Figure 5-10, additional non-crystallographic symmetry axes (dyad symbols in colors other than red) have been created from the original dimer axis (red) by translational lattice stacking. The single asymmetric unit of the p1 unit cell now contains two molecules—one entire NCS dimer forms the single motif.



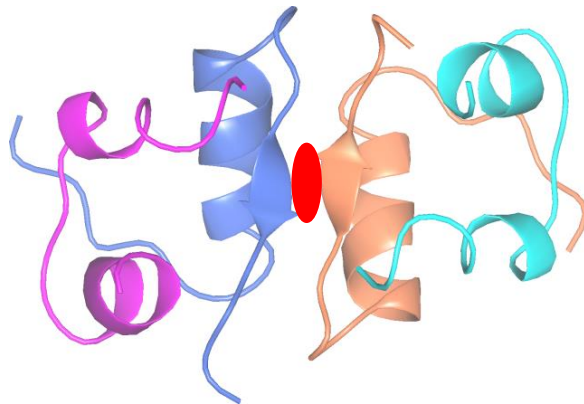
$$P'(xyz) = R P(xyz) + T(xyz)$$

Simmetria non-cristallografica

Nelle strutture proteiche la simmetria non-cristallografica è piuttosto comune, poiché spesso le proteine sono organizzate in oligomeri: (dimeri, trimetri...).

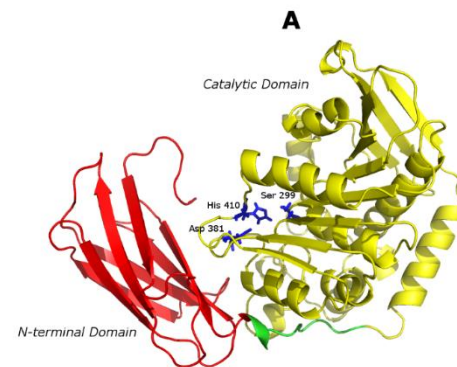
Le forme oligomeriche più comuni sono quelle che vanno da 2 monomeri a 6 monomeri.

Se i monomeri sono in relazione tra loro solo con simmetria locale (NCS), l'intero oligomero è contenuto nell'unità asimmetrica.



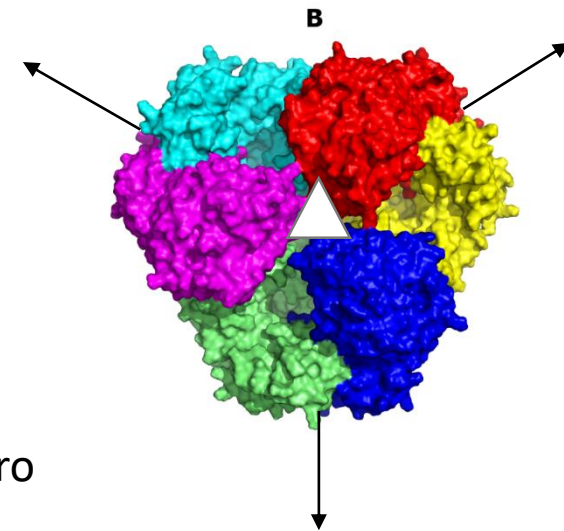
NCS di tipo 2

Un asse binario perpendicolare al piano

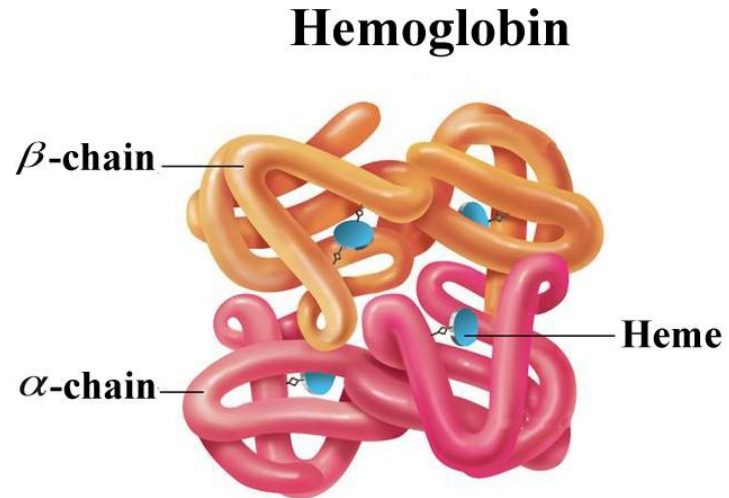


NCS di tipo 3_2 , un esamero nell'unità asimmetrica

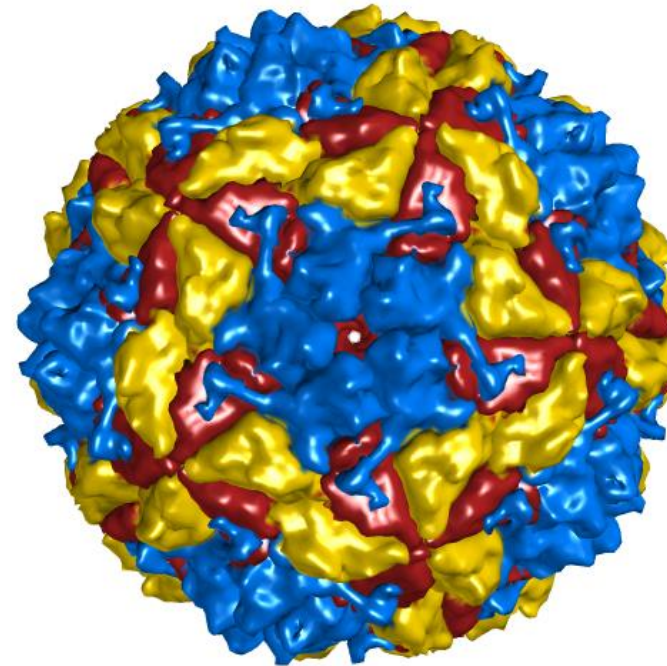
(asse ternario perpendicolare al piano e tre assi binari nel piano)



Simmetria Non-cristallografica



Picornavirus (capside)



NCS e density modification

La presenza di NCS può essere determinata dall'analisi del contenuto della cella unitaria (dal coefficiente di Matthews) che indica quante molecole sono (probabilmente) contenute nella cella unitaria.

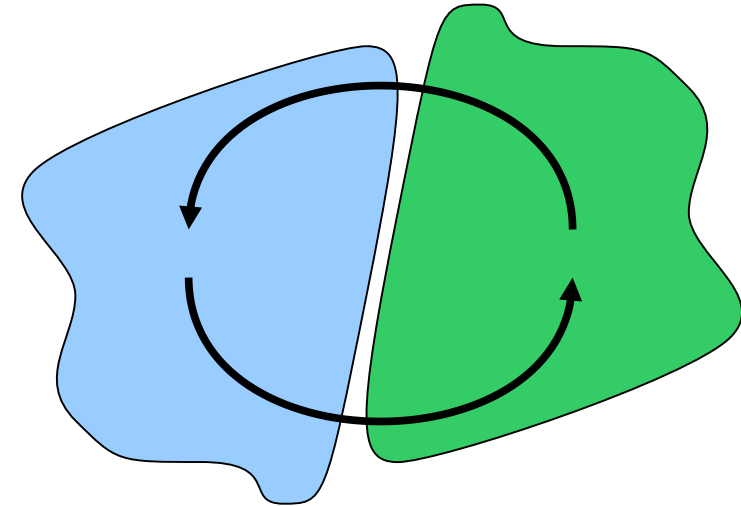
Nella fase di miglioramento della fase possiamo sfruttare l'NCS

*Se nell'unità asimmetrica esistono parti legate da simmetria locale (NCS), **queste parti possono essere mediate tra di loro (Averaging)**, quelle parti che non rispettano la simmetria locale sono eliminate.*

In pratica sfruttando la NCS imponiamo dei constraints sulla densità elettronica, **che deve obbedire nello spazio reale (mappa di densità elettronica) alla simmetria locale dovuta all'esistenza della NCS.**

L'utilizzo della NCS è **uno strumento molto potente** che determina un netto miglioramento della mappa di densità elettronica.

Più è elevato il numero di monomeri più grande è l'effetto sulla mappa.



Determinazione degli operatori di NCS

Un problema che si pone nell'uso dell'NCS dopo la prima 'fasatura' della mappa è la determinazione degli operatori di simmetria NCS (relazioni matematiche approssimate) che mettono in relazione i vari oligomeri tra loro.

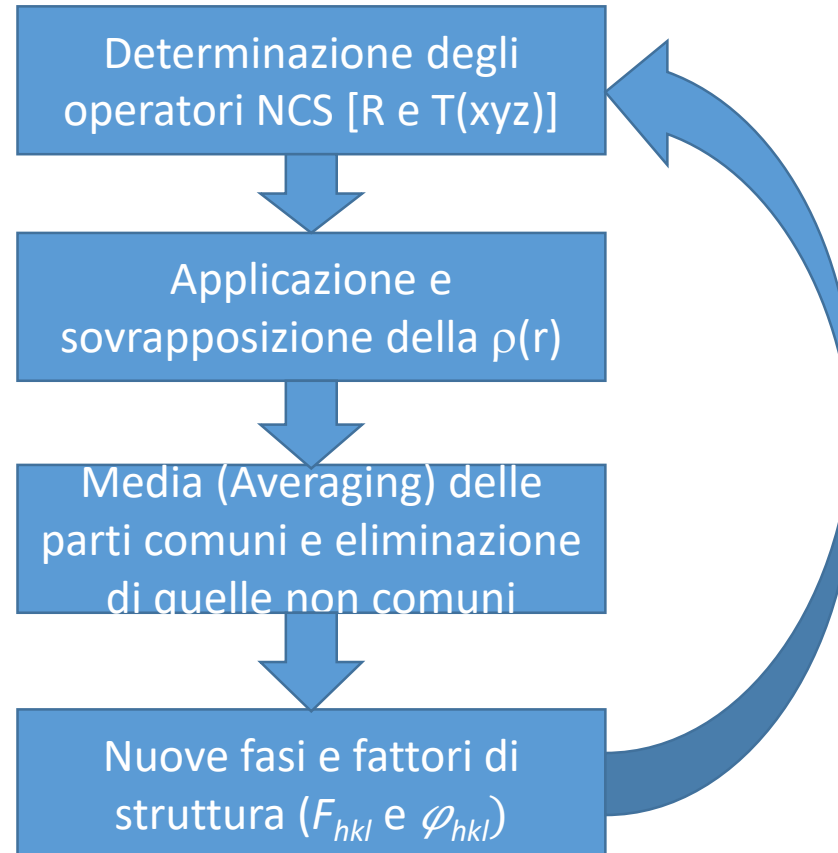
Queste relazioni possono essere determinate essenzialmente in due modi:

- **Dalle posizioni degli atomi pesanti**
- **Dal calcolo della funzione di Patterson** (si può calcolare con le sole intensità)

In entrambi i casi la simmetria non-cristallografica deve essere evidente, solo allora le relazioni matematiche vengono determinate.

Anche in questo caso spesso si procede iterativamente e in modo automatico per intervalli di risoluzione crescente.

Applicazione della NCS



Estensione delle fasi (Phase Extension)

Le fasi e la mappa iniziali sono spesso determinate ad una risoluzione più bassa di quella massima disponibile nel set di dati 'nativo' poiché i derivati isomorfi e le proteine contenenti Se-Met diffrangono a risoluzione più bassa di una equivalente struttura 'nativa'

Se si dispone di un secondo set di dati a più alta risoluzione, a partire da una prima mappa di densità elettronica calcolata da dati a risoluzione più bassa, si possono desumere le fasi dei fattori di struttura della proteina nativa aventi risoluzione più alta (phase extension) e non presenti nel set di dati utilizzato per ottenere le fasi iniziali e calcolare la prima mappa.

L'approccio seguito è lo stesso della Density Modification, dalla mappa di densità elettronica è possibile determinare le fasi dei fattori di struttura.

Funziona molto bene in combinazione con l'NCS.

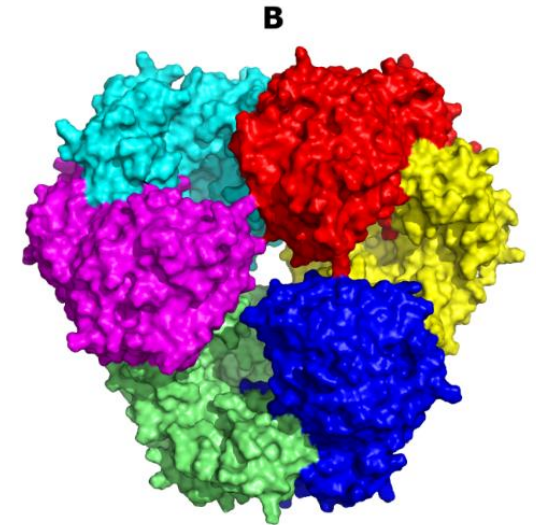
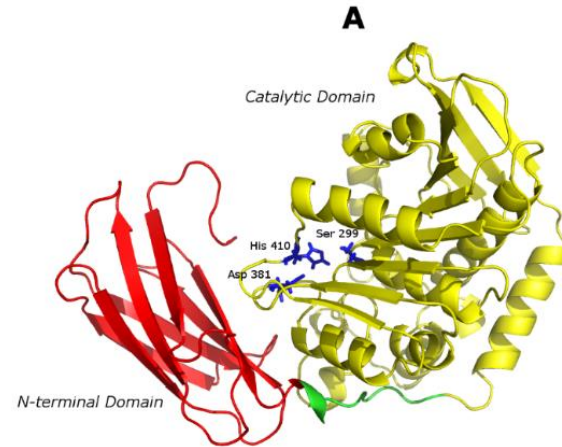
Viene applicata parallelamente alla Density Modification e in un certo senso è una tecnica di Density Modification poiché opera tra lo spazio reale (densità) e quello reciproco (fattori di struttura).

Applicazione della density modification: Axe7a

Applicazione della NCS

Applicazione: Axe7a (idrolasi)

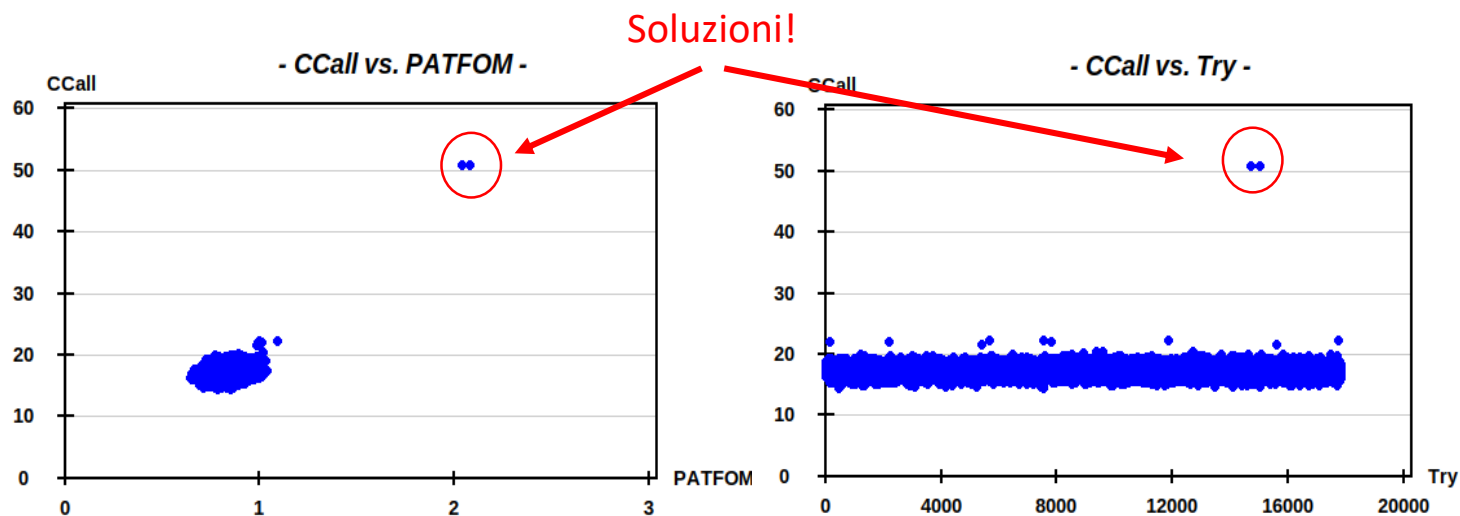
Monomero: ~ 40kDa (429 residui)
NCS: 6 (2574 residui totali)
Risoluzione: 3.2 Å (non altissima)
Metodo: SAD
Diffusore anom: Se
Num diff. Anom: 96 (16 per monomero)



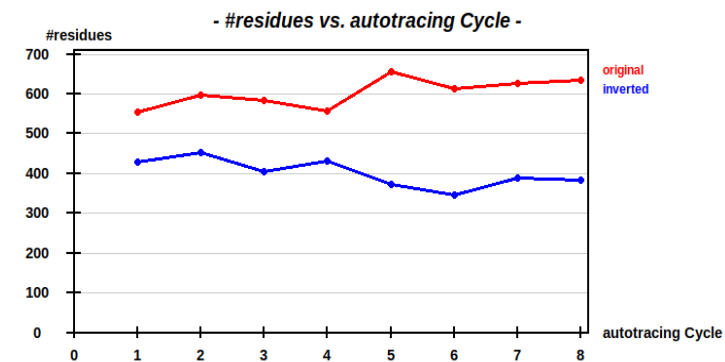
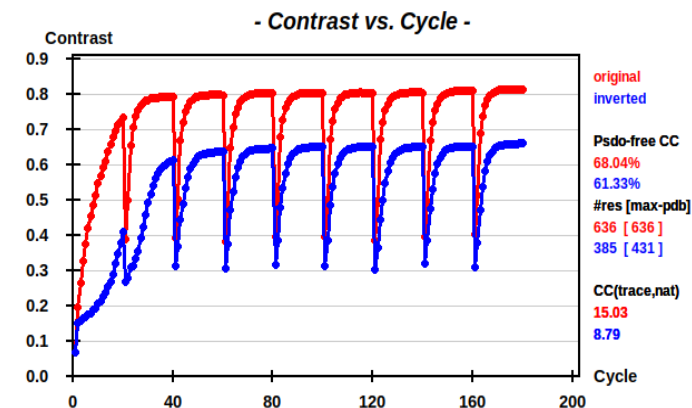
Struttura finale a 2.65 Å (con un altro dataset 'nativo')

Axe7A: Determinazione atomi pesanti e SAD

- Segnale anomalo debole (osservabile fino a 3.7 Å)
- La combinazione di Patterson e metodi diretti trova 83 atomi di selenio con occupazione tra 0.3 e 1
- Risoluzione del problema della fase tramite SAD
- **Le due strutture enantiomorfe sono distinguibili**



Determinazione posizione atomi Se

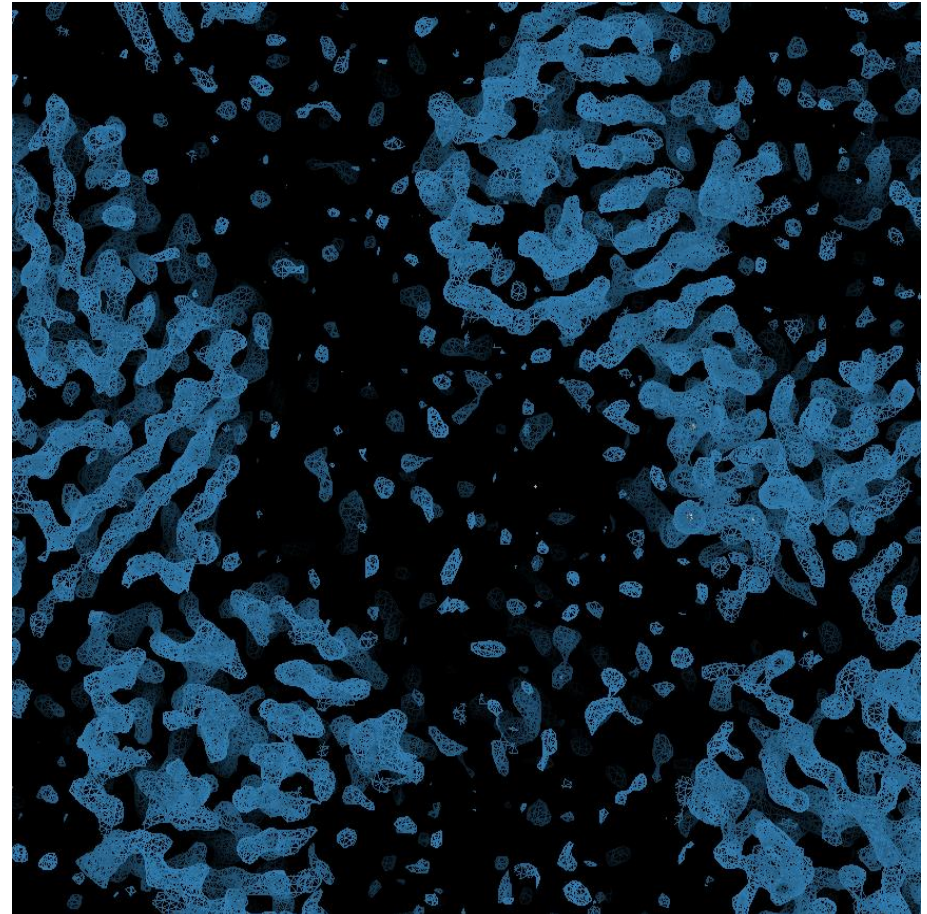
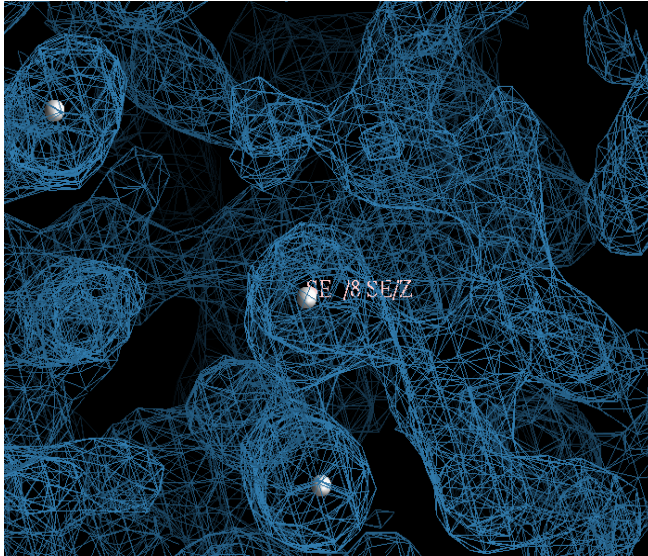


Distinzione enantiomorfi

Axe7A: Fasi e prima Mappa

- Posizione e occupazione degli atomi di Se migliorata
- Calcolo delle prime fasi

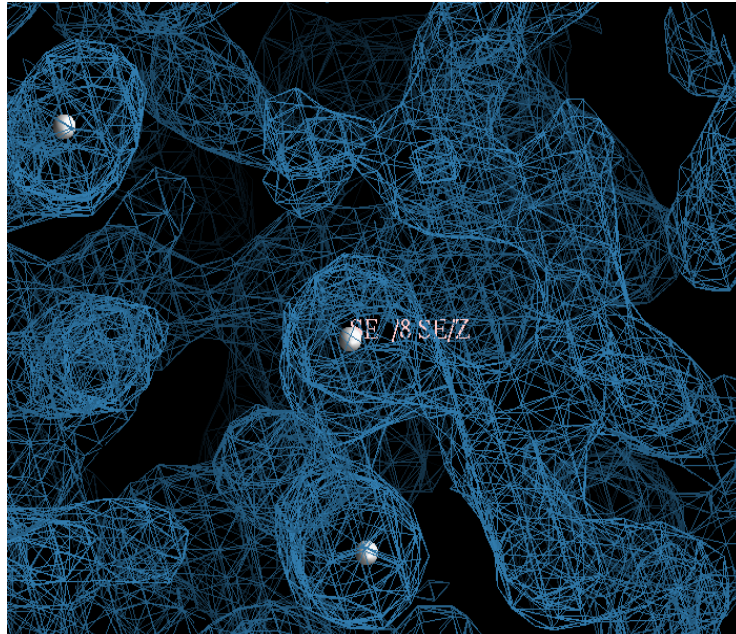
Nella mappa di densità si nota una distinzione tra lo spazio occupato dal solvente e quello occupato dalla proteina



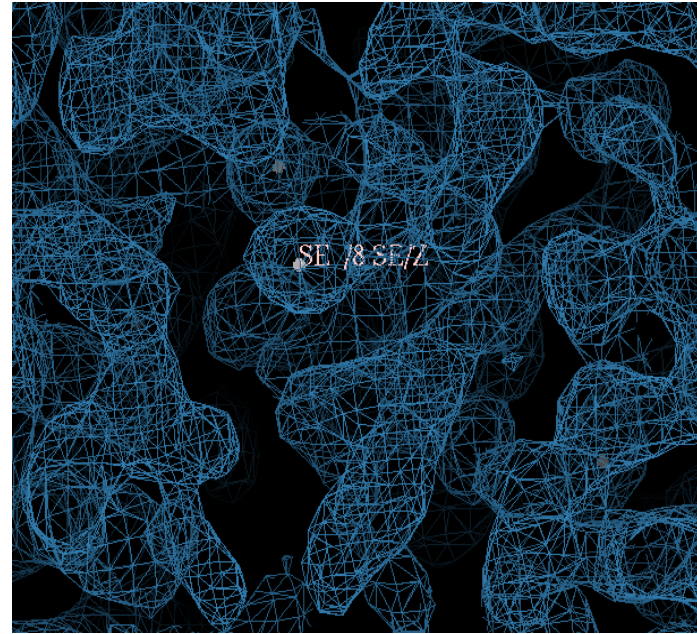
Axe7A: Density Modification

Applico density modification

- Solvent Flattening
- NCS Averaging (6 monomeri!)

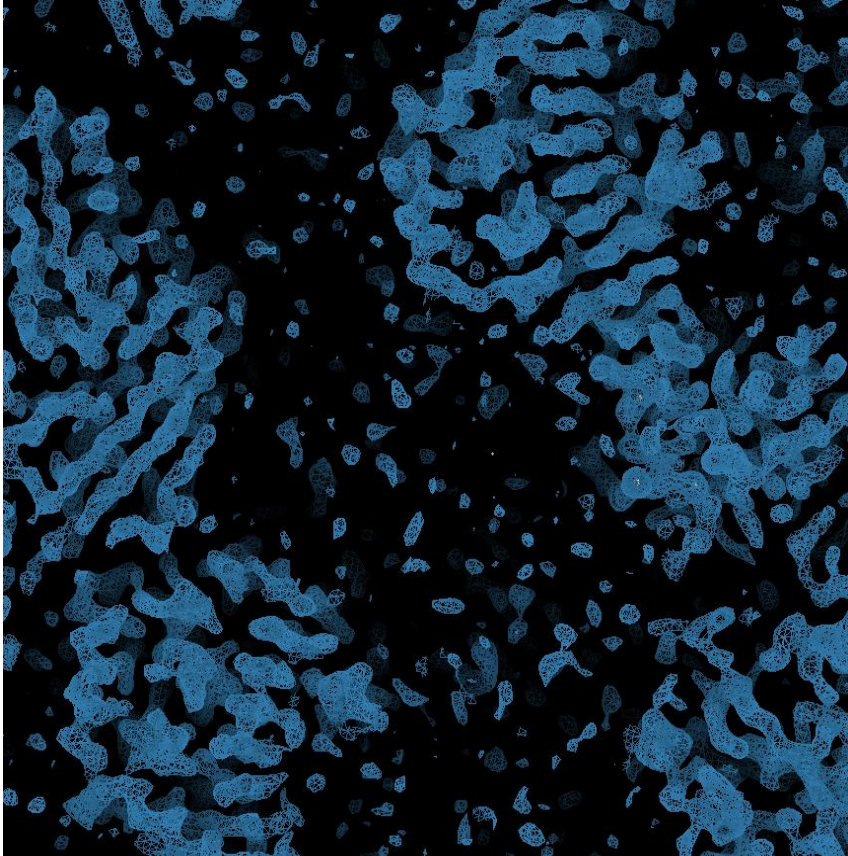


Pre-Density Modification

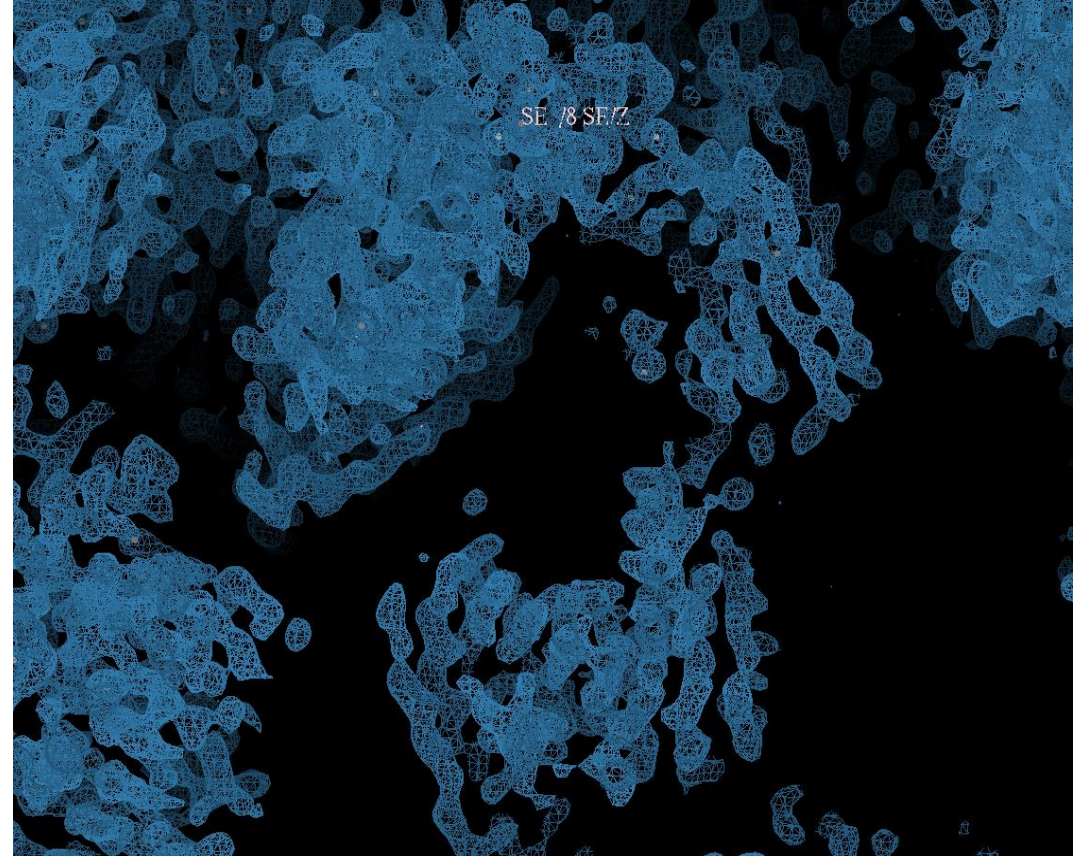


Post-Density Modification

Axe7A: Density Modification



Pre-DM



Post-DM

Costruzione del modello molecolare (Model Building)

Costruzione del Modello

Quando si dispone di una mappa di qualità adeguata si può **costruire il modello molecolare che meglio si adatta alla densità elettronica.**

Anche questo procedimento è iterativo e nel caso di determinazione delle fasi per mezzo di MIR e SAD/MAD si tratta di **costruire un po' alla volta il modello a partire dalla sequenza della proteina, dal tipo di atomo pesante e dalla forma della mappa.**

Questo procedimento è in larga parte automatizzato, tuttavia **la qualità della mappa, la sua risoluzione e altri fattori possono pesantemente influenzare l'esito del model building.**

Può accadere che solo una parte del modello venga costruito, (ad es: solo la Main Chain) o costruito in modo corretto (altre parti del modello molecolare sono costruite in modo non corretto). L'intervento umano è molto spesso necessario per correggere e/o completare il modello.

Questo procedimento è noto come **Automatic Model Building**

Automatic Model Building

La costruzione di un modello automatico si avvale di due proprietà fondamentali delle proteine:

- **Le proteine sono costituite di una catena lineare di aminoacidi**
- **Il numero di aminoacidi è limitato**

Viene anche usata una proprietà della densità elettronica:

- **La densità elettronica che descrive una catena polipeptidica deve essere continua (connessa)**

Queste proprietà sono utilizzate nel processo di model building automatico

Il model building automatico (Automatic Model Building) non ha una vera e propria teoria, quanto una serie di protocolli che si sono dimostrati efficaci nella pratica.

Possiamo dividerlo in due fasi:

- **Determinazione della Main Chain**
- **Costruzione delle Side Chain**

Determinazione automatica della Main Chain

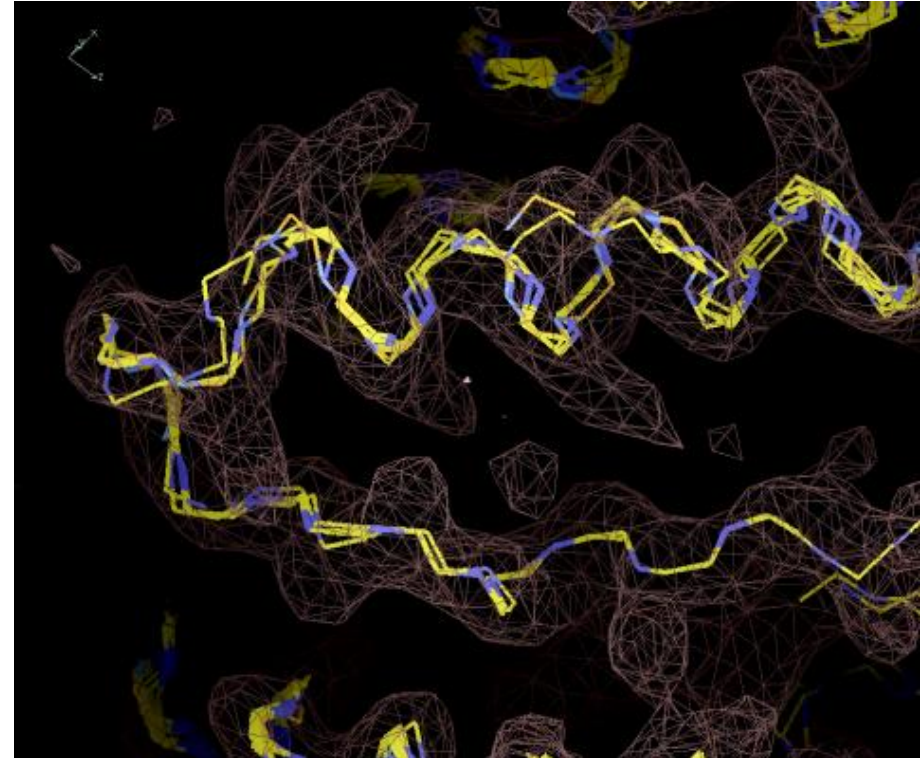
Sulla base della continuità della $\rho(r)$ è possibile posizionare dei segmenti di lunghezza adeguata che connettono i massimi della $\rho(r)$ e a partire da questi, stimare la posizione dei $C\alpha$.

Questa procedura è anche indicata come 'scheletrizzazione' (*skeletonization*).

In questo modo vengono costruiti dei frammenti di lunghezza variabile che poi vengono 'fusi' tra di loro, per generare dei frammenti aminoacidici (poli-alanina) lineari, rispondenti ai requisiti stereochimici di una catena polipeptidica.

Gli atomi inseriti vanno a contribuire al miglioramento delle fasi.

Nella fase di Autobuilding le fasi sperimentali si combinano con quelle derivanti dal modello molecolare in costruzione.



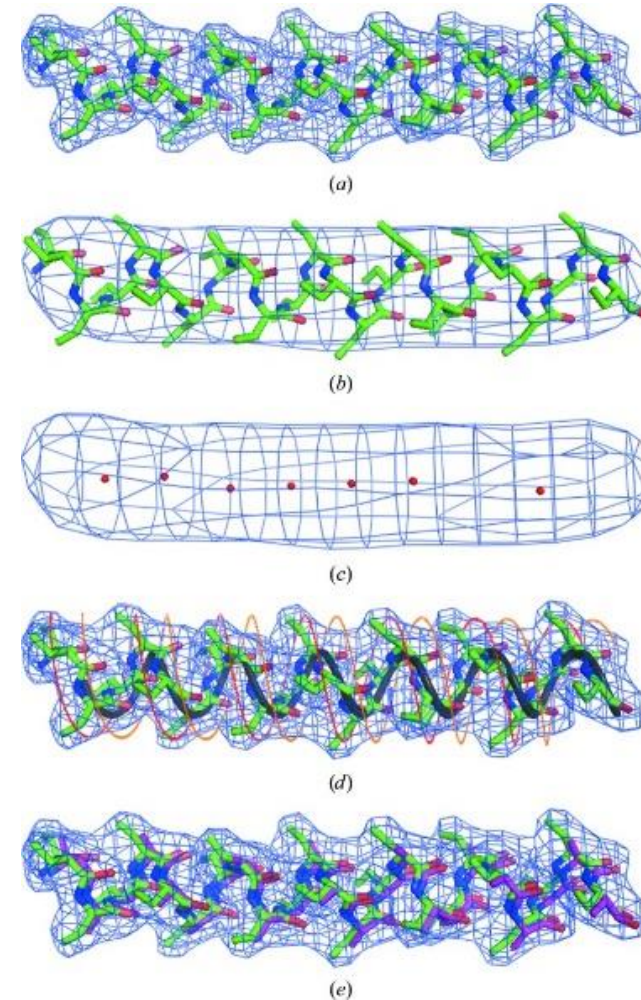
Insertimento di α -eliche e β -foglietti

Nelle procedure di Auto Building è inoltre possibile il riconoscimento della densità elettronica appartenente ad α -eliche.

E' una procedura che funziona bene anche a bassa risoluzione (4-6 Å).

A risoluzione intermedia (3-3.5 Å) è possibile determinare la direzione dell'elica.

La presenza di determinate strutture secondarie può essere predetta per via bioinformatica.



Docking

Il riconoscimento degli elementi di struttura secondaria nella mappa di densità, unito alla predizione delle strutture secondarie a partire dalla sequenza aminoacidica, permette l'assegnazione dei frammenti costruiti ad aminoacidi ben precisi, compatibilmente con la sequenza nota (**docking**).

Poiché la sequenza è nota, alcuni frammenti individuati nella mappa possono essere 'trasformati' in residui, compatibilmente con la sequenza aminoacidica.

Nel caso di mappe di densità a risoluzione relativamente elevata e di buona qualità si arriva a costruire correttamente più del 90% del modello complessivo.

AutoBuild

PHENIX Preferences Help Run Abort Save Xtriage Coot PyMOL

Configure AutoBuild_run_33_ Status Summary Model-building Structure status

Protein chains

Double-click on any residue or secondary structure element to zoom in on that region in Coot or PyMOL.

Select chain: A Clear selection ?

1 XXXXXXXXXXXXXXLAGELKEGSYVVIDGEPCRVVEIEKSKTGKHGSAKARI 50

51 VAVGVFDGGKRTLSPVDAQVEVPIIEKFTAQILSVSGDVIQLMDMRDYK 100

101 TIEVPMKYVEEEAKGRLAPGAEVEVWQILDRIYXXXXXXXXXXXX 150

151 KA 200

Non assegnati

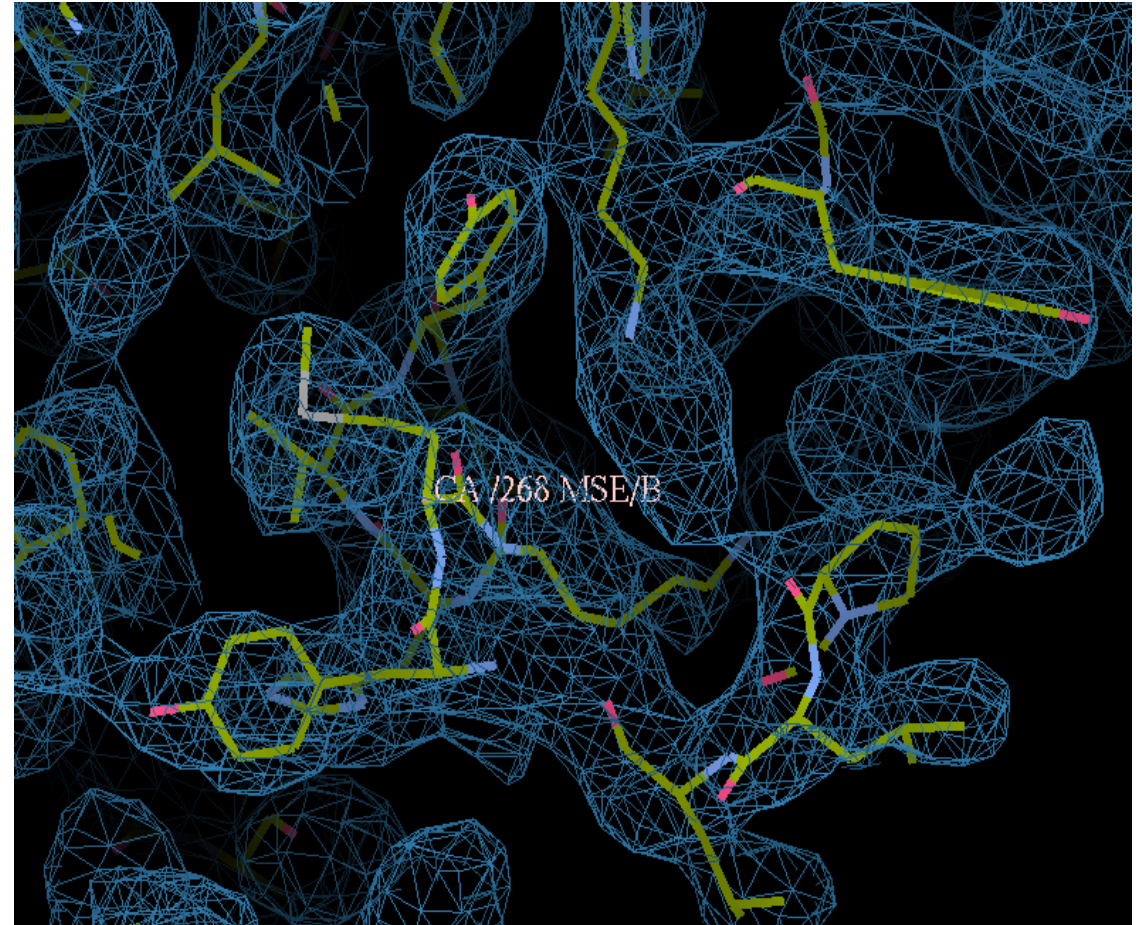
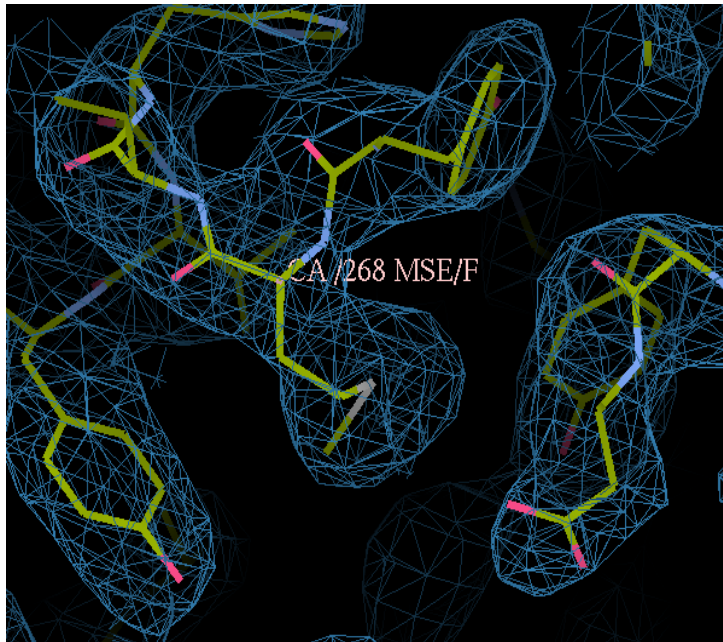
Autobuilding in azione: Axe7A

Axe7A: Autobuilding

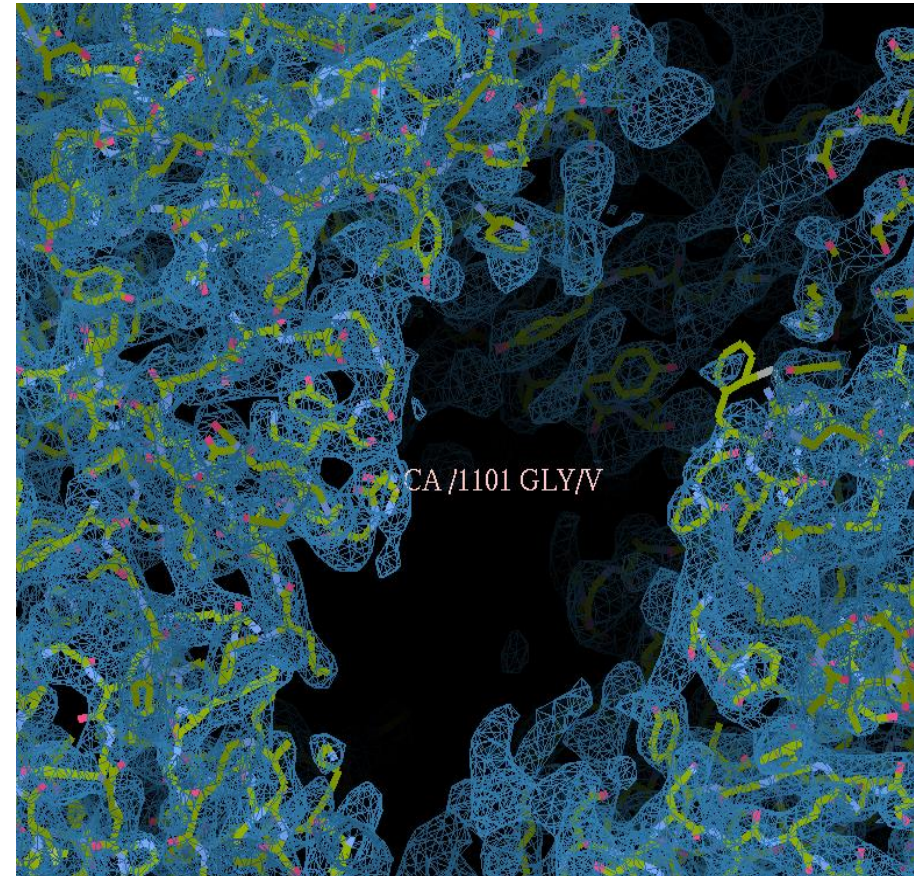
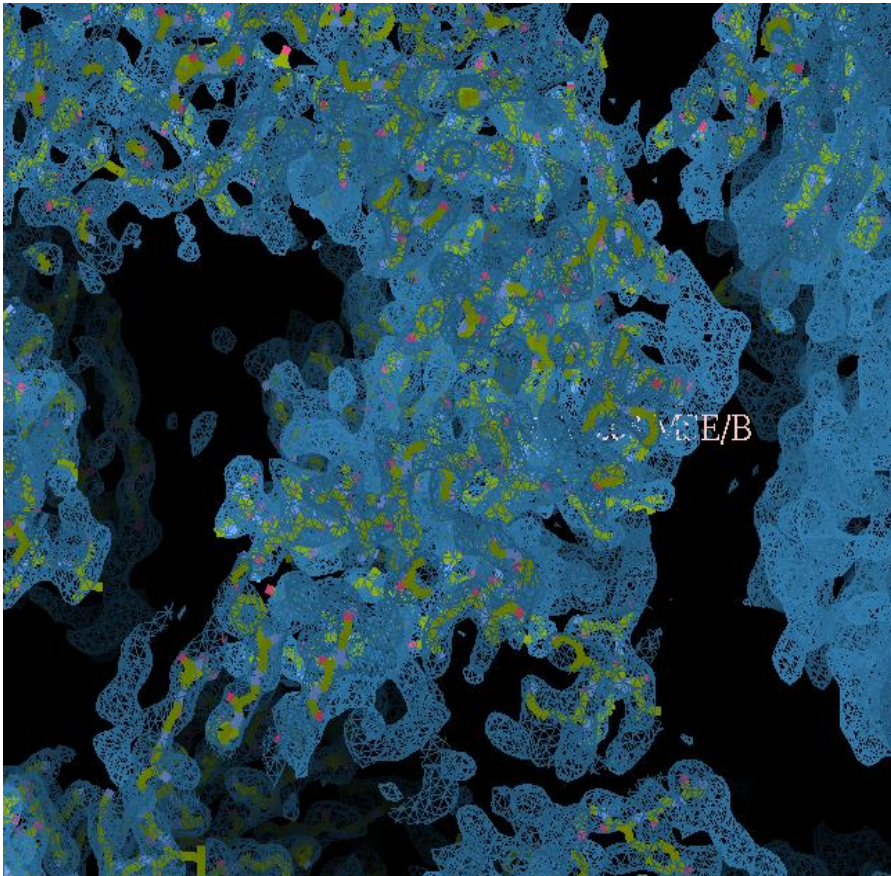
- Mappa: SAD/DM, 3.2 Å
- Numero di residui costruiti: ~1600 (2574)
- Numero di residui corretti: 40% della struttura

Sufficiente per costruire il modello finale!

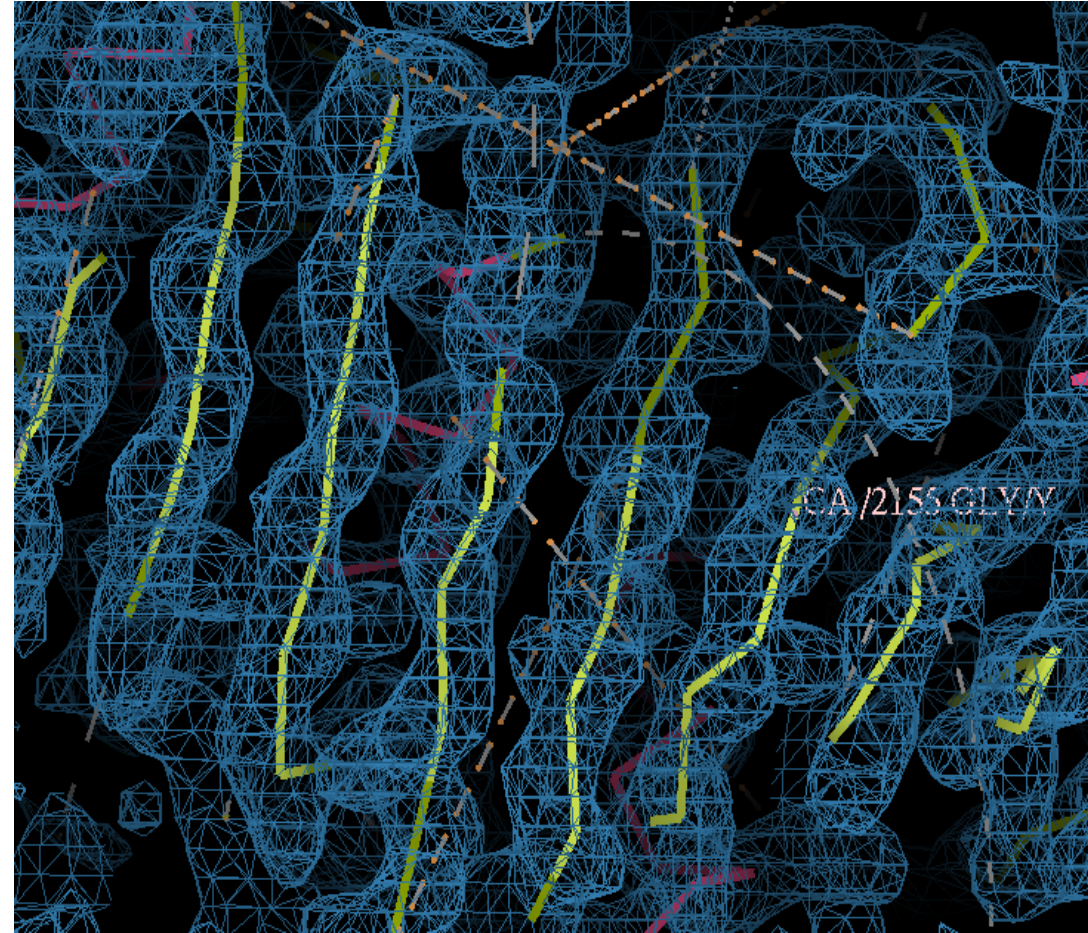
Se-Met



Axe7A: Autobuilding



Axe7A: α -eliche e β -foglietti



Axe7A: Docking

Non tutti i frammenti
posizionati erano in posizione
corretta.

Una parte minima della
sequenza è stata assegnata



Final model

R-work:	0.3257	R-free:	0.3930	CC:	0.67
Residues:	1622	Fragments:	205	Waters:	0