

Homology modelling

L'omology modeling delle proteine è il tipo di predizione di struttura terziaria più semplice ed affidabile. Viene richiesta soltanto una (o più) sequenze di riferimento su cui modellare la struttura. Le procedure sono semplici, ma la validazione è complessa come tutti gli altri tipi di predizione strutturale.

⇒ Costruire un modello è semplice.

⇒ Rifinire un modello è complesso.

⇒ Validare un modello è molto complesso, e serve una parte sperimentale.

Per l'homology modeling sono disponibili dei server dedicati all'automazione delle operazioni di base (raccolta degli omologhi strutturali, allineamenti, validazioni ecc.)



SWISS-MODEL

An Automated Comparative Protein Modelling Server

E' il server più utilizzato, permette di tracciare tutte le operazioni che svolge e di impostare alcuni parametri importanti per il risultato finale.

Vi si accede on-line, oppure

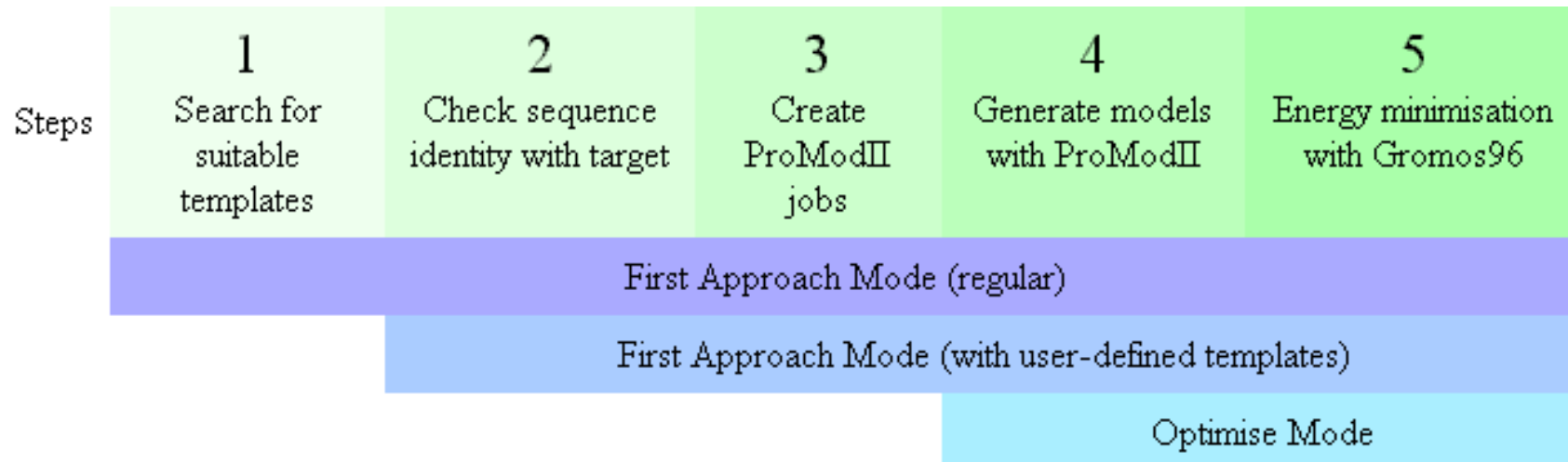
è interfacciato con Swiss PDB Viewer

che lancia le procedure base e poi permette di perfezionare i risultati grazie ai suoi strumenti di modeling.

Modelling requests:

- First Approach mode
- Alignment Interface
- Optimise (project) mode

La versione on-line suddivide il lavoro in cinque tappe principali: l'utente può posizionarsi in varie posizioni a seconda delle informazioni di cui dispone



```
Searching sequences of known 3D structures
Found 11DDF.pdb with P(N)=3.4e-59
Found 11CDF.pdb with P(N)=1.6e-13
Found 11EXT.pdb with P(N)=1.5e-09
Found 21EXT.pdb with P(N)=1.5e-09
Found 21NCF.pdb with P(N)=1.5e-09
Found 21TNR.pdb with P(N)=1.5e-09
Found 11NCF.pdb with P(N)=1.5e-09
```

Ricerca dei templati possibili

Viene utilizzato il programma BLASTP2 per cercare similarità di sequenza primaria contro la banca dati ExNRL-3D, contenente le controparti adattate per la ricerca blast delle di tutte le proteine contenute nella banca dati ExPDB, forma non ridondante e non annotata generata della PDB.

Visto che si utilizza BLAST, possono uscire più templati adatti perché :

- 1 - Esistono effettivamente più templati che hanno un'alta similarità **globale** con la sequenza di interesse.
- 2 - I templati trovati corrispondono a zone diverse della sequenza di interesse, che possiedono alte similarità **locali**.

Ricerca dei migliori templati

Vengono scelti tra tutti i modelli possibili quelli che hanno identità maggiore del 25% da allineamento pairwise struttura/sequenza e viene progettata la modellazione di tratti di sequenza lunghi più di 20 residui.

Grazie a questa operazione di ottimizzazione, vengono applicati i criteri dell'homology modelling.

```
11DDF.pdb: 100 % identity
11CDF.pdb: 37.33 % identity
11EXT.pdb: 26.4 % identity
21EXT.pdb: 26.4 % identity
21NCF.pdb: 26.4 % identity
21TNR.pdb: 26.4 % identity
11NCF.pdb: 26.4 % identity
```

```
Target Sequence: |=====|
11DDF.pdb        |
11CDF.pdb        |-----|
11EXT.pdb        |-----|
21EXT.pdb        |-----|
21NCF.pdb        |-----|
21TNR.pdb        |-----|
11NCF.pdb        |-----|
```

Due diversi domini per
la modellazione

Costruzione dei pre-modelli

Viene lanciato il programma ProModII su tutti i templati trovati. La sequenza delle operazioni svolte da ProModII offre, per ogni template, un modello strutturale ottimizzato della sequenza di interesse grazie agli allineamenti strutturali generati dal programma SIM.

```
ProModII trace log for Batch.1
=====

SPDBV: Loading Template
SPDBV: Loading Raw Sequence
SPDBV: Generating Structural Alignment
SPDBV: Aligning Raw Sequence
SPDBV: Refining Raw Sequence Alignment
SPDBV: Averaging Sidechains
SPDBV: Adding Missing Sidechains
SPDBV: Dumping Sequence Alignment
SPDBV: Dumping Preliminary Model
SPDBV: Done.
```

Ottimizzazioni energetiche

Viene applicato Gromos96 ad ogni struttura modellata, per valutare l'affidabilità energetica del modello e per effettuare la minimizzazione energetica.

Conversione formati Costruzione della topologia Aggiunta degli idrogeni Calcoli di dinamica molecolare

```
Gromos96 trace log for Batch.2
=====

Now running PROCS1 on file batch-procs0.dat ... Done.
Now running PROCS2 on file batch-procs1.dat ... Done.
Now running PROGMT on file batch-procs2.dat ... Done.
Now running PROGCH on file batch-procs2.dat ... Done.
Now running PROMD on file batch-progch.dat ... Done.
Now running PROMD on file batch-promd0.dat ... Done.
Detection of SS-Bonds within batch ...
  >SS-Bonds detection:  #1:  91 <-> 105
  >SS-Bonds detection:  #2:  66 <->  81
  >SS-Bonds detection:  #3:  69 <->  89
  >SS-Bonds detection:  #4:  25 <->  44
  >SS-Bonds detection:  #5:  47 <->  63
```

Alla fine viene prodotto un file contenente tutte le strutture utilizzate per la modellazione e la struttura incognita modellata (**TARGET**), con tutti gli allineamenti strutturali.

I risultati, compresa tutta la lista delle operazioni svolte dai vari programmi, vengono **spediti via e mail** e la struttura modellata si presenta come un progetto di Swiss PDB Viewer (nomefile.pdb).

Record type	Annotation
HEADER	Header line.
TITLE	User-supplied title or SWISS-PROT entry DE line.
EXPDTA	Always THEORETICAL MODEL.
JRNL	Please site these references in case of publication
REMARK 3	Brief description of modelling method and minimisation parameters.
REMARK 5	Description of the templates used to build the model.
REMARK 999	Description of model sequence with respect to corresponding SWISS-PROT entry.
SEQALI	Sequence alignment generated by ProModII.

La struttura modellata contiene anche informazioni riguardanti la **confidenza della predizione**. Nell'ultima colonna dei record ATOM del file PDB in genere sono contenuti i dati sulla mobilità dell'atomo ottenuta dai dati NMR o di cristallografia:

In questo caso il **B-FACTOR** non è sperimentale ma è ottenuto da un'analisi statistica della struttura modellata con tutti i template disponibili ed è legato a

- 1 - Numero dei template utilizzati per il modelling**
- 2 - Deviazione di posizione del modello dai template**
- 3 - Distance trap (la soglia di distanza considerata di inattendibilità)**

Modelling requests:

- [First Approach mode](#)
- [Alignment Interface](#)
- [Optimise \(project\) mode](#)

Se sono noti degli allineamenti rifiniti a mano e sicuramente attendibili è possibile utilizzare l'interfaccia con l'allineamento anziché il first approach.

Viene chiesto di incollare un allineamento, che deve contenere al suo interno la proteina da modellare e anche la struttura o le strutture di riferimento. In questo modo la modellazione parte da uno step successivo rispetto prima, saltando tutta la prima parte di ricerca dei template e di allineamento.

Alignment Input Format: Target sequence:

Paste your Alignment here:

Template sequence: Chain-ID:

Da cercare nel database ExPDB

Modelling requests:

- [First Approach mode](#)
- [Alignment Interface](#)
- [Optimise \(project\) mode](#)

Dopo un primo modelling, è possibile perfezionare a mano l'allineamento strutturale e salvarlo come progetto swiss-model in Swiss PDB Viewer.

Questo modo è il più raffinato e il più “pericoloso”: si utilizzano solo gli strumenti di refining del server Swiss-Model, ma il modello è sostanzialmente già fatto, quindi se è sbagliato (es. allineamento non corretto, templati sbagliati ecc.) rimarrà sbagliato!

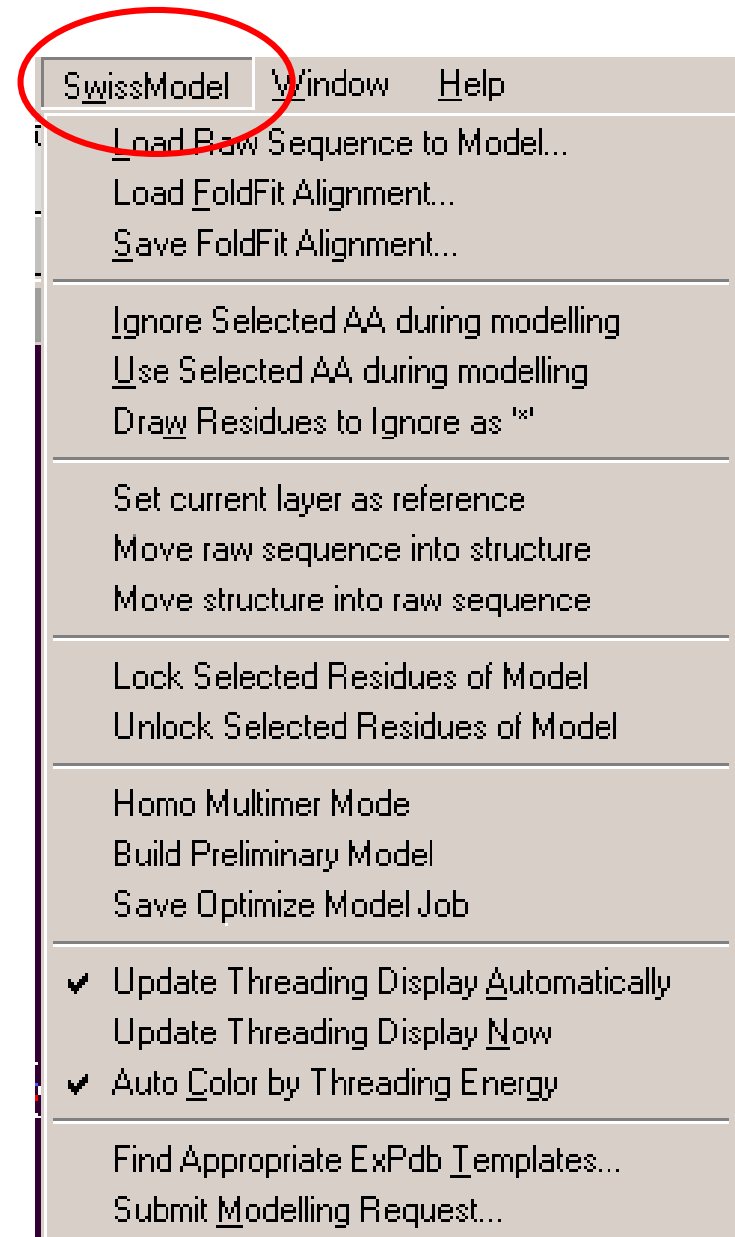
Da Swiss PDB Viewer è possibile lanciare direttamente una richiesta di modellazione ottimizzata, data che l'ottimizzazione è possibile con gli strumenti di modellazione fine del programma. Anche in Optimize mode i risultati vengono spediti via e-mail, e vanno riaperti in seguito.

Utilizzare Swiss PDB Viewer per il modelling

In generale, Swiss PDB Viewer viene considerato un programma per valutare la qualità di un modello strutturale e per il manual refining delle strutture.

In effetti però è possibile iniziare e portare a termine un progetto di modellazione interamente, senza utilizzare la versione on-line di SWISS-MODEL.

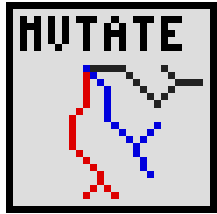
In alternativa, si effettua il modeling con il server e poi si apre il progetto che arriva via e-mail e si corregge a mano.



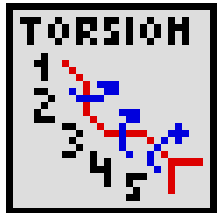
Valutare la qualità del modello

- 1 - Osservare il r.m.s.d. tra modello e template.
- 2 - Colorare tutto in base al B-factor (se c'è).
- 3 - Osservare il plot di Ramachandran.
- 4 - Selezionare gli amino acidi in collisione tra loro.
- 5 - Selezionare gli amino acidi in collisione con il backbone.
- 6 - Selezionare gli amino acidi con legami idrogeno impropri.
- 7 - Osservare le polarità dei residui esposti e sepolti.
- 8 - Calcolare l'energia libera.

Modificare la struttura del modello



Permette di modificare un residuo con un altro, ma anche di cambiare il **ROTAMERO** della catena laterale: in natura infatti ogni residuo presenta la catena laterale in un certo numero **finito** di disposizioni (rotameri) archiviate in una libreria apposita.



Permette di modificare tutti gli angoli di legame del residuo selezionato: se si cambiano gli angoli phi e psi naturalmente si sposta anche il backbone della proteina e bisogna decidere se può spostarsi l'N o il C terminale.

Per evitare gli spostamenti delle estremità si può rompere temporaneamente il legame peptidico, per poi effettuare le modifiche e la nuova ligazione.

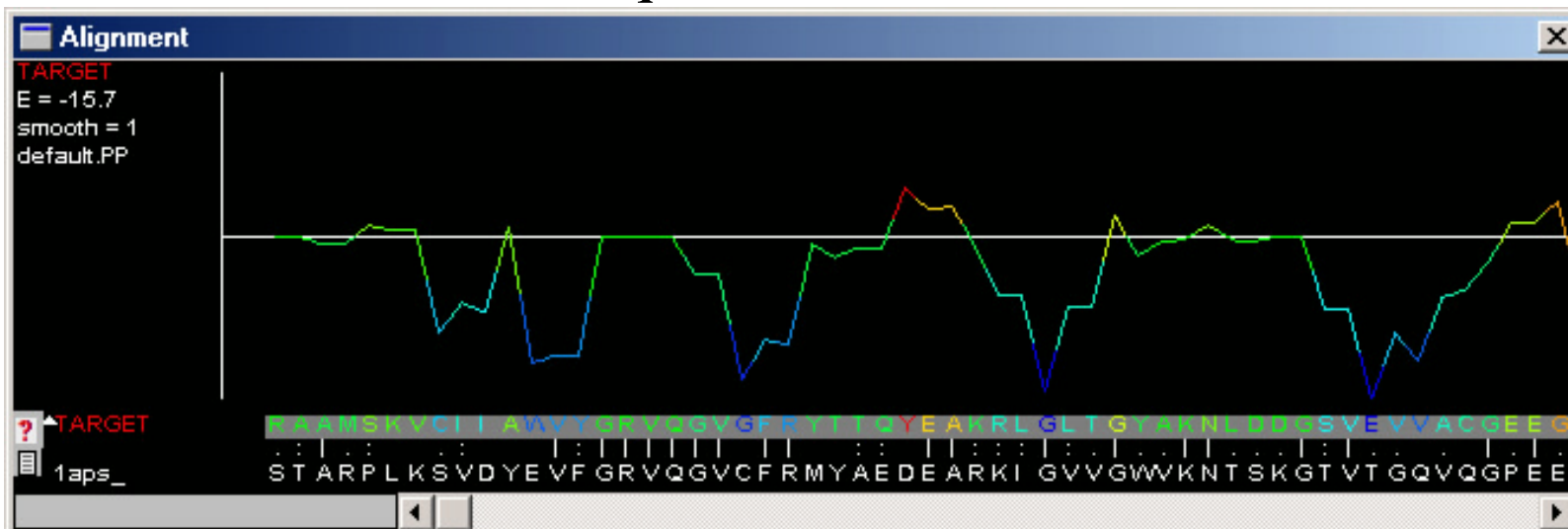
Perfezionamento dell'allineamento

La finestra Alignment mostra l'allineamento strutturale come è stato generato dal server di modelling. E' possibile modificare l'allineamento selezionando il residuo da spostare e:

- spacebar/del: inserimento/rimozione di un gap
- freccia dx/sx: spostamento del residuo

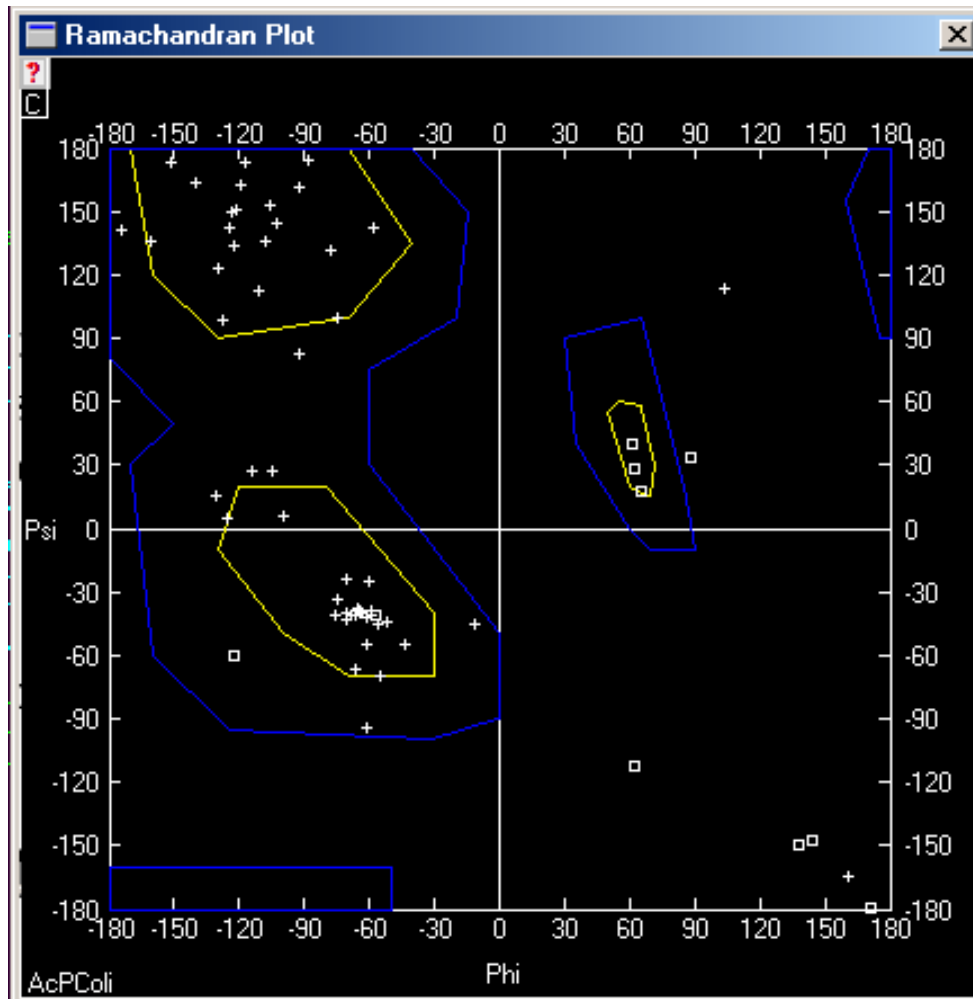
Ogni modifica si riflette anche sulla struttura.

Il plot riguarda la “threading energy”, cioè un valore che indica quanto è adatto il residuo a stare in quell'intorno di residui.



Ramachandran plot

Per ogni residui vengono plottati i rispettivi psi su phi: solo certe combinazioni sono permesse. Inoltre è possibile cambiare gli angoli trascinandoli nel grafico. Il backbone della proteina cambia!



Giallo: aree permesse
Blu: aree limite
Quadrati: glicine
Crocette: altri residui

Le glicine vanno trattate a parte parlando di angoli permessi: l'amino acido è così piccolo che non ha in pratica angoli vietati...

Riorientamento delle catene laterali

E' possibile riorientare le catene laterali in termini di rotameri dal menu Tools -> Fit Selected Sidechains



Quick and Dirty: mette il rotamero a minor score

Exhaustive search: prova tutti gli angoli possibili

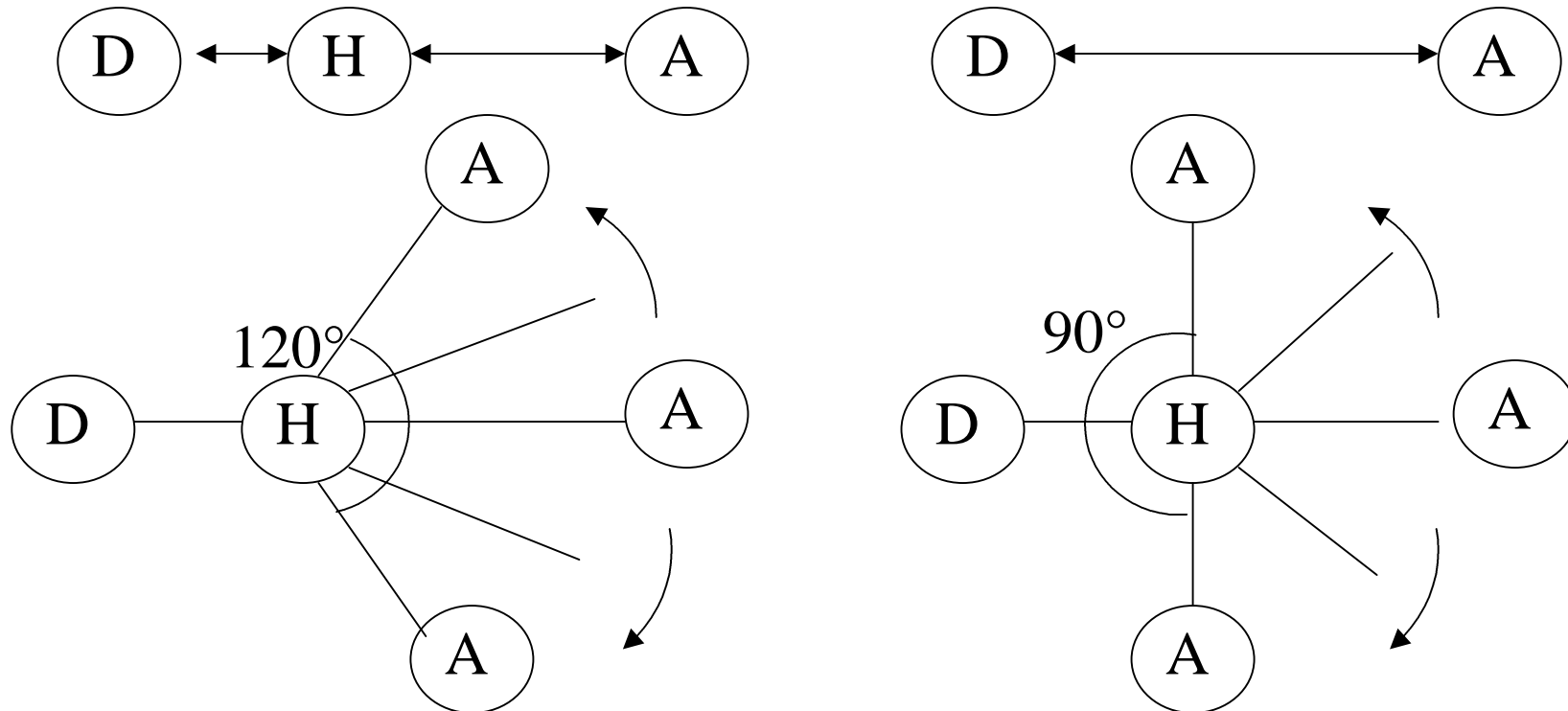
Simulated annealing: non funziona

Lo score (clash score) di uno spostamento è calcolato secondo:

- 4 x numero di scontri tra atomi del legame peptidico
- + 3 x numero di scontri con gli ossigeni del backbone
- + 2 x numero di scontri delle catene laterali
- numero di idrogeni
- 4 x numero di ponti disolfuro formati

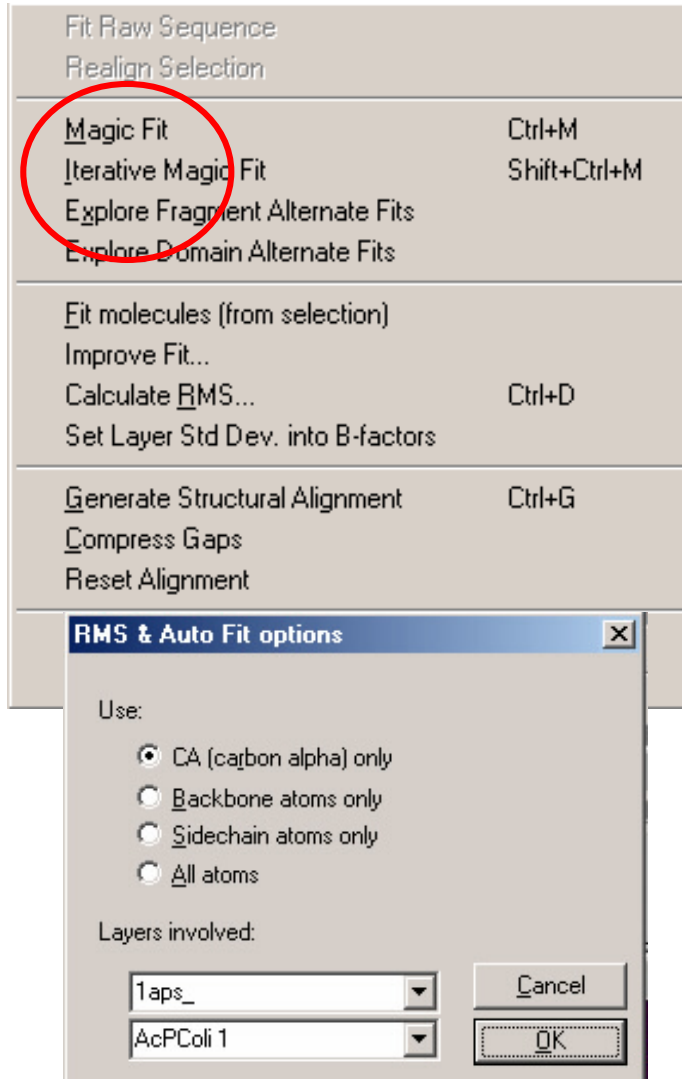
Ponti idrogeno

Possono essere “facilmente” localizzati osservando i tipi di atomi e le distanze tra di essi. Di default viene assegnato un ponte idrogeno se un donatore e un accettore si trovano tra 1.2 e 2.76 Å e con un angolo di 90° se non c'è l'idrogeno, tra 2.35 e 3.2 Å e 120° se c'è l'idrogeno.



Superimposizione delle strutture

L'allineamento tridimensionale si può effettuare secondo vari criteri:



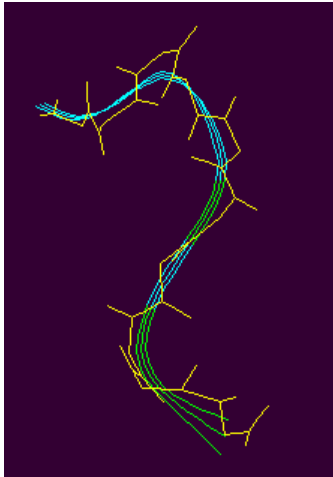
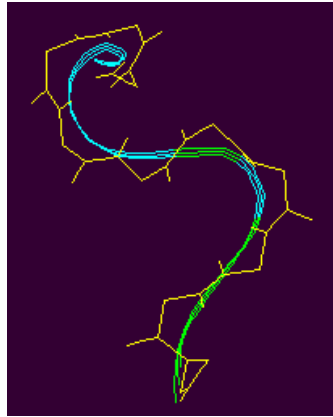
1 - **Magic fit**: Fa un allineamento pairwise con una PAM e poi adatta le strutture.

2 - **Iterative**: fa un magic, poi un improve, poi minimizza l'r.m.s.d.

3 - **Explore**: non considera la sequenza primaria, ma solo le coordinate degli atomi.

Le superimposizioni e l'r.m.s.d si possono applicare al backbone, ai C-alfa, alle catene laterali o a tutti gli atomi...

Rimodellazione dei loops



Molto spesso nei modelli le stime peggiori si fanno sulle zone più mobili (i loops). Esistono però delle banche dati dove sono scritti i vari modi di connettere strutture secondarie. E' possibile pertanto modellare i loops non sulla strutture vecchia, ma su un loop di questo database (**Scan**). Inoltre è possibile minimizzare l'energia libera del loop solo (**Build**).

Build

Build Loop...

Scan Loop Database...

clash score: 5

PP: -3.67

FF: 6659.9

C-N+

CA-C-N+

C-N+-CA+

-0.12

7.88

6.74

0.44

0.00

0.00

0.20

0.00

0.00

Threading energy

Campi di forza

Deviazioni

Cliccando sopra tutti i possibili loops generati vengono ordinati in base al criterio di scelta fatto, permettendo la scelta del loop.