REGISTRO LEZIONI

MASTER COPY
Tel. 388/9837745

8 Marzo 2011

9 Marzo 2011

14 Marzo 2011

16 Marzo 2011

21 Marzo 2011 En.

23 Marzo 2011 En.

28 Marzo 2011

30 Marzo 2011

4 Aprile 2011

6 Aprile 2011

11 Aprile 2011 Es.

13 Aprile 2011 Es.

18 Aprile 2011 En.

20 Aprile 2011 Es.

2 Maggio 2011

4 Maggio 2011

9 Maggio 2011

11 Maggio 2011 Es.

16 Maggio 2011

18 Maggio 2011

23 Maggio 2011

25 Maggio 2011

30 Maggio 2011

31 Maggio 2011

master COPY COPISTERIA 050/8312126 388/9837745

master
COPY
COPISTERIA
050/8312126 388/9837745

Indirizzo email:

a.bechini@ing.unipi.it

MASTER COPY

Tel. 388/9837745

8 marzo 2011

Questo corso presenta argomenti a cavallo tra la biologia e l'informatica, per cui ci sono problemi a livello concettuale e metodologico, questo perché queste discipline diverse hanno un modo di operare diverso. L'approccio in questi due settori è diverso. Si supponga di dover studiare un fenomeno biologico di interesse che di per sé sia rappresentabile in forma analitica come $y=f(x)$, qual è l'approccio del biologo? È un approccio esclusivamente osservativo e procede nell'analisi della realtà tramite esperimenti, cioè va a vedere quali sono i campioni, cerca di caratterizzare al meglio il fenomeno che gli si presenta davanti e a partire dai risultati procede per classificazioni.
In realtà il biologo in genere, non si focalizza su una potenziale definizione della funzione $f(x)$ ma procede per classificazioni. Questo perché, quando si ha a che fare con fenomeni biologici, ben difficilmente si riesce a esprimere un fenomeno sotto una precisa forma analitica, dato che i fenomeni biologici sono talmente complessi che non ha molto senso concentrarsi sulla $f(x)$.
Invece l'informatico ha un approccio di tipo diverso, in quanto l'informatico procede andando a sviluppare dei modelli astratti, su algoritmi che lavorano in un certo modo sui dati, quindi tramite l'utilizzo di programmi che possono lavorare sui dati ottenuti.
In informatica quello che si cerca di fare è applicare un approccio in cui si cerca di tirar fuori un modello della realtà, per cui invece di andare a concentrarsi sul dato specifico, ci si focalizza su come il dato debba essere elaborato secondo quale procedura e modelli. Quindi, quando si scrive un programma, si stabilisce una procedura la quale, a partire dal valore della 'x', permette di ottenere dei risultati. Per cui non si procede per campionamento ma avendo in mente un modello. Per cui per l'informatica, gli specifici valori di 'x' e 'y' non sono importanti.
Quindi mettere insieme il mondo della biologia e quello dell'informatica a volte può essere complicato. Per cui noi ci occuperemo di fenomeni biologici in cui è possibile andare a ricorrere a strumenti tipici dell'informatica.
Quindi si parte da una situazione in cui si osserva il fenomeno biologico e si cerca di astrarre dei modelli e delle procedure operative, per poter elaborare l'informazione che ci viene da questi esperimenti. Per cui dobbiamo lavorare con l'informazione biologica.
Questo può essere fatto a vari livelli, infatti possiamo individuare diversi livelli di astrazione nei sistemi biologici che andiamo ad analizzare e in particolare si può partire dal basso, andando a individuare nel GENI la locazione in cui è contenuta l'informazione che a noi interessa e su cui dobbiamo andare a operare. Il livello di astrazione superiore è quello delle BIOMOLECOLE, in cui prende forma l'informazione contenuta nei geni. Le biomolecole interagiranno tra di loro secondo determinati meccanismi, ed è importante andare a capire quali sono questi meccanismi e quali sono le forze fondamentali che regolano l'interazione tra molecole e come è possibile andare a studiare questa interazione da un punto di vista informatico. A livello superiore troviamo le CELLULE, poi la loro aggregazione ci forma i tessuti, e successivamente troviamo gli apparati e infine il corpo intero.

Per quanto riguarda i geni, e in generale le sequenze di DNA, la parte dell'informatica che può essere di aiuto è l'ANALISI DELLE SEQUENZE, cioè esistono tutta una serie di algoritmi che sono stati elaborati che ci permettono di tirar fuori tutta una serie di informazioni dalle sequenze e soprattutto nel confrontare varie sequenze di nucleotidi o aminoacidi.
A livello superiore troviamo la strutture molecolari, il moto delle molecole e la loro interazione (su tale argomento diremo molte cose). Quando si studiano questi fenomeni, ricorrere a un approccio

analitico è impensabile, perché la complessità del problema è tale che l'unico approccio possibile è quello SIMULATIVO. Nel procede con un approccio simulativo però, ci scontriamo con una quantità di problemi enorme, cioè con problemi computazionali, per cui dobbiamo riuscire a capire gli strumenti informatici fino a che punto possono essere utilizzati e quali sono gli approcci che devono essere utilizzati per poter ottenere dei risultati.
A livello ancora superiore possiamo sviluppare dei modelli di interazione tra cellule, e quindi applicare modelli simulativi di questi fenomeni.
Andando a livelli ancora superiore ci sono le varie problematiche di biomeccanica e poi tutta la parte di controllo per un eventuale chirurgia o robotica.

Noi ci focalizzeremo fondamentalmente sui primi due livelli, gli altri livelli non ci interessano.

Argomenti del corso

→ La prima cosa che vedremo è il linguaggio di programmazione Python.

→ Poi andremo a parlare in modo più specifico di algoritmi in generale, questo perché avremo a che fare con diversi tipi di algoritmi durante tutto il corso.

→ Poi vedremo una parte molto importante del corso che è quella dell'analisi delle 'sequenze' in generale, da un punto di vista informatico poco importa se abbiamo una sequenza di lettere quali ATCG (sequenza di nucleotidi) oppure qualsiasi altro insieme di simboli. È chiaro che poi quando andiamo ad applicare ciò che sviluppiamo, il nostro campo applicativo è quello biologico, per cui non ha senso approcciare il problema in modo totalmente teorico, ma dovremo vedere subito quali sono le applicazioni in campo biomedico. In prima istanza avremo a che fare con sequenze di DNA, successivamente potremo anche avere sequenze di RNA oppure sequenze di proteine.

→ Un altro argomento importante è la 'MODELLAZIONE MOLECOLARE', come avevamo detto prima, sul livello di astrazione direttamente superiore a quello dei geni, si trovano le biomolecole, le quali sono strutture complesse, per cui dovremo capire come andare a fare la loro modellazione. La loro modellazione viene fatta andando a ricorrere alla 'Molecular Mechanics', ovvero immaginiamo che le molecole siano assimilabili a dei corpi meccanici e dovremmo andare a capire come definire questi modelli. Una volta che li abbiamo definiti, ci interessa capire anche come queste molecole possono evolvere nel tempo, anche perché non sono oggetti statici, ma dinamici e inoltre qualsiasi attività biologica è legata al movimento di queste molecole. Il movimento può essere di una singola molecola o di un complesso di molecole, oppure di interazione tra più molecole. Per cui ci interessa capire come studiare il moto di queste molecole.
Tutte queste cose fanno parte di un'area di studio detta 'BIOLOGIA COMPUTAZIONE DI TIPO STRUTTURALE', in quanto ci si concentra sulle strutture delle molecole.

→ Un altro argomento è lo studio del 'Docking', cioè riuscire a capire come più molecole si incastrano l'una con l'altra. Questo è importante perché ci fa capire come avvengono le interazioni tra proteina e proteina, oppure tra particolari molecole più piccole, come quelle dei farmaci, e le proteine stesse. Lo studio di queste interazioni può essere fatto con i campioni, come normalmente viene fatto nella ricerca in vitro, oppure può essere fatto con l'ausilio di strumenti informatici, i quali cercano di capire qual è effettivamente la struttura delle molecole e come ci si può lavorare.

→ Un altro campo di interesse, ma che sicuramente noi non riusciremo a sviluppare, è detto 'System Biology', ovvero la modellazione delle interazioni tra le varie sostanze in ambito cellulare. Cioè si vuole cercare di capire come le varie sostanze effettivamente interagiscono tra di loro e possono portare a determinati fenomeni macroscopici a livello cellulare.

MASTER COPY

Tel. 388/9837745

Cosa ci prefiggiamo di ottenere da questo corso?
Speriamo di arrivare a capire quali sono le varie tematiche della bioinformatica, e arrivare a maneggiare strumenti software indispensabili per lo studio di problematiche biologiche. Inoltre questi strumenti spesso devono essere utilizzati in concerto, per ottenere determinati risultati, quindi bisogna saper mettere i vari strumenti utilizzati insieme.

Il quadro da cui partiamo è il seguente, possiamo partire dalle seguenti considerazioni: il dogma centrale della biologia molecole sono le sequenze di DNA. Quindi, stiamo cercando di capire qual è il flusso dell'informazione dal DNA alla funzione svolta a livello cellulare. Noi siamo interessati all'informazione che è reperibile ai diversi livelli, e a seconda della forma di questa informazione, lungo il percorso, dovremo utilizzare strumenti diversi.
Si parte dalla sequenza di DNA, la quale si sa che può essere replicata, attraverso la trascrizione otteniamo l'm-RNA, poi l'informazione che è ivi contenuta viene trasferita in modo da avere la traduzione in sequenze di proteine:

DNA → *trascrizione* → m-RNA → *traduzione* → PROTEINE → *folding* → STRUTTURA PROTEINE

L'informazione contenuta nell'm-RNA e nelle proteine è essenzialmente la stessa, però c'è qualcosa di strano, in quanto una volta che le sequenze di proteine sono state ottenute, queste catene si ripiegano attraverso il folding, per ottenere una macromolecola e la funzione di questa macromolecola dipende dalla struttura tridimensionale della molecola stessa. Quindi nella prima parte del flusso riportato sopra, possiamo andare a lavorare con degli strumenti che devono lavorare su sequenze di simboli, in vario modo. Nella parte che segue invece, dobbiamo applicare strumenti diversi in quanto abbiamo a che fare con strumenti tridimensionali e l'informazione sta nella forma di questi oggetti tridimensionali.
In realtà questa ultima parte del flusso subisce dei condizionamenti forti dall'ambiente in cui avviene, per cui per studiare come vanno le cose qui, non si può partire esclusivamente dall'analisi delle singole componenti della molecola, ma bisogna capire in quale contesto operativo questa molecola viene collocata.
Alla fine si dice che, dalla sequenza si arriva alla struttura, e la struttura determina la funzionalità della biomolecola. Questo quadro però non è così semplice, ma è molto più complesso, già a partire dall'inizio. Infatti non si passa tranquillamente dal DNA all'm-RNA, ma ci sono dei processi che intervengono, nel senso che prima si otterrà una catena attraverso la trascrizione, e per ottenere poi l'm-RNA si applica il meccanismo di splicing, in cui certe porzioni vengono prese ed eliminate. L'm-RNA poi potrà essere tradotto in proteina, ma anche in questo caso, il folding non è un processo lineare come è stato descritto sopra, ma è estremamente complicato. Prima si ha un precursore della proteina che una volta messo nel suo ambiente operativo, subirà delle modifiche particolare, e solo quando queste modifiche post-trasduzionali sono state terminate, effettivamente la proteina assumerà la sua struttura dalla quale poi sarà determinata la sua funzione. In questa fase può accadere di tutto, nel senso che ci può essere l'interazione con altre sostanze e non soltanto con il