

Dr.ssa Monica Scali

*Università degli Studi di Siena
Dipartimento di Scienze Ambientali
"G. Sarfatti"
Via P. A. Mattioli, 4
53100 Siena*



La proteomica e le prospettive del miglioramento genetico della vite



Analisi proteiche:

- ◊ impiego nella realizzazione di programmi di miglioramento genetico nella vite
- ◊ ottimo sistema di controllo sulla qualità del prodotto

Programma di lavoro:

Le proteine: cosa sono e a cosa servono;
Metodi per l'analisi delle proteine (elettroforesi mono e bidimensionali);
Caratterizzazione varietale mediante protein fingerprinting;
Stato di avanzamento della coltura in vitro della vite e analisi del pattern proteico.

La vite è la pianta da frutto più coltivata nel mondo e la più importante in termini economici. I suoi frutti sono utilizzati per molteplici usi (frutta fresca, succhi, frutta secca), ma soprattutto sono la base di prodotti di alto valore economico come il vino, il Vin Santo e la grappa.

Le cultivars di vite dedicate alla produzione di vino derivano dalla selezione di genotipi di origine antica, generati prevalentemente da incroci spontanei centinaia di anni fa.

Cultivar o varietà: prodotto di un singolo semenzale o di un individuo moltiplicati per via vegetativa. Durante questi cicli moltiplicativi possono comparire delle variazioni genetiche spontanee e la cultivar risulta allora costituita da una serie di cloni che mantengono lo stesso nome dato che il loro fenotipo è molto simile. Quando la variazione interessa un carattere tecnologico importante il clone viene riconosciuto come una forma differente della cultivar originale (es. Pinot bianco, Pinot grigio probabili mutazioni spontanee del Pinot nero,)





Un po' di botanica per capire la vite

Tutto si riconduce alla famiglia botanica delle *Vitaceae* a cui appartiene il genere *Vitis*, che si suddivide in due sottogeneri: *Muscadina* ($2n=40$) e *Vitis* (*ex Euvitis*) ($2n=38$) scarsamenti interfertili i quali contengono specie differenti.

La specie europea (appartenente a Euvitis) è la *Vitis vinifera* suddivisa a sua volta in due sottospecie, la *sativa* (quella coltivata) e la *silvestris* (significa selvatica) da cui si sono originate diverse migliaia di varietà coltivate per produrre i vini migliori.

Se la *Vitis vinifera* è fondamentale per la produzione della tipologia di uva, a bacca bianca e a bacca rossa, altrettanto importanti sono altre specie, soprattutto di origine americana: questi generi, resistenti alla fillossera, hanno infatti salvato la viticoltura europea e vengono utilizzati come portinnesti.

Variabilità relativamente grande tra i vitigni coltivati di molti caratteri della pianta (foglie, germogli, acini, grappoli)

Ampelografia

Descrizione oggettiva (differenze morfologiche) delle caratteristiche dei vitigni in modo da renderne possibile la caratterizzazione varietale.

Limiti:

Analisi ristretta al periodo vegetativo e non riguarda i portinnesti di un vigneto

Influenza sul fenotipo di condizioni ambientali, nutrizionali e sanitarie

15000 cultivar di vite esistenti nelle collezioni di tutto il mondo (difficile identificazione e distinzione morfologiche)

Richiesta persone esperte

Sviluppo di nuove tecniche

Isoenzimi

DNA
Protein fingerprinting



Requisito essenziale per la moderna industria viti-vinicola

Identificazione affidabile dei materiali viticoli

E' indubbio che un corretto riconoscimento dei materiali di moltiplicazione contribuisca a garantire la qualità nella filiera viti-vinicola

in Italia curiosità crescente degli operatori verso i servizi di Identificazione varietale

Marcatore molecolari (MS) per l'identificazione di cultivar, verifica di sinonimia e omonimia, ricostruzione dei pedigree, studi delle relazioni genetiche tra i vitigni, origine geografica dei vitigni. AFLP potenzialità per individuare differenze minori tra genotipi come quelle clonali

Sempre più frequente che il profilo di marcatori molecolari sia inserito nella documentazione che accompagna la domanda di iscrizione di un vitigno al Catalogo Nazionale.



Basi fisiologiche, ormonali e molecolari di molti processi dello sviluppo della vite sono poco conosciute nonostante la loro importanza per la qualità delle produzioni.

Analisi dell'espressione genica

Contribuisce a fornire nuove conoscenze su alcune vie metaboliche importanti e sulle reazioni della pianta all'attacco di patogeni

Questo approccio si svilupperà notevolmente nel prossimo futuro con la sempre crescente disponibilità di nuove informazioni di sequenza della vite la sequenza del suo intero genoma e le nuove piattaforme per l'analisi dell'espressione genica potranno essere sfruttate appieno (microarray)

Programmi di sequenziamento del genoma della vite



International Grape Genome Program

Obiettivo principale: sequenziare il genoma della vite (IGGP).

Aspettative: le nuove conoscenze sul genoma aiuteranno a conoscere a fondo la biologia della vite, la viticoltura e l'enologia. Francia, Sud Africa, USA, Italia, Spagna, Cile, Australia, Germania



Grapevine genome organisation



l'institut national
de la recherche agronomique



Università



ANALYSIS OF THE GRAPE GENOME

ROBERT HENRY¹, Effie Ablett¹, Slade Lee², Maurizio Rossetto², George Seaton², Michael Staley²

¹ Centre for Plant Conservation Genetics, Southern Cross University, PO Box 157, Lismore NSW 2480, Australia

² Australian Agriculture Research Institute, Southern Cross University, PO Box 157, Lismore NSW 2480, Australia

The grape genome is relatively small (approximately 500Mbp) making it an attractive target for genomics. The grape genome is thus similar in size to the rice genome but is composed of a larger number of small chromosomes amenable to physical mapping. Grapes are a woody perennial and thus differ from other species that have been analyzed in detail.

We have sequenced cDNA libraries from vegetative tissues and fruits to generate an EST database. Analysis of the first 5000 sequences (2500 from each library) indicated that most ESTs obtained have homology to known genes. The clones had homology to plant genes (68%) and to genes from other organisms (14%).

Approximately 18% did not show homology to any sequence of known function in the databases.

The two libraries had 275 genes in common. Redundancy was 42% in the leaf library and 39% in the berry library. Functional analysis may identify genes controlling many unique traits in this.

Libro bianco IGGP spiega in dettaglio gli obiettivi della ricerca, i benefici e i vantaggi

International Grape Genome Program

www.vitaceae.org

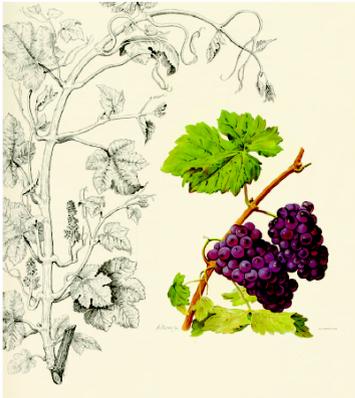


TABLE OF CONTENTS

Summary	1
I. Introduction	2
II. General Scientific Goals	4
III. Specific Objectives	4
1. Genome Analysis	4
2. Functional Analysis	7
3. Resource Centers	7
4. Bibliometrics	8
5. Workshops and Symposia	9
6. Social and Legal Issues	9
7. Organization and Implementation of the Multinational Grape Genome Research Program	10
Appendix 1. Steering Committee Members	11
Appendix 2. List of Current Working Groups and Current Plans	13
Appendix 3. Status of Grapevine Genome Analysis in Several Countries as of January 2002	14

A note about the cover: The image on the cover is from *Angledigraphis* by P. Fuchs & J. Fournier published at the beginning of the 18th century between the years 1741-45. These magnificent volumes contained state-of-the-art knowledge on the biology of the grapevine and the varieties. In the 17th century, one hundred years later, a new chapter in the understanding of grapevine biology is about to begin through the use of molecular technologies to unravel the structure of the grapevine genome.

White Page pdf

July 2002

Volume 1.1

1



Scopo del sequenziamento del genoma vite:

Capire le basi genetiche e molecolari di tutti i processi biologici rilevanti per la vite. Questo è fondamentale per sfruttare in modo efficiente le risorse biologiche della vite nello sviluppo di nuove cultivars con migliori qualità e ridotti costi economici ed ambientali.

Queste conoscenze sono inoltre fondamentali per lo sviluppo di nuovi sistemi diagnostici per vigneti e cantina.

Tratti considerati di primario interesse sono la resistenza a stress biotici (patogeni) ed abiotici, tratti qualitativi per uva da tavola e da vino, tratti riproduttivi in grado di determinare una migliore resa.



Il rapido sviluppo di tecnologie ed applicazioni genetiche degli ultimi anni e il recente sequenziamento del genoma di *Arabidopsis* e riso hanno fornito i mezzi e l'informazione comparativa per una caratterizzazione dettagliata del genoma della vite.

L'informazione funzionale derivata da *Arabidopsis* offre inoltre un sistema modello per le analisi funzionali dei geni della vite.

Sarà possibile identificare ed analizzare da un punto di vista funzionale geni di vite implicati nel controllo di importanti tratti qualitativi, facilitando il miglioramento della vite per caratteri importanti, quali la resistenza alle malattie o la qualità della bacca.

L'analisi dell'espressione genica permetterà di capire come la pianta interagisce e risponde all'ambiente circostante ed alle pratiche culturali adottate.



La vite è particolarmente adatta agli studi di genomica. Si tratta di una pianta diploide che può essere facilmente incrociata. Il suo genoma è piccolo di circa 500 Mb, equivalente a 3-4 volte il genoma di *Arabidopsis*.

La vite può essere considerata come una pianta rappresentativa di un gruppo di piante decidue, legnose e perenni che producono frutti polposi contenenti composti del metabolismo secondario responsabili del colore, della fragranza e dell'aroma.



Un articolo pubblicato su Science (2002, Apr 5;296(5565):79-92) ha riportato la notizia del sequenziamento del genoma nucleare di *Oryza sativa* ssp. *indica*, la sottospecie più diffusa e coltivata in Cina. Il genoma di riso è di circa 440 Mb con un numero di geni variabile tra 46000 e 55000.



PIANTA MODELLO

Il Progetto Genoma *Arabidopsis thaliana*, che ha avuto inizio nel 1996 ed ha coinvolto 15 laboratori di Unione Europea, Stati Uniti e Giappone, ha portato alla pubblicazione nel 1999 della sequenza dei cromosomi 2 e 4, e si è concluso nel dicembre 2000 con la pubblicazione della sequenza completa dei cromosomi 1, 3 e 5. *Arabidopsis thaliana* (anche denominata arabide comune) è una pianta annuale che cresce nei sentieri o sui muri dei giardini e fa parte della stessa famiglia del cavolo e della senape.

Perché *Arabidopsis*
Di piccole dimensioni.
Ciclo vitale corto.

Prodigiosa produzione di semi.
Facile da propagare in laboratorio.

Con un genoma relativamente piccolo, circa 125 Mb. I genomi di mais e grano sono di 2.500 e 16.000 Mb rispettivamente.





Un obiettivo fondamentale da tener sempre presente nel corso del programma "genoma vite" è lo sviluppo e l'uso di metodi in grado di definire le funzioni geniche sia a livello molecolare che di organismo.

E' necessario poter disporre di sistemi di analisi in grado di aiutare a capire le funzioni molecolari dei geni, la regolazione e l'interazione genica.

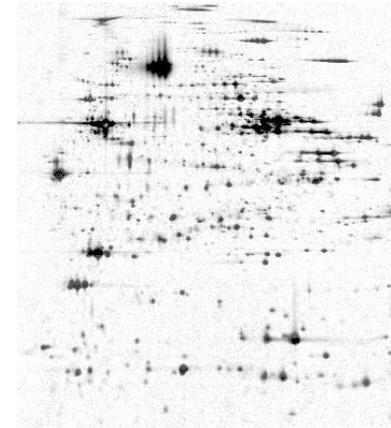
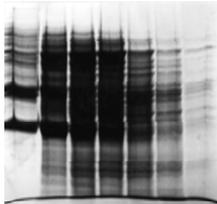
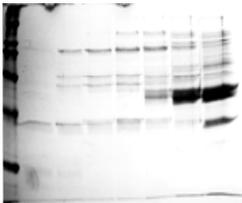
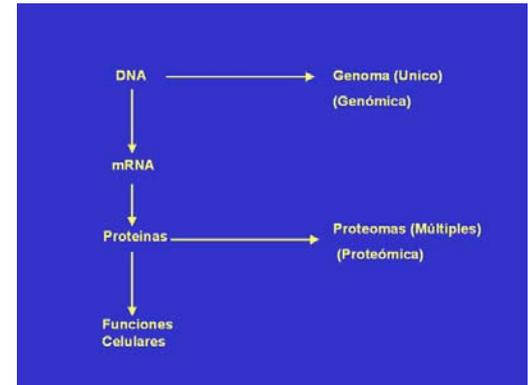
Questi includono la generazione e la collezione di mutanti e di varianti naturali, lo sviluppo di sistemi di trasformazione transiente e stabile, il "gene knockout" mediante RNAi o "Virus induced silencing".

Lo sviluppo e l'uso di "EST microarrays" è considerato uno dei più importanti steps per correlare l'espressione genica con specifici processi biologici.

Le informazioni di espressione genica devono essere complementate da studi di **interazione proteina-proteina**, analisi sistematica del **proteoma** e dei **profili metabolici**.



Genomica e Proteomica





Cosa sono le proteine?



Le **proteine** sono sostanze organiche indispensabili, composte da Carbonio, Idrogeno, Ossigeno e Azoto che le nostre cellule utilizzano per crescere e riprodursi.

Esistono diversi tipi di proteine, ma se osserviamo le loro molecole ci accorgiamo che risultano formate dall'unione di molecole più piccole chiamate **aminoacidi**.

Venti aa sono coinvolti nella sintesi proteica e a seconda del numero e dell'ordine con cui essi si dispongono, si hanno diversi tipi di proteine. Dei venti aminoacidi **otto** sono considerati essenziali per il nostro organismo e possiamo reperirli solamente dagli alimenti.

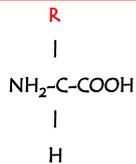


Gli aminoacidi essenziali sono 8:
leucina, isoleucina, lisina, metionina, fenilalanina, treonina, triptofano e valina.

Aminoacidi non essenziali: *alanina, acido aspartico, cistina, acido glutammico, glicina, prolina, serina, tiroxina, cisteina, arginina, istidina.*



Gli enzimi sono costituiti da proteine e la maggior parte delle proteine, ma non tutte, sono enzimi. Le proteine sono costituite da una o più catene polipeptidiche ciascuna composta solitamente da centinaia di monomeri, gli aminoacidi. Ognuno dei venti aminoacidi proteici è codificato dal codice genetico.



Tutti gli aminoacidi sono caratterizzati da un gruppo carbossilico COOH (o COO⁻) ed uno amminico NH₂ (o NH⁺) legati ad un atomo di C (Carbonio) al quale sono legati anche un atomo di H e la parte restante della molecola, R, che differisce per ciascuno degli aminoacidi.



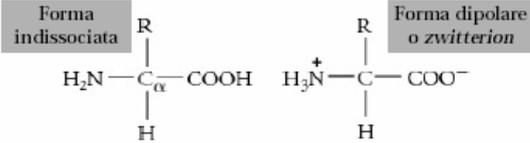
		Second letter					
		U	C	A	G		
First letter	U	UUU Phenylalanine UUC UUA Leucine UUG	UCU Serine UCC UCA UCG	UAU Tyrosine UAC UAA Stop codon UAG Stop codon	UGU Cysteine UGC UGA Stop codon UGG Tryptophan	U C A G	
	C	CUU Leucine CUC CUA CUG	CCU Proline CCC CCA CCG	CAU Histidine CAC CAA CAG	CGU Arginine CGC CGA CGG	U C A G	
	A	AUU Isoleucine AUC AUA AUG Methionine; initiation codon	ACU Threonine ACC ACA ACG	AAU Asparagine AAC AAA AAG	AGU Serine AGC AGA AGG	U C A G	
	G	GUU Valine GUC GUA GUG	GCU Alanine GCC GCA GCG	GAU Aspartic acid GAC GAA GAG	GGU Glycine GGC GGA GGG	U C A G	

Ogni aa corrisponde ad una tripletta di basi presente nel codice genetico



Struttura delle proteine

Catene lineari di aminoacidi in sequenza, numero e combinazioni variabili

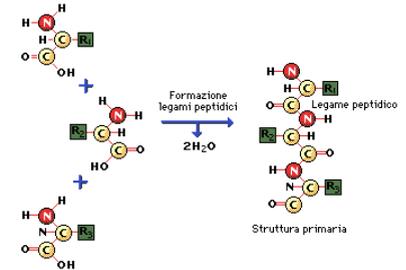
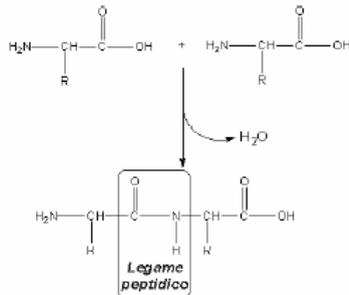


Gli aminoacidi possono unirsi fra loro con legami peptidici tra il gruppo amminico di un aminoacido e il gruppo carbossilico dell'aminoacido successivo con eliminazione di H₂O, (condensazione) formando un dipeptide, un tripeptide o una catena polipeptidica a seconda del numero di aminoacidi coinvolti, che, salvo eccezioni, non supera i 300.

Il peso molecolare medio di un aminoacido è 120 e quello dei polipeptidi varia da 10.000 a oltre 1.000.000 D. Nei polipeptidi le cariche di ciascun aminoacido dei gruppi -NH₂ e -COOH coinvolti nel legame peptidico si annullano tranne che per un gruppo amminico iniziale ed uno carbossilico terminale.



Legame peptidico



La struttura primaria di una proteina consiste in una lunga sequenza di aminoacidi legati per mezzo del cosiddetto **legame peptidico**: il gruppo carbossilico di un aminoacido (un atomo di carbonio, due di ossigeno e uno di idrogeno) si lega al gruppo amminico di quello adiacente (un atomo di azoto e due di idrogeno) con la liberazione di una molecola d'acqua.



In una proteina la sequenza lineare degli aminoacidi legati fra loro ne rappresenta la **struttura primaria**.

Le possibili combinazioni nelle sequenze sono innumerevoli; in una catena di 300 monomeri con tutti i 20 aminoacidi sono 20 elevato a 300, un numero superiore al numero di tutti gli atomi esistenti.

In natura, pur con proteine di oltre 1000 unità aminoacidiche, il numero delle combinazioni, anche se elevatissimo, è ben distante dal massimo delle possibilità (nelle piante il numero delle singole proteine è stato stimato in circa 40.000).

Questa ampia possibilità di combinazioni è anche corresponsabile della biodiversità specifica, cioè, in definitiva, dell'esistenza di un numero elevatissimo di specie animali e vegetali.

La quantità di determinati aminoacidi in una proteina determina il numero di cariche su questa e spesso la sua funzionalità.



Per ciascuna proteina esiste un punto isoelettrico (a pH intermedio) in corrispondenza del quale le cariche positive e quelle negative si compensano e la solubilità della proteina stessa è minima (Ciò permette la precipitazione delle proteine con opportuni solventi).

Una proteina può essere formata da una o più catene polipeptidiche; in alcune proteine sono presenti catene polipeptidiche identiche, in altre due o più catene diverse. Il peso molecolare di un **polipeptide** è compreso in genere tra **15.000 e 100.000 D**, ma una **proteina** può avere anche un peso di **diversi milioni di D**.



Classificazione delle proteine

Le proteine possono essere classificate in in:

1. Proteine **semplici** se costituite solo da aminoacidi.
2. **Coniugate** se alla proteina è legato un gruppo non proteico, indicato con il termine di gruppo prostetico; a loro volta le proteine coniugate in base alla natura chimica del gruppo prostetico sono distinte nelle seguenti classi:

Glicoproteine, se il gruppo prostetico è uno zucchero.

Lipoproteine, se è un lipide.

Nucleoproteine, le proteine sono complessate con acidi nucleici.

Emoproteine, la frazione non proteica è il gruppo eme, in questo caso l'esempio ci è dato dall'emoglobina

Metallo proteine, contenenti ioni metallici.

Fosfoproteine, il gruppo prostetico è l'acido fosforico.



Classificazione delle proteine in base alla solubilità

◊Insolubili in ambienti acquosi
*Fibrose e strutturali (collagene, elastina, keratina)
*Proteine di membrana (pompe ioniche)

◊Globulari e solubili in ambienti acquosi

*Catalizzatori (enzimi)
*Deposito (ferritina, caseina)
*Trasporto (emoglobina, sieroalbumina)
*Contrattili (miosina, actina)
*Protezione (anticorpi, immunoglobuline, fibrinogeno)
*Tossine
*Trasmettitori (ormoni, recettori)



La **struttura primaria** di una proteina ne determina anche la sua conformazione tridimensionale ed il ruolo nella cellula ed è strettamente collegabile alla sequenza dei nucleotidi nei geni che la codificano.

La modificazione anche di un solo aminoacido nella struttura primaria di una proteina può provocare gravissime alterazioni nell'organismo in cui avviene (la grave malattia dell'anemia falciforme è dovuta alla sostituzione di un residuo di acido glutammico con uno di valina in posizione 6 fra i 146 aminoacidi di una delle quattro catene dell'emoglobina).

Ciò non è però la regola; in molte proteine si possono verificare numerose sostituzioni di aminoacidi senza che ne venga compromessa minimamente la funzionalità; naturalmente, tali sostituzioni avvengono in regioni non critiche delle catene. Nelle catene polipeptidiche esistono pertanto alcuni aminoacidi relativamente "più importanti", che sono quelli che determinano la conformazione tridimensionale delle proteine (le quali sono molecole rigide e più compatte rispetto alla o alle catene lineari dalle quali sono formate).



Livelli di organizzazione strutturale delle proteine

Alla grande varietà di funzioni delle proteine corrisponde una grande varietà di strutture tridimensionali. Ogni proteina presenta diversi livelli di organizzazione che si integrano originando la sua conformazione tridimensionale specifica (proteina allo stato nativo).

Alcune proteine sono costituite da una unica catena polipeptidica ripiegata su se stessa, altre possono essere costituite da due o più catene identiche o meno.

Struttura primaria: è data dalla sequenza aminoacidica nella catena polipeptidica.

Ile Asn Gln Gly Phe Asp Leu Leu Arg Ser Gly



Nelle proteine si può riconoscere anche una struttura secondaria risultante dalle interazioni fra alcuni aminoacidi vicini fra loro e una struttura terziaria determinata dalle interazioni fra aminoacidi relativamente distanti.

Energie di legame nelle proteine - legami forti

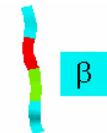
- ◊ Legami peptidici (100 kcal/mole)
 - ◊ Legami disolfuro o S-S (50 kcal/mole)
- Non vengono rotti con l'ebollizione, ma solo con l'azione prolungata di acidi o basi concentrate

Gli enzimi proteolitici sono in grado di rompere selettivamente tali legami.



Raramente le catene polipeptidiche si dispongono nello spazio assumendo una conformazione distesa, tendendo piuttosto a richiudersi su se stesse dando luogo a due tipi di conformazioni, chiamate rispettivamente **alfa-elica** e **struttura-beta**: nella prima la catena polipeptidica forma una spirale od un'elica, mentre nella seconda la catena polipeptidica può esistere in una forma più distesa per cui si ripiega su se stessa in modo che due parti della catena giacciono una accanto all'altra decorrendo parallele o antiparallele (come un foglio pieghettato).

L'alfa-elica e la struttura-beta rappresentano un grado intermedio di organizzazione strutturale che costituisce la struttura *secondaria* della proteina.





Struttura terziaria: è data dalla combinazione di più regioni ad alfa-elica e/o beta-foglietto collegate tra loro da segmenti che formano delle anse, le regioni ad ansa.

Le regioni ad ansa costituiscono in genere il sito funzionale della proteina: il sito attivo di un enzima o il sito di legame di una proteina di trasporto o di un anticorpo.

Poiché la funzione di una proteina è correlata alla sua conformazione tridimensionale, sono utilizzati dei simboli convenzionali per le strutture secondarie che permettono una rappresentazione semplificata della forma della proteina; questi simboli sono cilindri e nastri attorcigliati per le α -eliche, frecce per le strutture beta e stringhe irregolari per le regioni ad ansa.

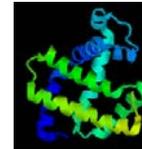
Ripiegamenti e spiralizzazioni della struttura terziaria sono determinati non solo da ponti H, ma anche da ponti disolfuri (S-S), legami ionici elettrostatici e forze idrofobiche.

Le proteine presentano dei motivi strutturali dati dalla combinazione di strutture secondarie consecutive.

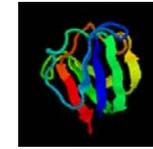
I motivi strutturali si riscontrano in proteine diverse dove hanno un significato funzionale analogo.

I domini possono essere formati interamente da alfa-eliche (**dominio tutto alfa**), interamente da foglietti-beta (**dominio tutto beta**) o da un insieme di alfa-eliche e beta-foglietti (**dominio alfa/beta**).

Dominio alfa



Dominio beta

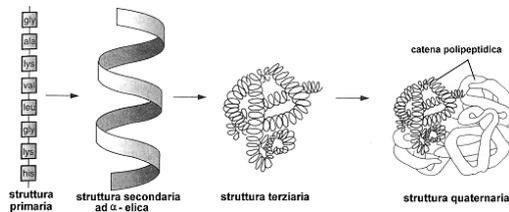


Alfa - beta



Molte proteine, infine, contengono più di una catena polipeptidica: i singoli polipeptidi si piegano a formare una struttura secondaria o terziaria, ma si legano pure uno all'altro formando la cosiddetta **struttura quaternaria**.

Struttura quaternaria: la proteina è formata da più catene polipeptidiche (subunità) unite con lo stesso tipo di legami che stabilizzano la struttura terziaria.



Livelli di organizzazione delle proteine. Struttura primaria, secondaria, terziaria e quaternaria.

In base alla conformazione esterna, si possono distinguere due categorie principali di proteine **globulari** e **fibrose**.

Gli **enzimi** sono proteine **globulari**.

La conformazione di una proteina, in particolare quella termodinamicamente più stabile, derivante dalle sue strutture (I, II, III e IV) è essenziale per la sua funzionalità e nella cellula è determinata, oltre che dalla sequenza aminoacidica, dalle condizioni di pH, temperatura, dalla forza ionica e dalle caratteristiche di eventuali molecole interagenti con la proteina stessa.

Gli aminoacidi maggiormente idrofobi tendono a concentrarsi verso l'interno della proteina, al contrario di quelli maggiormente idrofili.