

Predizione della struttura delle proteine

Predire la struttura secondaria

Le strutture secondarie sono determinate in primo luogo dalla struttura primaria che ha dei vincoli di disposizione nello spazio.

Mediamente il 50% dei residui in una proteina si trovano in conformazione strutturata classica (alpha o beta) o altre conformazioni atipiche ma ugualmente strutturate.

Le strutture beta-strand è molto contesto dipendente, essendo influenzata da beta-strands vicini che originano dei foglietti, quindi è molto difficile da predire.

Le strutture alpha-elica sono strettamente dipendenti dal susseguirsi dei residui nella catena, quindi sono più facili da predire.

La struttura secondaria è una caratteristica del BACKBONE delle proteine, non delle catene laterali.

Metodo statistico di Chou & Fasman e GOR

Si basa sulla banca dati PDB Brookhaven, che raccoglie tutte le strutture proteiche note.

- 1- Osservazione della frequenza di ogni residuo in una struttura in ogni entry del database.
- 2 - Costruzione di una tabella di probabilità per ogni residuo in ogni struttura.
- 3 - Applicazione di un algoritmo di scorrimento a finestra per trovare i nuclei di strutturazione, prima alpha, poi beta, poi loops.
- 4 - Allungamento dei nuclei nelle due direzioni per trovare le regioni strutturate

=> affidabilità: 64%

Name	P (a)	P (b)	P (turn)	f (i)	f (i+1)	f (i+2)	f (i+3)
Alanine	142	83	66	0.06	0.076	0.035	0.058
Arginine	98	93	95	0.070	0.106	0.099	0.085
Aspartic Acid	101	54	146	0.147	0.110	0.179	0.081
Asparagine	67	89	156	0.161	0.083	0.191	0.091
Cysteine	70	119	119	0.149	0.050	0.117	0.128
Glutamic Acid	151	037	74	0.056	0.060	0.077	0.064
Glutamine	111	110	98	0.074	0.098	0.037	0.098
Glycine	57	75	156	0.102	0.085	0.190	0.152
Histidine	100	87	95	0.140	0.047	0.093	0.054
Isoleucine	108	160	47	0.043	0.034	0.013	0.056
Leucine	121	130	59	0.061	0.025	0.036	0.070
Lysine	114	74	101	0.055	0.115	0.072	0.095
Methionine	145	105	60	0.068	0.082	0.014	0.055
Phenylalanine	113	138	60	0.059	0.041	0.065	0.065
Proline	57	55	152	0.102	0.301	0.034	0.068
Serine	77	75	143	0.120	0.139	0.125	0.106
Threonine	83	119	96	0.086	0.108	0.065	0.079
Tryptophan	108	137	96	0.077	0.013	0.064	0.167
Tyrosine	69	147	114	0.082	0.065	0.114	0.125
Valine	106	170	50	0.062	0.048	0.028	0.053

Agadir

*An algorithm to
predict
the helical
content of
peptides*

Non predice la struttura secondaria, ma analizza una sequenza alla ricerca di zone ad alta probabilità di formare una alfa elica.

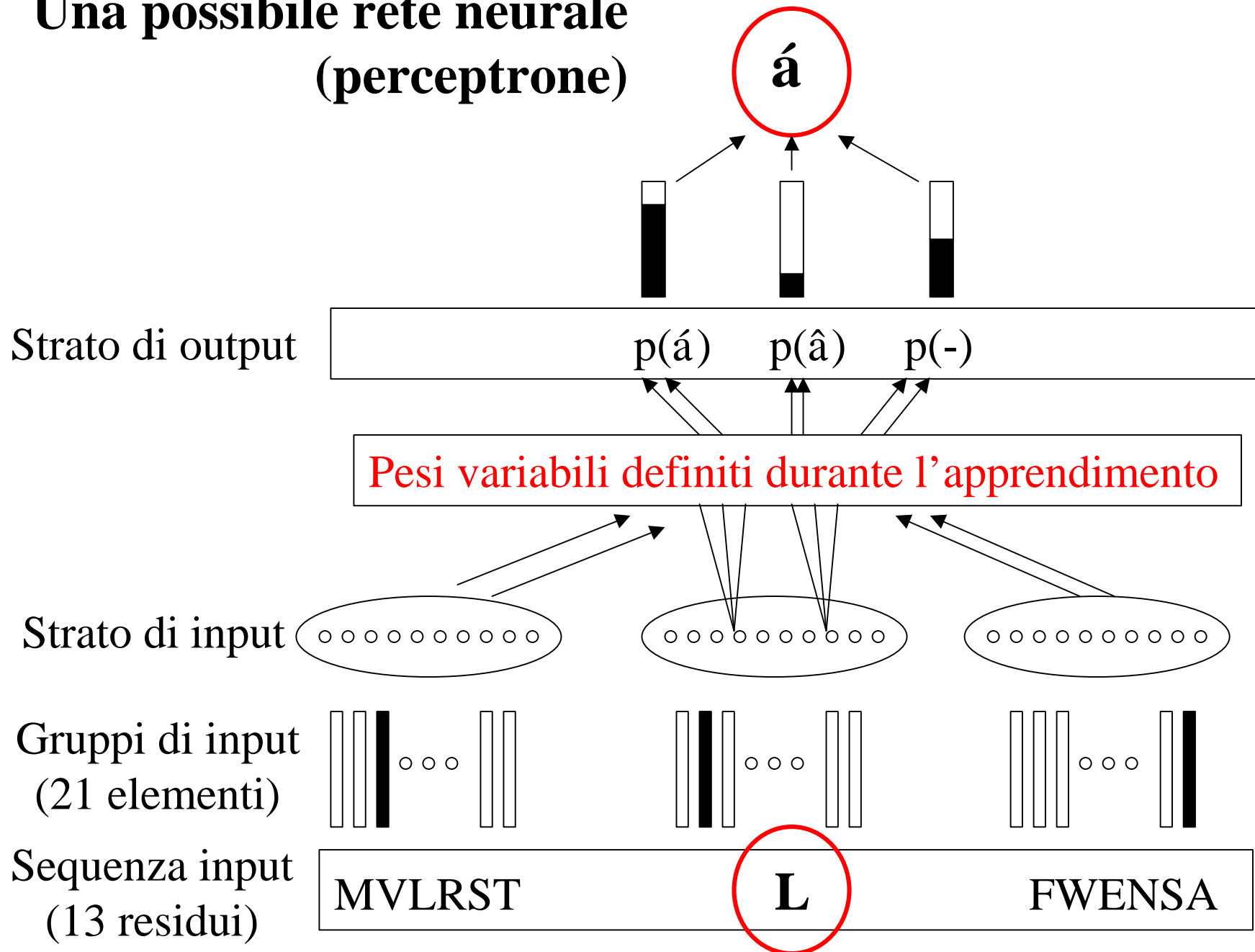
L'approccio non è statistico, ma

TERMODINAMICO, consentendo una simulazione di strutturazione alfa *in silicio*

Parameter	Value	Allowed Range
Initial pH	<input type="text" value="1.00"/>	[1.00 - 14.00]
Final pH	<input type="text" value="14.00"/>	[1.00 - 14.00]
Temperature	<input type="text" value="278"/>	[273K - 400K]
Ionic Strength	<input type="text" value="0.100"/>	[0.0 - 1.0]

50,	E,	1.4,
51,	E,	3.1,
52,	G,	3.7,
53,	Q,	16.0,
54,	V,	17.5,
55,	E,	18.3,
56,	K,	18.7,
57,	L,	19.1,
58,	M,	18.7,
59,	Q,	17.6,
60,	W,	17.1,
61,	L,	15.9,
62,	K,	9.0,
63,	S,	3.9,
64,	G,	2.0,
65,	G,	0.0,

Una possibile rete neurale (perceptrone)



PHD (Profile network from HeiDelberg)

E' basato sul concetto che se si multiallineano proteine simili, si ottengono delle conservazioni che rispettano la conservazione della struttura.

1 - La singola query viene mandata contro Swiss-Prot per trovare proteine simili

2 - Le proteine trovate vengono multiallineate

3 - Viene generato un profilo, che entra in una rete neurale (PHD)

I risultati che si ottengono sono sottoposti ad una analisi statistica per valutare l'attendibilità delle predizioni per ogni residuo.

=> affidabilità: dal 72% fino al 90%

Le due fasi principali di PHD

```

UNK_212560      SKVCIIAWVY GRVQGVGFRY TTQYEAKRLG LTGYAKNLDD GSVEVVACGE
b_ACYP_ECOLIO  SKVCIIAWVY GRVQGVGFRY TTQYEAKRLG LTGYAKNLDD GSVEVVACGE
b_ACYP_MYCTU1  ..VRLTAWVH GWVQGVGFRW WTRCRALELG LTGYAANHAD GRVLVVAQGP
b_ACYP_ARCFU2  ..IALEIYVS GNVQGVGFRY FTRRVARELG IKGYVKNLDP GRVYIYAVGE
b_ACYM_RABIT3  .....VF GRVQGVCFRM YTEGEAKKIG VVGWVKNTSK GTVTGQVQGP
b_ACYM_HORSE4  .....VF GRVQGVCFRM YAEDEARKIG VVGWVKNTSK GTVTGQVQGP
b_ACYP_DROME5  .....VF GRVQGVNFRR HALRKAKTLG LRGWCMN SSR GTVKGYIEGR
b_Y713_METJA6  .....VK GIVQGVGFRP FVYRIAKKNN LKGYVKNMGN Y.VEILIEGK
b_YE05_METJA7  .....IY GRIQHVGFRE RIENLGHALG IDGIVYNYED GTVRILANFP
b_HYPF_ECOLI8  .....IR GKVQGVGFRP FVWQLAQQLN LHGDVCNDGD G.VEVRLR..

```

```

UNK_212560      EGQVEKLMQW LKSGGPRSAR VERVLSEPHH PSGELTDFRI R
b_ACYP_ECOLIO  EGQVEKLMQW LKSGGPRSAR VERVLSEPHH PSGELTDFRI R
b_ACYP_MYCTU1  RAACQKLLQL LQG.DTTPGR VAKVVADWSQ STEQITGFSE R
b_ACYP_ARCFU2  ELTLDKFLSA VKSGPP.... . . . . . . . . . . . . . . . . .
b_ACYM_RABIT3  EDKVNSMKSU LSKVGPSSR IDR..... . . . . . . . . . . . . .
b_ACYM_HORSE4  EEKVNSMKSU LSKVGPSSR IDR..... . . . . . . . . . . . . .
b_ACYP_DROME5  PAEMDVMKEW LRTTGSPSS IEKV..... . . . . . . . . . . . . .
b_Y713_METJA6  KEDIRNFIND LKNKKPPLSR IDKLDIEEIK GIEEFDDFYI .
b_YE05_METJA7  SERKKKLFK. . . . . . . . . . . . . . . . . . . . . . . . .
b_HYPF_ECOLI8  .EDPETFLVQ LYQHCPLAR IDSVREPF I WSQLPTEFTI R

```

```

          10          20          30          40          50          60          70
          |          |          |          |          |          |          |
SKVCIIAWVYGRVQGVGFRYTTQYEAKRLGLTGYAKNLDDGSVEVVACGEEGQVEKLMQWLKSGGPRSAR
CcEEEEEEEEeeccccCcHHHHHHHHCCeeEEEEeCCCCEEEEEEcCChHHHHHHHHHHhCCCCCc
VERVLSEPHHPSGELTDFRI
eccccccCCCCccccceEeC

```


JPRED

Lancia automaticamente molti programmi per la predizione della struttura secondaria e poi crea una sequenza consenso pesata.

SeqName	PHD	DSC	PREDATOR	Mulpred	NNSSP	Zpred	Consensus	Quality	Length	Class	Fold
1FXIA	78.12	83.33	72.92	76.04	77.08	58.33	81.25	3.12	96	Alpha and beta (a+b)	beta-Grasp

```

OrigSeq      : 1-----11-----21-----31-----41-----51-----61-----71-----81-----91----- :
              ASYKVTLKTPDGDNVITVPDDEYILDVAEEGLDLPYSCRAGACSTCAGKLVSGPAPDEDQSFLDDDQIQAGYILTCVAYPTGDCVIETHKEEALY :OrigSeq
cons         : --EEEEEE--EE--HHHHHHHH-----EEEE-----HHH--EEEEEE--EEEE----- :cons
dsc          : --EEEEEE--EE--HHHHHHHH-----EEEEEEEE-----EEEEEE--EEEE----- :dsc
mul          : --EEEEEE--EE--HHHHHHHH-----EEE-----H--HHHEEEEEEE--EEHH----- :mul
nnssp       : --EEEEEE--EEEE--HHHHHHHH-----EEEEEE-----EEEEEE--EEEE--EE :nnssp
phd         : --EEEEEE--EEEE--HHHHHHHHH--E-----EEEE-----HHH--EEEEEE--EEEE----- :phd
pred        : --EEEEEE--EEEE--HHHHHHHH-----EEEE-----HHHHHHHEEEEE----- :pred
zpred       : HHEEEEE--EE--HHHHHHHHHH-----EEEEEEEEEE-----HHHHH--EEEE--EEHHHHHHHH :zpred
  
```

```

dssp        : --EEEEEE--EEEEEE--HHHHHHHH-----EEEEEE--E-----EEHHH-E--EEEE----- :dssp
define     : EEEEEEE--EEEEEE--HHHHHHHHHEEEEE-----EEEEEE-----EEEE--EEEE--EEEE----- :define
stride     : -EEEEEE--EEEEEE--HHHHHHHH-----EEEEEE-----EEE-----EEEE----- :stride
  
```

Database di strutture secondarie

DSSP (Kabsh & Sander): generato tramite il programma DSSP che ricostruisce le strutture secondarie da un file PDB in cui sono contenute informazioni tridimensionali, analizzando i ponti idrogeno.

HSSP (Sander & Schneider): nasce dall'allineamento di protine ad alta similarità, a cui viene fatto corrispondere un profilo di struttura secondaria comune. Esiste una entry hssp per ogni struttura proteica nota.

L'importanza di questi database sta nel fatto che molti metodi per la predizione li consultano per cercare delle similarità.



E' la banca dati delle strutture tridimensionali delle proteine

Si effettuano per lo più ricerche di tipo testuale, e il motore di ricerca indica quali sono le strutture trovate

Search the Archive

Enter a [PDB ID](#) or keyword

KEY: = Download compressed (GNU zipped) PDB file = View PDB file = Structure viewing options

<input type="checkbox"/> 1APS	Deposited: 20-Feb-1991 Exp. Method: NMR	{ EXPLORE }
Title	Three-dimensional structure of acylphosphatase. Refinement and structure analysis.	
Classification	Hydrolase(Acting On Acid Anhydrides)	
Compound	Acylphosphatase (E.C. 3.6.1.7) (NMR, 5 Structures)	
<input type="checkbox"/> 1GXI	Deposited: 11-Apr-2002 Exp. Method: X-ray Diffraction Resolution: 1.30 Å	{ EXPLORE }
Title	Hydrogenase Maturation Protein Hypf "Acylphosphatase-Like" N-Terminal Domain (Hypf-Acp) In Complex With Sulfate	
Classification	Phosphatase	
Compound	Mol_Id: 1 ; Molecule : Hydrogenase Maturation Protein Hypf; Chain : A; Fragment : Acylphosphatase-Like Domain, Residues 1-91; Engineered : Yes	

Le entries PDB: i file .pdb

Un file PDB è un file di testo in cui sono riportate le coordinate di tutti gli atomi di una proteina, oltre a numerose altre annotazioni di varia natura.

```
HEADER      ACYLPHOSPHATASE                               08-NOV-96      2ACY
TITLE       ACYL-PHOSPHATASE (COMMON TYPE) FROM BOVINE TESTIS
COMPND      MOL_ID: 1;
COMPND      2 MOLECULE: ACYLPHOSPHATASE;
COMPND      3 CHAIN: NULL;
COMPND      4 SYNONYM: ACP;
COMPND      5 EC: 3.6.1.7;
COMPND      6 BIOLOGICAL_UNIT: MONOMER
SOURCE      MOL_ID: 1;
SOURCE      2 ORGANISM_SCIENTIFIC: BOS TAURUS;
SOURCE      3 ORGANISM_COMMON: BOVINE;
SOURCE      4 ORGAN: TESTIS;
SOURCE      5 CELLULAR_LOCATION: CYTOPLASM
KEYWDS      ACYLPHOSPHATASE, PHOSPHORIC MONOESTER HYDROLASE
EXPDTA      X-RAY DIFFRACTION
AUTHOR      M.M.G.M.THUNNISSEN,P.NORDLUND
```

Nel file PDB possono essere inclusi dati di struttura secondaria, ma non è per questo che sono stati fatti

Esistono delle sezioni precise del file in cui inserire certe informazioni

```

HELIX      1  H1 PHE      22  LYS      32  1                                     11
HELIX      2  H2 ALA      55  THR      67  1                                     13
SHEET      1  S1 5 HIS      74  VAL      85  0
SHEET      2  S1 5 LEU       6  LYS      16 -1  N  ASP      10  O  HIS      81
SHEET      3  S1 5 VAL      47  PRO      54 -1  N  GLY      53  O  ILE       7
SHEET      4  S1 5 VAL      36  ASN      41 -1  N  VAL      36  O  GLN      52
SHEET      5  S1 5 PHE      94  VAL      97  1  N  VAL      97  O  ASN      41
TURN       1  T1 PHE      22  THR       5          TYPE II '
TURN       2  T2 PHE      14  VAL      17          TYPE IV
TURN       3  T3 VAL      17  VAL      20          TYPE IV
TURN       4  T4 THR      42  GLY      45          TYPE I
TURN       5  T5 SER      70  SER      73          TYPE I

```

Manuale dei campi dei files PDB

Tag	n° atomo	Tipo atomo	Residuo	n° residuo	Coordinata X	Coordinata Y	Coordinata Z	Flessibilità		
ATOM	1	N	ALA	1	13.196	6.710	0.033	1.00	80.57	N
ATOM	2	CA	ALA	1	11.875	6.755	-0.583	1.00	78.70	C
ATOM	3	C	ALA	1	10.858	6.436	0.487	1.00	77.39	C
ATOM	4	O	ALA	1	11.101	5.567	1.314	1.00	80.16	O
ATOM	5	CB	ALA	1	11.811	5.706	-1.691	1.00	79.51	C
ATOM	6	N	GLU	2	9.727	7.120	0.496	1.00	66.64	N
ATOM	7	CA	GLU	2	8.741	6.765	1.493	1.00	65.50	C
ATOM	8	C	GLU	2	7.398	6.494	0.835	1.00	65.06	C
ATOM	9	O	GLU	2	7.316	6.375	-0.396	1.00	62.92	O
ATOM	10	CB	GLU	2	8.748	7.578	2.788	1.00	66.75	C
ATOM	11	CG	GLU	2	9.232	6.756	3.974	1.00	67.60	C
ATOM	12	CD	GLU	2	9.572	7.604	5.177	1.00	100.00	C
ATOM	13	OE1	GLU	2	8.631	8.227	5.725	1.00	100.00	O
ATOM	14	OE2	GLU	2	10.759	7.641	5.577	1.00	100.00	O
ATOM	15	N	GLY	3	6.346	6.331	1.627	1.00	57.60	N
ATOM	16	CA	GLY	3	5.059	6.045	1.024	1.00	54.68	C
ATOM	17	C	GLY	3	4.967	4.552	0.630	1.00	51.19	C

I programmi per la visualizzazione delle strutture tridimensionali delle proteine interpretano i files PDB posizionando “sfere” di diametro adeguato nelle posizioni descritte dal files, posizionando i legami secondo i criteri imposti dal tipo di residuo (es. tutti i carboni alpha (CA) saranno legati insieme, ogni residuo avrà i suoi atomi legati tra loro ecc.).

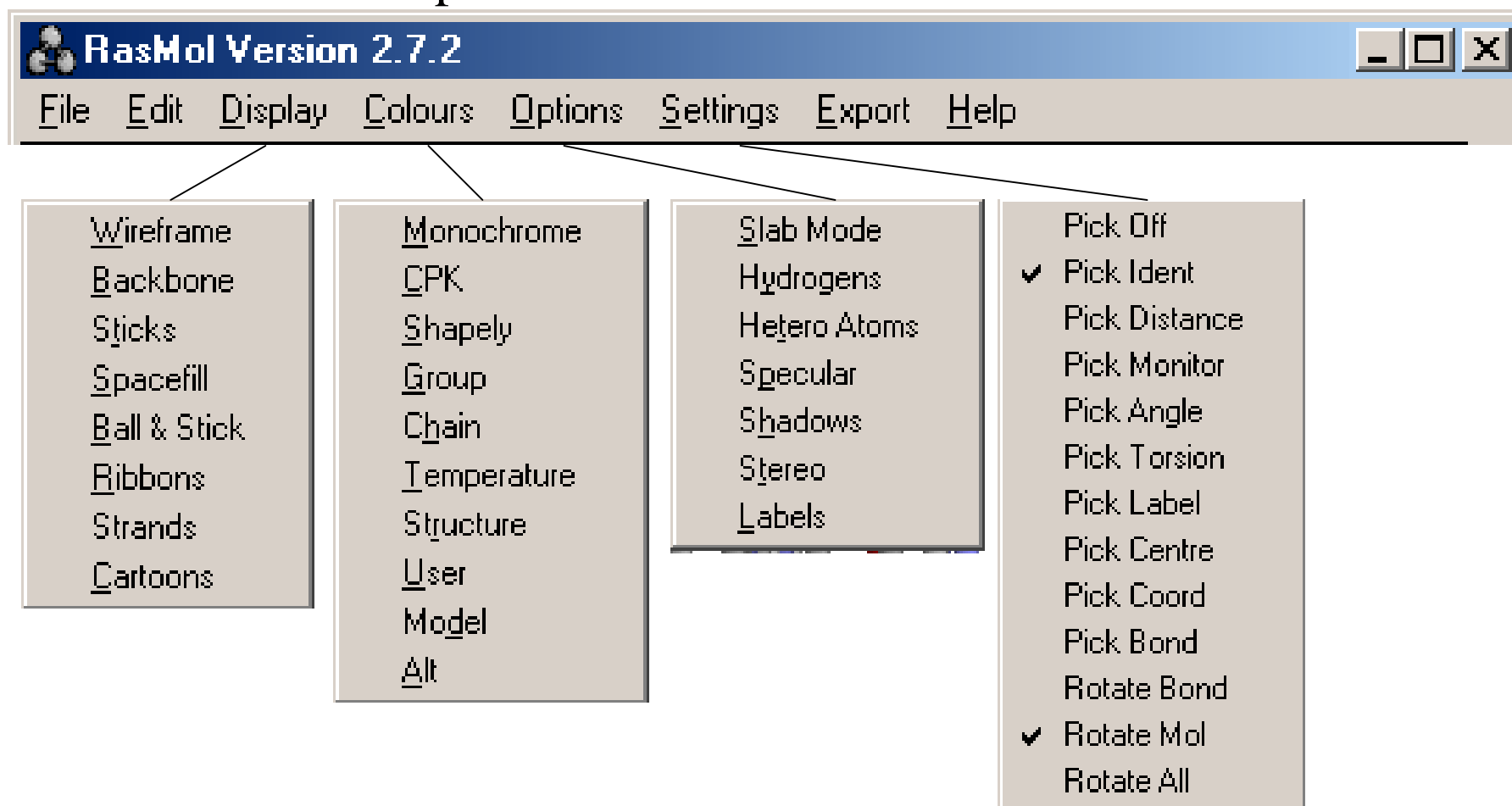
I programmi più utilizzati per la visualizzazione tridimensionale delle proteine sono

[RasMol](#) : per la visualizzazione e l’osservazione delle proteine

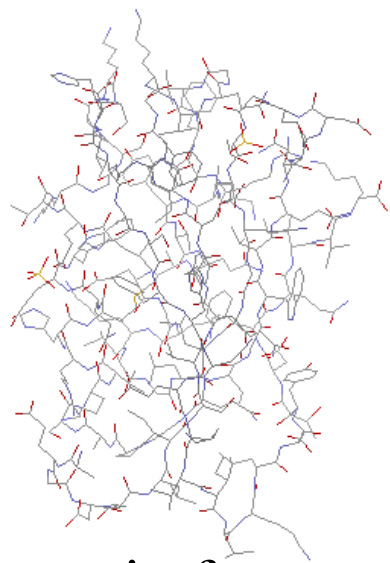
[Deep View – Swiss PDB Viewer](#): per una analisi dettagliata e per apportare modificazioni alla proteina

RasMol

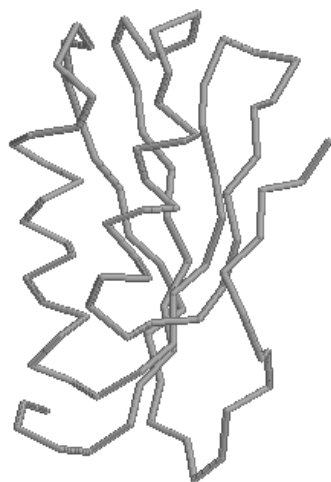
Presenta 2 finestre: la linea di comando e la parte grafica, in cui viene effettuato il rendering delle strutture. Si possono effettuare moltissime operazioni con il mouse senza scrivere nulla, ma molte altre non sono disponibili.



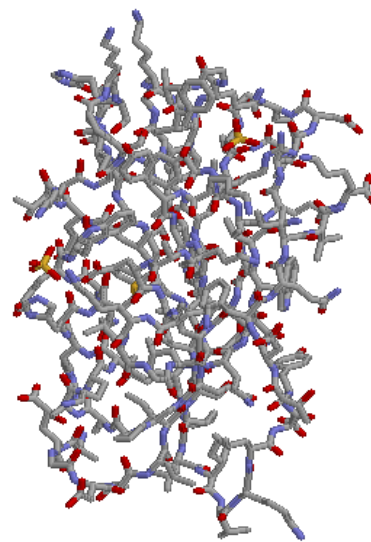
Display



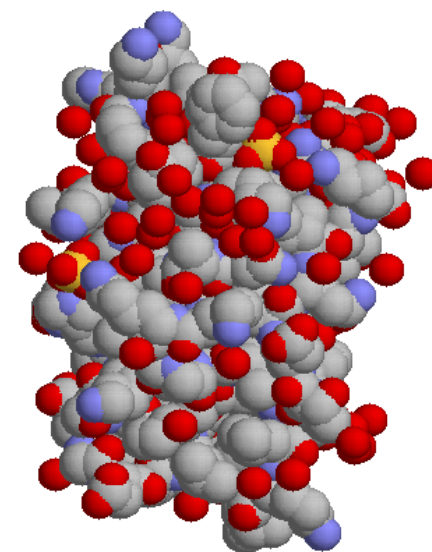
wireframe



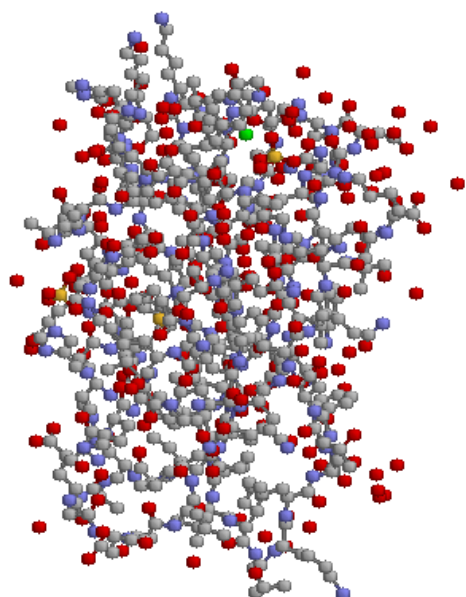
backbone



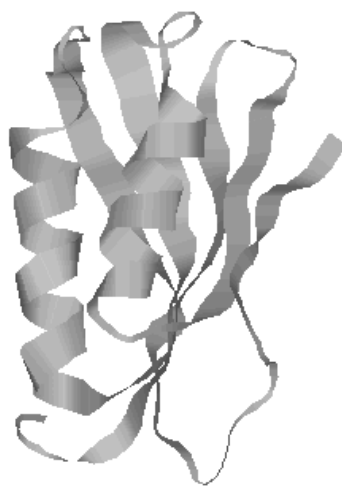
sticks



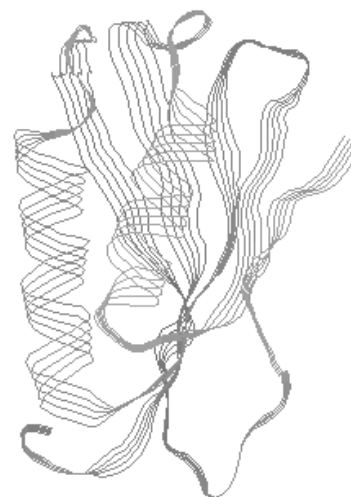
spacefill



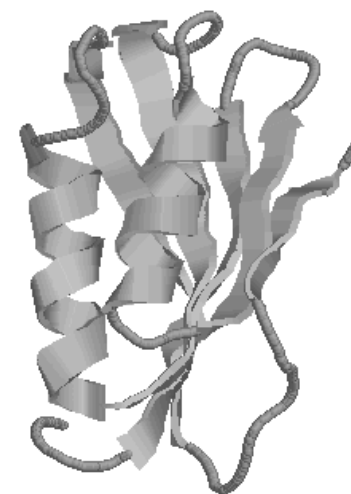
ball&stick



ribbons



strands



cartoon

Commands

<u>backbone</u>	<u>background</u>	<u>cartoon</u>	<u>centre</u>
<u>clipboard</u>	<u>colour</u>	<u>connect</u>	<u>cpk</u>
<u>define</u>	<u>dots</u>	<u>echo</u>	<u>english</u>
<u>exit</u>	<u>hbonds</u>	<u>help</u>	<u>label</u>
<u>load</u>	<u>monitor</u>	<u>pause</u>	<u>print</u>
<u>quit</u>	<u>refresh</u>	<u>renumber</u>	<u>reset</u>
<u>restrict</u>	<u>ribbons</u>	<u>rotate</u>	<u>save</u>
<u>script</u>	<u>select</u>	<u>set</u>	<u>show</u>
<u>slab</u>	<u>source</u>	<u>spacefill</u>	<u>spanish</u>
<u>ssbonds</u>	<u>star</u>	<u>stereo</u>	<u>strands</u>
<u>structure</u>	<u>trace</u>	<u>translate</u>	<u>wireframe</u>
<u>write</u>	<u>zap</u>	<u>zoom</u>	

Si possono caricare degli "script", cioè una lista di operazioni da svolgere in sequenza scritti in un unico file di testo con l'opportuna sintassi.

Le selezioni si effettuano indicando o il numero dei residui o utilizzando dei gruppi predefiniti (es. `select 10-20` seleziona i residui da 10 a 20, `select 10,20` seleziona solo i residui 10 e 20, `select arg` seleziona tutte le arginine, `select alpha` seleziona tutte le alpha eliche).

Combinando `select`, `restrict` e `colour` è possibile mettere in rilievo particolari sezioni della molecola.