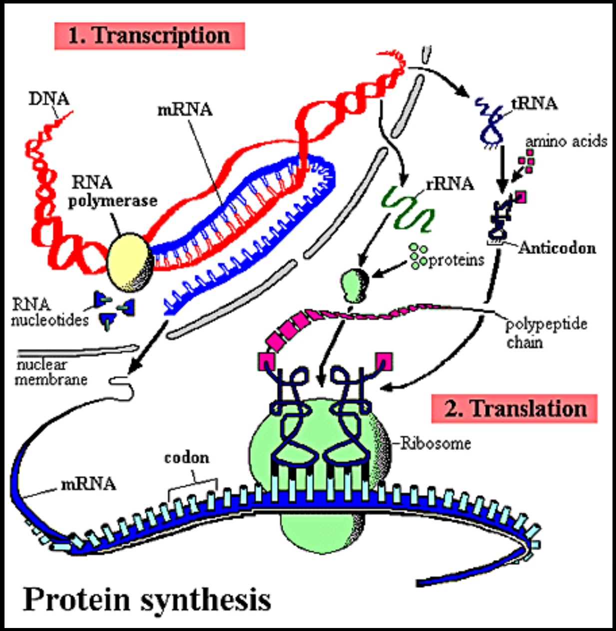
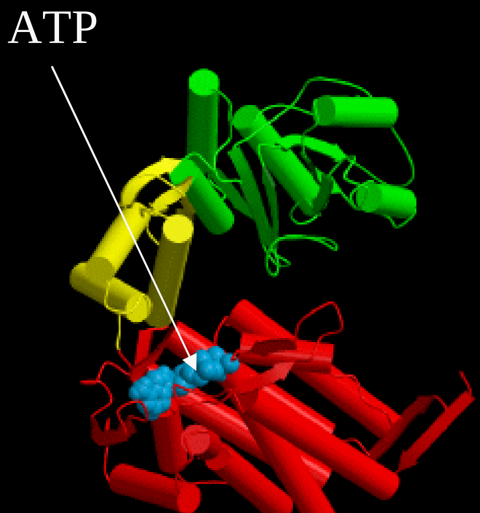
Bioinformática: Integración de información estructural y evolutiva para el estudio de la función de las proteínas.

ESCUELA COMPLUTENSE DE VERANO. Madrid.

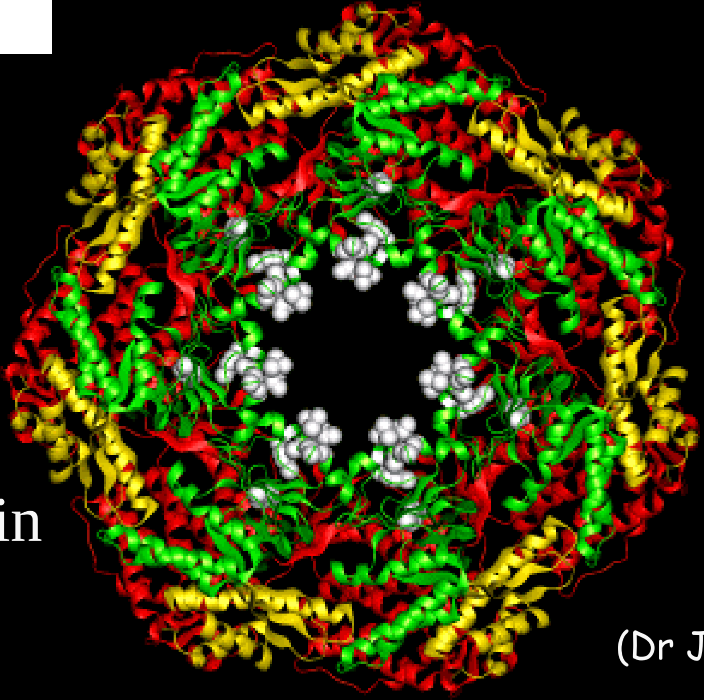
Paulino Gómez-Puertas

JuLio 2005





subunit

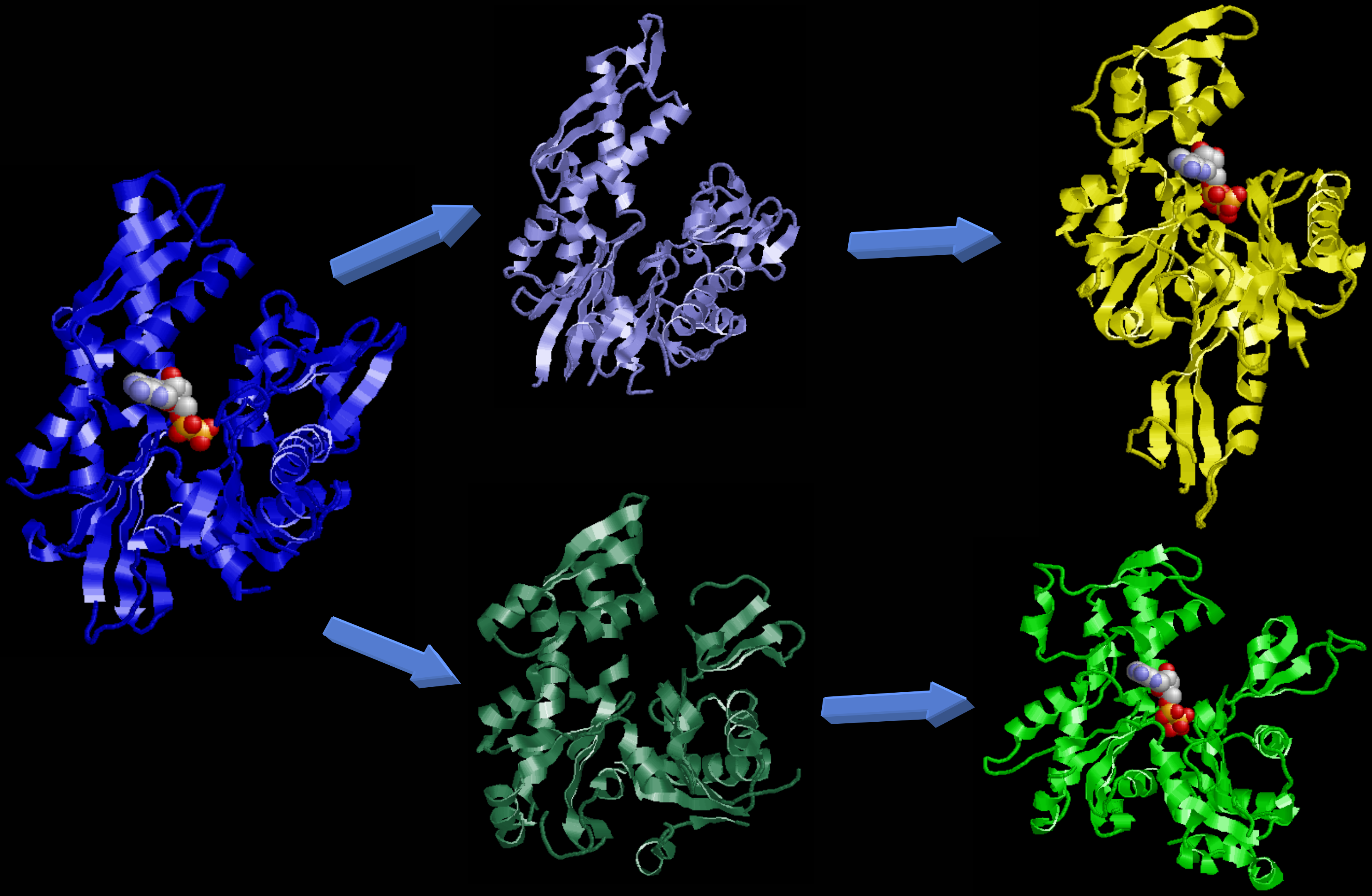


heptamer

Molecular chaperonin
GroEL

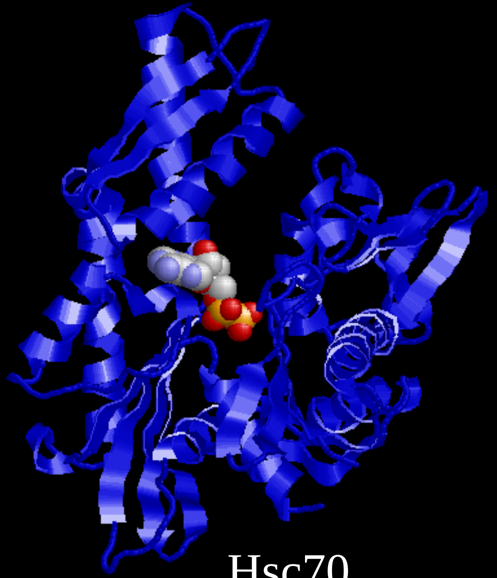
(Dr Jianpeng Ma, Harvard Univ.)

protein structure evolution

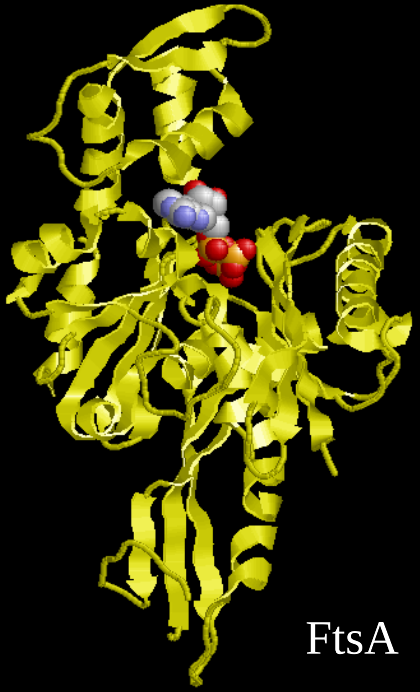


DnaK

Actin

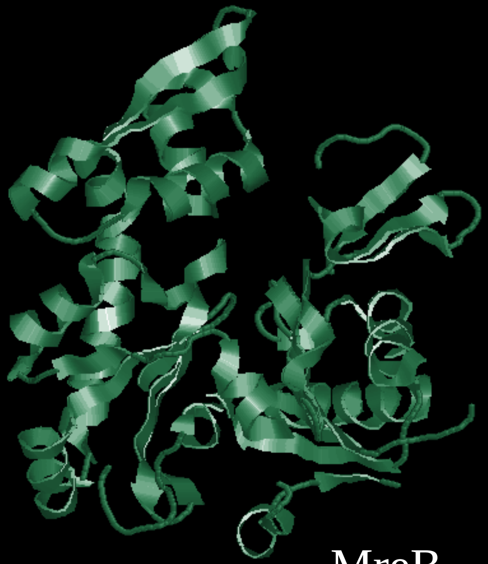


Hsc70

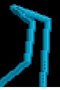


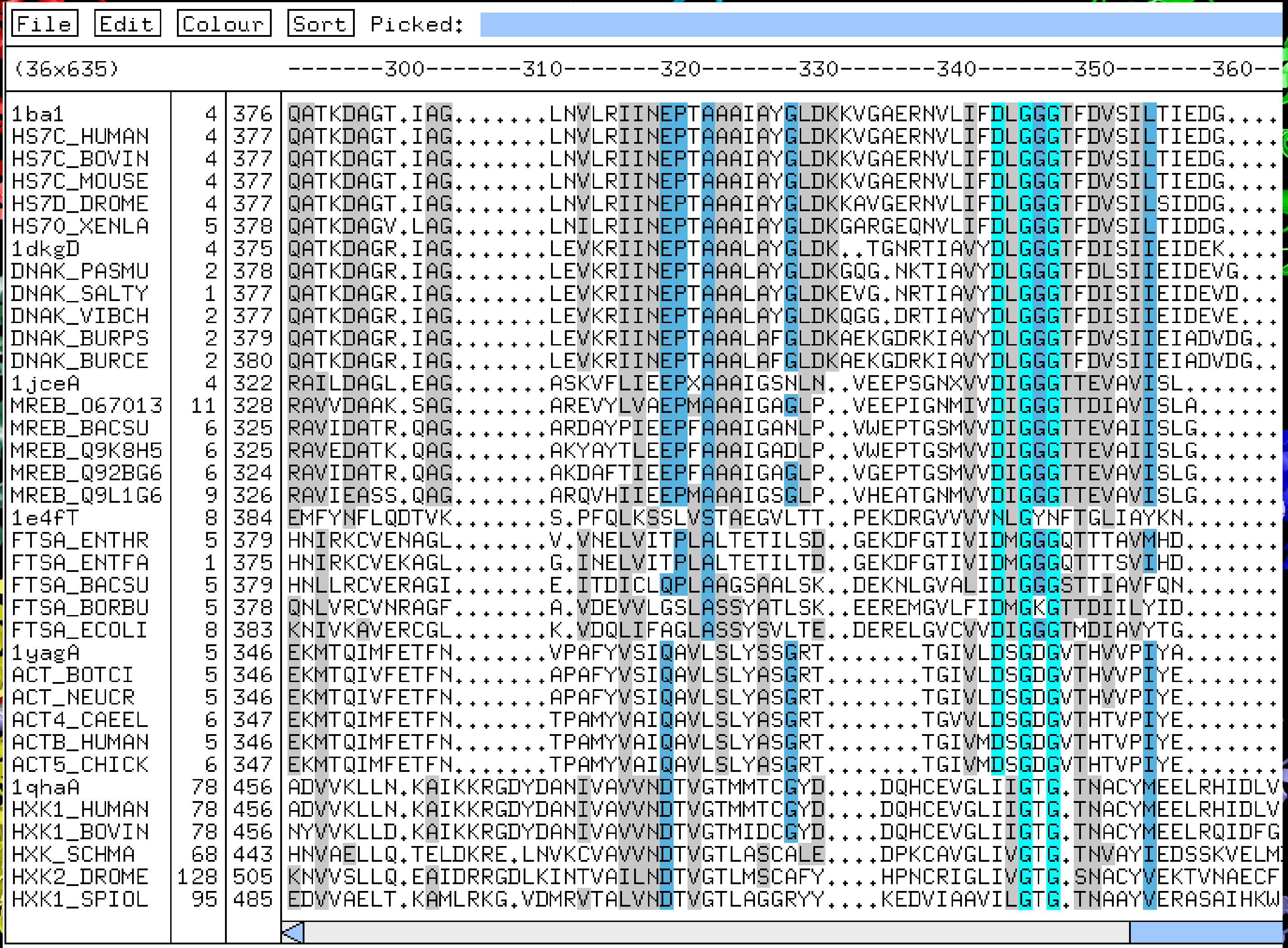
FtsA

Hexokinase



MreB





```
(36x635)          -------300-------310-------320-------330-------340-------350-------360--

1ba1          4  376 QATKDAGT.IAG.......LNVLRIINEPTAAAIAYGLDKKVGAERNVLIFDLGGGTFDVSILTIEDG....
HS7C_HUMAN    4  377 QATKDAGT.IAG.......LNVLRIINEPTAAAIAYGLDKKVGAERNVLIFDLGGGTFDVSILTIEDG....
HS7C_BOVIN    4  377 QATKDAGT.IAG.......LNVLRIINEPTAAAIAYGLDKKVGAERNVLIFDLGGGTFDVSILTIEDG....
HS7C_MOUSE    4  377 QATKDAGT.IAG.......LNVLRIINEPTAAAIAYGLDKKVGAERNVLIFDLGGGTFDVSILTIEDG....
HS7D_DROME    4  377 QATKDAGT.IAG.......LNVLRIINEPTAAAIAYGLDKKAVGERNVLIFDLGGGTFDVSILSIDDG....
HS70_XENLA    5  378 QATKDAGV.LAG.......LNILRIINEPTAAAIAYGLDKGARGEQNVLIFDLGGGTFDVSILTIDDG....
1dkgD         4  375 QATKDAGR.IAG.......LEVKRIINEPTAAALAYGLDK..TGNRTIAVYDLGGGTFDISIIEIDEK....
DNAK_PASMU    2  378 QATKDAGR.IAG.......LEVKRIINEPTAAALAYGLDKGQG.NKTIAVYDLGGGTFDLSIIEIDEVG...
DNAK_SALTY    1  377 QATKDAGR.IAG.......LEVKRIINEPTAAALAYGLDKEVG.NRTIAVYDLGGGTFDISIIEIDEVD...
DNAK_VIBCH    2  377 QATKDAGR.IAG.......LEVKRIINEPTAAALAYGLDKQGG.DRTIAVYDLGGGTFDISIIEIDEVE...
DNAK_BURPS    2  379 QATKDAGR.IAG.......LEVKRIINEPTAAALAFGLDKAEKGDRKIAVYDLGGGTFDVSIIEIADVDG..
DNAK_BURCE    2  380 QATKDAGR.IAG.......LEVKRIINEPTAAALAFGLDKAEKGDRKIAVYDLGGGTFDVSIIEIADVDG..
1jceA         4  322 RAILDAGL.EAG.......ASKVFLIEEPXAAAIGSNLN..VEEPSGNXVVDIGGGTTEVAVISL.......
MREB_O67013  11  328 RAVVDAAK.SAG.......AREVYLVAEPMAAAIGAGLP..VEEPIGNMIVDIGGGTTDIAVISLA......
MREB_BACSU    6  325 RAVIDATR.QAG.......ARDAYPIEEPFAAAIGANLP..VWEPTGSMVVDIGGGTTEVAIISLG......
MREB_Q9K8H5   6  325 RAVEDATK.QAG.......AKYAYTLEEPFAAAIGADLP..VWEPTGSMVVDIGGGTTEVAIISLG......
MREB_Q92BG6   6  324 RAVIDATR.QAG.......AKDAFTIEEPFAAAIGAGLP..VGEPTGSMVVDIGGGTTEVAVISLG......
MREB_Q9L1G6   9  326 RAVIEASS.QAG.......ARQVHIIEEPMAAAIGSGLP..VHEATGNMVVDIGGGTTEVAVISLG......
1e4fT         8  384 EMFYNFLQDTVK.......S.PFQLKSSLVSTAEGVLTT..PEKDRGVVVVNLGYNFTGLIAYKN.......
FTSA_ENTHR    5  379 HNIRKCVENAGL.......V.VNELVITPLALTETILSD..GEKDFGTIVIDMGGGQTTTAVMHD.......
FTSA_ENTFA    1  375 HNIRKCVEKAGL.......G.INELVITPLALTETILTD..GEKDFGTIVIDMGGGQTTTSVIHD.......
FTSA_BACSU    5  379 HNLLRCVERAGI.......E.ITDICLQPLAAGSAALSK..DEKNLGVALIDIGGGSTTIAVFQN.......
FTSA_BORBU    5  378 QNLVRCVNRAGF.......A.VDEVVLGSLASSYATLSK..EEREMGVLFIDMGKGTTDIILYID.......
FTSA_ECOLI    8  383 KNIVKAVERCGL.......K.VDQLIFAGLASSYSVLTE..DERELGVCVVDIGGGTMDIAVYTG.......
1yagA         5  346 EKMTQIMFETFN.......VPAFYVSIQAVLSLYSSGRT.......TGIVLDSGDGVTHVVPIYA.......
ACT_BOTCI     5  346 EKMTQIVFETFN.......APAFYVSIQAVLSLYASGRT.......TGIVLDSGDGVTHVVPIYE.......
ACT_NEUCR     5  346 EKMTQIVFETFN.......APAFYVSIQAVLSLYASGRT.......TGIVLDSGDGVTHVVPIYE.......
ACT4_CAEEL    6  347 EKMTQIMFETFN.......TPAMYVAIQAVLSLYASGRT.......TGVVLDSGDGVTHTVPIYE.......
ACTB_HUMAN    5  346 EKMTQIMFETFN.......TPAMYVAIQAVLSLYASGRT.......TGIVMDSGDGVTHTVPIYE.......
ACT5_CHICK    6  347 EKMTQIMFETFN.......TPAMYVAIQAVLSLYASGRT.......TGIVMDSGDGVTHTVPIYE.......
1qhaA        78  456 ADVVKLLN.KAIKKRGDYDANIVAVVNDTVGTMMTCGYD....DQHCEVGLIIGTG.TNACYMEELRHIDLV
HXK1_HUMAN   78  456 ADVVKLLN.KAIKKRGDYDANIVAVVNDTVGTMMTCGYD....DQHCEVGLIIGTG.TNACYMEELRHIDLV
HXK1_BOVIN   78  456 NYVVKLLD.KAIKKRGDYDANIVAVVNDTVGTMIDCGYD....DQHCEVGLIIGTG.TNACYMEELRQIDFG
HXK_SCHMA    68  443 HNVAELLQ.TELDKRE.LNVKCVAVVNDTVGTLASCALE....DPKCAVGLIVGTG.TNVAYIEDSSKVELM
HXK2_DROME  128  505 KNVVSLLQ.EAIDRRGDLKINTVAILNDTVGTLMSCAFY....HPNCRIGLIVGTG.SNACYVEKTVNAECF
HXK1_SPIOL   95  485 EDVVAELT.KAMLRKG.VDMRVTALVNDTVGTLAGGRYY....KEDVIAAVILGTG.TNAAYVERASAIHKW
```

Structural alignment

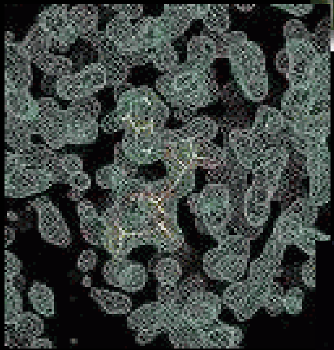
Bioinformática: Integración de información estructural y evolutiva para el estudio de la función de las proteínas.

- Predicción de estructura de proteínas.
- Extracción de información evolutiva.
- Predicción de interacciones:

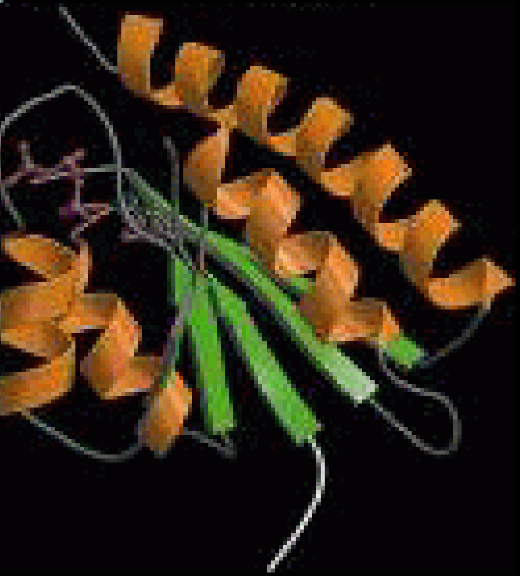
proteína - proteína

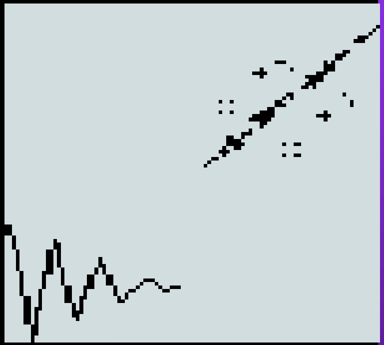
proteína-ligando

Protein structure determination

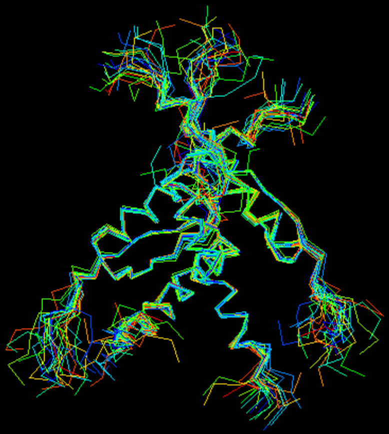


Xray crystallography

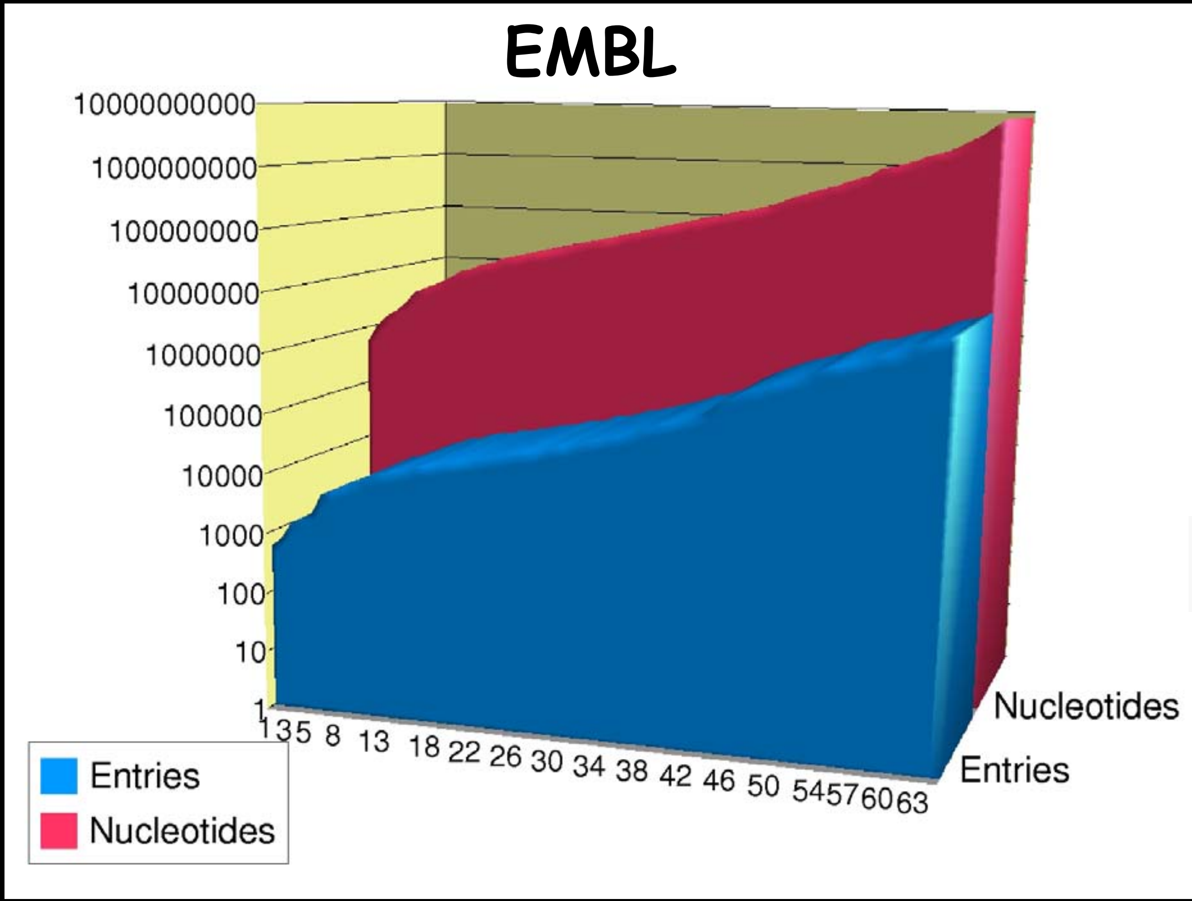


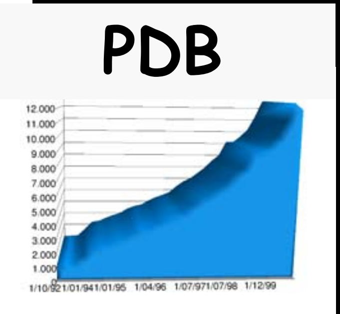


NMR



By D. Devos

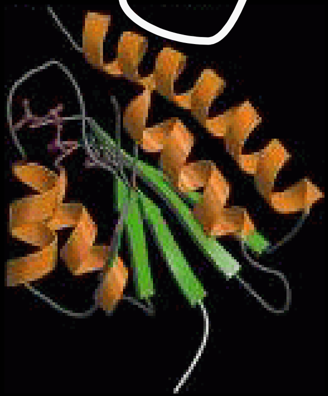

EMBL
10000000000
1000000000
100000000
10000000
1000000
100000
10000
1000
100
10
1
1 3 5 8 13 18 22 26 30 34 38 42 46 50 54 57 60 63
Nucleotides
Entries
Entries
Nucleotides


PDB
12.000
11.000
10.000
9.000
8.000
7.000
6.000
5.000
4.000
3.000
2.000
1.000
0
1/10/92 1/01/94 1/01/95 1/04/96 1/07/97 1/07/98 1/12/99

Protein structure prediction

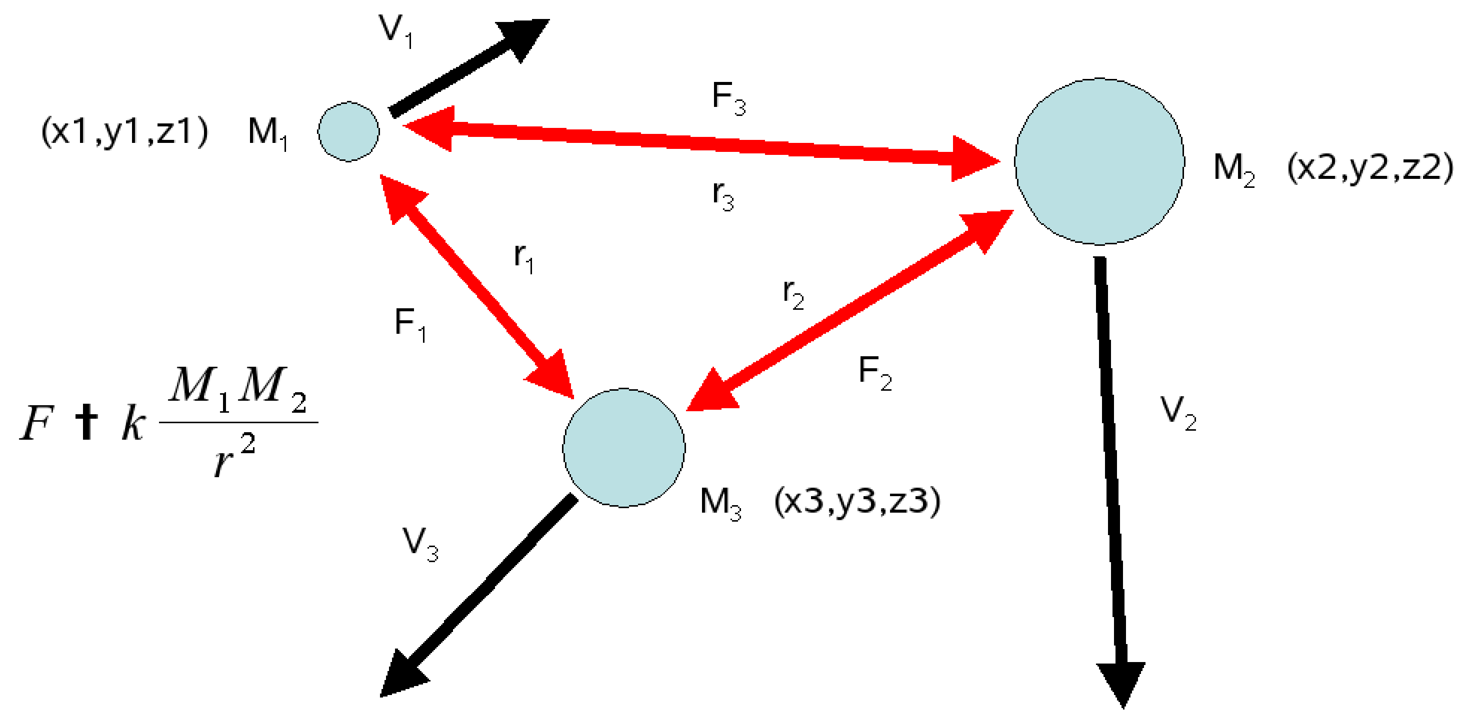
Physical principles: not in the next few years?

MAKEFGIPAA
VAGTVLNVVE
AGGWVTTIVS
ILTAVGSGGL
SLLAAAGRES
IKAYLKKEIK
KKGKRAVIAW

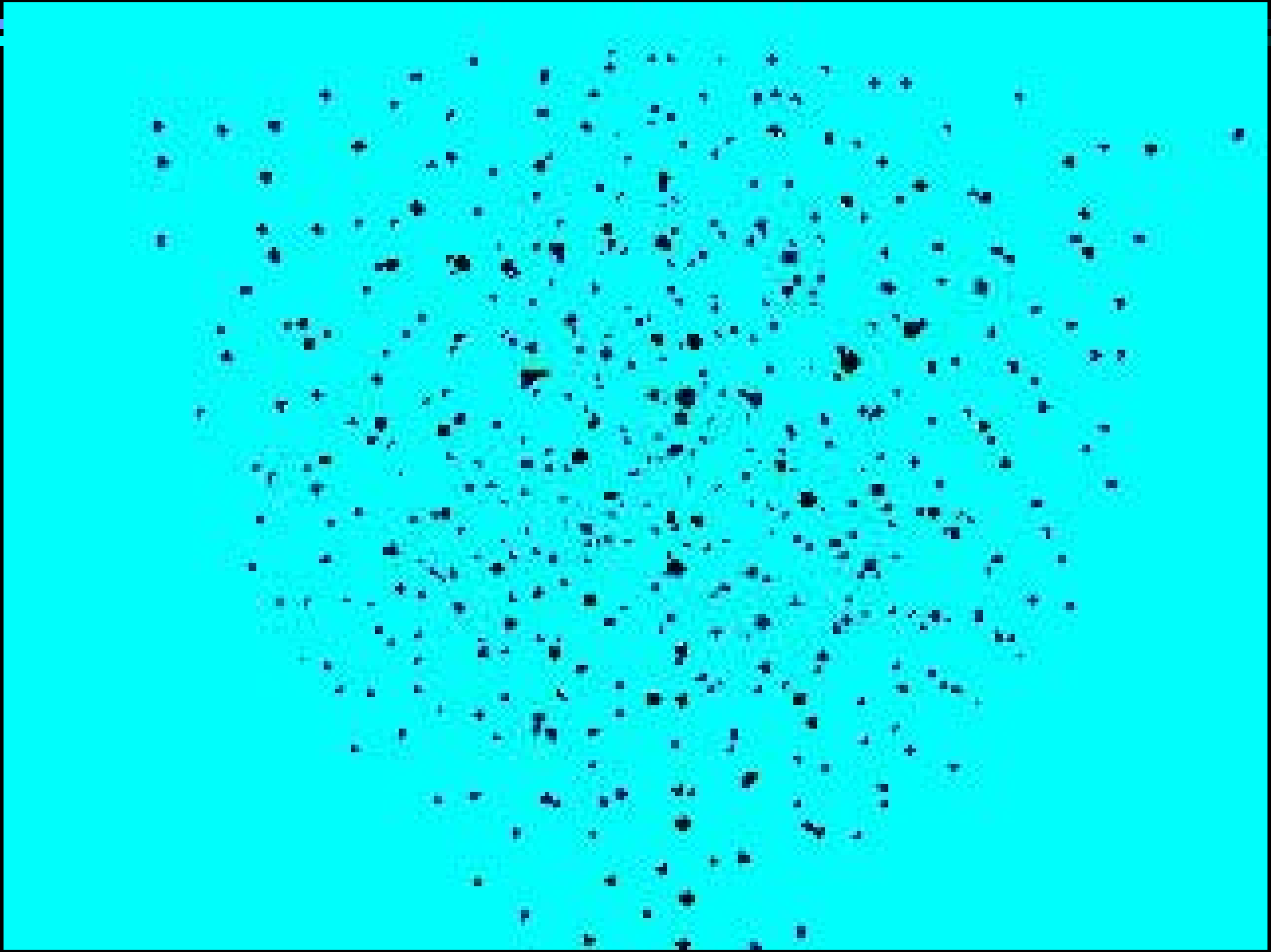


Molecular dynamics simulation.

ab initio methods

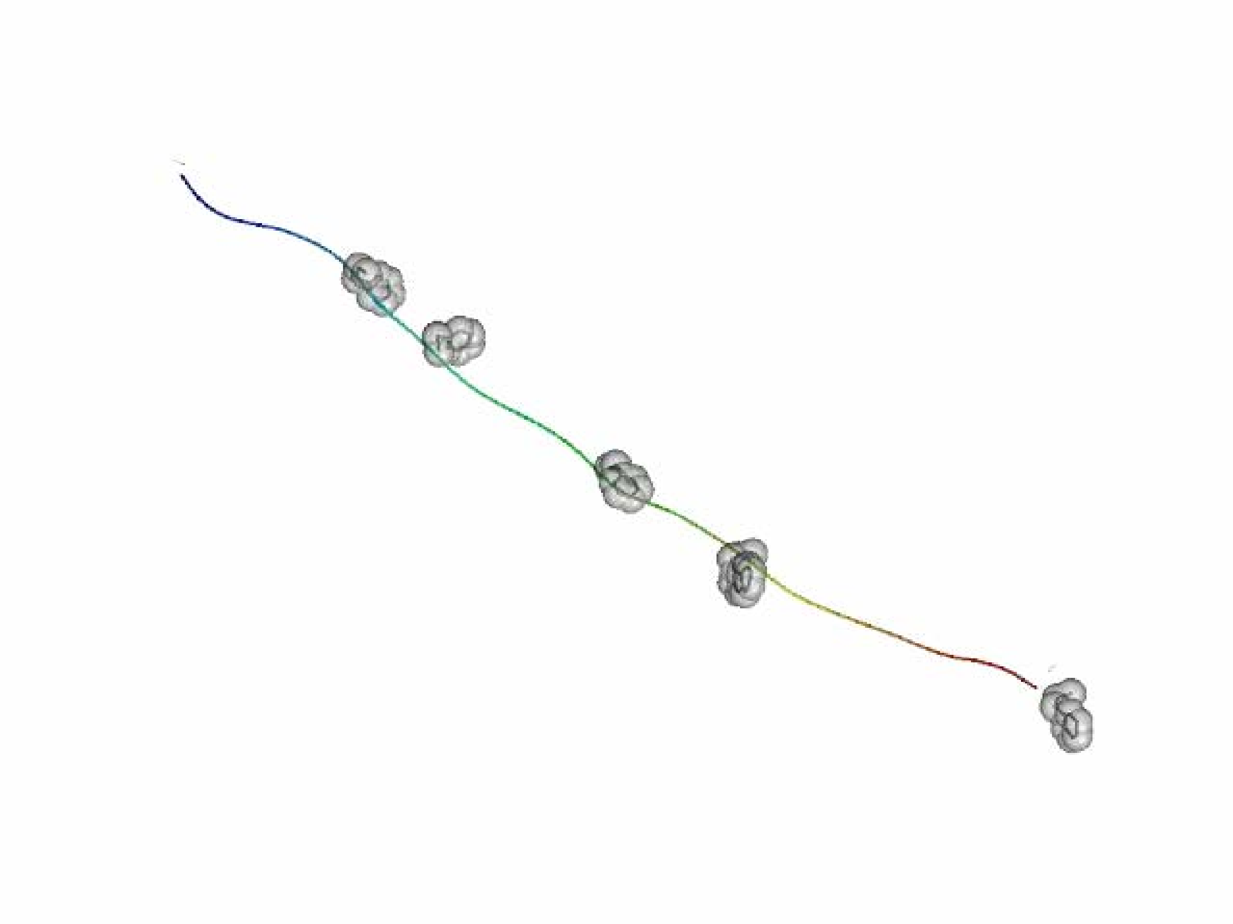


$$F \dagger k \frac{M_1 M_2}{r^2}$$



Alain Leppinette. CAB.

Folding@home: Simulations of the villin headpiece



3 Phe

1 Trp

1 Phe

The folding time is on the order of 10 micro-seconds.

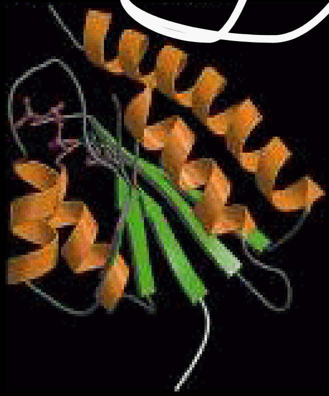
CA(2+)-REGULATED ACTIN-BINDING PROTEIN.

VILLIN CONSISTS OF A LARGE CORE FRAGMENT, THE AMINO-TERMINAL PORTION, AND A SMALL HEADPIECE, THE CARBOXYL-TERMINAL PORTION. THE HEADPIECE BINDS F-ACTIN STRONGLY IN BOTH THE PRESENCE AND ABSENCE OF CALCIUM.

MAJOR COMPONENT OF MICROVILLI OF INTESTINAL EPITHELIAL CELLS.

Protein structure prediction

MAKEFGIPAA
VAGTVLNVVE
AGGWVTTIVS
ILTAVGSGGL
SLLAAAGRES
IKAYLKKEIK
KKGKRAVIAW



Physical principles: not in the next few years?

Informatics (copying from known cases)

- Homology modeling

```
 Score = 85.1 bits (208), Expect = 6e-19
  Identities = 27/56 (48%), Positives = 42/56 (74%), Gaps = 1/56 (1%)

Query: 2  FIAIYDYKAETEEDLTIK KGEKLEIIEKEGD-WWKAKAIGSGEIGYIPANYIAAAE 56
              F+A+YDY+A TE+DL+     KGEK +I+       WW+A+++       +GE GYIP+NY+A  +
Sbjct: 8   FVALYDYEARTEDDLSFHKGEKFQILNSSEGDWWEARSLTTGETGYIPSNYVAPVD 63
```

- Threading

MAKEFGIPAA
VAGTVLNVVE
AGGWVTTIV
S ILTAVGSGGL
SLLAAAGRES
IKAYLKKEIK
KKGKRAVIAW

PDB

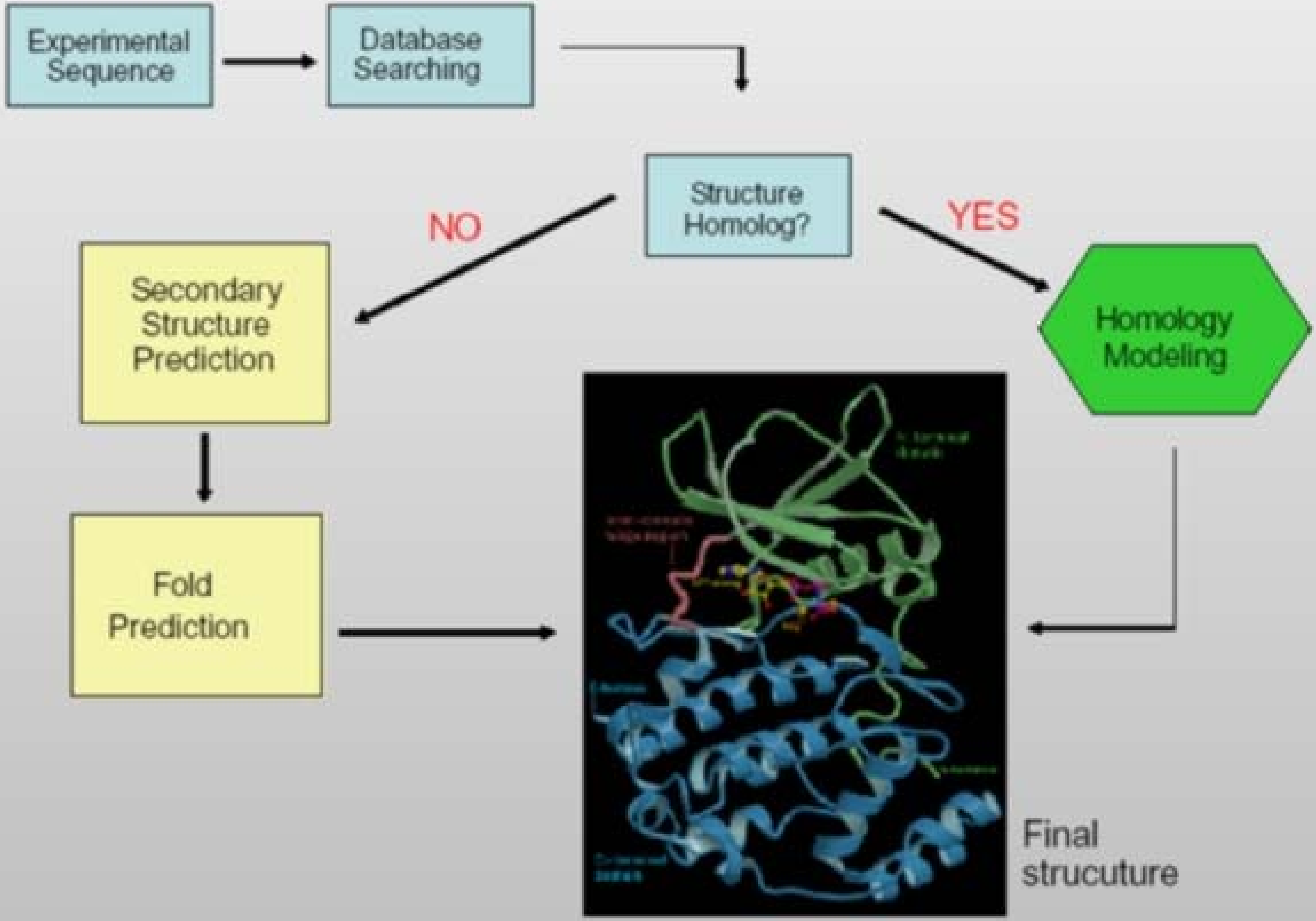
Fold ranking

Pseudo-Energy level?

By D. Devos

Paulino Gomez-Puertas
BioInfo Unit. CBM

Protein structure prediction. Flow chart.



Experimental Sequence

Database Searching

Structure Homolog?

NO

YES

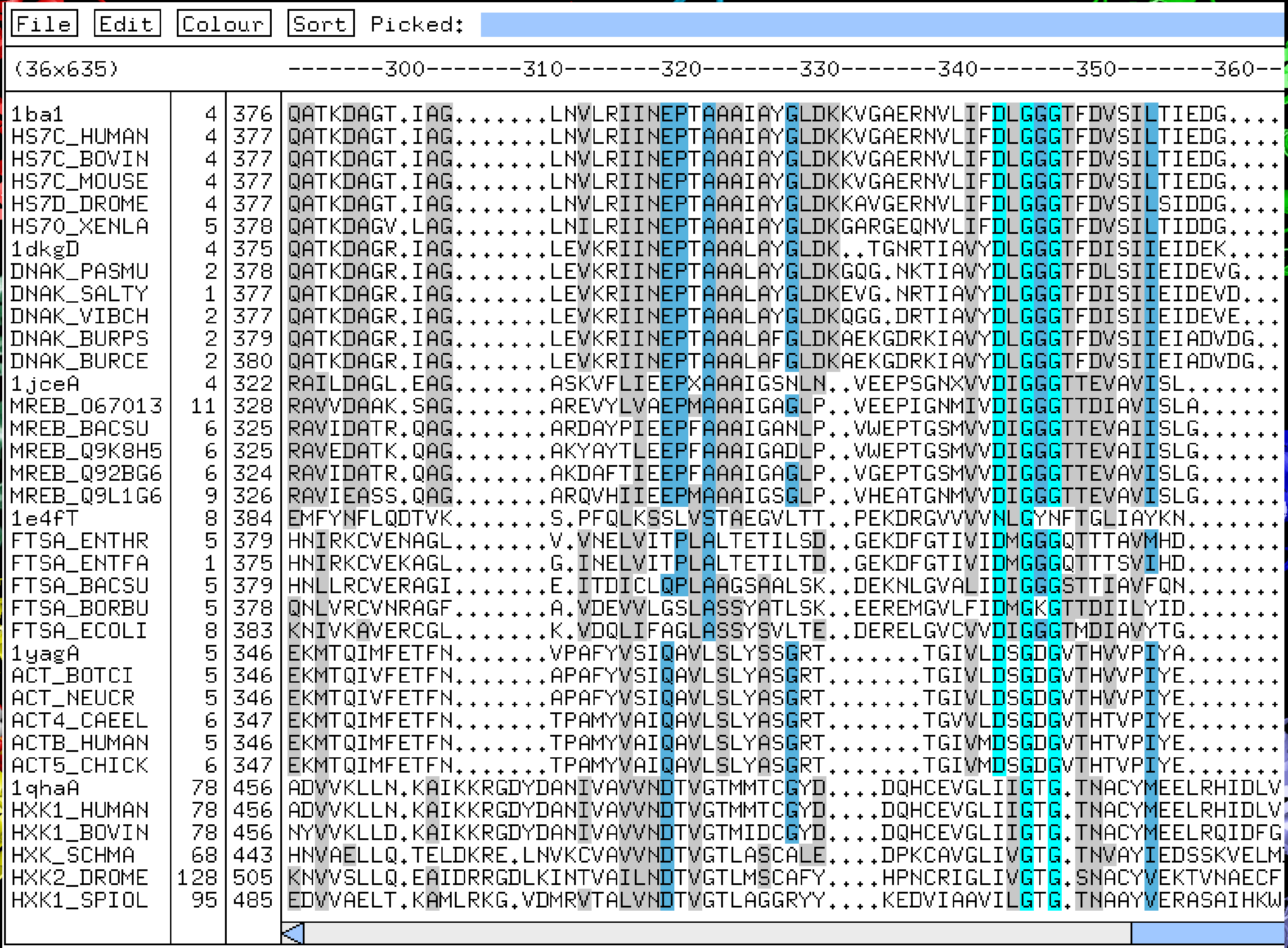
Secondary Structure Prediction

Fold Prediction

Homology Modeling

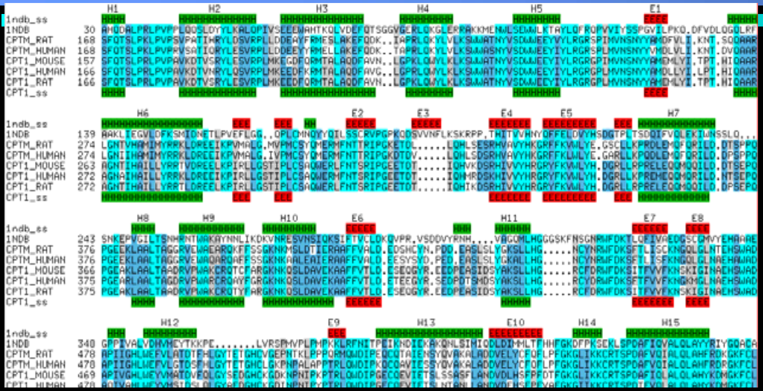
Final strucuture

Predicción de estructura de proteínas. Modelado por homología. Threading.



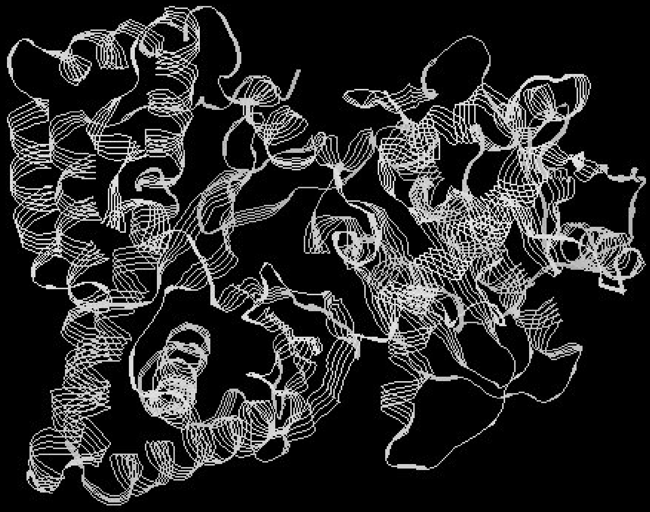
```
(36x635)                   -------300-------310-------320-------330-------340-------350-------360--
1ba1            4  376  QATKDAGT.IAG.......LNVLRIINEPTAAAIAYGLDKKVGAERNVLIFDLGGGTFDVSILTIEDG....
HS7C_HUMAN      4  377  QATKDAGT.IAG.......LNVLRIINEPTAAAIAYGLDKKVGAERNVLIFDLGGGTFDVSILTIEDG....
HS7C_BOVIN      4  377  QATKDAGT.IAG.......LNVLRIINEPTAAAIAYGLDKKVGAERNVLIFDLGGGTFDVSILTIEDG....
HS7C_MOUSE      4  377  QATKDAGT.IAG.......LNVLRIINEPTAAAIAYGLDKKVGAERNVLIFDLGGGTFDVSILTIEDG....
HS7D_DROME      4  377  QATKDAGT.IAG.......LNVLRIINEPTAAAIAYGLDKKAVGERNVLIFDLGGGTFDVSILSIDDG....
HS70_XENLA      5  378  QATKDAGV.LAG.......LNILRIINEPTAAAIAYGLDKGARGEQNVLIFDLGGGTFDVSILTIDDG....
1dkgD           4  375  QATKDAGR.IAG.......LEVKRIINEPTAAALAYGLDK..TGNRTIAVYDLGGGTFDISIIEIDEK....
DNAK_PASMU      2  378  QATKDAGR.IAG.......LEVKRIINEPTAAALAYGLDKGQG.NKTIAVYDLGGGTFDLSIIEIDEVG...
DNAK_SALTY      1  377  QATKDAGR.IAG.......LEVKRIINEPTAAALAYGLDKEVG.NRTIAVYDLGGGTFDISIIEIDEVD...
DNAK_VIBCH      2  377  QATKDAGR.IAG.......LEVKRIINEPTAAALAYGLDKQGG.DRTIAVYDLGGGTFDISIIEIDEVE...
DNAK_BURPS      2  379  QATKDAGR.IAG.......LEVKRIINEPTAAALAFGLDKAEKGDRKIAVYDLGGGTFDVSIIEIADVDG..
DNAK_BURCE      2  380  QATKDAGR.IAG.......LEVKRIINEPTAAALAFGLDKAEKGDRKIAVYDLGGGTFDVSIIEIADVDG..
1jceA           4  322  RAILDAGL.EAG.......ASKVFLIEEPXAAAIGSNLN..VEEPSGNXVVDIGGGTTEVAVISL.......
MREB_O67013    11  328  RAVVDAAK.SAG.......AREVYLVAEPMAAAIGAGLP..VEEPIGNMIVDIGGGTTDIAVISLA......
MREB_BACSU      6  325  RAVIDATR.QAG.......ARDAYPIEEPFAAAIGANLP..VWEPTGSMVVDIGGGTTEVAIISLG......
MREB_Q9K8H5     6  325  RAVEDATK.QAG.......AKYAYTLEEPFAAAIGADLP..VWEPTGSMVVDIGGGTTEVAIISLG......
MREB_Q92BG6     6  324  RAVIDATR.QAG.......AKDAFTIEEPFAAAIGAGLP..VGEPTGSMVVDIGGGTTEVAVISLG......
MREB_Q9L1G6     9  326  RAVIEASS.QAG.......ARQVHIIEEPMAAAIGSGLP..VHEATGNMVVDIGGGTTEVAVISLG......
1e4fT           8  384  EMFYNFLQDTVK.......S.PFQLKSSLVSTAEGVLTT..PEKDRGVVVVNLGYNFTGLIAYKN.......
FTSA_ENTHR      5  379  HNIRKCVENAGL.......V.VNELVITPLALTETILSD..GEKDFGTIVIDMGGGQTTTAVMHD.......
FTSA_ENTFA      1  375  HNIRKCVEKAGL.......G.INELVITPLALTETILTD..GEKDFGTIVIDMGGGQTTTSVIHD.......
FTSA_BACSU      5  379  HNLLRCVERAGI.......E.ITDICLQPLAAGSAALSK..DEKNLGVALIDIGGGSTTIAVFQN.......
FTSA_BORBU      5  378  QNLVRCVNRAGF.......A.VDEVVLGSLASSYATLSK..EEREMGVLFIDMGKGTTDIILYID.......
FTSA_ECOLI      8  383  KNIVKAVERCGL.......K.VDQLIFAGLASSYSVLTE..DERELGVCVVDIGGGTMDIAVYTG.......
1yagA           5  346  EKMTQIMFETFN.......VPAFYVSIQAVLSLYSSGRT.......TGIVLDSGDGVTHVVPIYA.......
ACT_BOTCI       5  346  EKMTQIVFETFN.......APAFYVSIQAVLSLYASGRT.......TGIVLDSGDGVTHVVPIYE.......
ACT_NEUCR       5  346  EKMTQIVFETFN.......APAFYVSIQAVLSLYASGRT.......TGIVLDSGDGVTHVVPIYE.......
ACT4_CAEEL      6  347  EKMTQIMFETFN.......TPAMYVAIQAVLSLYASGRT.......TGVVLDSGDGVTHTVPIYE.......
ACTB_HUMAN      5  346  EKMTQIMFETFN.......TPAMYVAIQAVLSLYASGRT.......TGIVMDSGDGVTHTVPIYE.......
ACT5_CHICK      6  347  EKMTQIMFETFN.......TPAMYVAIQAVLSLYASGRT.......TGIVMDSGDGVTHTVPIYE.......
1qhaA          78  456  ADVVKLLN.KAIKKRGDYDANIVAVVNDTVGTMMTCGYD....DQHCEVGLIIGTG.TNACYMEELRHIDLV
HXK1_HUMAN     78  456  ADVVKLLN.KAIKKRGDYDANIVAVVNDTVGTMMTCGYD....DQHCEVGLIIGTG.TNACYMEELRHIDLV
HXK1_BOVIN     78  456  NYVVKLLD.KAIKKRGDYDANIVAVVNDTVGTMIDCGYD....DQHCEVGLIIGTG.TNACYMEELRQIDFG
HXK_SCHMA      68  443  HNVAELLQ.TELDKRE.LNVKCVAVVNDTVGTLASCALE....DPKCAVGLIVGTG.TNVAYIEDSSKVELM
HXK2_DROME    128  505  KNVVSLLQ.EAIDRRGDLKINTVAILNDTVGTLMSCAFY....HPNCRIGLIVGTG.SNACYVEKTVNAECF
HXK1_SPIOL     95  485  EDVVAELT.KAMLRKG.VDMRVTALVNDTVGTLAGGRYY....KEDVIAAVILGTG.TNAAYVERASAIHKW
```

Homologs: proteins sharing a common ancestor

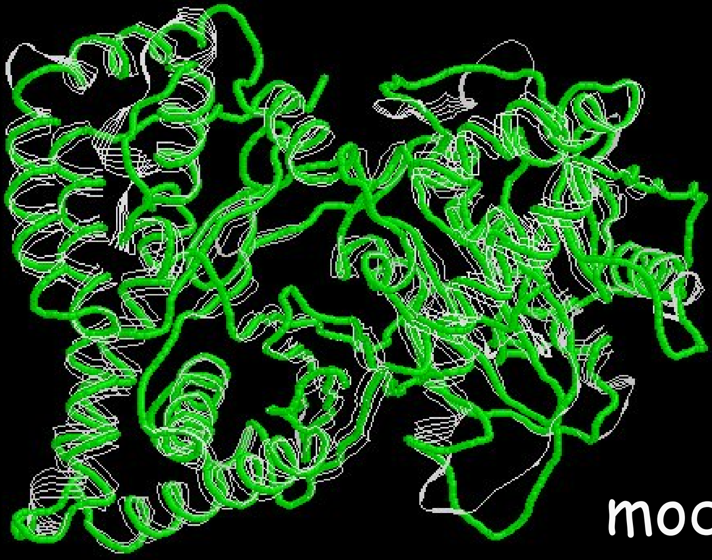


```
AHPLTSDFGGHTERDLHA
|| |||: ||||| |:||
AHTLTSEGGGHTEADVHA
```

alignment



template



model

Classical homology modeling algorithm

Generate alignment query - target

Replace conserved amino acid side-chains

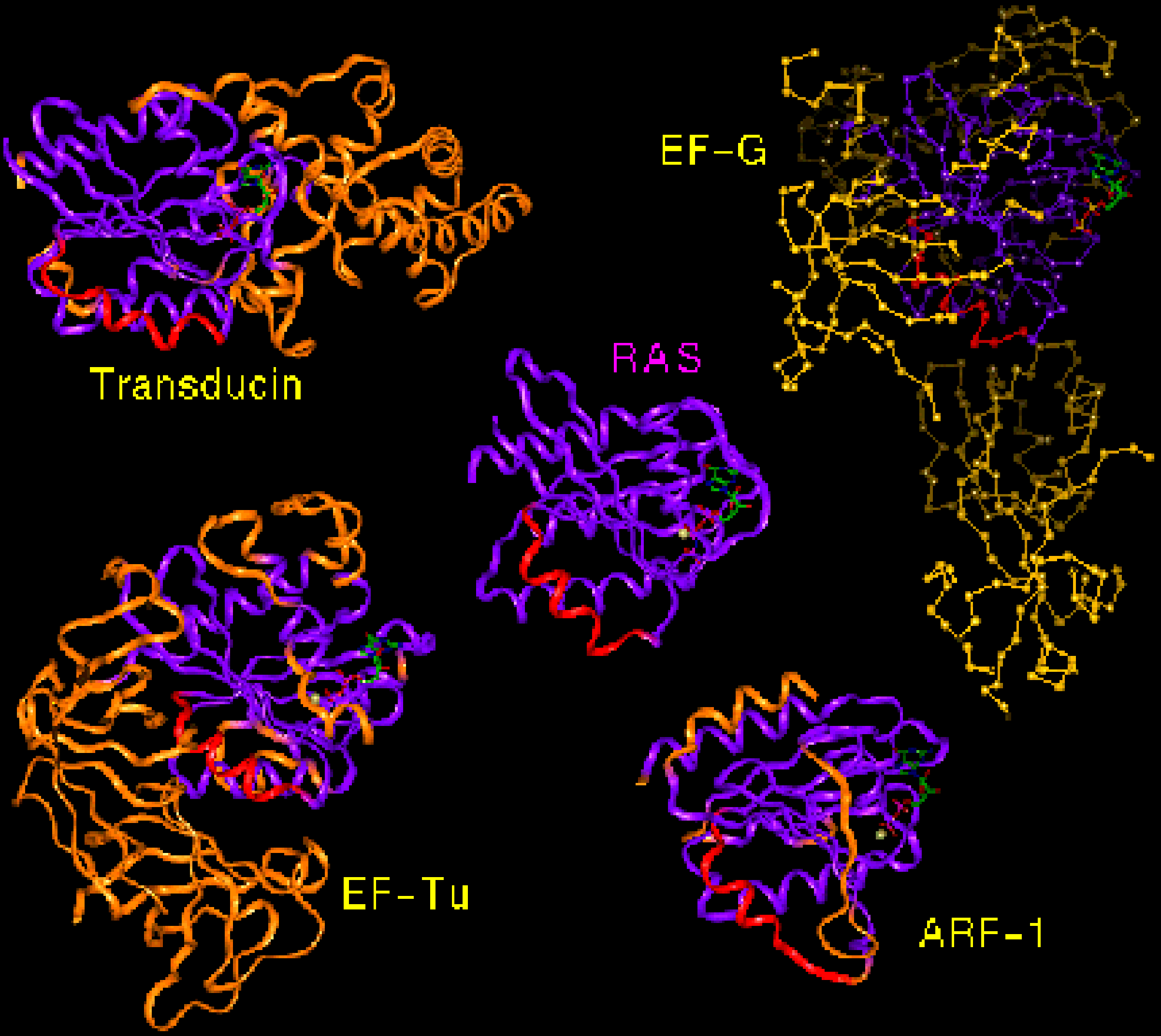
Replace other amino acids

- use most common rotamers
- keep as many atoms in positions as possible

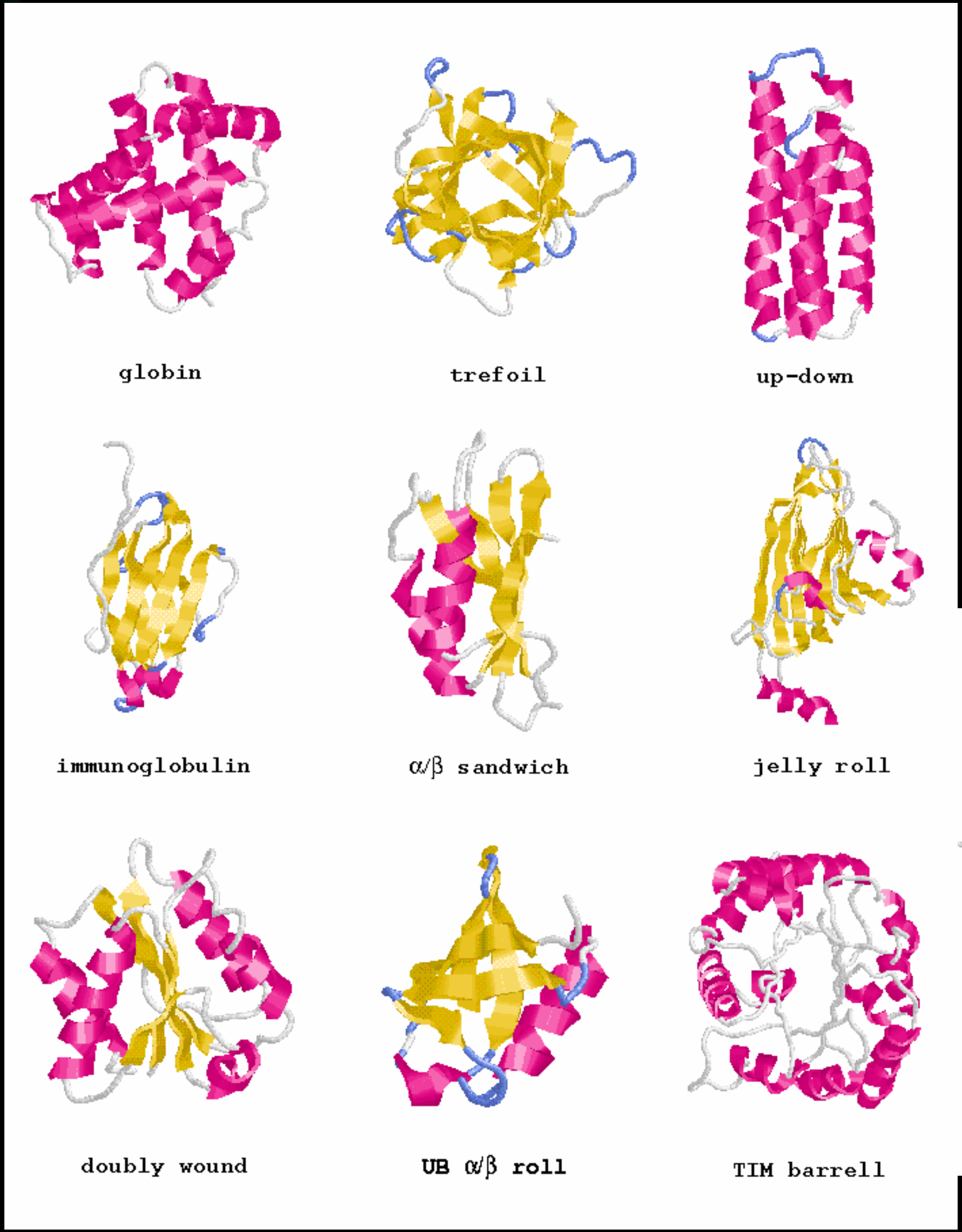
Model loops regions (insertions and deletions)

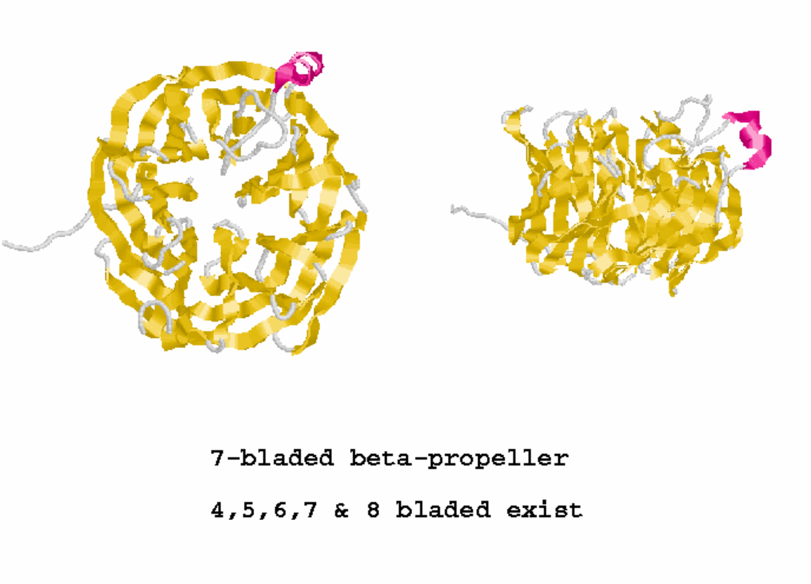
Optimize packing

Quality check

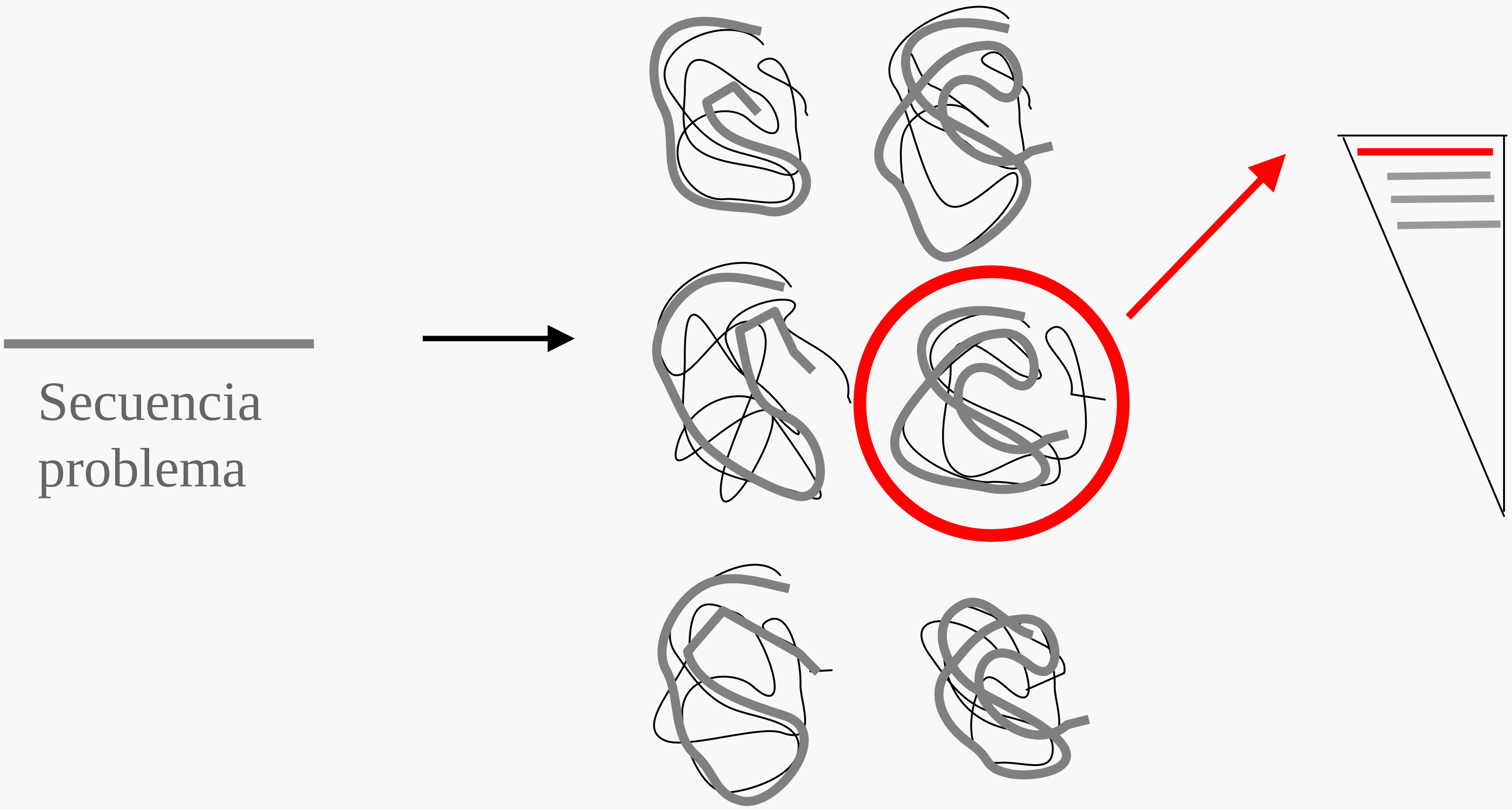

EF-G
Transducin
RAS
EF-Tu
ARF-1

Superfolds
(Orengo et al.)





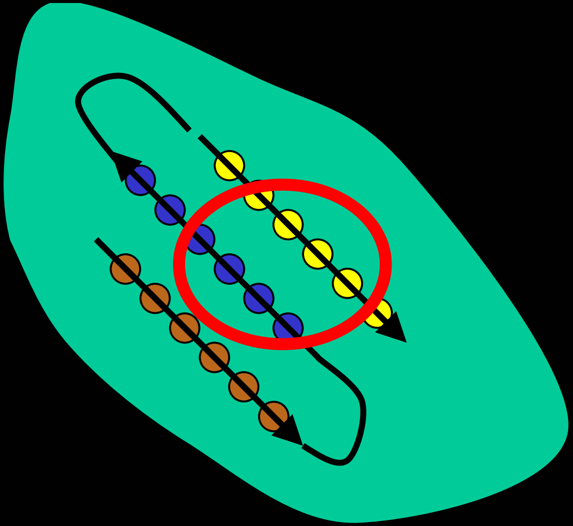
Algoritmos de *threading*. General.



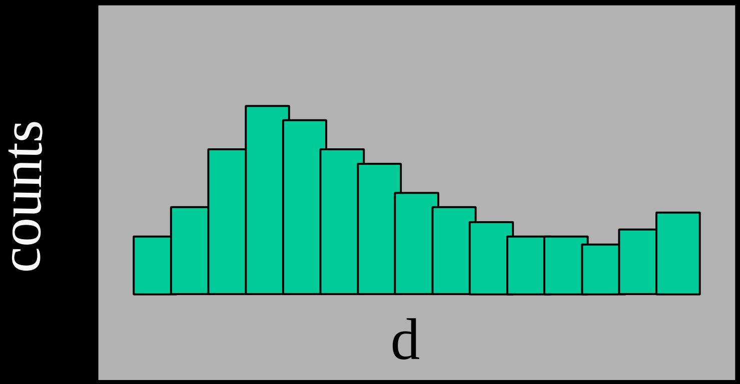
Secuencia problema

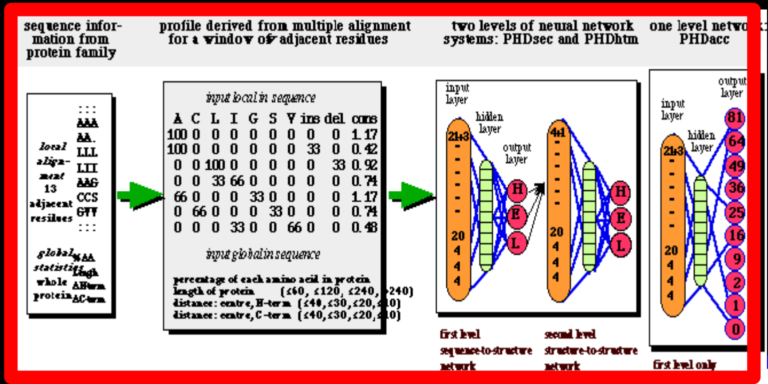
Algoritmos de *threading*

Potenciales de contacto / Coincidencia de estructura secundaria y accesibilidad



Count pairs of each residue type at different separations



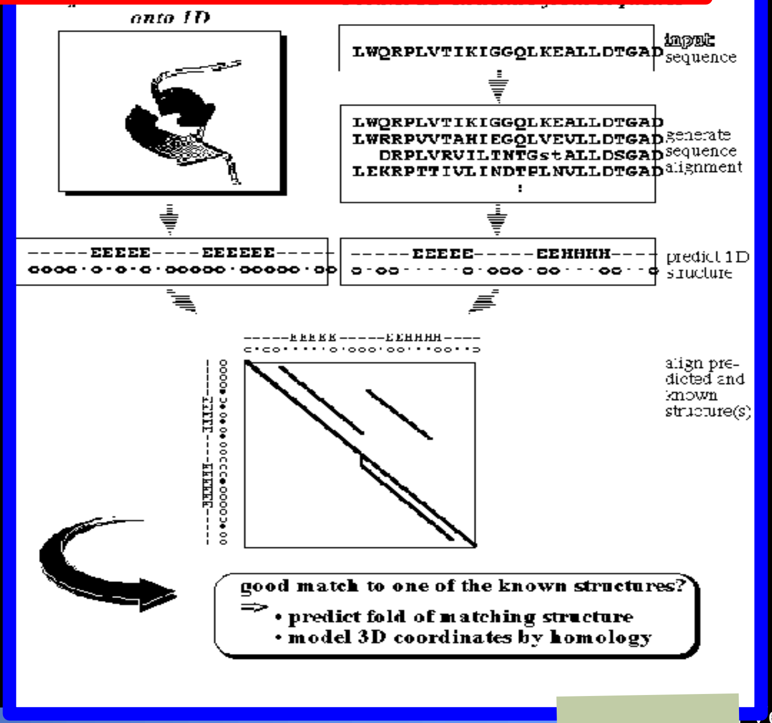


input local in sequence

A	C	L	I	G	S	V	ins	del	cons
100	0	0	0	0	0	0	0	0	1.17
100	0	0	0	0	0	0	33	0	0.42
0	0	100	0	0	0	0	0	33	0.92
0	0	33	66	0	0	0	0	0	0.74
66	0	0	0	33	0	0	0	0	1.17
0	66	0	0	0	33	0	0	0	0.74
0	0	0	33	0	0	66	0	0	0.48

input global in sequence

percentage of each amino acid in protein
length of protein (≤60, ≤120, ≤240, >240)
distance: centre, N-term (≤40, ≤30, ≤20, ≤10)
distance: centre, C-term (≤40, ≤30, ≤20, ≤10)



```
LWQRPLVTIKIGGQLKEALLDTGAD

LWQRPLVTIKIGGQLKEALLDTGAD
LWRRPVVTAHIEGQLVEVLLDTGAD
  DRPLVRVILTNTGstALLDSGAD
LEKRPTTIVLINDTPLNVLLDTGAD
            :

-----EEEEE----EEEEEE-----
-----EEEEE-----EEHHHH----
```

good match to one of the known structures?
⇒
- predict fold of matching structure
- model 3D coordinates by homology

Jones, 1992; Sippl, 1995

Bioinformática: Integración de información estructural y evolutiva para el estudio de la función de las proteínas.

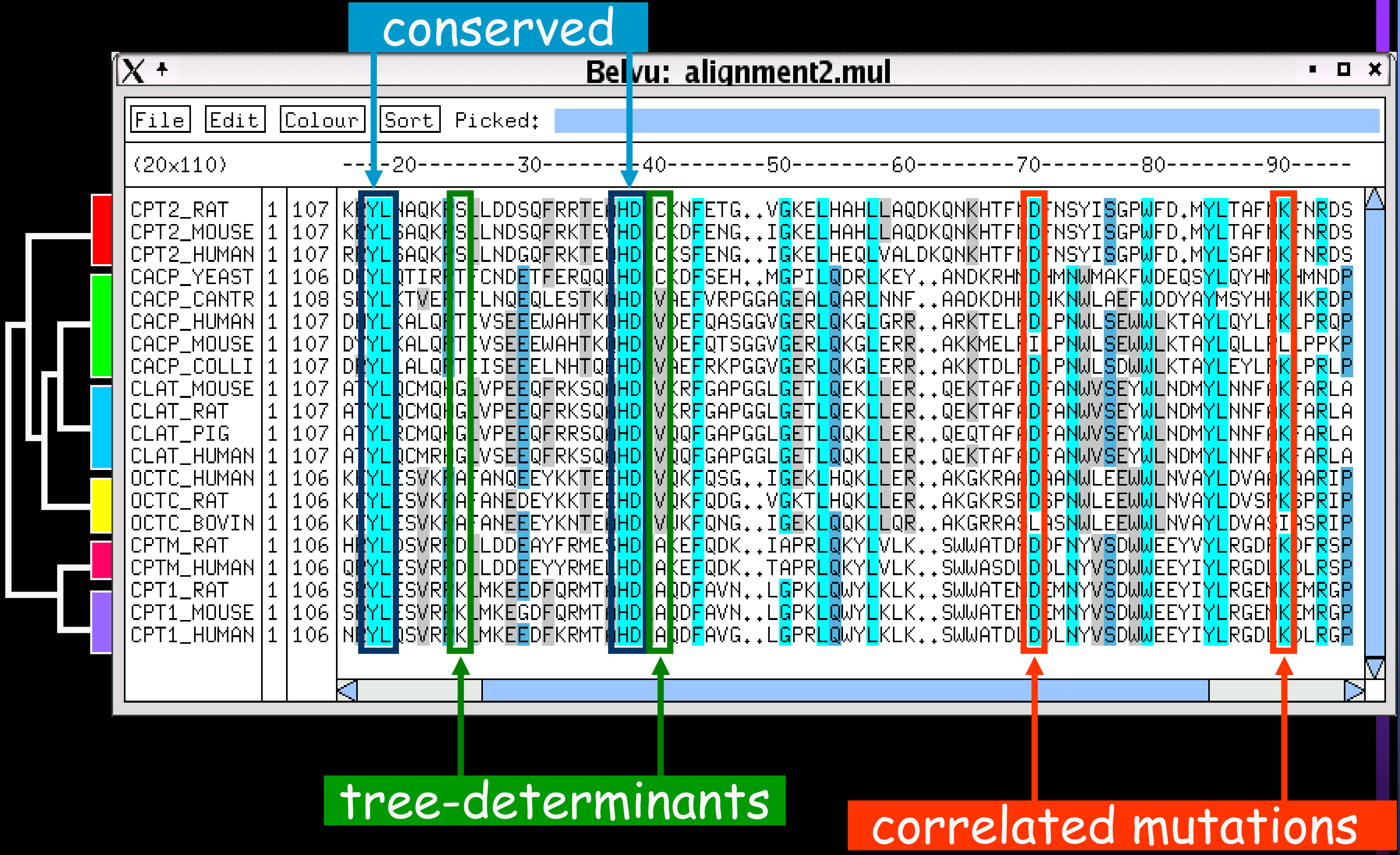
- Predicción de estructura de proteínas.
- Extracción de información evolutiva.
- Predicción de interacciones:

  proteína - proteína

  proteína-ligando

1/2

Integration of Evolutive Information



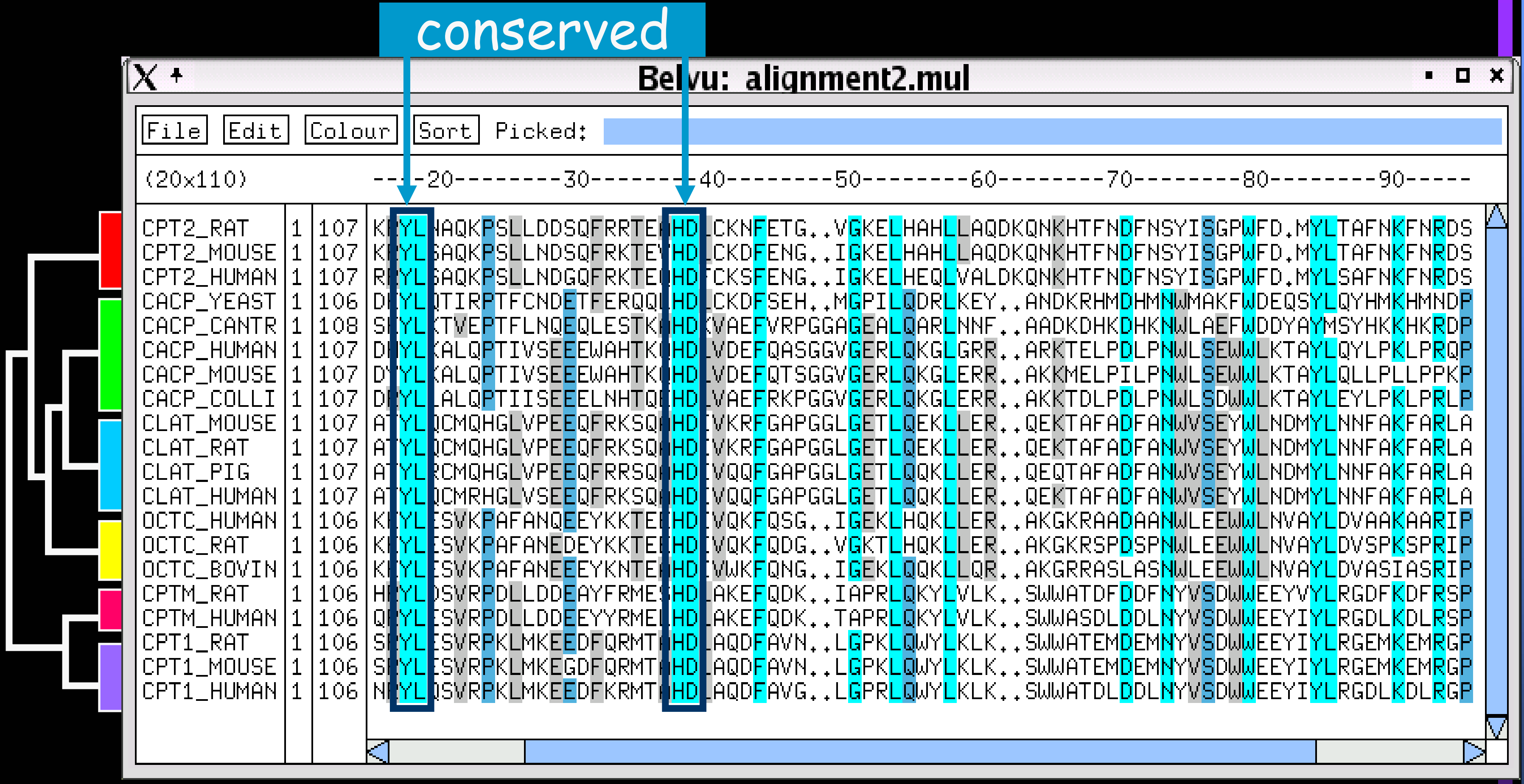
conserved

tree-determinants

correlated mutations

Information extracted from multiple sequence alignments

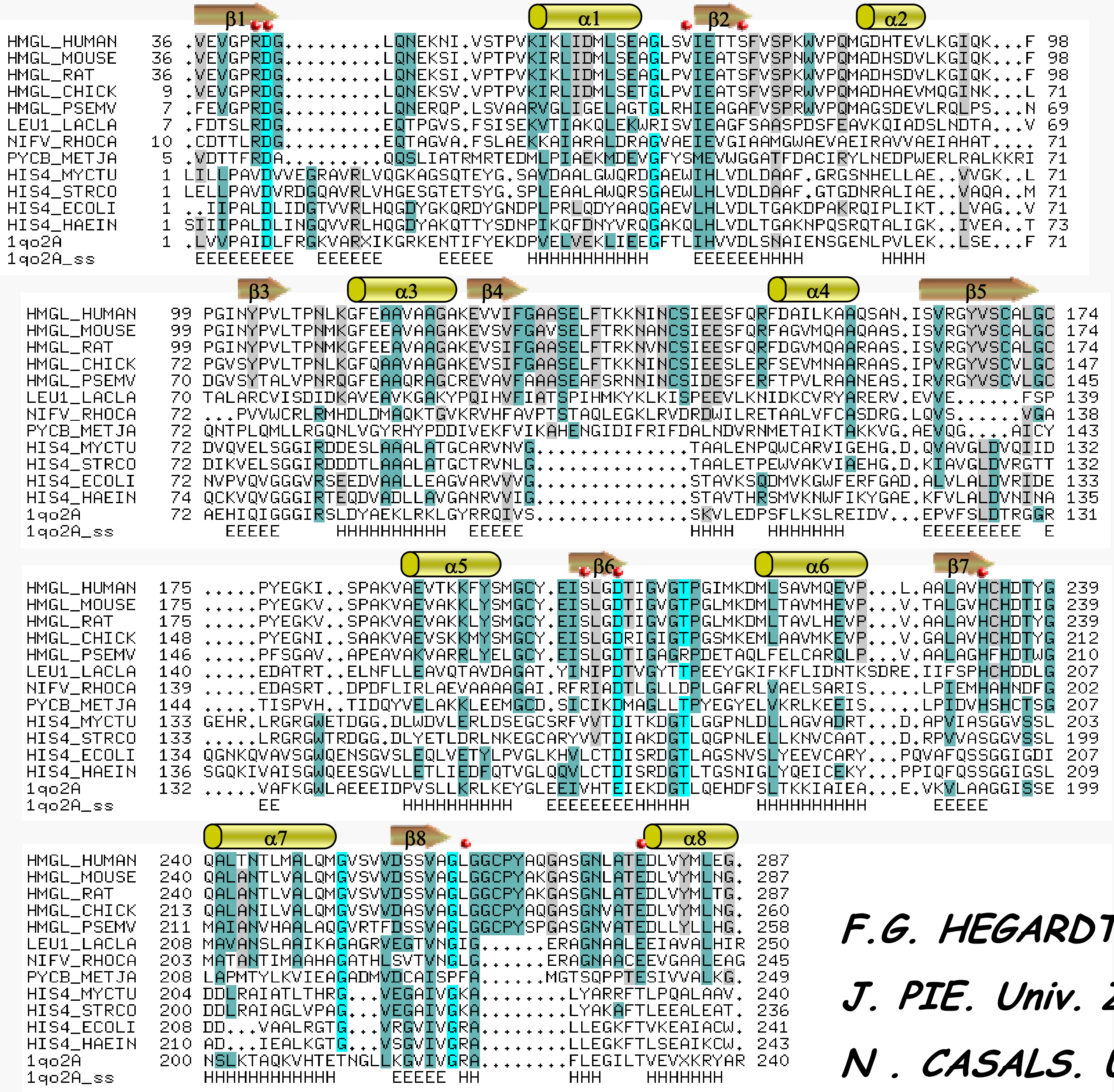
Integration of Evolutive Information



```
(20x110)          ----20--------30--------40--------50--------60--------70--------80--------90-----
CPT2_RAT    1 107 KPYLNAQKPSLLDDSQFRRTEKHDLCKNFETG..VGKELHAHLLAQDKQNKHTFNDFNSYISGPWFD.MYLTAFNKFNRDS
CPT2_MOUSE  1 107 KPYLGAQKPSLLNDSQFRKTEYHDLCKDFENG..IGKELHAHLLAQDKQNKHTFNDFNSYISGPWFD.MYLTAFNKFNRDS
CPT2_HUMAN  1 107 RPYLGAQKPSLLNDGQFRKTEQHDFCKSFENG..IGKELHEQLVALDKQNKHTFNDFNSYISGPWFD.MYLSAFNKFNRDS
CACP_YEAST  1 106 DKYLQTIRPTFCNDETFERQQLHDLCKDFSEH..MGPILQDRLKEY..ANDKRHMDHMNWMAKFWDEQSYLQYHMKHMNDP
CACP_CANTR  1 108 SKYLKTVEPTFLNQEQLESTKAHDKVAEFVRPGGAGEALQARLNNF..AADKDHKDHKNWLAEFWDDYAYMSYHKKHKRDP
CACP_HUMAN  1 107 DHYLKALQPTIVSEEEWAHTKQHDLVDEFQASGGVGERLQKGLGRR..ARKTELPDLPNWLSEWWLKTAYLQYLPKLPRQP
CACP_MOUSE  1 107 DYYLKALQPTIVSEEEWAHTKQHDLVDEFQTSGGVGERLQKGLERR..AKKMELPILPNWLSEWWLKTAYLQLLPLLPPKP
CACP_COLLI  1 107 DPYLLALQPTIISEEELNHTQEHDLVAEFRKPGGVGERLQKGLERR..AKKTDLPDLPNWLSDWWLKTAYLEYLPKLPRLP
CLAT_MOUSE  1 107 ATYLQCMQHGLVPEEQFRKSQKHDIVKRFGAPGGLGETLQEKLLER..QEKTAFADFANWVSEYWLNDMYLNNFAKFARLA
CLAT_RAT    1 107 ATYLQCMQHGLVPEEQFRKSQKHDIVKRFGAPGGLGETLQEKLLER..QEKTAFADFANWVSEYWLNDMYLNNFAKFARLA
CLAT_PIG    1 107 ATYLRCMQHGLVPEEQFRRSQKHDIVQQFGAPGGLGETLQQKLLER..QEQTAFADFANWVSEYWLNDMYLNNFAKFARLA
CLAT_HUMAN  1 107 ATYLQCMRHGLVSEEQFRKSQKHDIVQQFGAPGGLGETLQQKLLER..QEKTAFADFANWVSEYWLNDMYLNNFAKFARLA
OCTC_HUMAN  1 106 KHYLESVKPAFANQEEYKKTEEHDIVQKFQSG..IGEKLHQKLLER..AKGKRAADAANWLEEWWLNVAYLDVAAKAARIP
OCTC_RAT    1 106 KHYLESVKPAFANEDEYKKTEEHDIVQKFQDG..VGKTLHQKLLER..AKGKRSPDSPNWLEEWWLNVAYLDVSPKSPRIP
OCTC_BOVIN  1 106 KHYLESVKPAFANEEEYKNTEKHDIVWKFQNG..IGEKLQQKLLQR..AKGRRASLASNWLEEWWLNVAYLDVASIASRIP
CPTM_RAT    1 106 HPYLDSVRPDLLDDEAYFRMESHDLAKEFQDK..IAPRLQKYLVLK..SWWATDFDDFNYVSDWWEEYVYLRGDFKDFRSP
CPTM_HUMAN  1 106 QPYLESVRPDLLDDEEYYRMELHDLAKEFQDK..TAPRLQKYLVLK..SWWASDLDDLNYVSDWWEEYIYLRGDLKDLRSP
CPT1_RAT    1 106 SPYLESVRPKLMKEEDFQRMTAHDLAQDFAVN..LGPKLQWYLKLK..SWWATEMDEMNYVSDWWEEYIYLRGEMKEMRGP
CPT1_MOUSE  1 106 SPYLESVRPKLMKEGDFQRMTAHDLAQDFAVN..LGPKLQWYLKLK..SWWATEMDEMNYVSDWWEEYIYLRGEMKEMRGP
CPT1_HUMAN  1 106 NPYLQSVRPKLMKEEDFKRMTAHDLAQDFAVG..LGPRLQWYLKLK..SWWATDLDDLNYVSDWWEEYIYLRGDLKDLRGP
```

Information extracted from multiple sequence alignments

Hydroxymethylglutaryl-CoA lyase (HMGL)



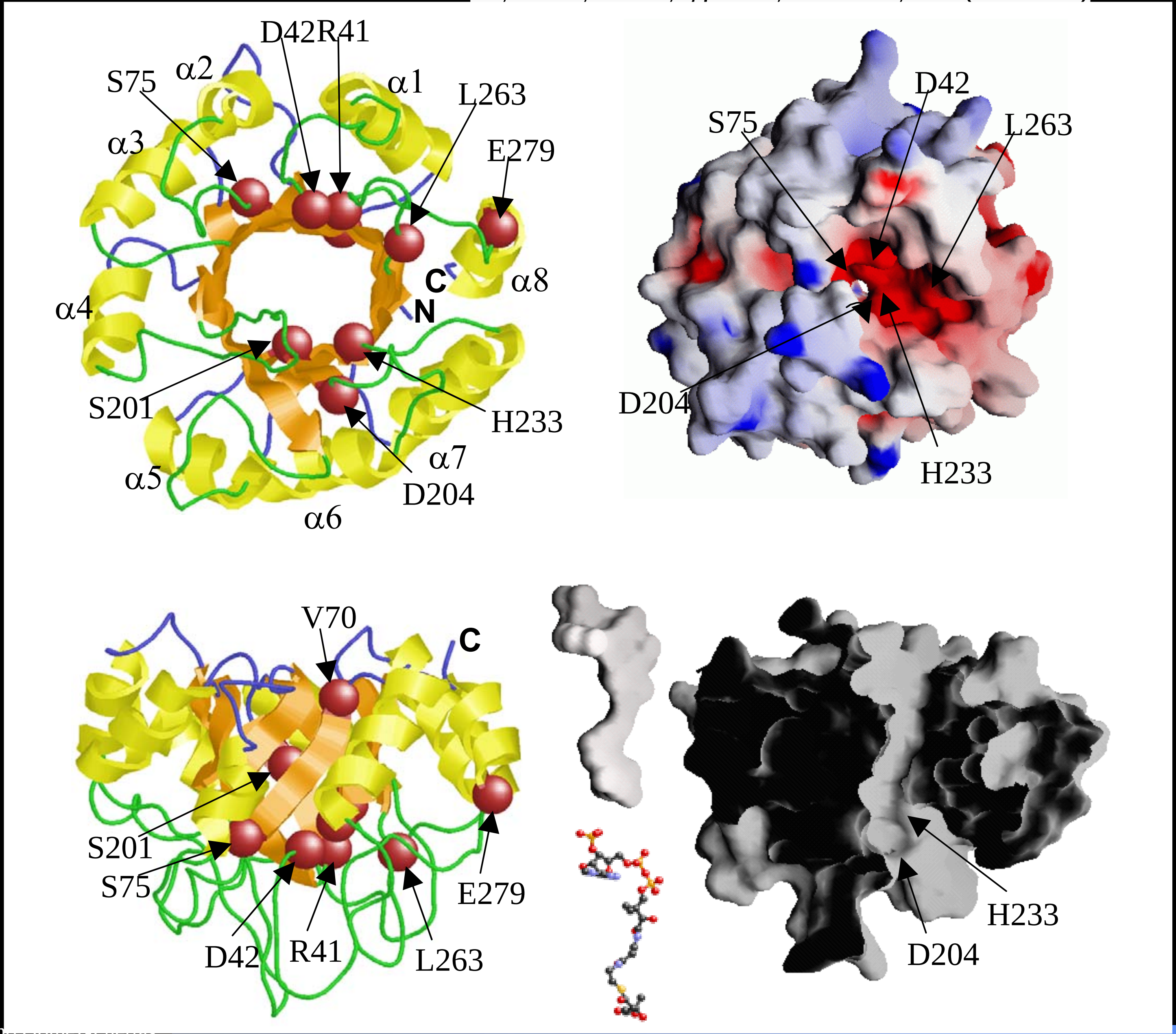
```
HMGL_HUMAN   36 .VEVGPRDG.........LQNEKNI.VSTPVKIKLIDMLSEAGLSVIETTSFVSPKWVPQMGDHTEVLKGIQK...F 98
HMGL_MOUSE   36 .VEVGPRDG.........LQNEKSI.VPTPVKIRLIDMLSEAGLPVIEATSFVSPNWVPQMADHSDVLKGIQK...F 98
HMGL_RAT     36 .VEVGPRDG.........LQNEKSI.VPTPVKIKLIDMLSEAGLPVIEATSFVSPKWVPQMADHSDVLKGIQK...F 98
HMGL_CHICK    9 .VEVGPRDG.........LQNEKSV.VPTPVKIRLIDMLSETGLPVIEATSFVSPRWVPQMADHAEVMQGINK...L 71
HMGL_PSEMV    7 .FEVGPRDG.........LQNERQP.LSVAARVGLIGELAGTGLRHIEAGAFVSPRWVPQMAGSDEVLRQLPS...N 69
LEU1_LACLA    7 .FDTSLRDG.........EQTPGVS.FSISEKVTIAKQLEKWRISVIEAGFSAASPDSFEAVKQIADSLNDTA...V 69
NIFV_RHOCA   10 .CDTTLRDG.........EQTAGVA.FSLAEKKAIARALDRAGVAEIEVGIAAMGWAEVAEIRAVVAEIAHAT.... 71
PYCB_METJA    5 .VDTTFRDA.........QQSLIATRMRTEDMLPIAEKMDEVGFYSMEVWGGATFDACIRYLNEDPWERLRALKKRI 71
HIS4_MYCTU    1 LILLPAVDVVEGRAVRLVQGKAGSQTEYG.SAVDAALGWQRDGAEWIHLVDLDAAF.GRGSNHELLAE..VVGK..L 71
HIS4_STRCO    1 LELLPAVDVRDGQAVRLVHGESGTETSYG.SPLEAALAWQRSGAEWLHLVDLDAAF.GTGDNRALIAE..VAQA..M 71
HIS4_ECOLI    1 ..IIPALDLIDGTVVRLHQGDYGKQRDYGNDPLPRLQDYAAQGAEVLHLVDLTGAKDPAKRQIPLIKT..LVAG..V 71
HIS4_HAEIN    1 SIIIPALDLINGQVVRLHQGDYAKQTTYSDNPIKQFDNYVRQGAKQLHLVDLTGAKNPQSRQTALIGK..IVEA..T 73
1qo2A         1 .LVVPAIDLFRGKVARXIKGRKENTIFYEKDPVELVEKLIEEGFTLIHVVDLSNAIENSGENLPVLEK..LSE...F 71
1qo2A_ss         EEEEEEEEE  EEEEEE     EEEEE  HHHHHHHHHHH     EEEEEEHHHH       HHHH

HMGL_HUMAN   99 PGINYPVLTPNLKGFEAAVAAGAKEVVIFGAASELFTKKNINCSIEESFQRFDAILKAAQSAN.ISVRGYVSCALGC 174
HMGL_MOUSE   99 PGINYPVLTPNMKGFEEAVAAGAKEVSVFGAVSELFTRKNANCSIEESFQRFAGVMQAAQAAS.ISVRGYVSCALGC 174
HMGL_RAT     99 PGINYPVLTPNMKGFEEAVAAGAKEVSIFGAASELFTRKNVNCSIEESFQRFDGVMQAARAAS.ISVRGYVSCALGC 174
HMGL_CHICK   72 PGVSYPVLTPNLKGFQAAVAAGAKEVSIFGAASELFTKKNINCSIEESLERFSEVMNAARAAS.IPVRGYVSCVLGC 147
HMGL_PSEMV   70 DGVSYTALVPNRQGFEAAQRAGCREVAVFAAASEAFSRNNINCSIDESFERFTPVLRAANEAS.IRVRGYVSCVLGC 145
LEU1_LACLA   70 TALARCVISDIDKAVEAVKGAKYPQIHVFIATSPIHMKYKLKISPEEVLKNIDKCVRYARERV.EVVE......FSP 139
NIFV_RHOCA   72 ...PVVWCRLRMHDLDMAQKTGVKRVHFAVPTSTAQLEGKLRVDRDWILRETAALVFCASDRG.LQVS......VGA 138
PYCB_METJA   72 QNTPLQMLLRGQNLVGYRHYPDDIVEKFVIKAHENGIDIFRIFDALNDVRNMETAIKTAKKVG.AEVQG....AICY 143
HIS4_MYCTU   72 DVQVELSGGIRDDESLAAALATGCARVNVG..............TAALENPQWCARVIGEHG.D.QVAVGLDVQIID 132
HIS4_STRCO   72 DIKVELSGGIRDDDTLAAALATGCTRVNLG..............TAALETPEWVAKVIAEHG.D.KIAVGLDVRGTT 132
HIS4_ECOLI   72 NVPVQVGGGVRSEEDVAALLEAGVARVVVG..............STAVKSQDMVKGWFERFGAD.ALVLALDVRIDE 133
HIS4_HAEIN   74 QCKVQVGGGIRTEQDVADLLAVGANRVVIG..............STAVTHRSMVKNWFIKYGAE.KFVLALDVNINA 135
1qo2A        72 AEHIQIGGGIRSLDYAEKLRKLGYRRQIVS..............SKVLEDPSFLKSLREIDV...EPVFSLDTRGGR 131
1qo2A_ss          EEEEE     HHHHHHHHHH  EEEEE               HHHH   HHHHHHH       EEEEEEEEE  E

HMGL_HUMAN  175 .....PYEGKI..SPAKVAEVTKKFYSMGCY.EISLGDTIGVGTPGIMKDMLSAVMQEVP...L.AALAVHCHDTYG 239
HMGL_MOUSE  175 .....PYEGKV..SPAKVAEVAKKLYSMGCY.EISLGDTIGVGTPGLMKDMLTAVMHEVP...V.TALGVHCHDTIG 239
HMGL_RAT    175 .....PYEGKV..SPAKVAEVAKKLYSMGCY.EISLGDTIGVGTPGLMKDMLTAVLHEVP...V.AALAVHCHDTYG 239
HMGL_CHICK  148 .....PYEGNI..SAAKVAEVSKKMYSMGCY.EISLGDRIGIGTPGSMKEMLAAVMKEVP...V.GALAVHCHDTYG 212
HMGL_PSEMV  146 .....PFSGAV..APEAVAKVARRLYELGCY.EISLGDTIGAGRPDETAQLFELCARQLP...V.AALAGHFHDTWG 210
LEU1_LACLA  140 .....EDATRT..ELNFLLEAVQTAVDAGAT.YINIPDTVGYTTPEEYGKIFKFLIDNTKSDRE.IIFSPHCHDDLG 207
NIFV_RHOCA  139 .....EDASRT..DPDFLIRLAEVAAAAGAI.RFRIADTLGLLDPLGAFRLVAELSARIS....LPIEMHAHNDFG 202
PYCB_METJA  144 .....TISPVH..TIDQYVELAKKLEEMGCD.SICIKDMAGLLTPYEGYELVKRLKEEIS....LPIDVHSHCTSG 207
HIS4_MYCTU  133 GEHR.LRGRGWETDGG.DLWDVLERLDSEGCSRFVVTDITKDGTLGGPNLDLLAGVADRT...D.APVIASGGVSSL 203
HIS4_STRCO  133 .....LRGRGWTRDGG.DLYETLDRLNKEGCARYVVTDIAKDGTLQGPNLELLKNVCAAT...D.RPVVASGGVSSL 199
HIS4_ECOLI  134 QGNKQVAVSGWQENSGVSLEQLVETYLPVGLKHVLCTDISRDGTLAGSNVSLYEEVCARY...PQVAFQSSGGIGDI 207
HIS4_HAEIN  136 SGQKIVAISGWQEESGVLLETLIEDFQTVGLQQVLCTDISRDGTLTGSNIGLYQEICEKY...PPIQFQSSGGIGSL 209
1qo2A       132 .....VAFKGWLAEEEIDPVSLLKRLKEYGLEEIVHTEIEKDGTLQEHDFSLTKKIAIEA...E.VKVLAAGGISSE 199
1qo2A_ss             EE           HHHHHHHHHH   EEEEEEEEEHHHHH     HHHHHHHHHH      EEEEE

HMGL_HUMAN  240 QALTNTLMALQMGVSVVDSSVAGLGGCPYAQGASGNLATEDLVYMLEG. 287
HMGL_MOUSE  240 QALANTLVALQMGVSVVDSSVAGLGGCPYAKGASGNLATEDLVYMLNG. 287
HMGL_RAT    240 QALANTLVALQMGVSVVDSSVAGLGGCPYAKGASGNLATEDLVYMLTG. 287
HMGL_CHICK  213 QALANILVALQMGVSVVDASVAGLGGCPYAQGASGNVATEDLVYMLNG. 260
HMGL_PSEMV  211 MAIANVHAALAQGVRTFDSSVAGLGGCPYSPGASGNVATEDLLYLLHG. 258
LEU1_LACLA  208 MAVANSLAAIKAGAGRVEGTVNGIG......ERAGNAALEEIAVALHIR 250
NIFV_RHOCA  203 MATANTIMAAHAGATHLSVTVNGLG......ERAGNAACEEVGAALEAG 245
PYCB_METJA  208 LAPMTYLKVIEAGADMVDCAISPFA......MGTSQPPTESIVVALKG. 249
HIS4_MYCTU  204 DDLRAIATLTHRG...VEGAIVGKA........LYARRFTLPQALAAV. 240
HIS4_STRCO  200 DDLRAIAGLVPAG...VEGAIVGKA........LYAKAFTLEEALEAT. 236
HIS4_ECOLI  208 DD...VAALRGTG...VRGVIVGRA........LLEGKFTVKEAIACW. 241
HIS4_HAEIN  210 AD...IEALKGTG...VSGVIVGRA........LLEGKFTLSEAIKCW. 243
1qo2A       200 NSLKTAQKVHTETNGLLKGVIVGRA........FLEGILTVEVXKRYAR 240
1qo2A_ss        HHHHHHHHHHHH     EEEEE HH        HHH    HHHHHHH
```

F.G. HEGARDT. UB

J. PIE. Univ. Zar.

N . CASALS. UIC.

Hydroxymethylglutaryl-CoA lyase (HMGL)

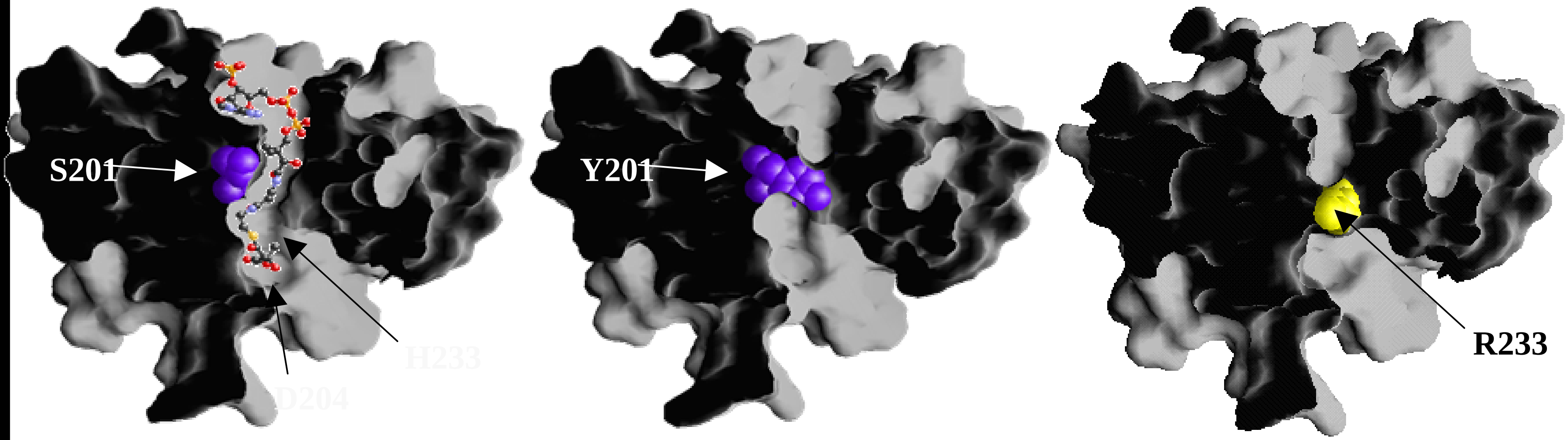


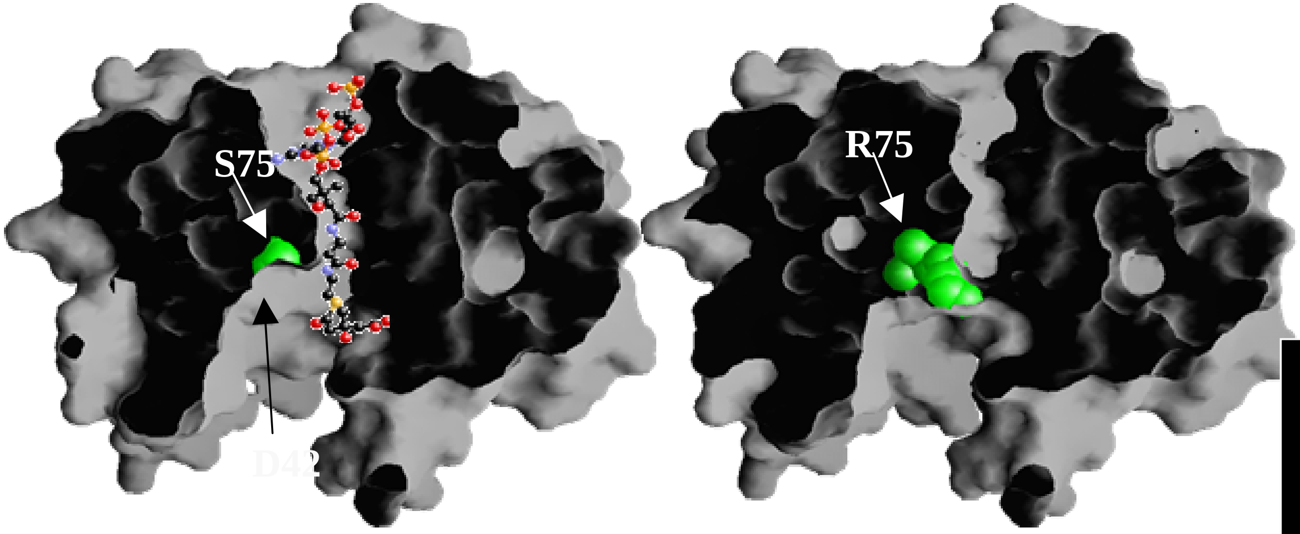
FG. Hegardt

J. Pie

N. Casals

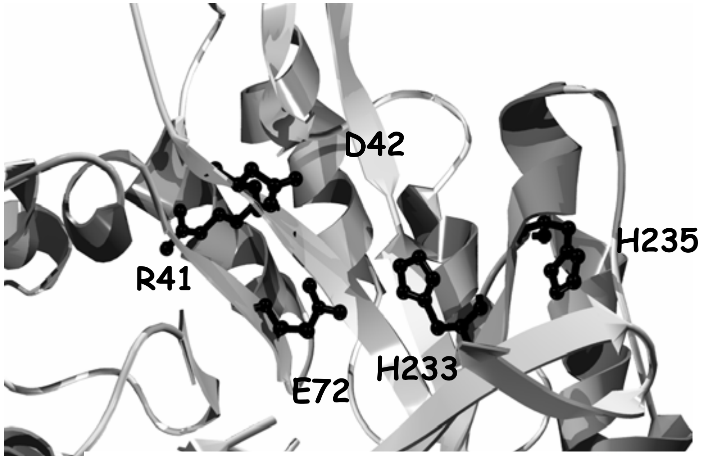
Hydroxymethylglutaryl-CoA lyase (HMGL)

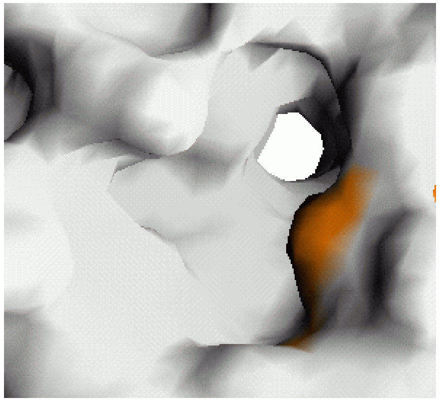


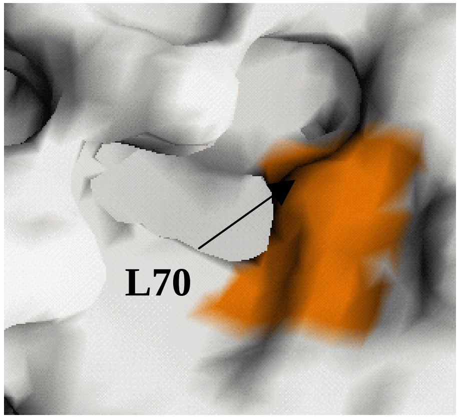


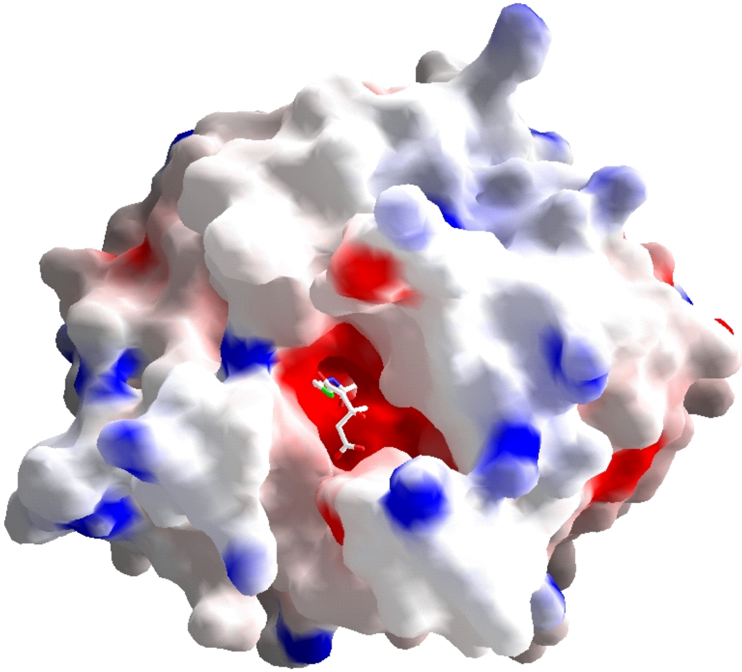
FG. Hegardt
J. Pie
N. Casals

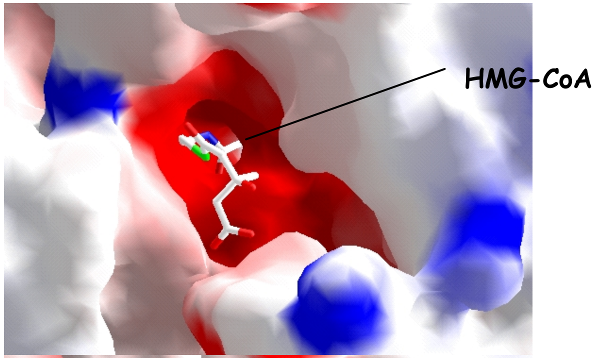
Hydroxymethylglutaryl-CoA lyase (HMGL)

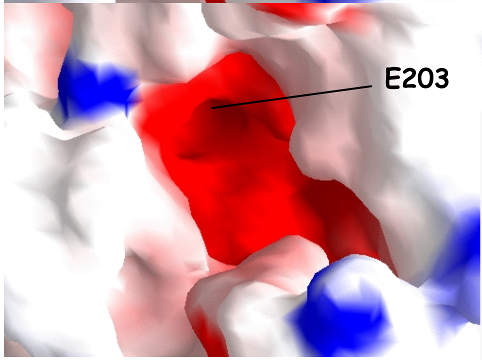










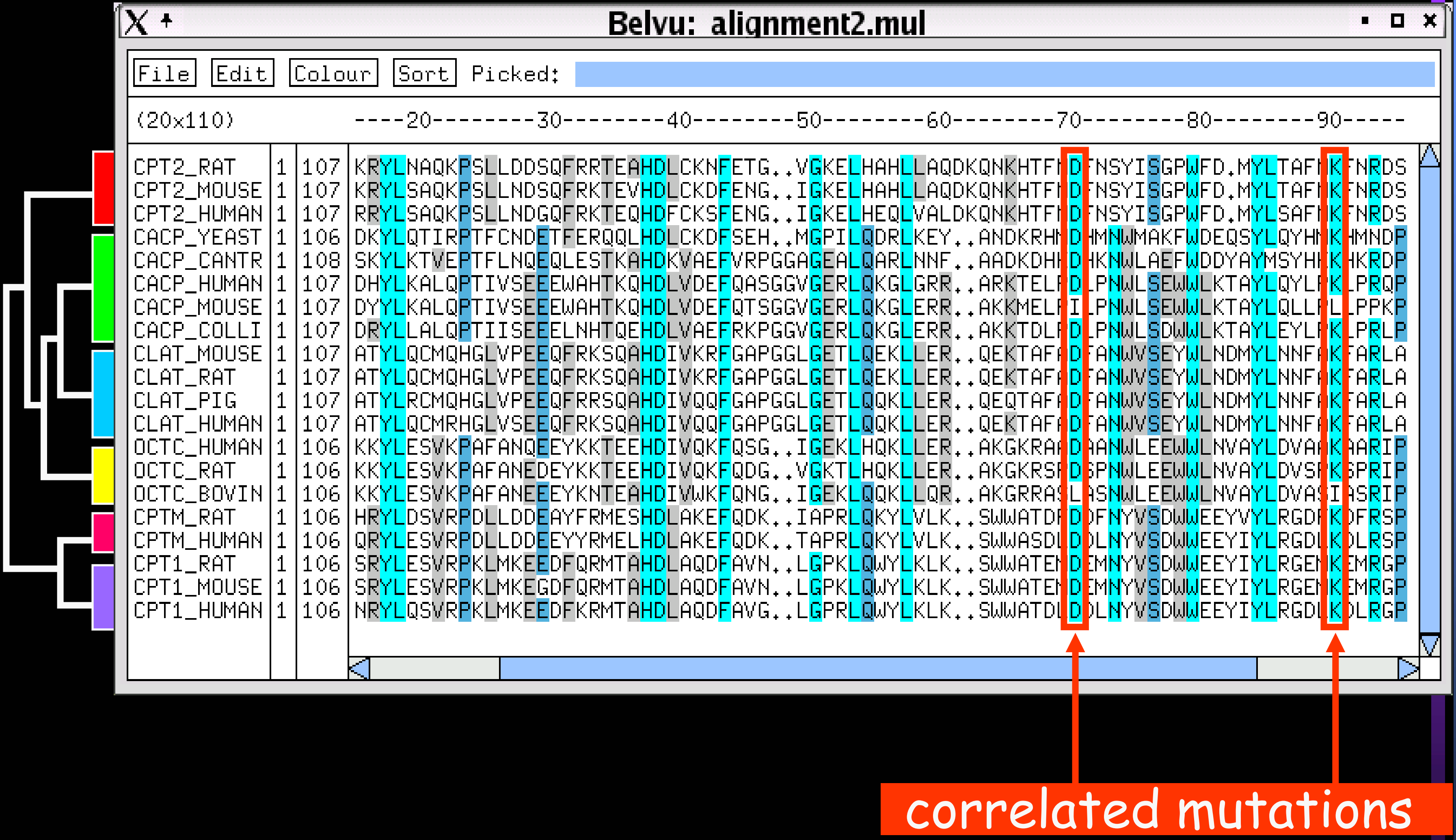


FG. Hegardt

J. Pie

N. Casals

Integration of Evolutive Information

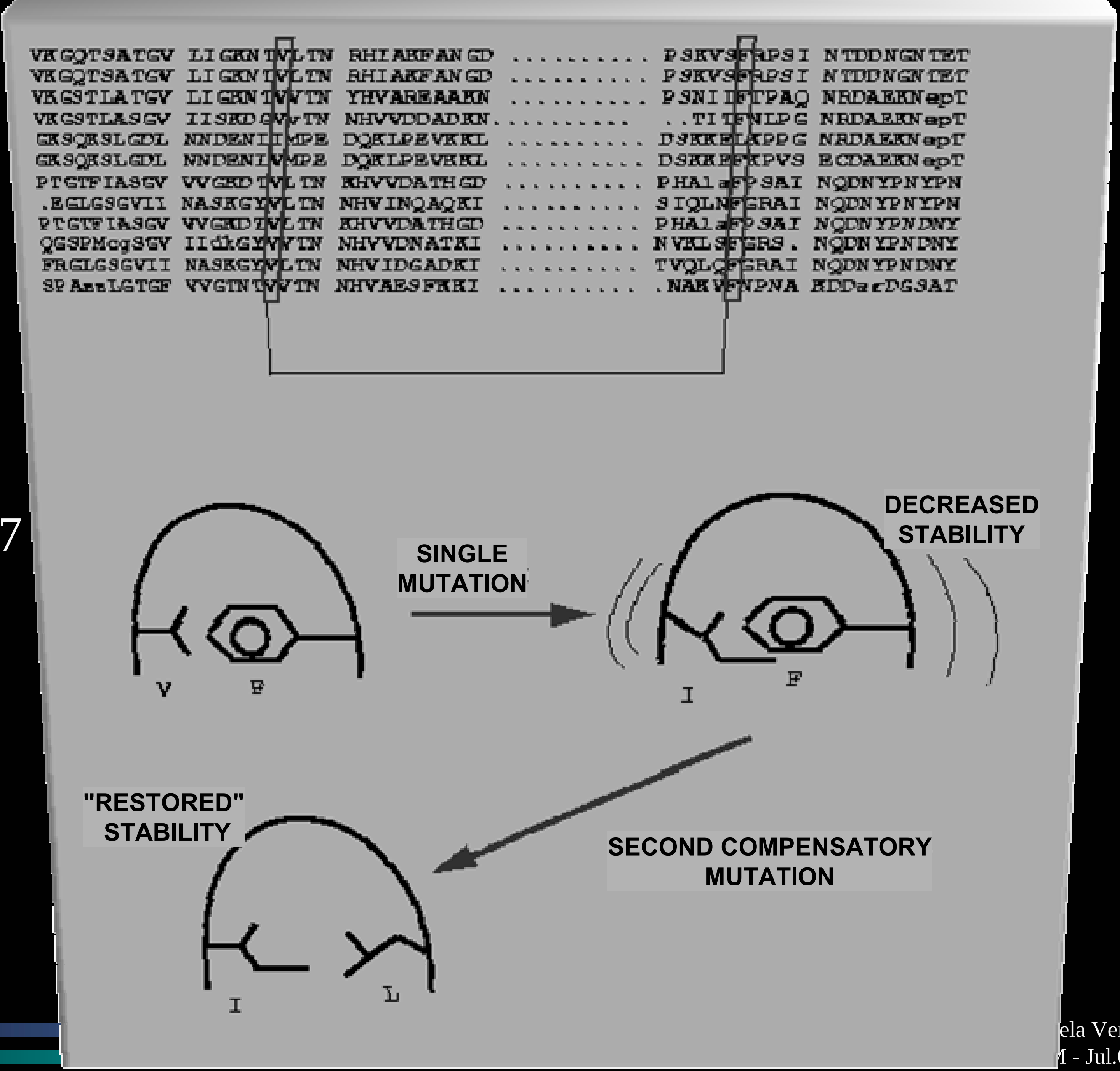


```
(20x110)            ----20--------30--------40--------50--------60--------70--------80--------90-----
CPT2_RAT    |1|107 |KRYLNAQKPSLLDDSQFRRTEAHDLCKNFETG..VGKELHAHLLAQDKQNKHTFNDFNSYISGPWFD.MYLTAFNKFNRDS
CPT2_MOUSE  |1|107 |KRYLSAQKPSLLNDSQFRKTEVHDLCKDFENG..IGKELHAHLLAQDKQNKHTFNDFNSYISGPWFD.MYLTAFNKFNRDS
CPT2_HUMAN  |1|107 |RRYLSAQKPSLLNDGQFRKTEQHDFCKSFENG..IGKELHEQLVALDKQNKHTFNDFNSYISGPWFD.MYLSAFNKFNRDS
CACP_YEAST  |1|106 |DKYLQTIRPTFCNDETFERQQLHDLCKDFSEH..MGPILQDRLKEY..ANDKRHNDHMNWMAKFWDEQSYLQYHNKHMNDP
CACP_CANTR  |1|108 |SKYLKTVEPTFLNQEQLESTKAHDKVAEFVRPGGAGEALQARLNNF..AADKDHKDHKNWLAEFWDDYAYMSYHKKHKRDP
CACP_HUMAN  |1|107 |DHYLKALQPTIVSEEEWAHTKQHDLVDEFQASGGVGERLQKGLGRR..ARKTELPDLPNWLSEWWLKTAYLQYLPKLPRQP
CACP_MOUSE  |1|107 |DYYLKALQPTIVSEEEWAHTKQHDLVDEFQTSGGVGERLQKGLERR..AKKMELPILPNWLSEWWLKTAYLQLLPLLPPKP
CACP_COLLI  |1|107 |DRYLLALQPTIISEEELNHTQEHDLVAEFRKPGGVGERLQKGLERR..AKKTDLPDLPNWLSDWWLKTAYLEYLPKLPRLP
CLAT_MOUSE  |1|107 |ATYLQCMQHGLVPEEQFRKSQAHDIVKRFGAPGGLGETLQEKLLER..QEKTAFADFANWVSEYWLNDMYLNNFAKFARLA
CLAT_RAT    |1|107 |ATYLQCMQHGLVPEEQFRKSQAHDIVKRFGAPGGLGETLQEKLLER..QEKTAFADFANWVSEYWLNDMYLNNFAKFARLA
CLAT_PIG    |1|107 |ATYLRCMQHGLVPEEQFRRSQAHDIVQQFGAPGGLGETLQQKLLER..QEQTAFADFANWVSEYWLNDMYLNNFAKFARLA
CLAT_HUMAN  |1|107 |ATYLQCMRHGLVSEEQFRKSQAHDIVQQFGAPGGLGETLQQKLLER..QEKTAFADFANWVSEYWLNDMYLNNFAKFARLA
OCTC_HUMAN  |1|106 |KKYLESVKPAFANQEEYKKTEEHDIVQKFQSG..IGEKLHQKLLER..AKGKRAADAANWLEEWWLNVAYLDVAAKAARIP
OCTC_RAT    |1|106 |KKYLESVKPAFANEDEYKKTEEHDIVQKFQDG..VGKTLHQKLLER..AKGKRSPDSPNWLEEWWLNVAYLDVSPKSPRIP
OCTC_BOVIN  |1|106 |KKYLESVKPAFANEEEYKNTEAHDIVWKFQNG..IGEKLQQKLLQR..AKGRRASLASNWLEEWWLNVAYLDVASIASRIP
CPTM_RAT    |1|106 |HRYLDSVRPDLLDDEAYFRMESHDLAKEFQDK..IAPRLQKYLVLK..SWWATDPDDFNYVSDWWEEYVYLRGDPKDFRSP
CPTM_HUMAN  |1|106 |QRYLESVRPDLLDDEEYYRMELHDLAKEFQDK..TAPRLQKYLVLK..SWWASDLDDLNYVSDWWEEYIYLRGDLKDLRSP
CPT1_RAT    |1|106 |SRYLESVRPKLMKEEDFQRMTAHDLAQDFAVN..LGPKLQWYLKLK..SWWATENDEMNYVSDWWEEYIYLRGENKEMRGP
CPT1_MOUSE  |1|106 |SRYLESVRPKLMKEGDFQRMTAHDLAQDFAVN..LGPKLQWYLKLK..SWWATENDEMNYVSDWWEEYIYLRGENKEMRGP
CPT1_HUMAN  |1|106 |NRYLQSVRPKLMKEEDFKRMTAHDLAQDFAVG..LGPRLQWYLKLK..SWWATDLDDLNYVSDWWEEYIYLRGDLKDLRGP
```

Information extracted from multiple sequence alignments

Correlated Mutations

Pazos et al.
J. Mol. Biol., 1997



Correlated mutations: CorrMut & PlotCorr

```
         i         j
1 ....S.........P......
2 ....V.........L......
3 ....L.........M......
```

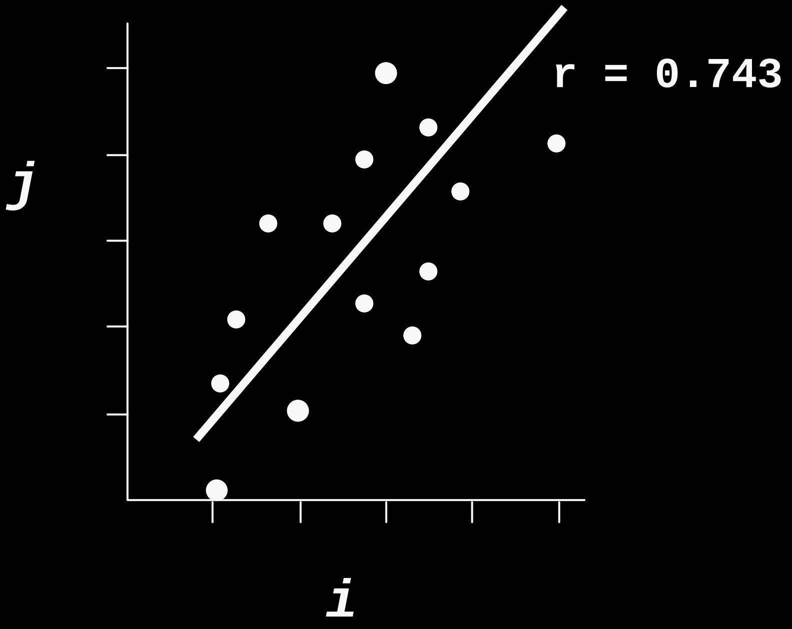
	1	2	3
1	-	SV	SL
2	-	-	VL
3	-	-	-

	1	2	3
1	-	PL	PM
2	-	-	LM
3	-	-	-

	1	2	3
1	-	1	2
2	-	-	3
3	-	-	-

	1	2	3
1	-	0	1
2	-	-	5
3	-	-	-

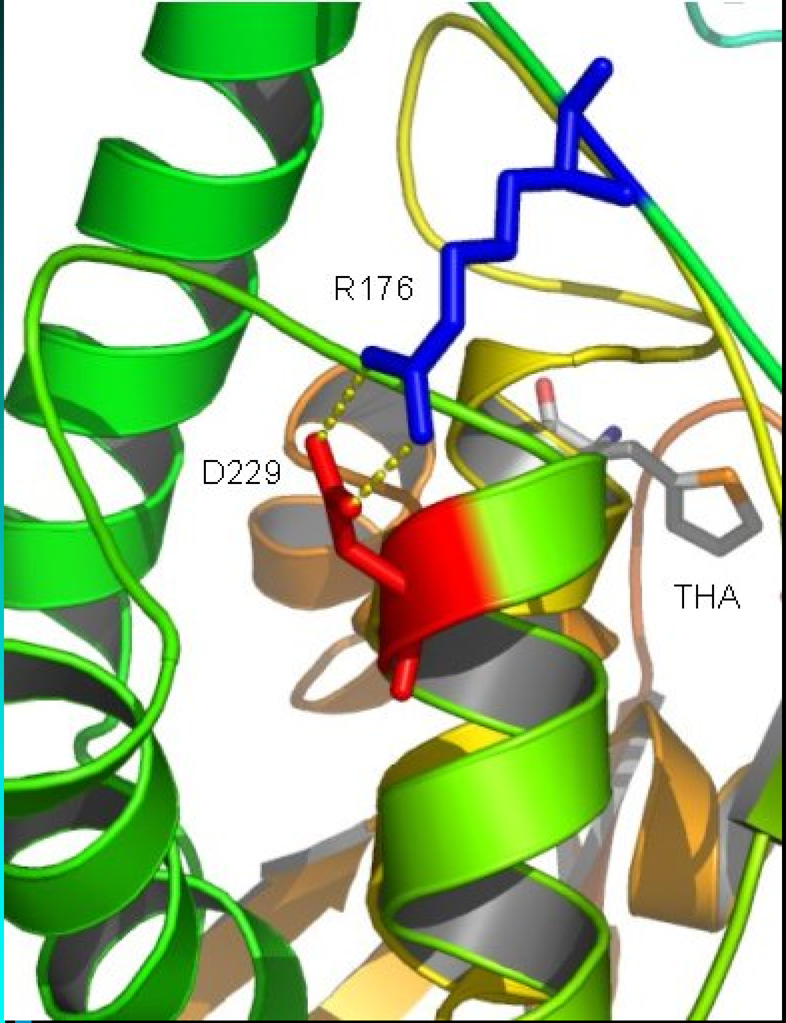
McLachlan matrix



Goebel, Sander, Schneider & Valencia. Proteins, 1994
Olmea & Valencia. Folding & Design, 1997
Pazos, Olmea & Valencia. CABIOS, 1997

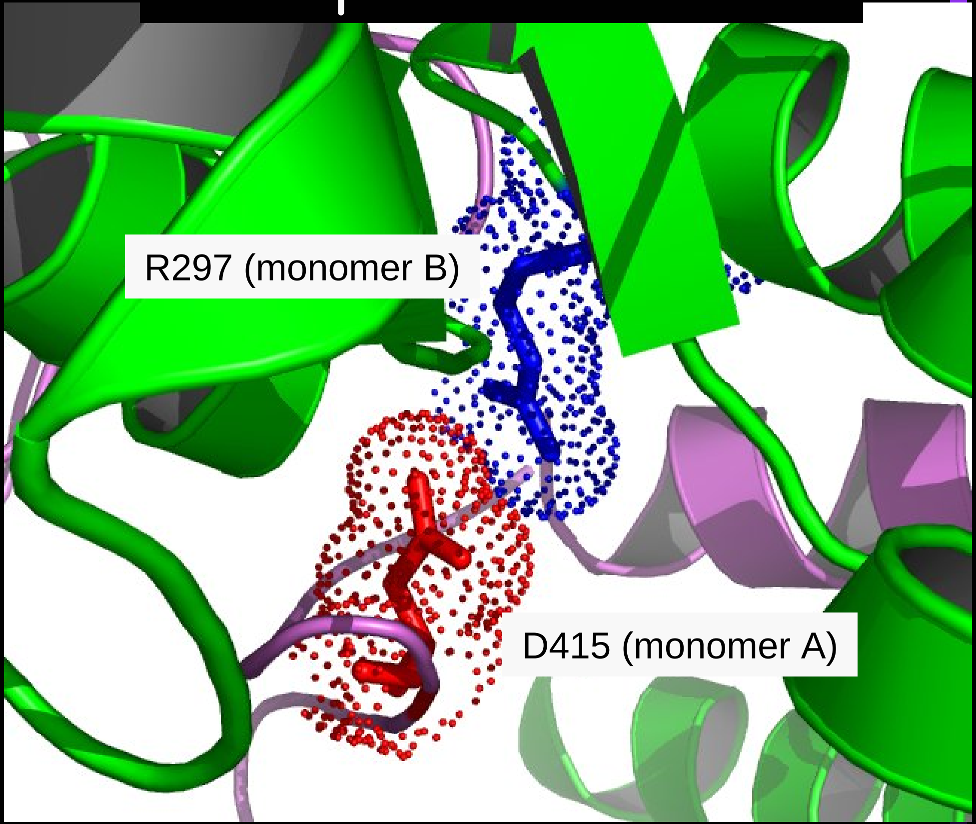
Correlated mutations

Phenylalanine hydroxylase (**PAH**) homotetramer.

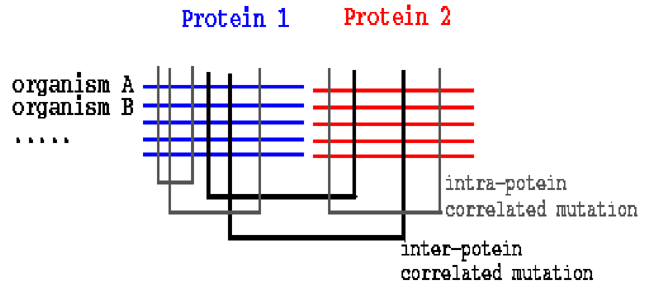


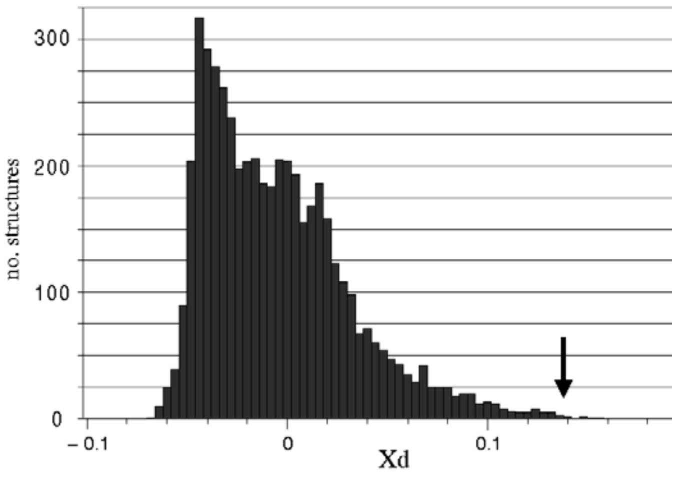
intra-protein contacts

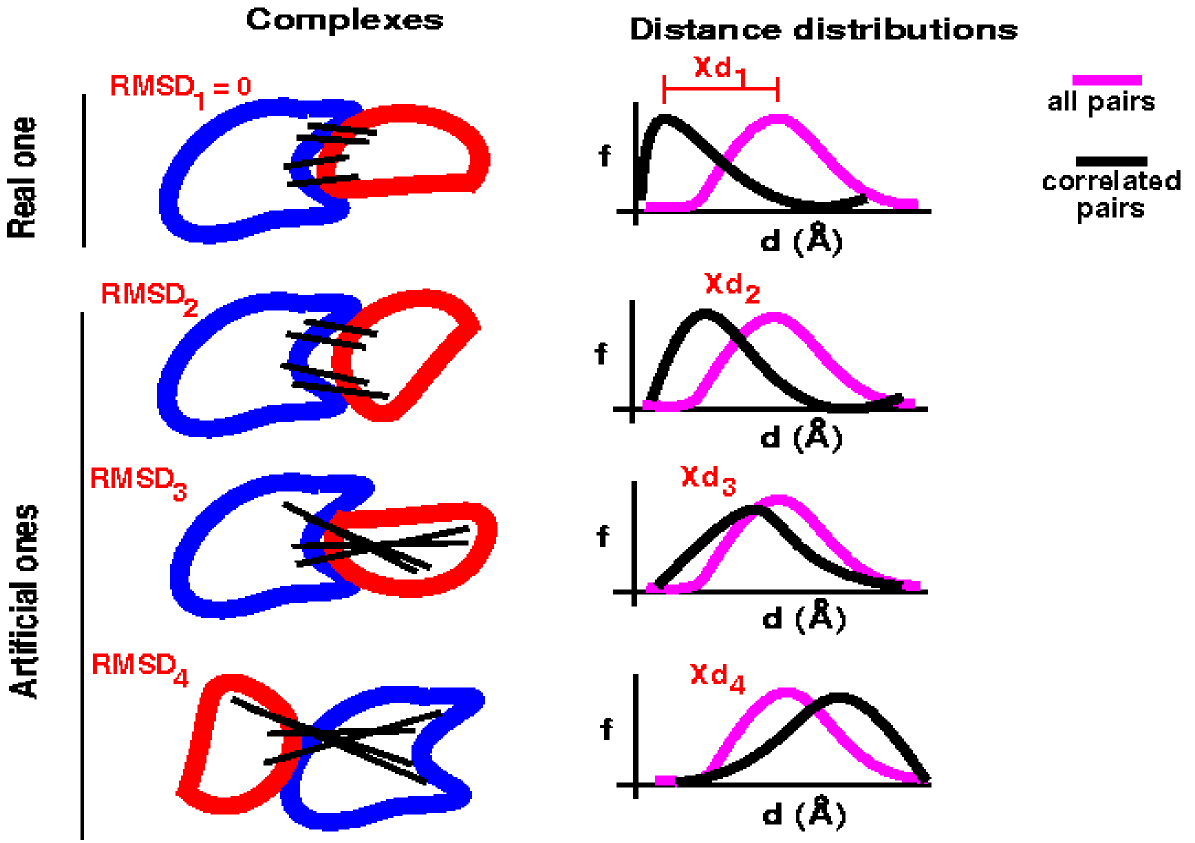
inter-protein contacts



Docking and inter-protein correlated mutations

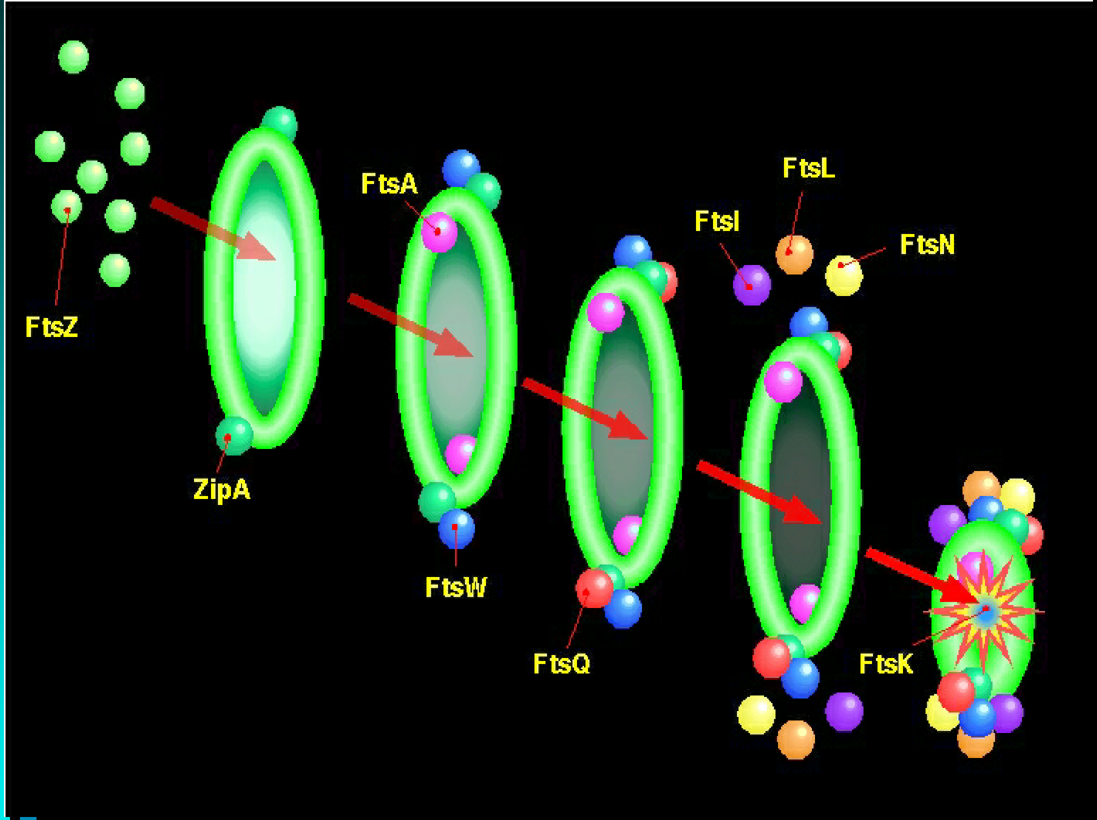






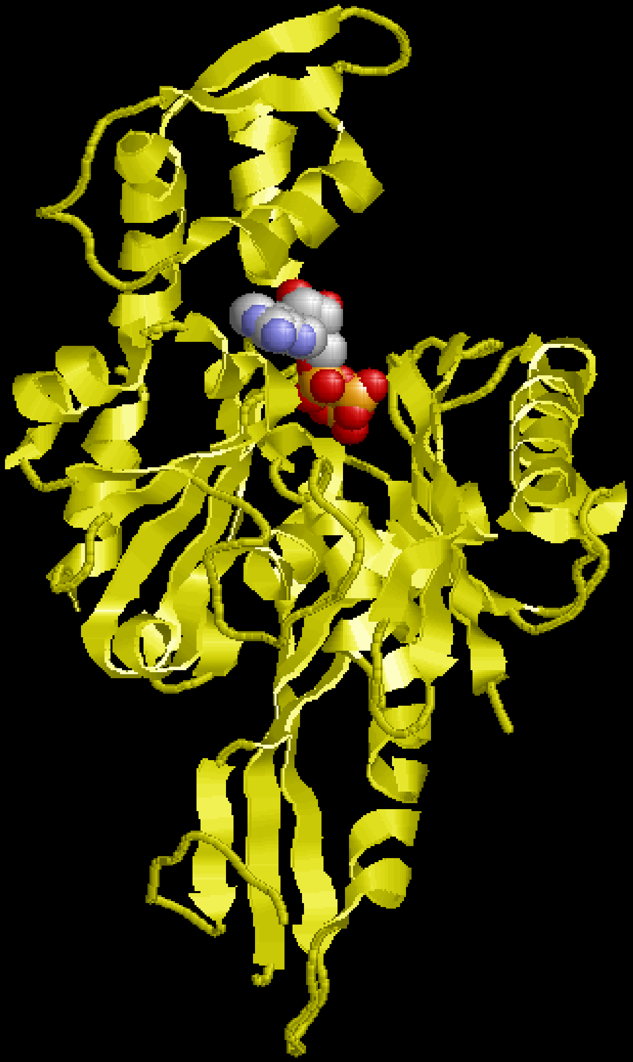
$$Xd = \sum_{j=1}^{j=n} \frac{P_{jc} - P_{ja}}{d_j \cdot n}$$

Pazos et al. J. Mol. Biol., 1997

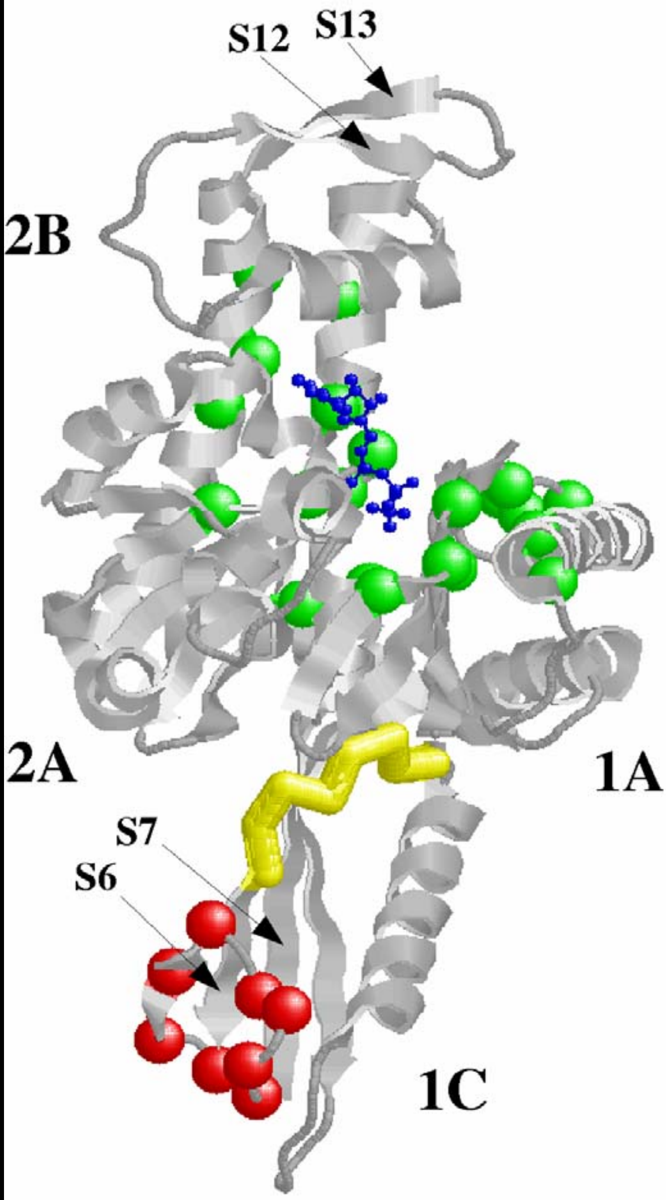


Bacterial cell septation

M. VICENTE. CNB. Madrid.

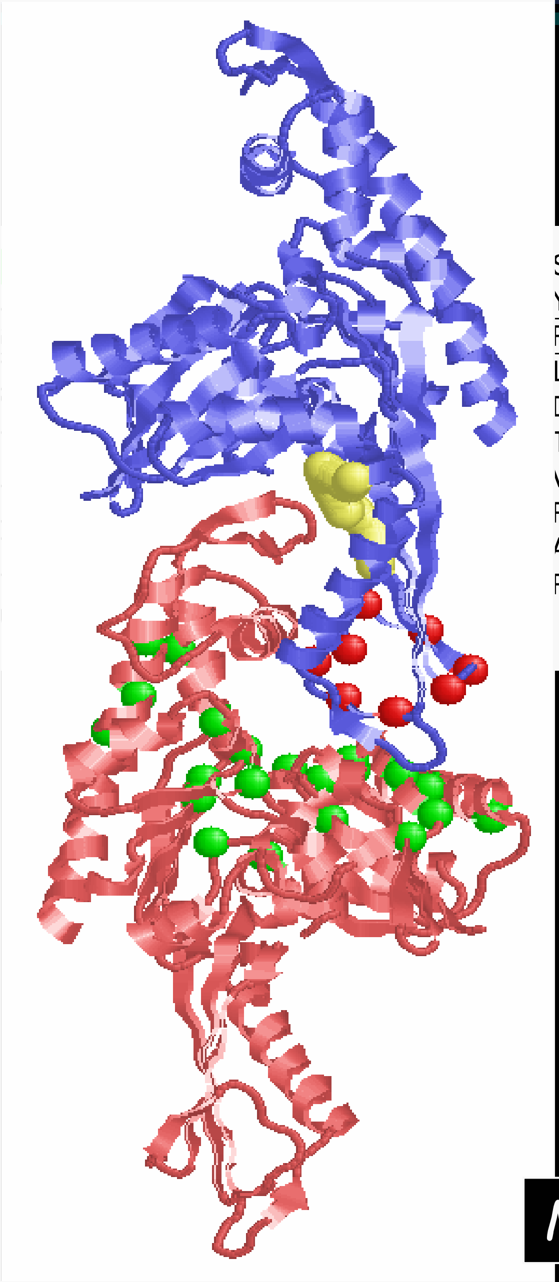


FtsA



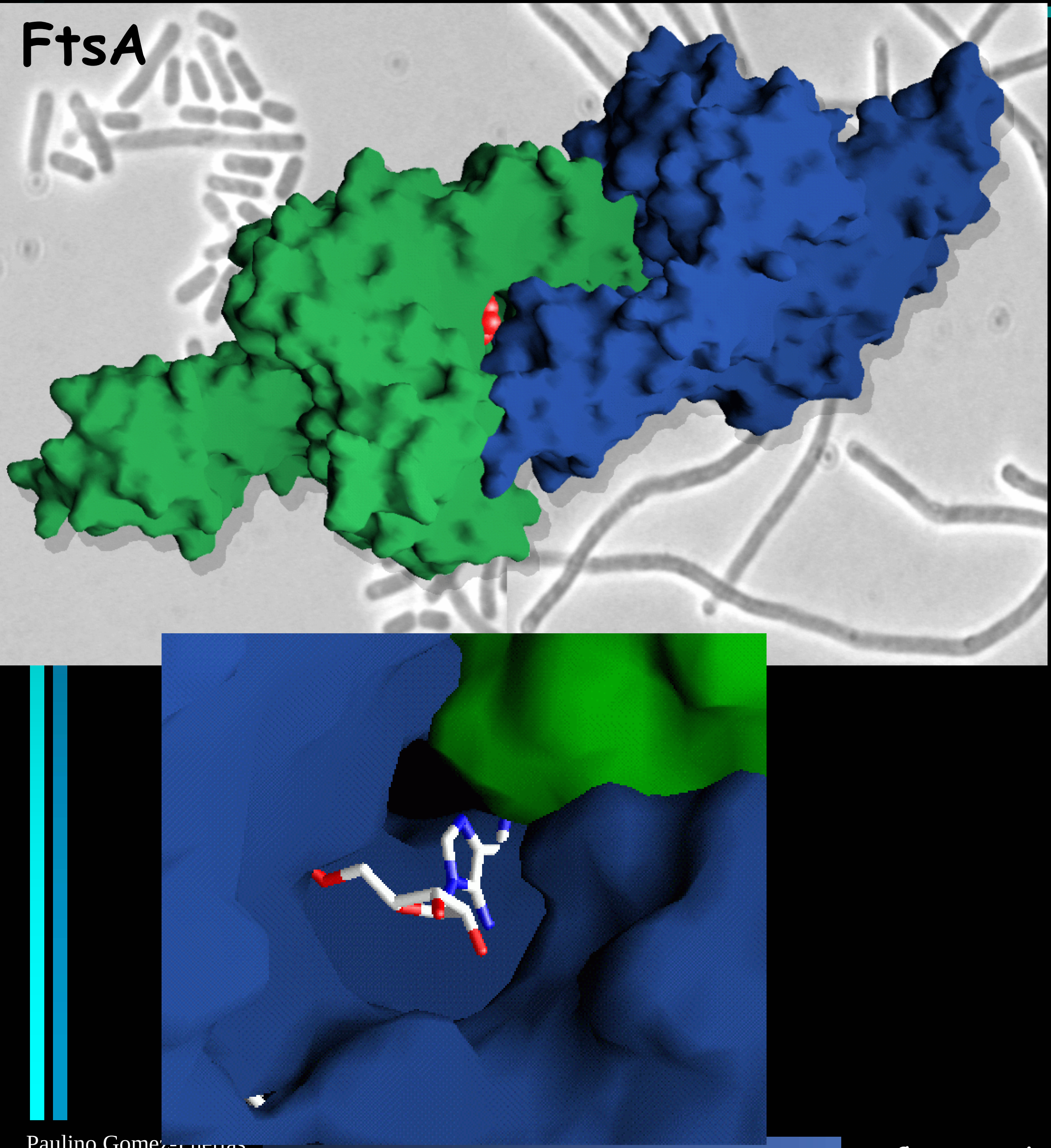
A_ECOLI

IKATDRKLVVGLE**I** S 40
P**S**R**GMD**KGGVNDLE Y 80
ALSGK**H**ISCQNEIG R 120
RDEH**RVLHVIPQ**EY L 160
TCHNDMAKNIVKAV D 200
RELGVCVVDIG**G**GT T 240
DIAYAFGTPPSDAE V 280
GRPPRSLQRQTLAE R 320
QGVKHHLAAGIV**L**T A 360
LNITGLTDYAQEPY R 400
TASVGSWIKRLNSW 420

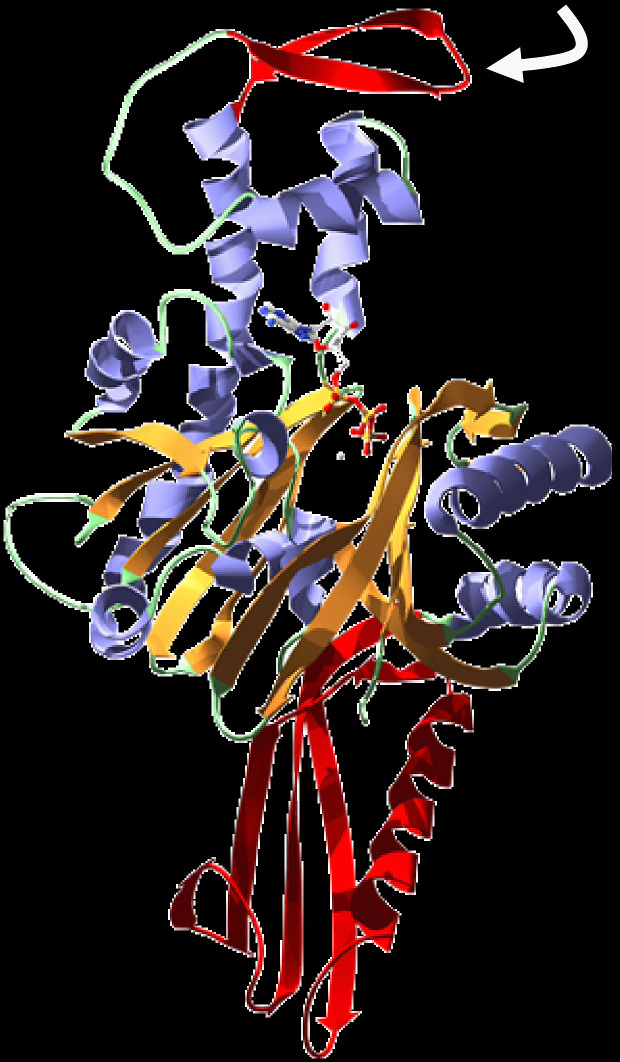


M. Vicente

Carettoni, Gomez-Puertas, et al. Proteins, 2003.



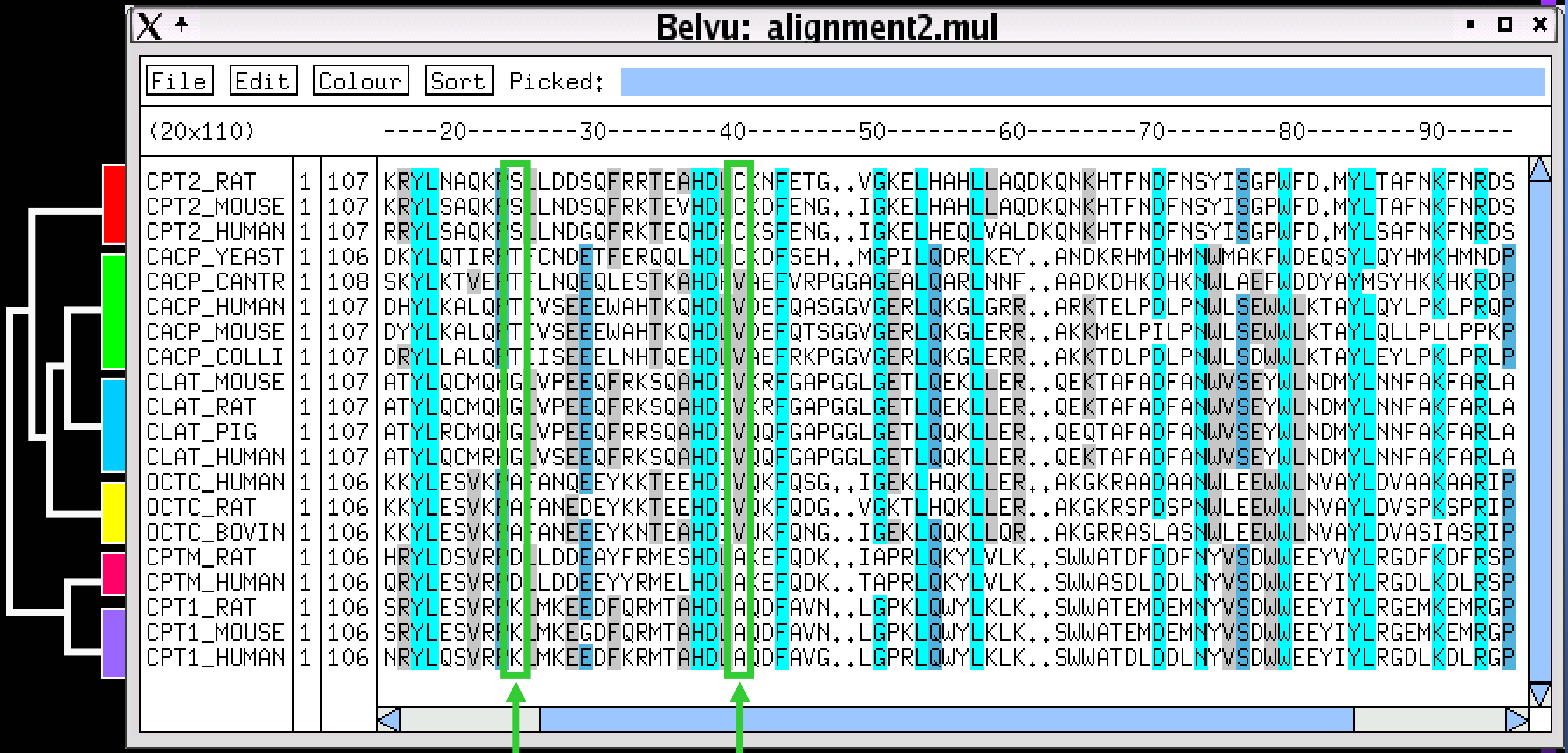
PROPOSED MODIFICATIONS:
SUCCESS!!



Rico et al - M. Vicente's Lab

Carettoni, Gomez-Puertas, et al. Proteins, 2003.

Integration of Evolutive Information

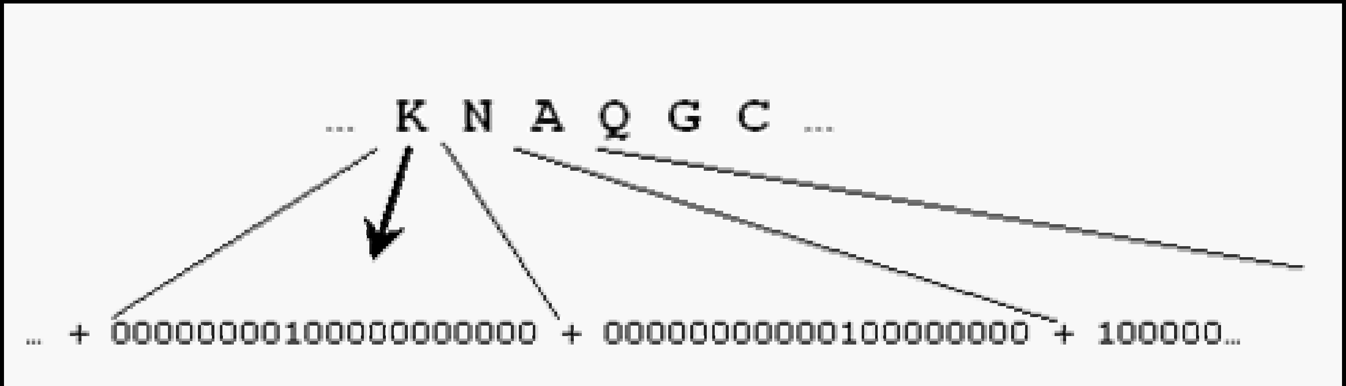


```
(20x110)                     ----20--------30--------40--------50--------60--------70--------80--------90-----
CPT2_RAT    | 1 | 107 | KRYLNAQKPSLLDDSQFRRTEAHDLCKNFETG..VGKELHAHLLAQDKQNKHTFNDFNSYISGPWFD.MYLTAFNKFNRDS
CPT2_MOUSE  | 1 | 107 | KRYLSAQKPSLLNDSQFRKTEVHDLCKDFENG..IGKELHAHLLAQDKQNKHTFNDFNSYISGPWFD.MYLTAFNKFNRDS
CPT2_HUMAN  | 1 | 107 | RRYLSAQKPSLLNDGQFRKTEQHDFCKSFENG..IGKELHEQLVALDKQNKHTFNDFNSYISGPWFD.MYLSAFNKFNRDS
CACP_YEAST  | 1 | 106 | DKYLQTIRPTFCNDETFERQQLHDLCKDFSEH..MGPILQDRLKEY..ANDKRHMDHMNWMAKFWDEQSYLQYHMKHMNDP
CACP_CANTR  | 1 | 108 | SKYLKTVEPTFLNQEQLESTKAHDKVAEFVRPGGAGEALQARLNNF..AADKDHKDHKNWLAEFWDDYAYMSYHKKHKRDP
CACP_HUMAN  | 1 | 107 | DHYLKALQPTIVSEEEWAHTKQHDLVDEFQASGGVGERLQKGLGRR..ARKTELPDLPNWLSEWWLKTAYLQYLPKLPRQP
CACP_MOUSE  | 1 | 107 | DYYLKALQPTIVSEEEWAHTKQHDLVDEFQTSGGVGERLQKGLERR..AKKMELPILPNWLSEWWLKTAYLQLLPLLPPKP
CACP_COLLI  | 1 | 107 | DRYLLALQPTIISEEELNHTQEHDLVAEFRKPGGVGERLQKGLERR..AKKTDLPDLPNWLSDWWLKTAYLEYLPKLPRLP
CLAT_MOUSE  | 1 | 107 | ATYLQCMQHGLVPEEQFRKSQAHDIVKRFGAPGGLGETLQEKLLER..QEKTAFADFANWVSEYWLNDMYLNNFAKFARLA
CLAT_RAT    | 1 | 107 | ATYLQCMQHGLVPEEQFRKSQAHDIVKRFGAPGGLGETLQEKLLER..QEKTAFADFANWVSEYWLNDMYLNNFAKFARLA
CLAT_PIG    | 1 | 107 | ATYLRCMQHGLVPEEQFRRSQAHDIVQQFGAPGGLGETLQQKLLER..QEQTAFADFANWVSEYWLNDMYLNNFAKFARLA
CLAT_HUMAN  | 1 | 107 | ATYLQCMRHGLVSEEQFRKSQAHDIVQQFGAPGGLGETLQQKLLER..QEKTAFADFANWVSEYWLNDMYLNNFAKFARLA
OCTC_HUMAN  | 1 | 106 | KKYLESVKPAFANQEEYKKTEEHDIVQKFQSG..IGEKLHQKLLER..AKGKRAADAANWLEEWWLNVAYLDVAAKAARIP
OCTC_RAT    | 1 | 106 | KKYLESVKPAFANEDEYKKTEEHDIVQKFQDG..VGKTLHQKLLER..AKGKRSPDSPNWLEEWWLNVAYLDVSPKSPRIP
OCTC_BOVIN  | 1 | 106 | KKYLESVKPAFANEEEYKNTEAHDIVWKFQNG..IGEKLQQKLLQR..AKGRRASLASNWLEEWWLNVAYLDVASIASRIP
CPTM_RAT    | 1 | 106 | HRYLDSVRPDLLDDEAYFRMESHDLAKEFQDK..IAPRLQKYLVLK..SWWATDFDDFNYVSDWWEEYVYLRGDFKDFRSP
CPTM_HUMAN  | 1 | 106 | QRYLESVRPDLLDDEEYYRMELHDLAKEFQDK..TAPRLQKYLVLK..SWWASDLDDLNYVSDWWEEYIYLRGDLKDLRSP
CPT1_RAT    | 1 | 106 | SRYLESVRPKLMKEEDFQRMTAHDLAQDFAVN..LGPKLQWYLKLK..SWWATEMDEMNYVSDWWEEYIYLRGEMKEMRGP
CPT1_MOUSE  | 1 | 106 | SRYLESVRPKLMKEGDFQRMTAHDLAQDFAVN..LGPKLQWYLKLK..SWWATEMDEMNYVSDWWEEYIYLRGEMKEMRGP
CPT1_HUMAN  | 1 | 106 | NRYLQSVRPKLMKEEDFKRMTAHDLAQDFAVG..LGPRLQWYLKLK..SWWATDLDDLNYVSDWWEEYIYLRGDLKDLRGP
```

tree-determinants

Information extracted from multiple sequence alignments

Sequence Space

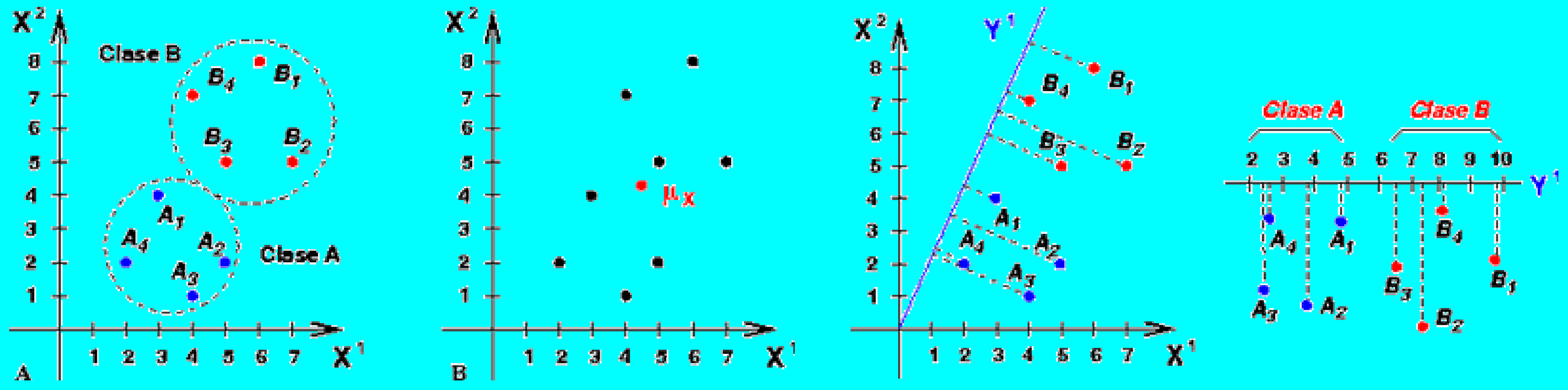


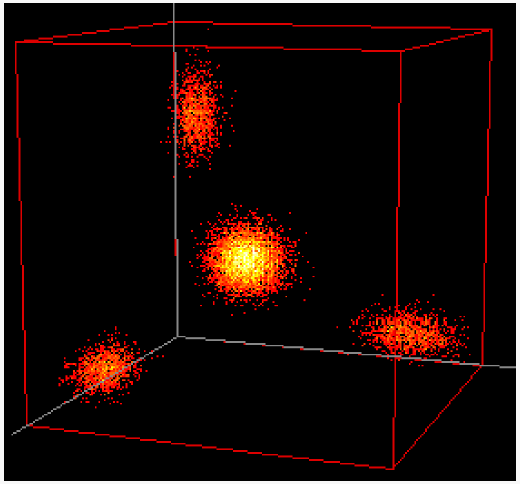
```
                 Residue 1                Residue 2         ...       Residue n
Secuence 1 00000100000000000000 00000000000000010000 ... 00000000100000000000
Secuence 2 00000100000000000000 00000000010000000000 ... 00000000000001000000
Secuence 1 00000000010000000000 00000000000000010000 ... 00001000000000000000
...
Secuence x 00000001000000000000 00001000000000000000 ... 00000000100000000000
```

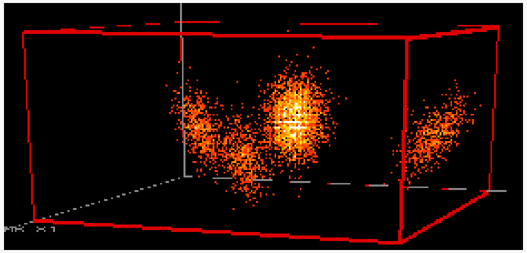
Sequencespace
(Casari, Sander, Valencia, Nature. Str. Bio. 95)

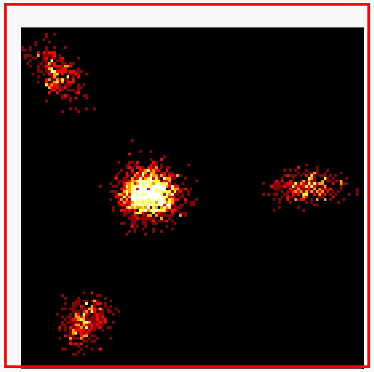
PRINCIPAL COMPONENTS ANALYSIS

2D → 1D



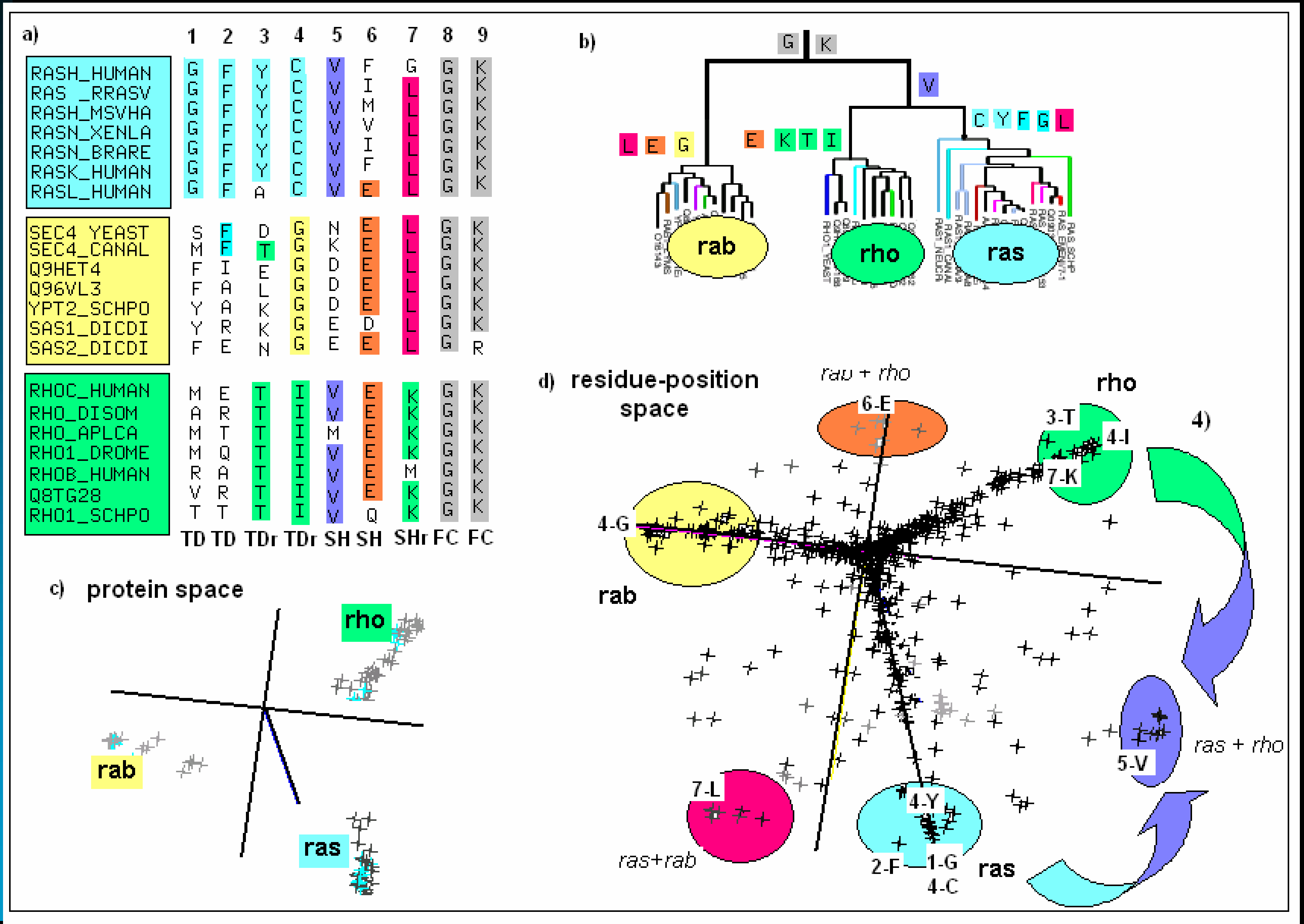






3D → 2D

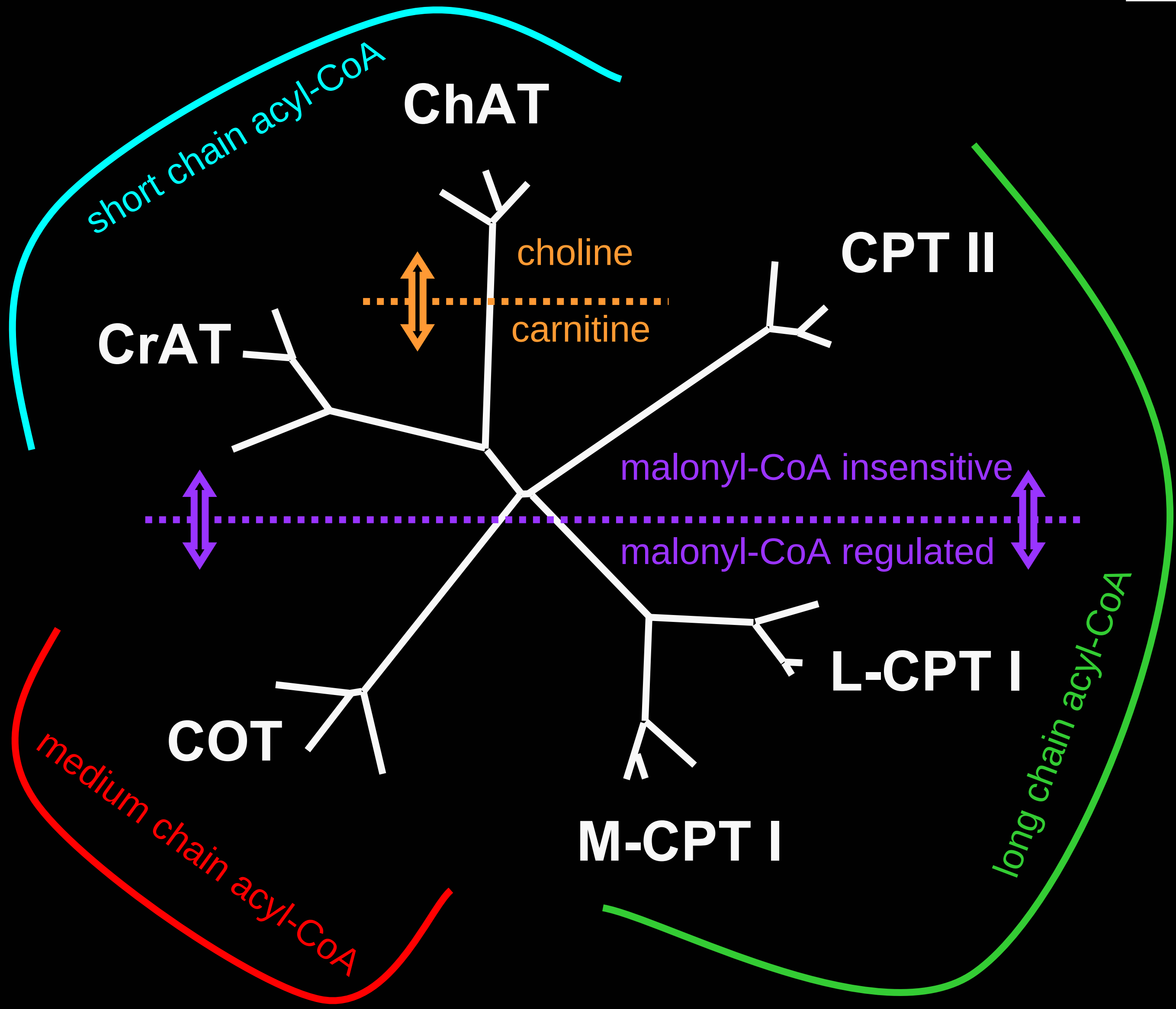
predicting key positions from MSA



	1	2	3	4	5	6	7	8	9
RASH_HUMAN	G	F	Y	C	V	F	G	G	K
RAS _RRASV	G	F	Y	C	V	I	L	G	K
RASH_MSVHA	G	F	Y	C	V	M	L	G	K
RASN_XENLA	G	F	Y	C	V	V	L	G	K
RASN_BRARE	G	F	Y	C	V	I	L	G	K
RASK_HUMAN	G	F	Y	C	V	F	L	G	K
RASL_HUMAN	G	F	A	C	V	E	L	G	K
SEC4 YEAST	S	F	D	G	N	E	L	G	K
SEC4_CANAL	M	F	T	G	K	E	L	G	K
Q9HET4	F	I	E	G	D	E	L	G	K
Q96VL3	F	A	L	G	D	E	L	G	K
YPT2_SCHPO	Y	A	K	G	D	E	L	G	K
SAS1_DICDI	Y	R	K	G	E	D	L	G	K
SAS2_DICDI	F	E	N	G	E	E	L	G	R
RHOC_HUMAN	M	E	T	I	V	E	K	G	K
RHO_DISOM	A	R	T	I	V	E	K	G	K
RHO_APLCA	M	T	T	I	M	E	K	G	K
RHO1_DROME	M	Q	T	I	V	E	K	G	K
RHOB_HUMAN	R	A	T	I	V	E	M	G	K
Q8TG28	V	R	T	I	V	E	K	G	K
RHO1_SCHPO	T	T	T	I	V	Q	K	G	K
	TD	TD	TDr	TDr	SH	SH	SHr	FC	FC

carnitine/choline acyl transferases

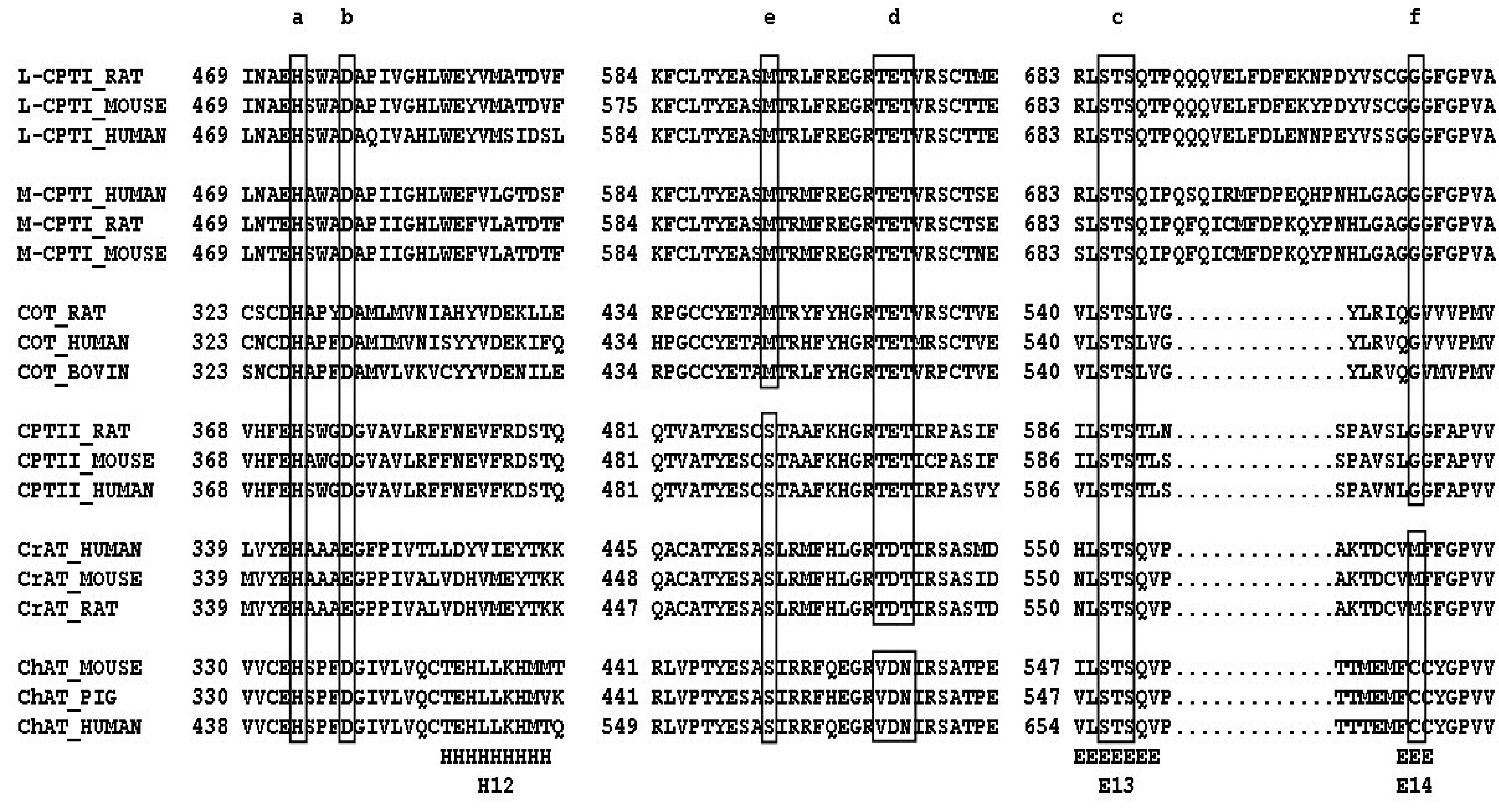
F.G. Hegardt



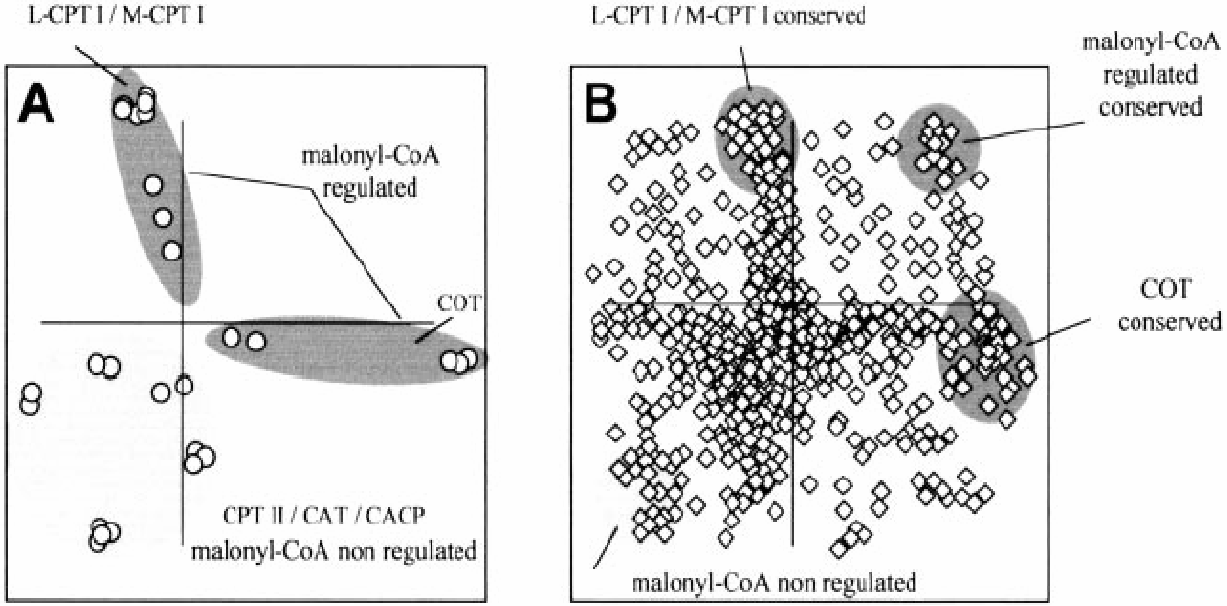
Carnitine-Choline: Thr/Glu/Thr vs. Val/Asp/Asn

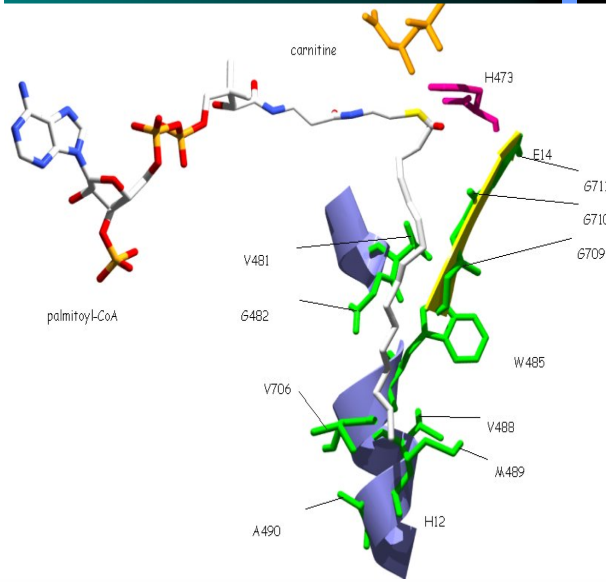
Malonyl-CoA regulation: Met vs. Ser

Short vs. Long substrate: Gly vs. Met



```
                                   a    b                                              e                d                           c                     f

L-CPTI_RAT    469 INAEHSWADAPIVGHLWEYVMATDVF   584 KFCLTYEASMTRLFREGRTETVRSCTME   683 RLSTSQTPQQQVELFDFEKNPDYVSCGGFGPVA
L-CPTI_MOUSE  469 INAEHSWADAPIVGHLWEYVMATDVF   575 KFCLTYEASMTRLFREGRTETVRSCTTE   683 RLSTSQTPQQQVELFDFEKYPDYVSCGGFGPVA
L-CPTI_HUMAN  469 LNAEHSWADAQIVAHLWEYVMSIDSL   584 KFCLTYEASMTRLFREGRTETVRSCTTE   683 RLSTSQTPQQQVELFDLENNPEYVSSGGFGPVA

M-CPTI_HUMAN  469 LNAEHAWADAPIIGHLWEFVLGTDSF   584 KFCLTYEASMTRMFREGRTETVRSCTSE   683 RLSTSQIPQSQIRMFDPEQHPNHLGAGGFGPVA
M-CPTI_RAT    469 LNTEHSWADAPIIGHLWEFVLATDTF   584 KFCLTYEASMTRMFREGRTETVRSCTSE   683 SLSTSQIPQFQICMFDPKQYPNHLGAGGFGPVA
M-CPTI_MOUSE  469 LNTEHSWADAPIIGHLWEFVLATDTF   584 KFCLTYEASMTRMFREGRTETVRSCTNE   683 SLSTSQIPQFQICMFDPKQYPNHLGAGGFGPVA

COT_RAT       323 CSCDHAPYDAMLMVNIAHYVDEKLLE   434 RPGCCYETAMTRYFYHGRTETVRSCTVE   540 VLSTSLVG.............YLRIQGVVVPMV
COT_HUMAN     323 CNCDHAPFDAMIMVNISYYVDEKIFQ   434 HPGCCYETAMTRHFYHGRTETMRSCTVE   540 VLSTSLVG.............YLRVQGVVVPMV
COT_BOVIN     323 SNCDHAPFDAMVLVKVCYYVDENILE   434 RPGCCYETAMTRLFYHGRTETVRPCTVE   540 VLSTSLVG.............YLRVQGVMVPMV

CPTII_RAT     368 VHFEHSWGDGVAVLRFFNEVFRDSTQ   481 QTVATYESCSTAAFKHGRTETIRPASIF   586 ILSTSTLN............SPAVSLGGFAPVV
CPTII_MOUSE   368 VHFEHAWGDGVAVLRFFNEVFRDSTQ   481 QTVATYESCSTAAFKHGRTETICPASIF   586 ILSTSTLS............SPAVSLGGFAPVV
CPTII_HUMAN   368 VHFEHSWGDGVAVLRFFNEVFKDSTQ   481 QTVATYESCSTAAFKHGRTETIRPASVY   586 VLSTSTLS............SPAVNLGGFAPVV

CrAT_HUMAN    339 LVYEHAAAEGFPIVTLLDYVIEYTKK   445 QACATYESASLRMFHLGRTDTIRSASMD   550 HLSTSQVP............AKTDCVMFFGPVV
CrAT_MOUSE    339 MVYEHAAAEGPPIVALVDHVMEYTKK   448 QACATYESASLRMFHLGRTDTIRSASID   550 NLSTSQVP............AKTDCVMFFGPVV
CrAT_RAT      339 MVYEHAAAEGPPIVALVDHVMEYTKK   447 QACATYESASLRMFHLGRTDTIRSASTD   550 NLSTSQVP............AKTDCVMSFGPVV

ChAT_MOUSE    330 VVCEHSPFDGIVLVQCTEHLLKHMMT   441 RLVPTYESASIRRFQEGRVDNIRSATPE   547 ILSTSQVP............TTMEMFCCYGPVV
ChAT_PIG      330 VVCEHSPFDGIVLVQCTEHLLKHMVK   441 RLVPTYESASIRRFHEGRVDNIRSATPE   547 VLSTSQVP............TTMEMFCCYGPVV
ChAT_HUMAN    438 VVCEHSPFDGIVLVQCTEHLLKHMTQ   549 RLVPTYESASIRRFQEGRVDNIRSATPE   654 VLSTSQVP............TTTEMFCCYGPVV
                                    HHHHHHHHH                                                       EEEEEEE                EEE
                                       H12                                                            E13                  E14
```





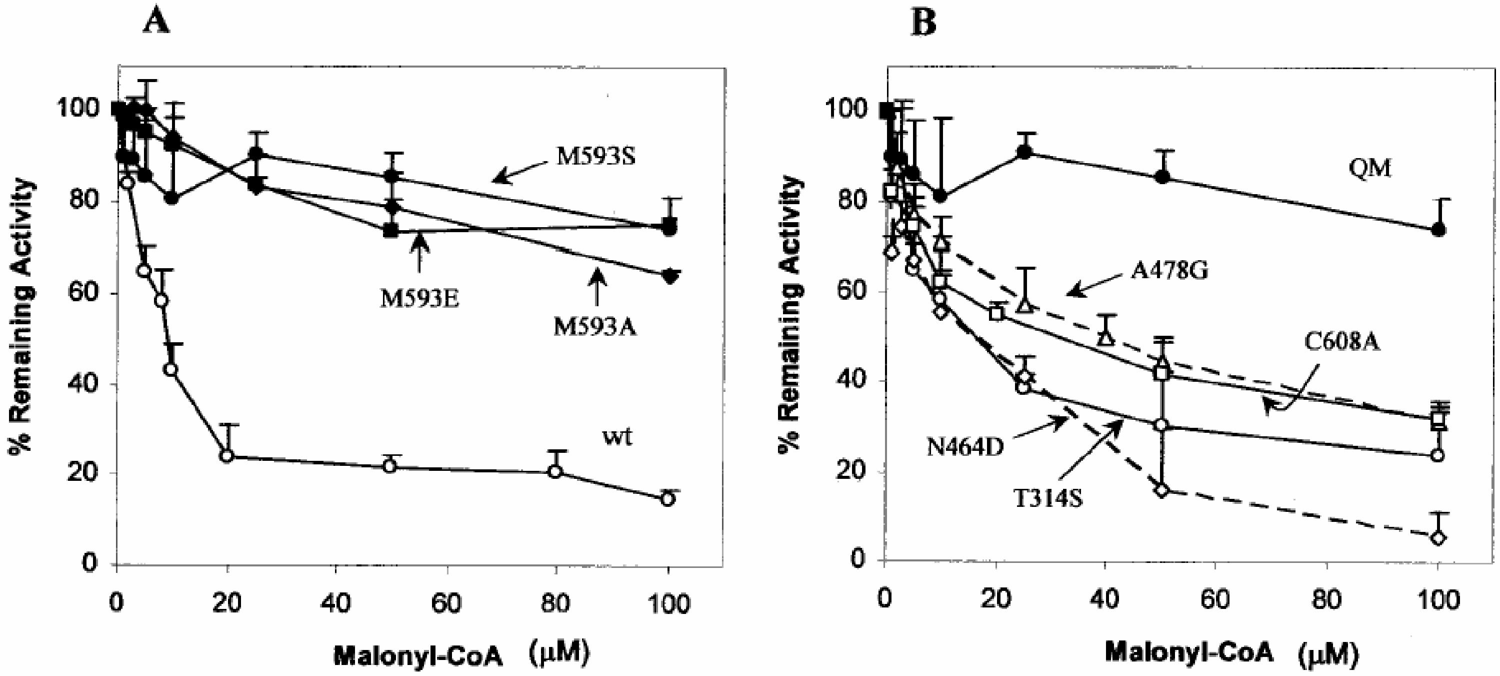
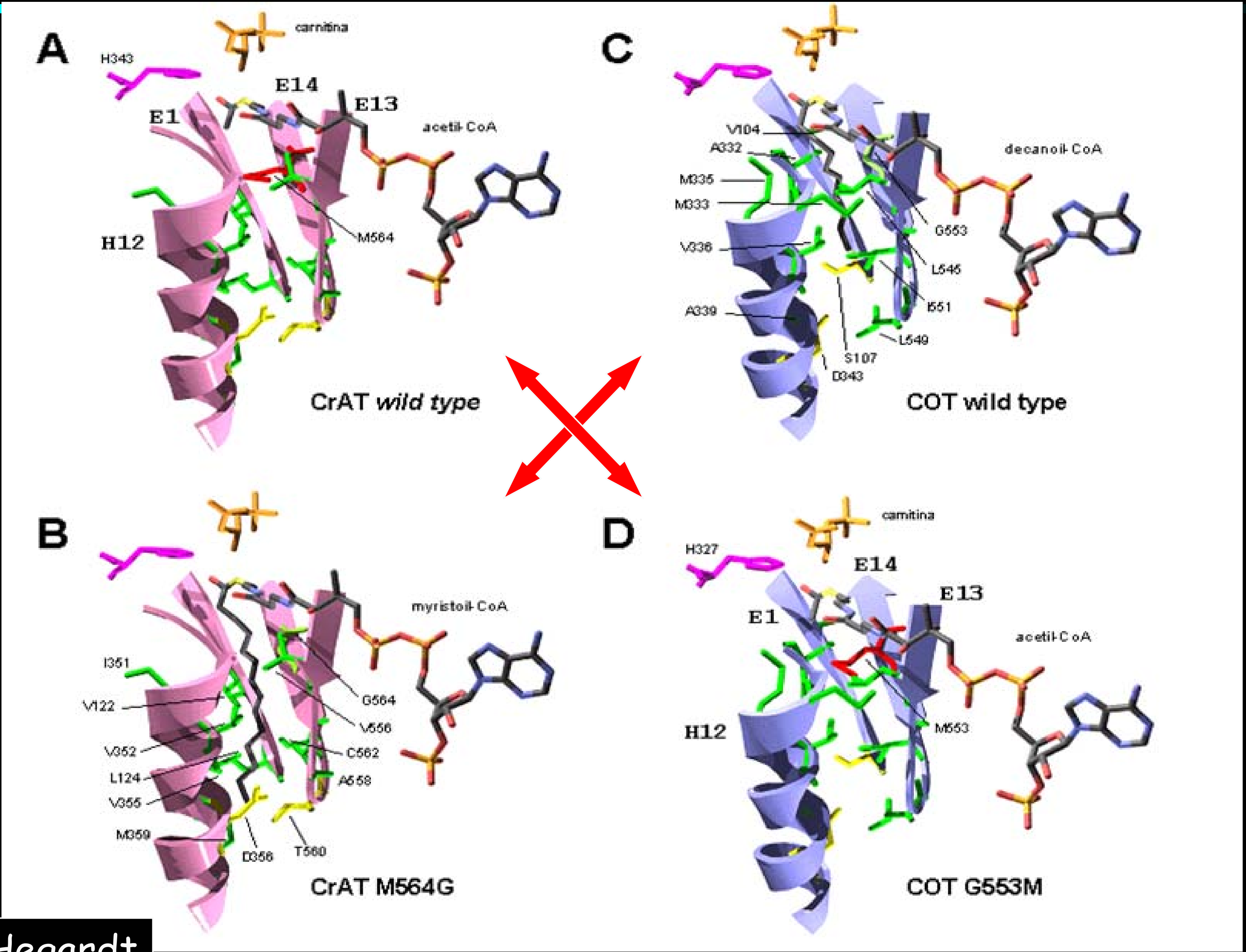
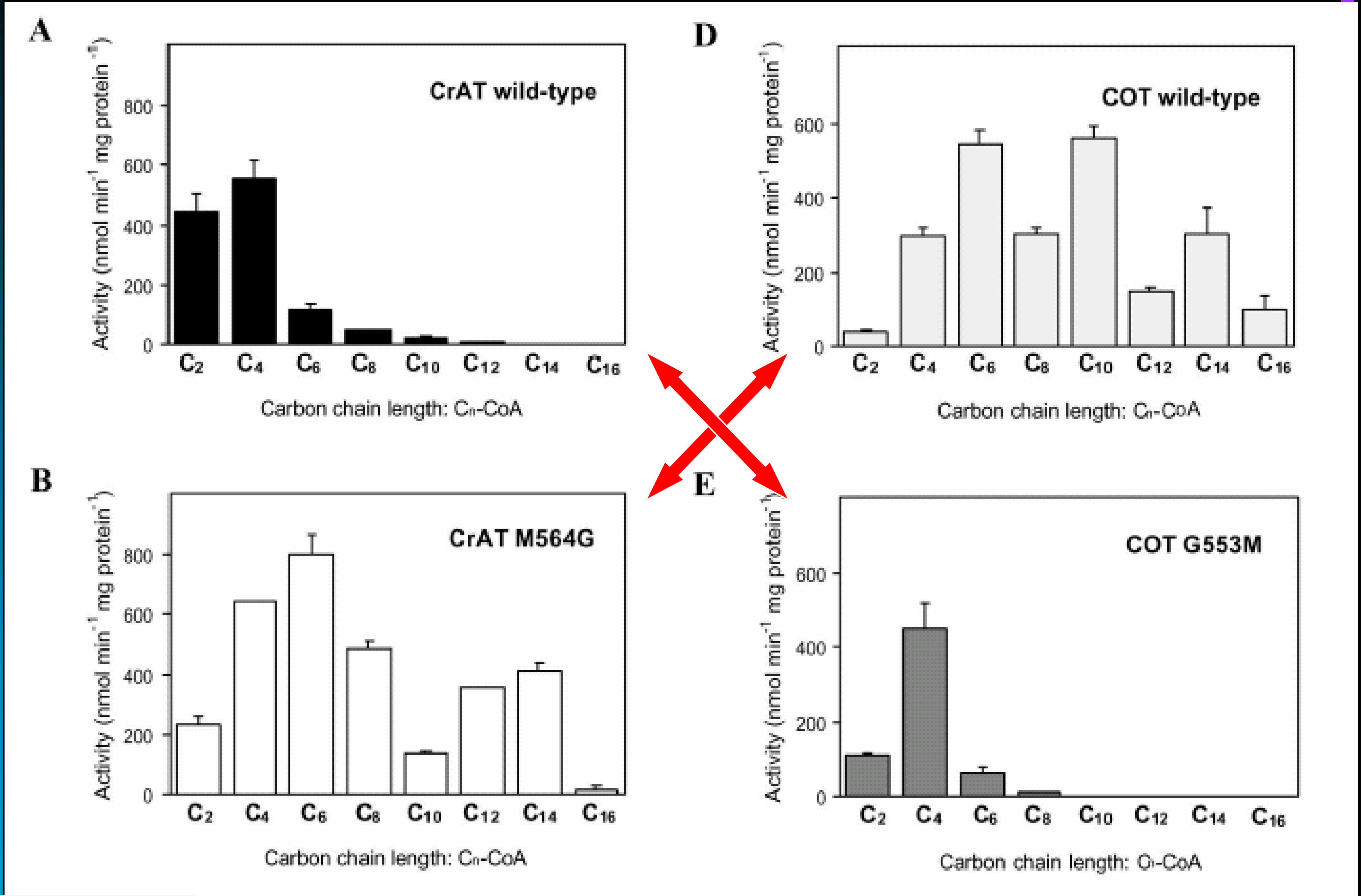


FIG. 4. **Effect of malonyl-CoA on the activity of yeast overexpressed L-CPT I (wild type) and several mutants.** *A*, L-CPT I wil

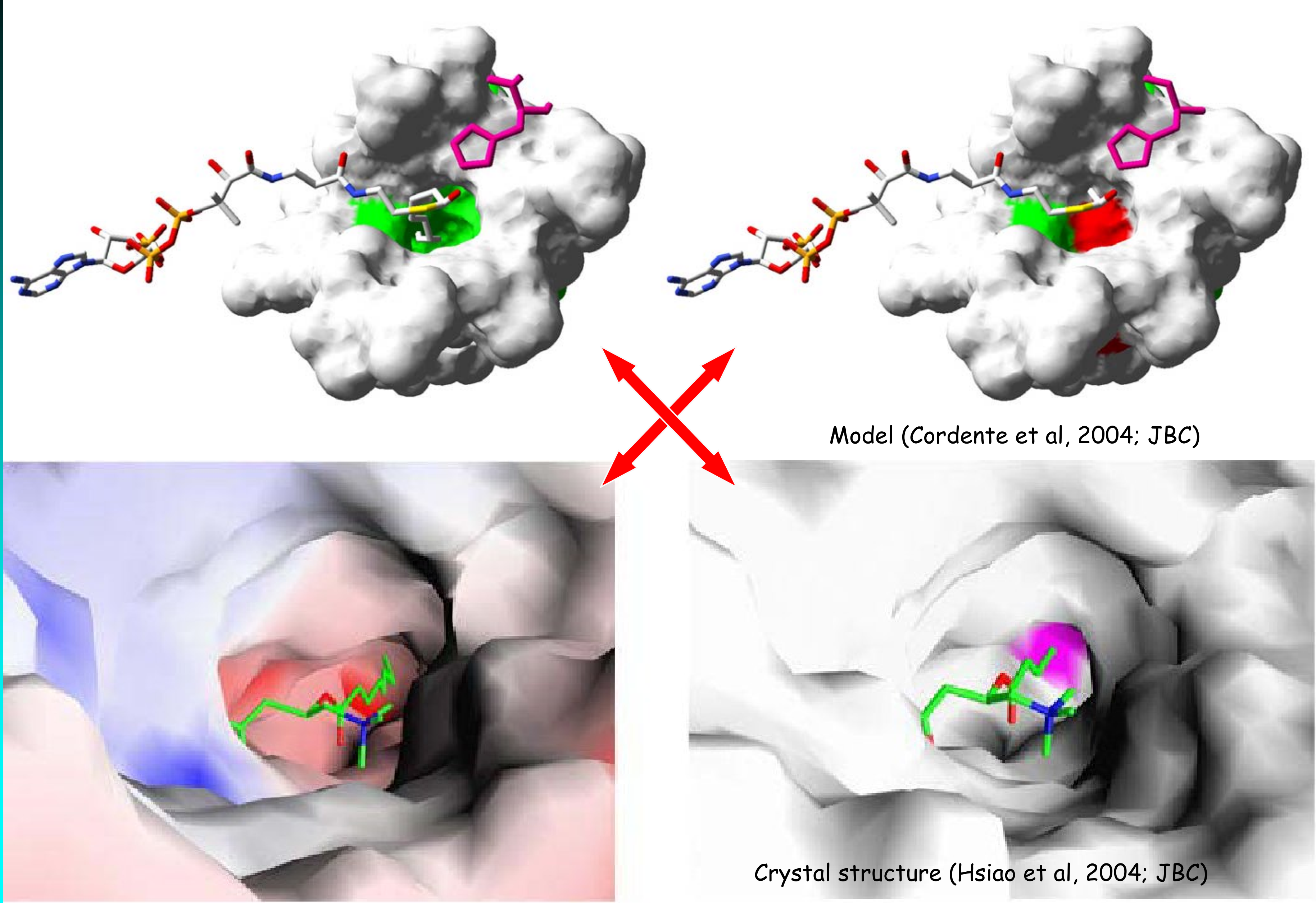
FG Hegardt

(Cordente, López-Viñas et al., JBC 2004)





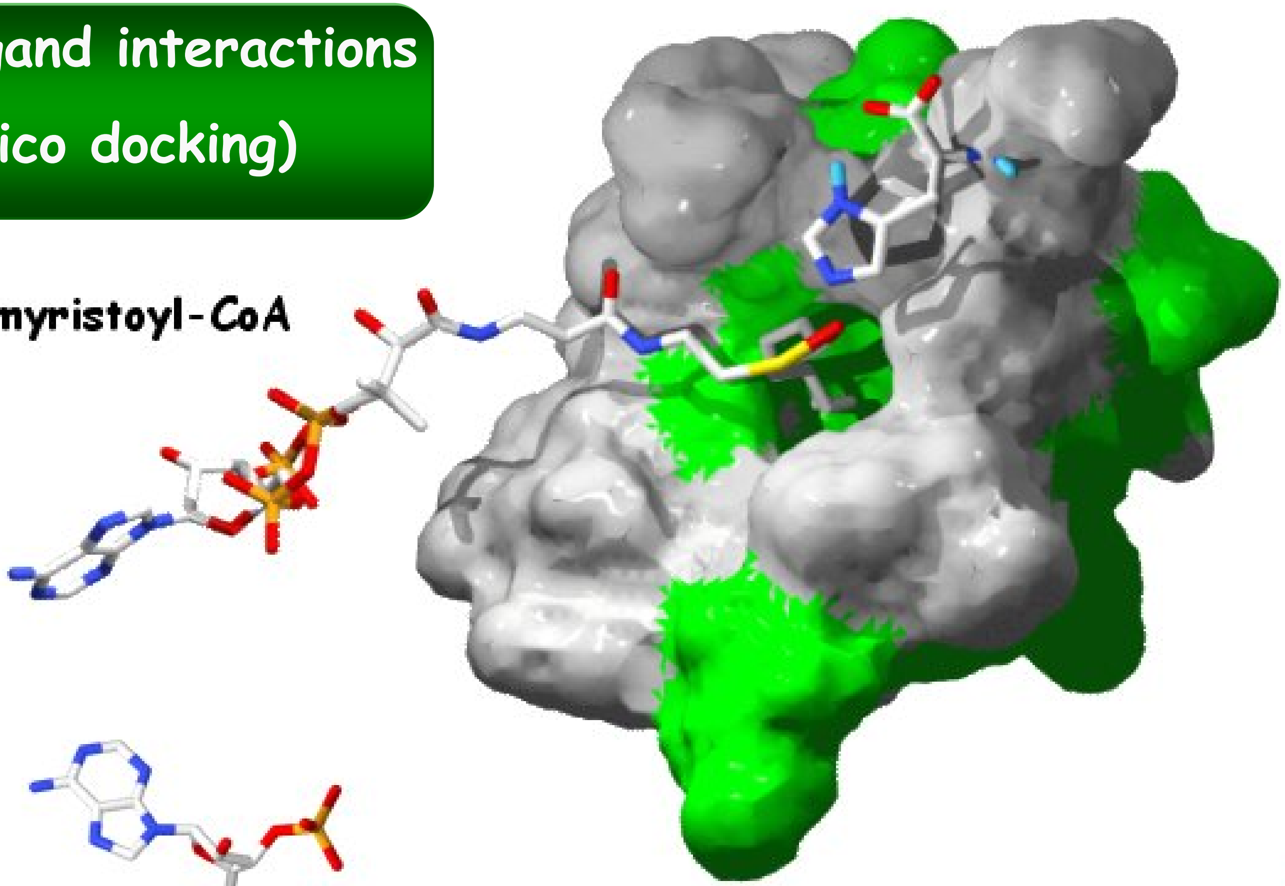
(Cordente, López-Viñas et al., JBC 2004)



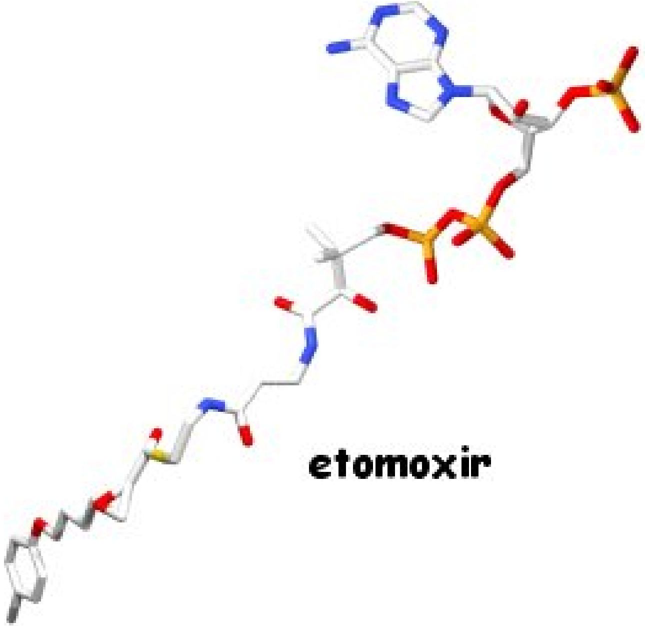
Model (Cordente et al, 2004; JBC)

Crystal structure (Hsiao et al, 2004; JBC)

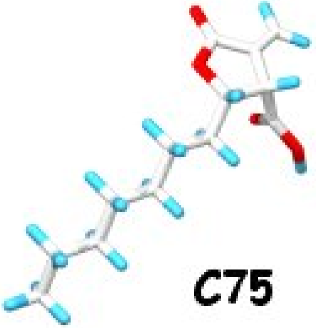
protein-ligand interactions
(in silico docking)



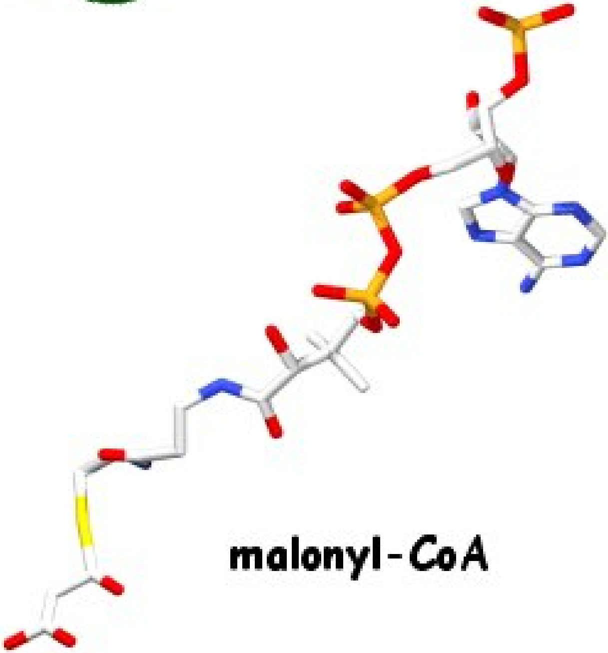
myristoyl-CoA



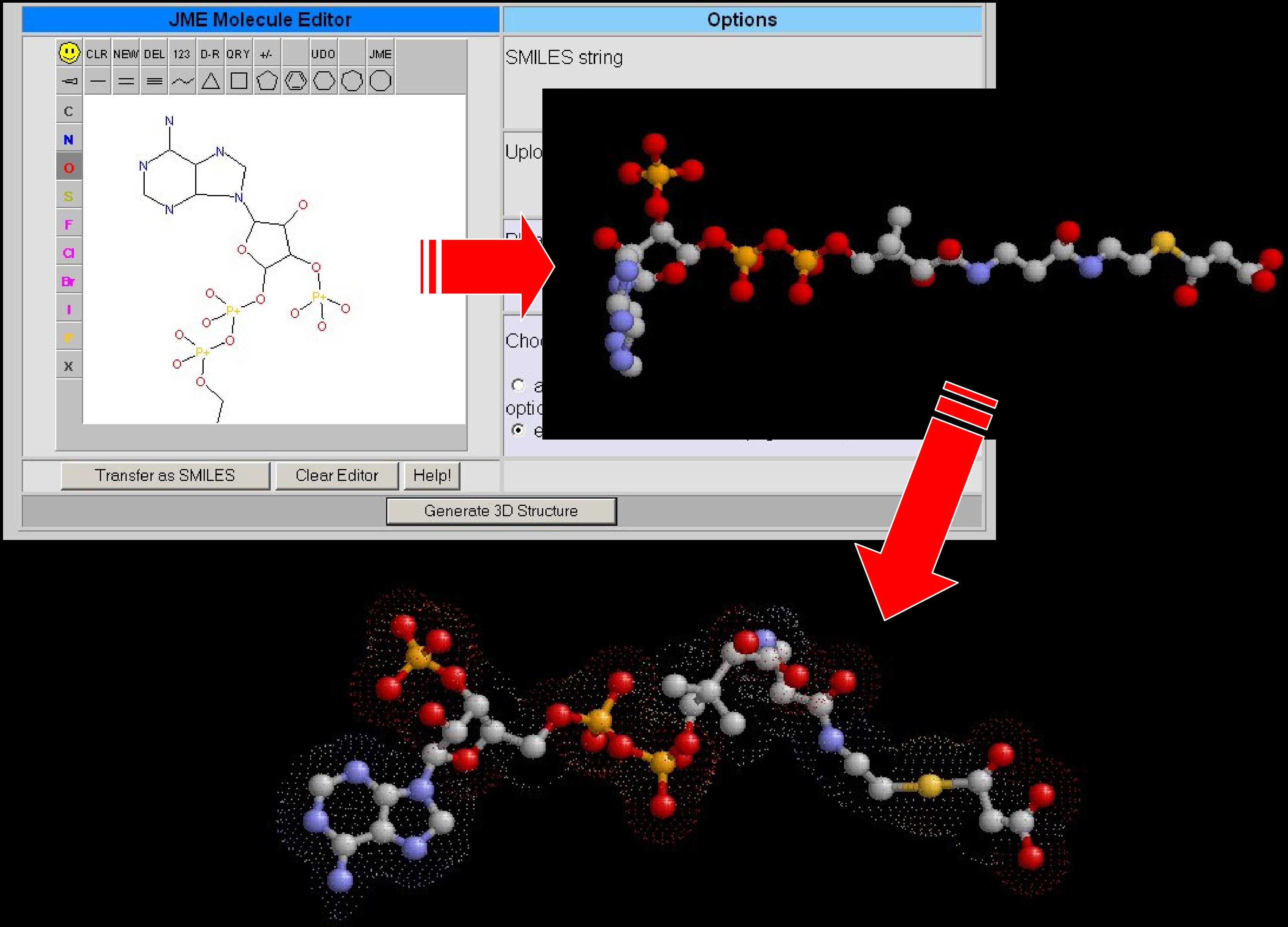
etomoxir



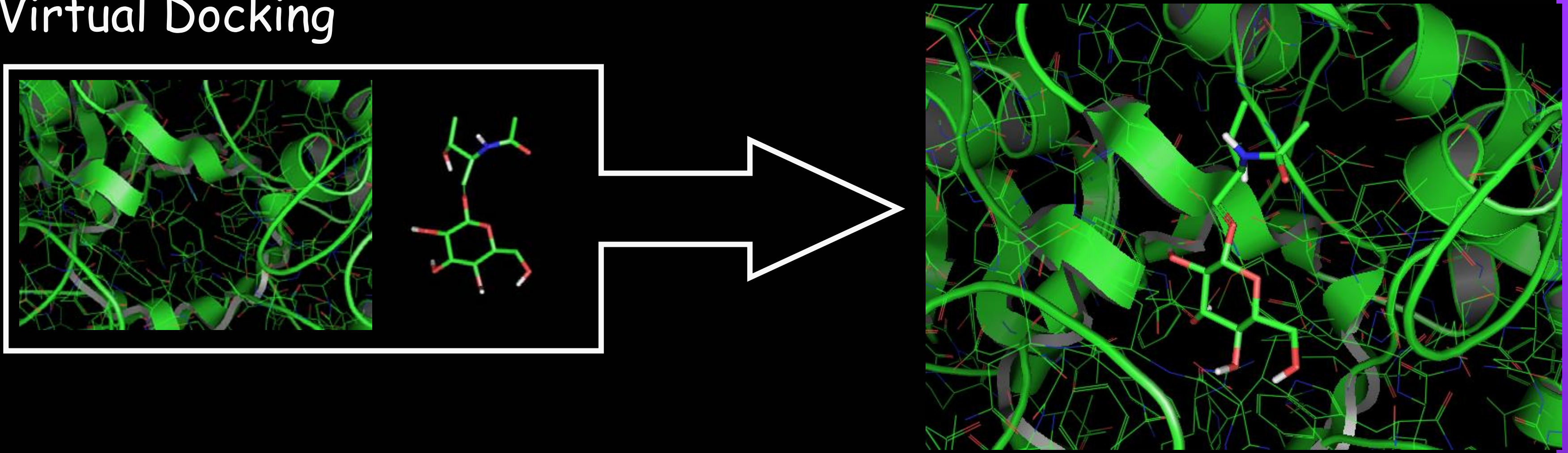
C75

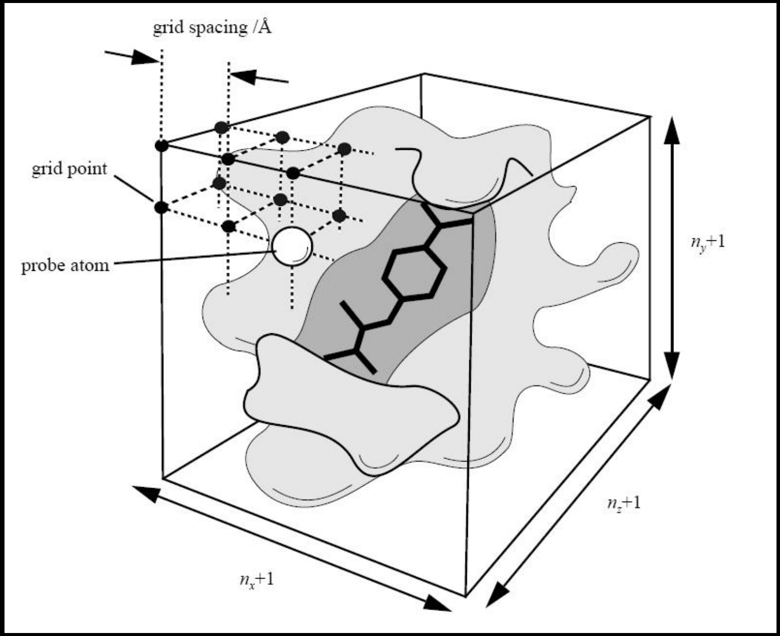


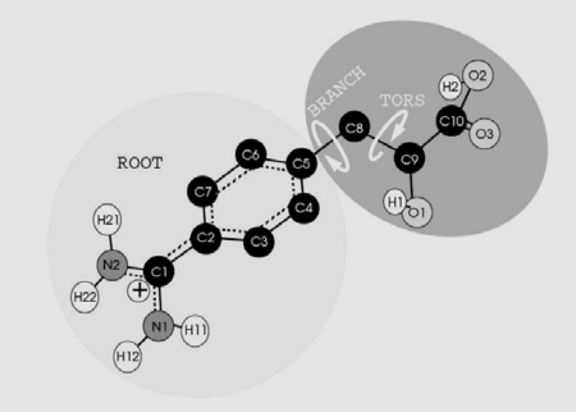
malonyl-CoA

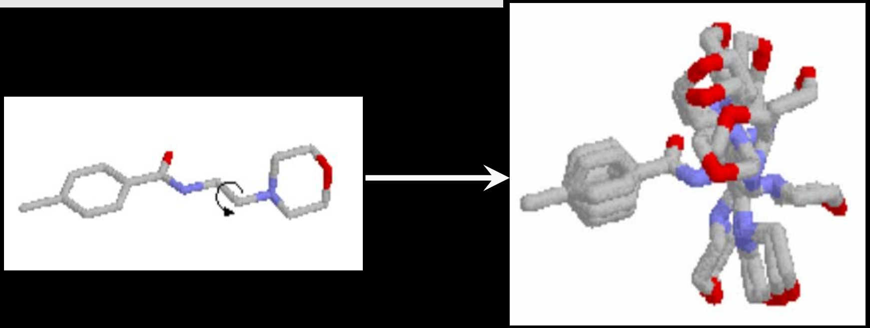

JME Molecule Editor
CLR
NEW
DEL
123
D-R
QRY
+/-
UDO
JME
C
N
O
S
F
Cl
Br
I
P
X
Options
SMILES string
Uplo
Choo
a
optic
e
Transfer as SMILES
Clear Editor
Help!
Generate 3D Structure

Virtual Docking

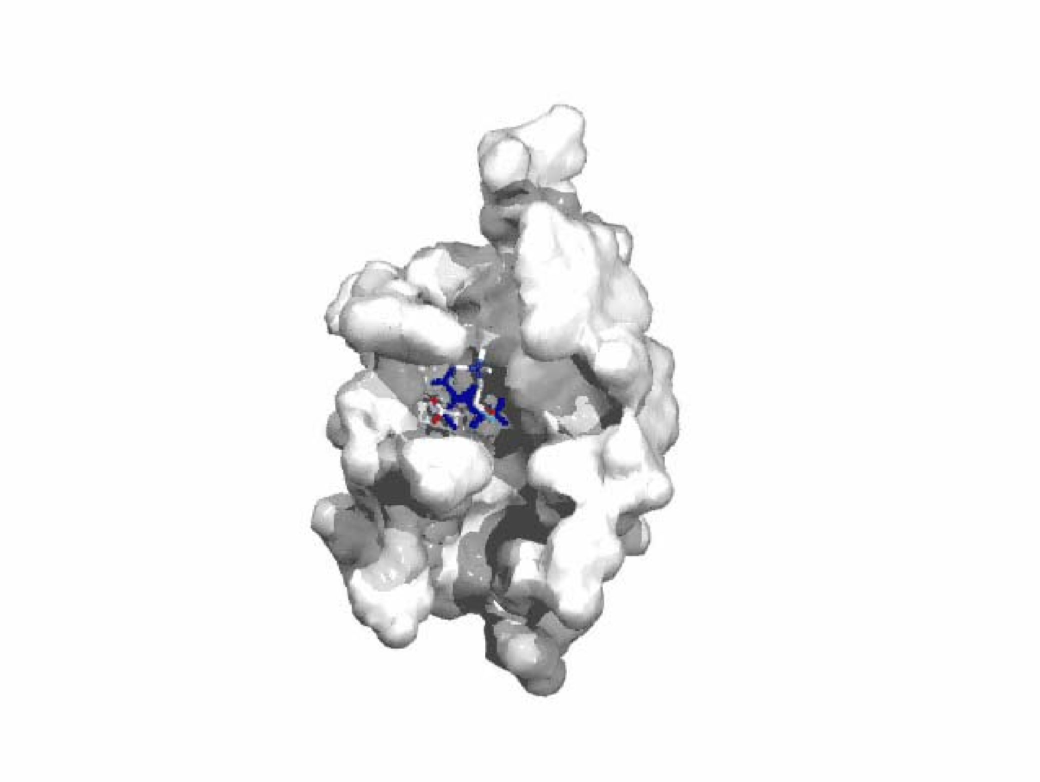


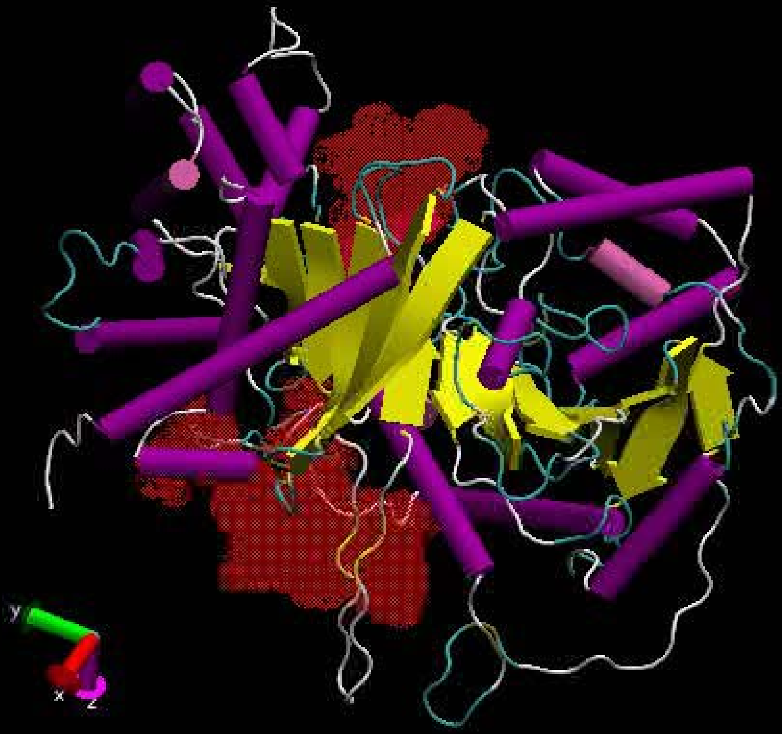




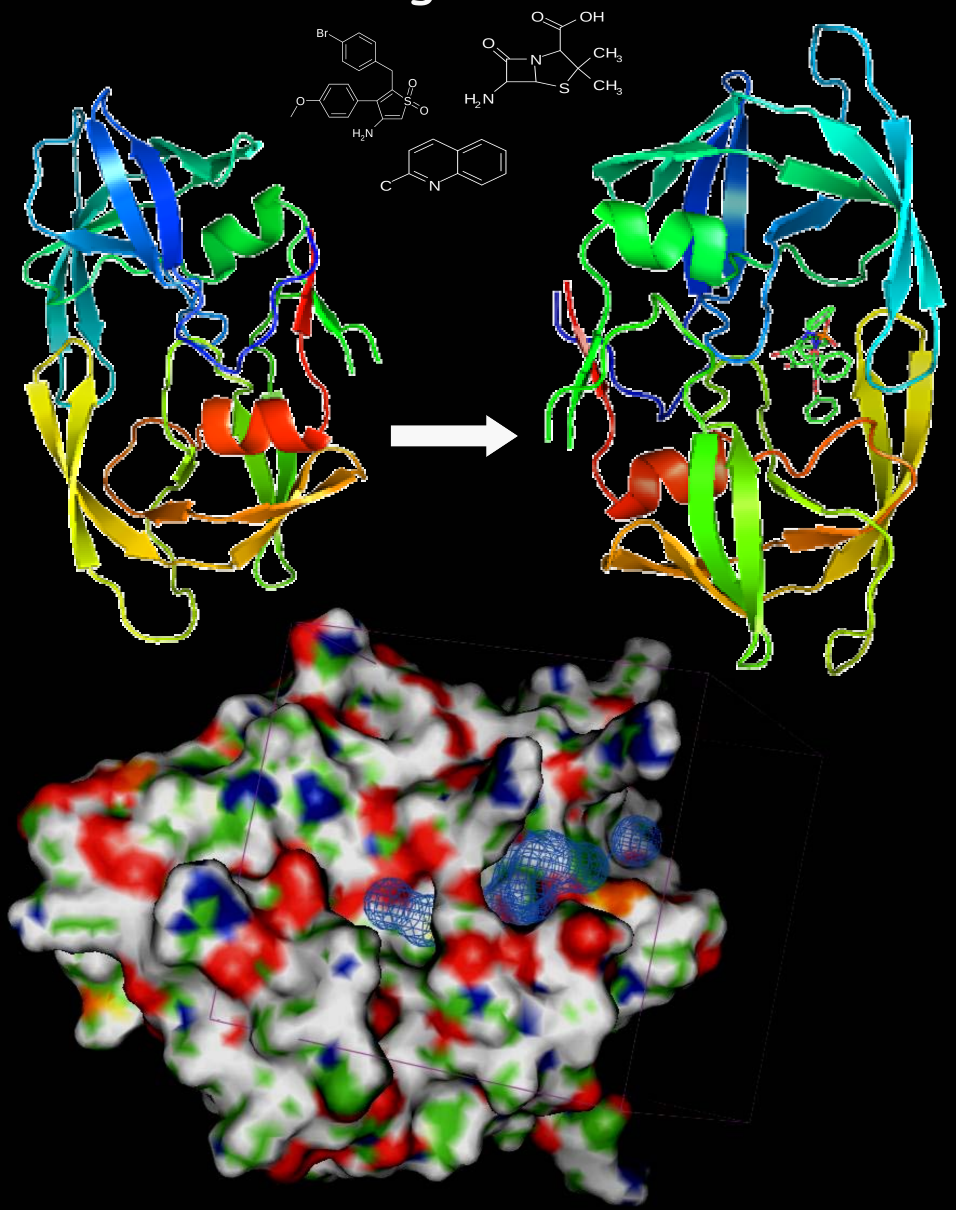


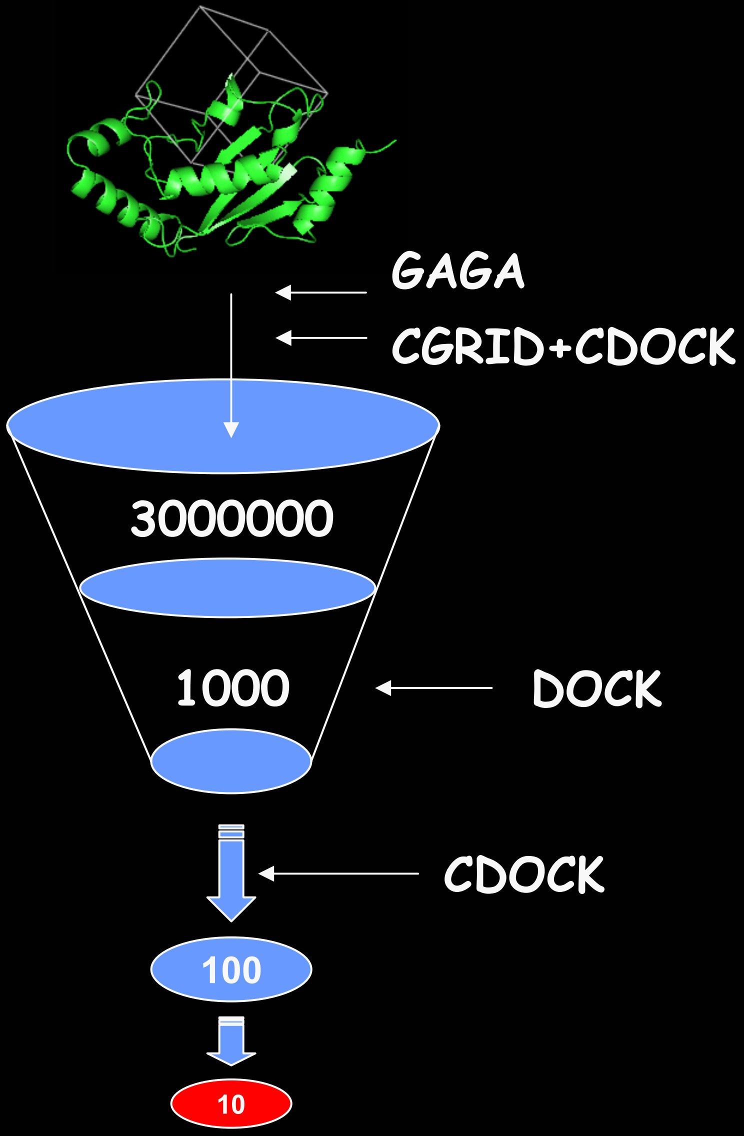
Virtual Docking





Virtual Screening




GAGA

CGRID+CDOCK

3000000

1000

DOCK

CDOCK

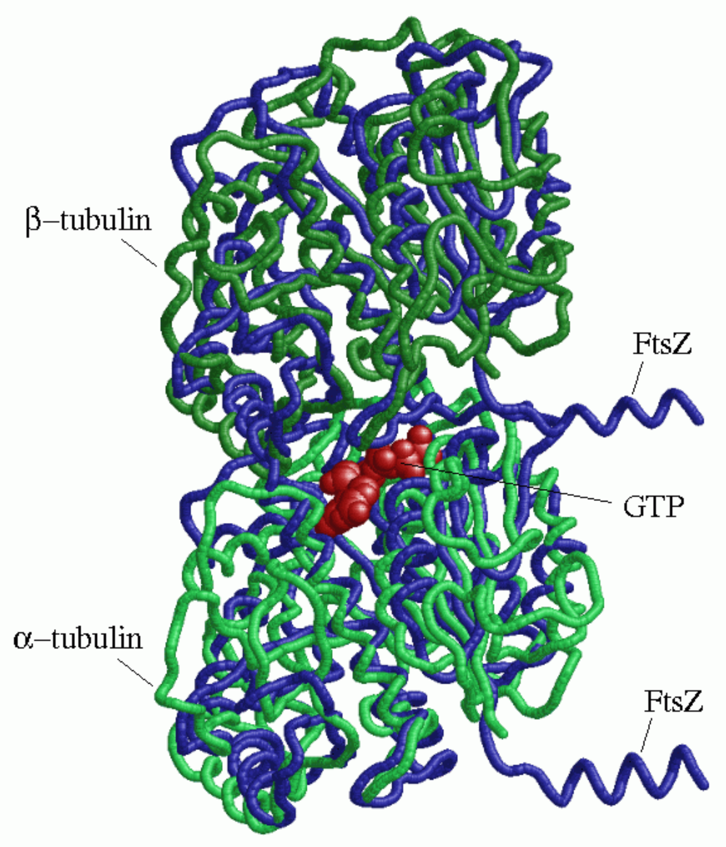
100

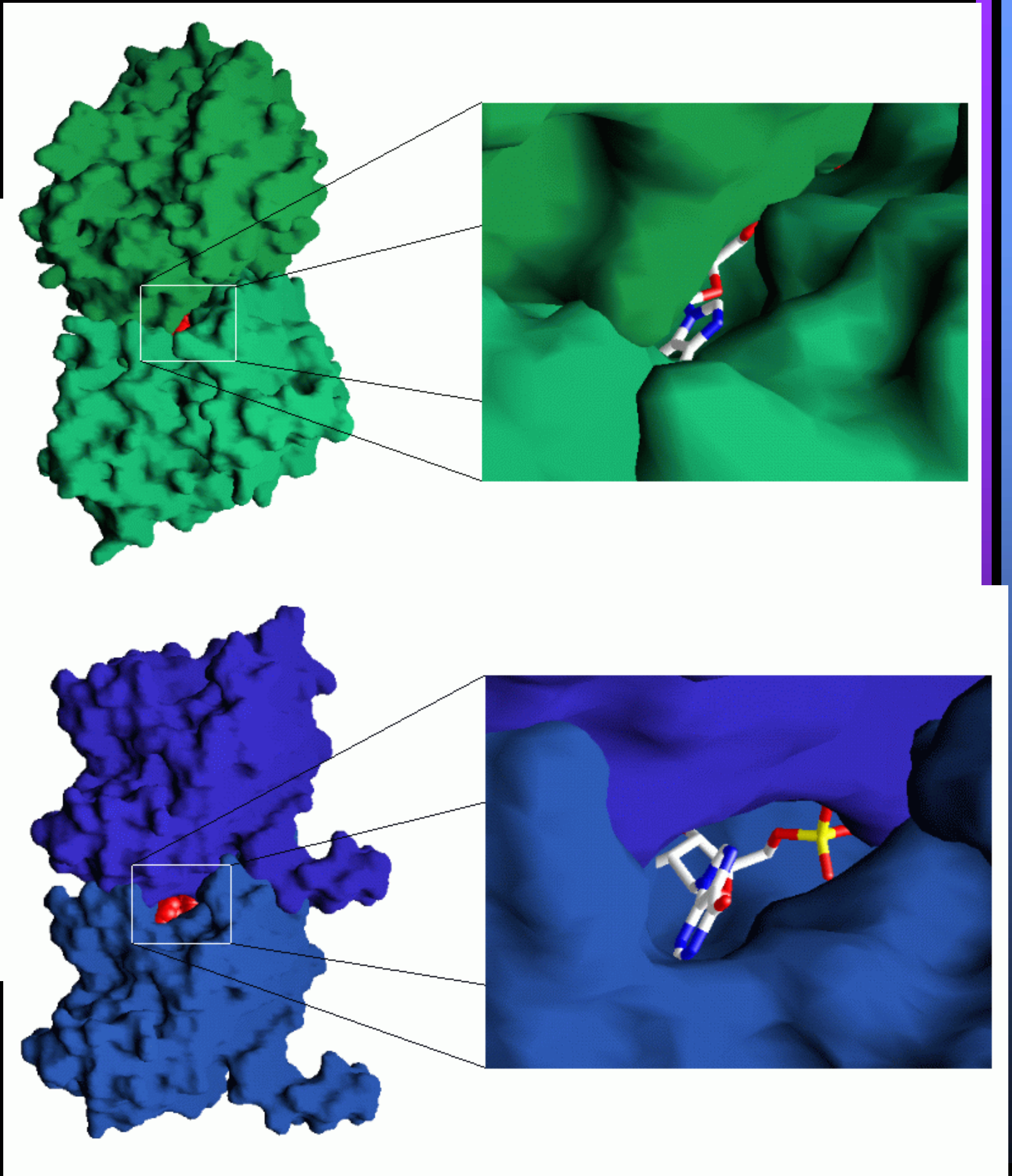
10

protein-protein interactions

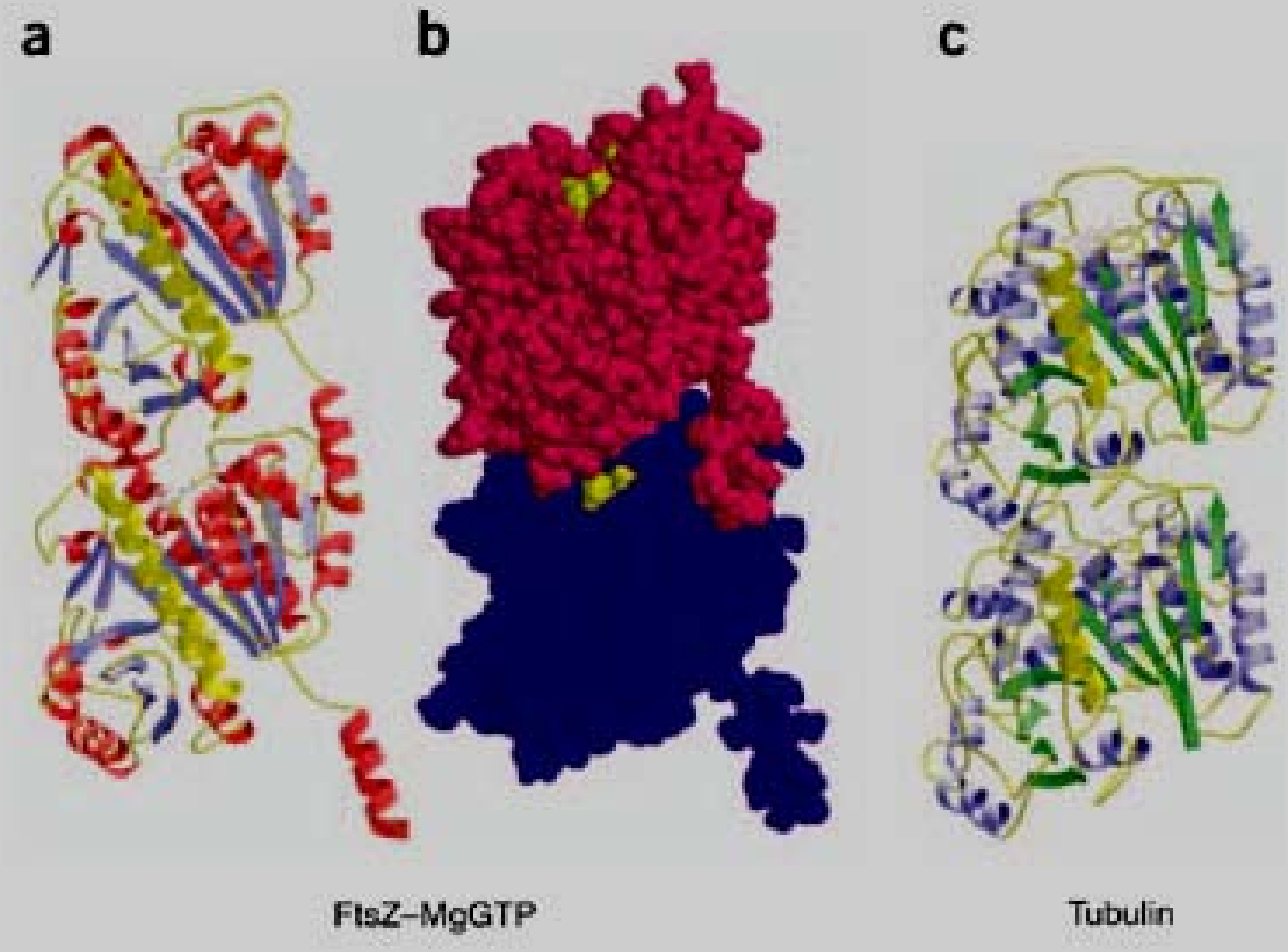
(in silico docking)

homology-based docking

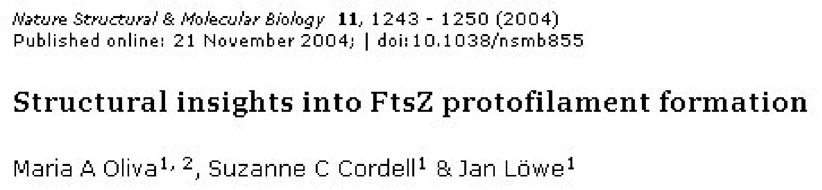




M. Vicente



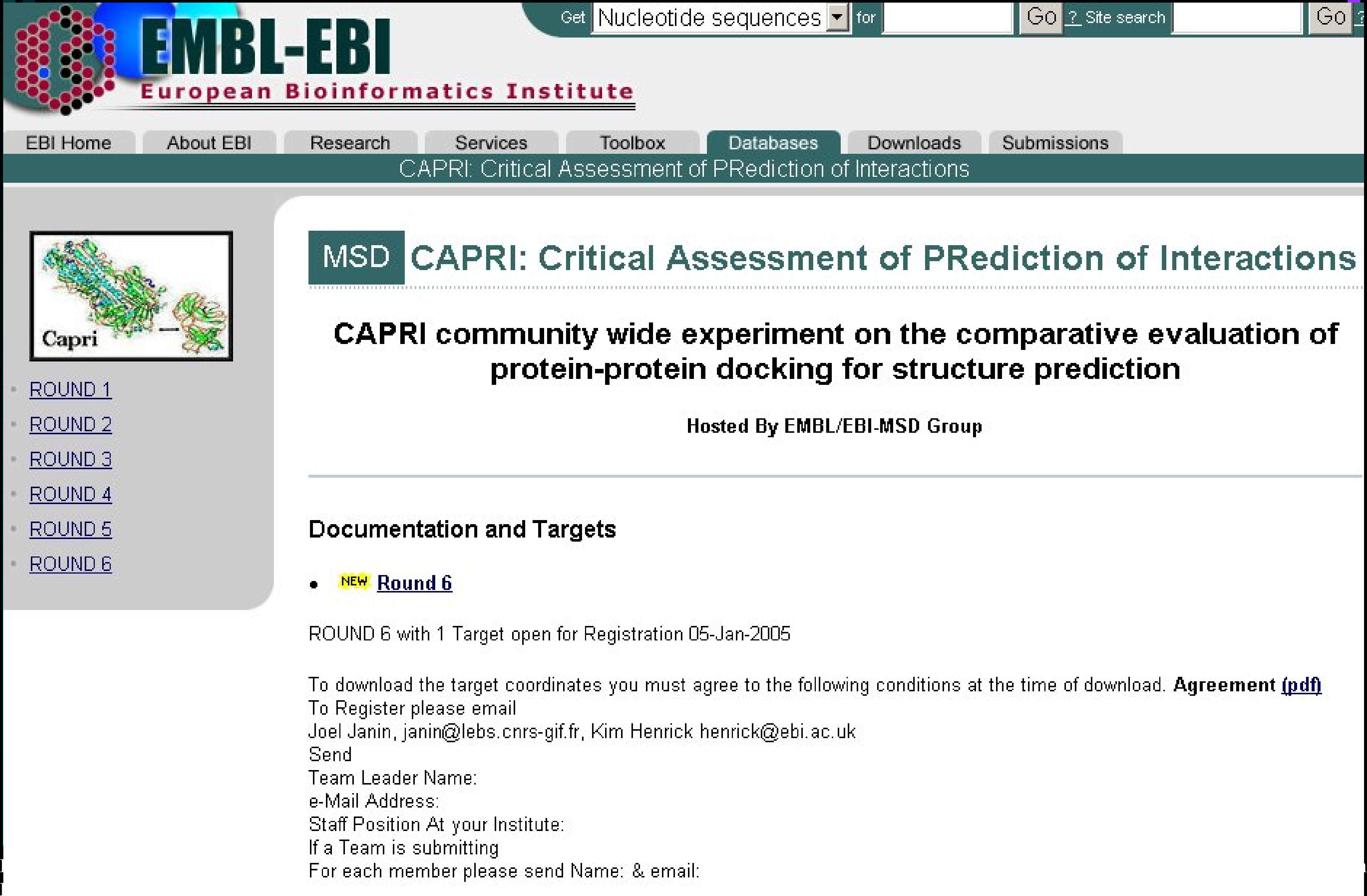
M. Vicente


Nature Structural & Molecular Biology **11**, 1243 - 1250 (2004)
Published online: 21 November 2004; | doi:10.1038/nsmb855

Structural insights into FtsZ protofilament formation

Maria A Oliva[1, 2], Suzanne C Cordell[1] & Jan Löwe[1]

http://capri.ebi.ac.uk



CAPRI: Critical Assessment of PRediction of Interactions

CAPRI community wide experiment on the comparative evaluation of protein-protein docking for structure prediction

Hosted By EMBL/EBI-MSD Group

Documentation and Targets

- NEW **Round 6**

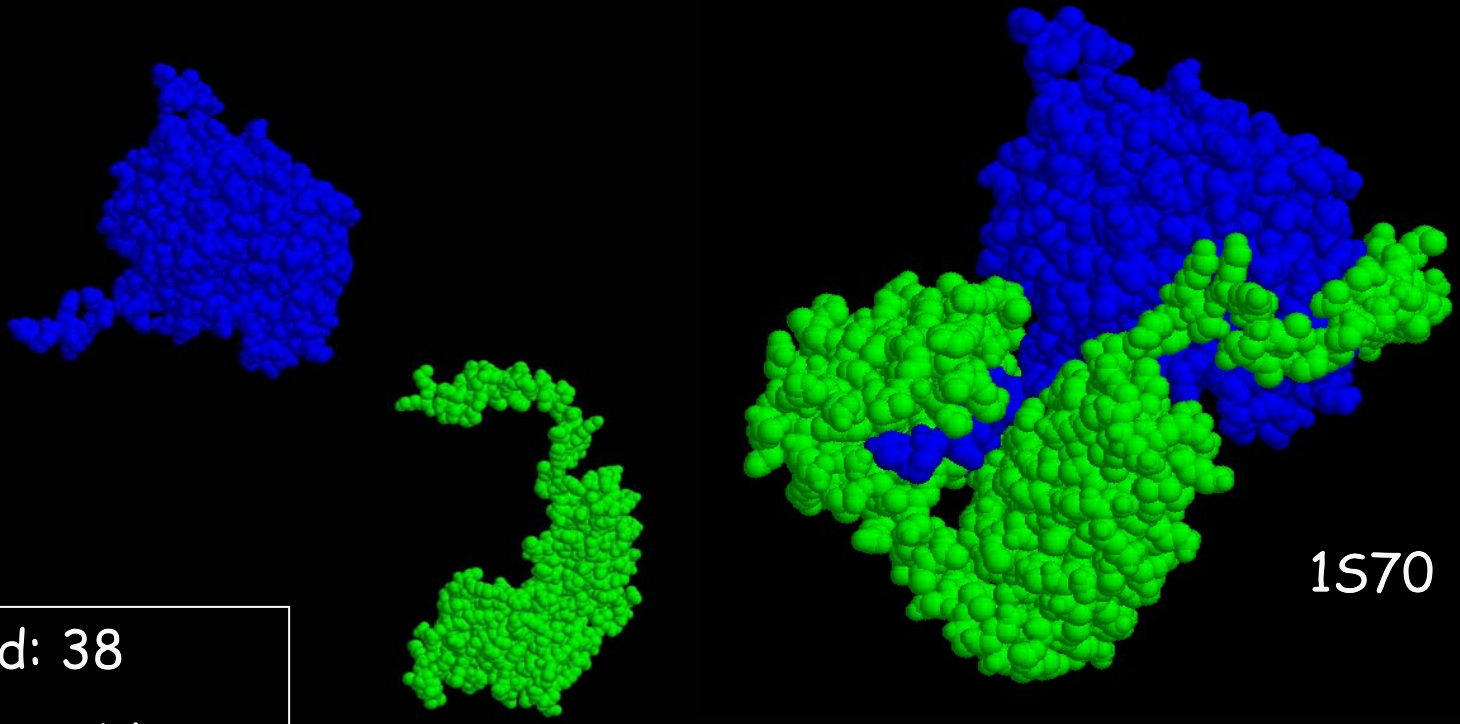
ROUND 6 with 1 Target open for Registration 05-Jan-2005

To download the target coordinates you must agree to the following conditions at the time of download. **Agreement (pdf)**
To Register please email
Joel Janin, janin@lebs.cnrs-gif.fr, Kim Henrick henrick@ebi.ac.uk
Send
Team Leader Name:
e-Mail Address:
Staff Position At your Institute:
If a Team is submitting
For each member please send Name: & email:

CAPRI ROUND 5

Target 14: PROTEIN SERINE/THREONINE PHOSPHATASE-1 complex with MYOSIN PHOSPHATASE TARGETING SUBUNIT 1

From: Roberto Dominguez (Boston Biomedical Research Institute)

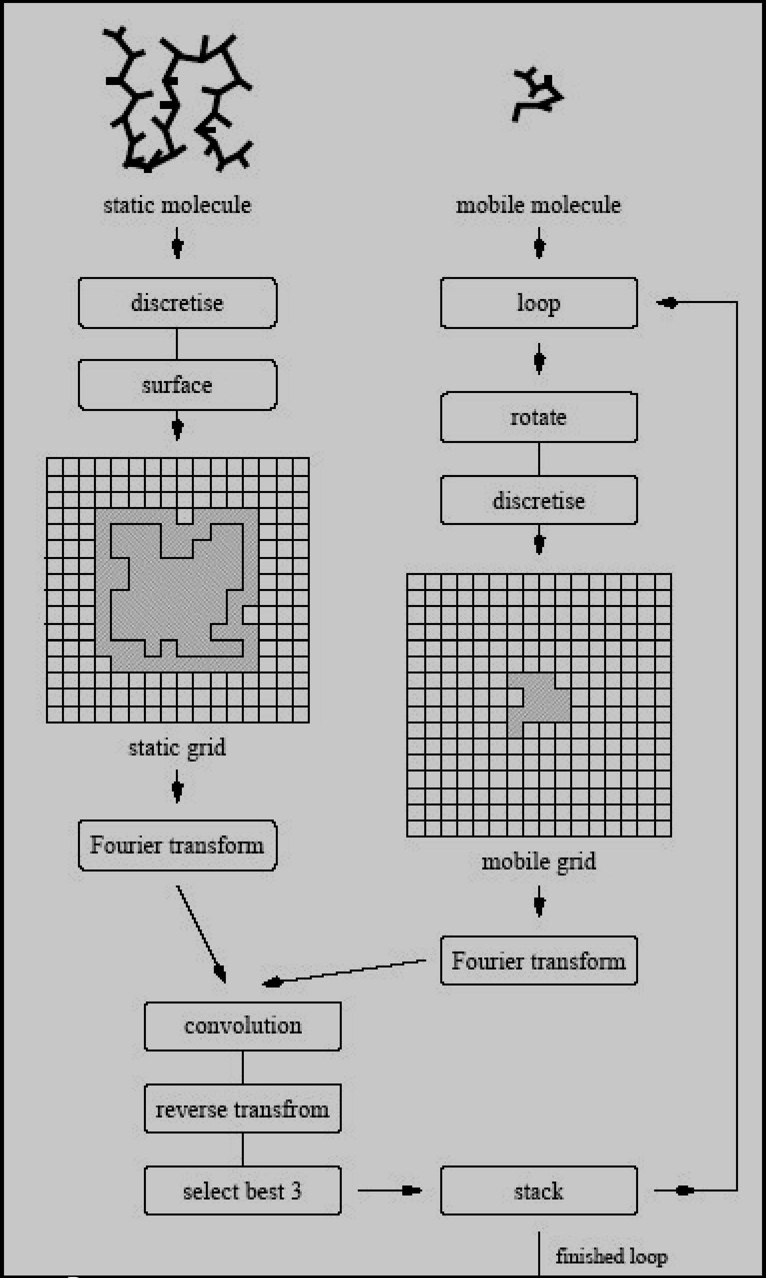


Good: 38

acceptable: 32

scoring:

- RMSD between correct and predicted positions of ligand.
- Correct location of contact surfaces both in receptor and ligand.
- Correct prediction of residue contacts.



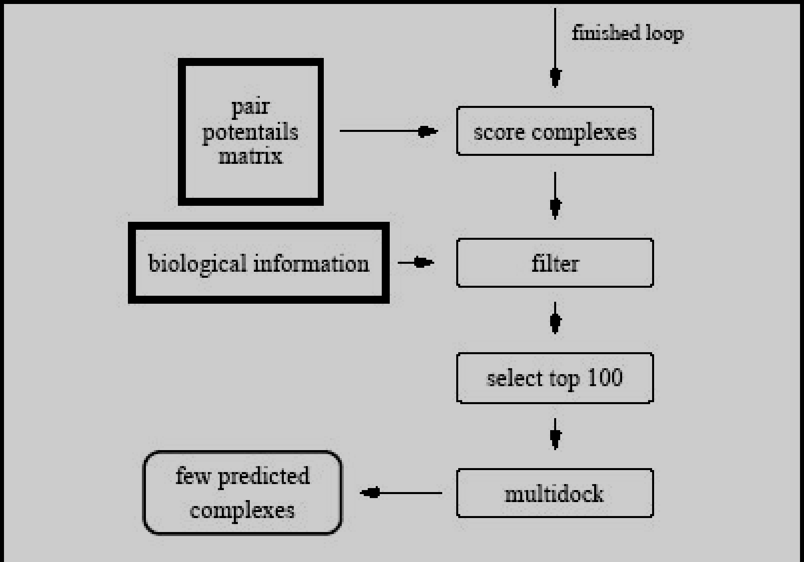


Figure 1: Flow diagram of overall docking method

3D-Dock. M Sternberg

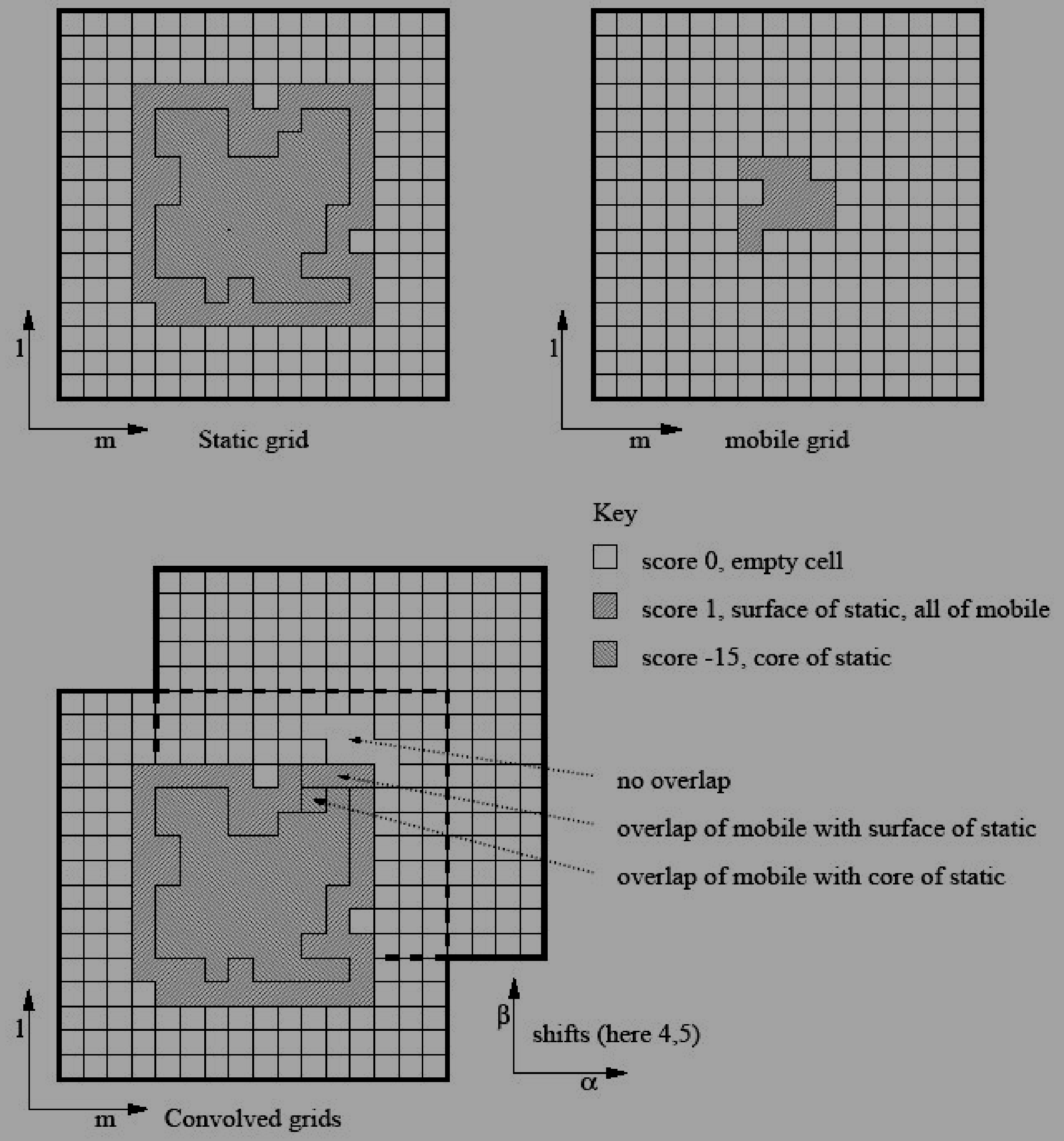
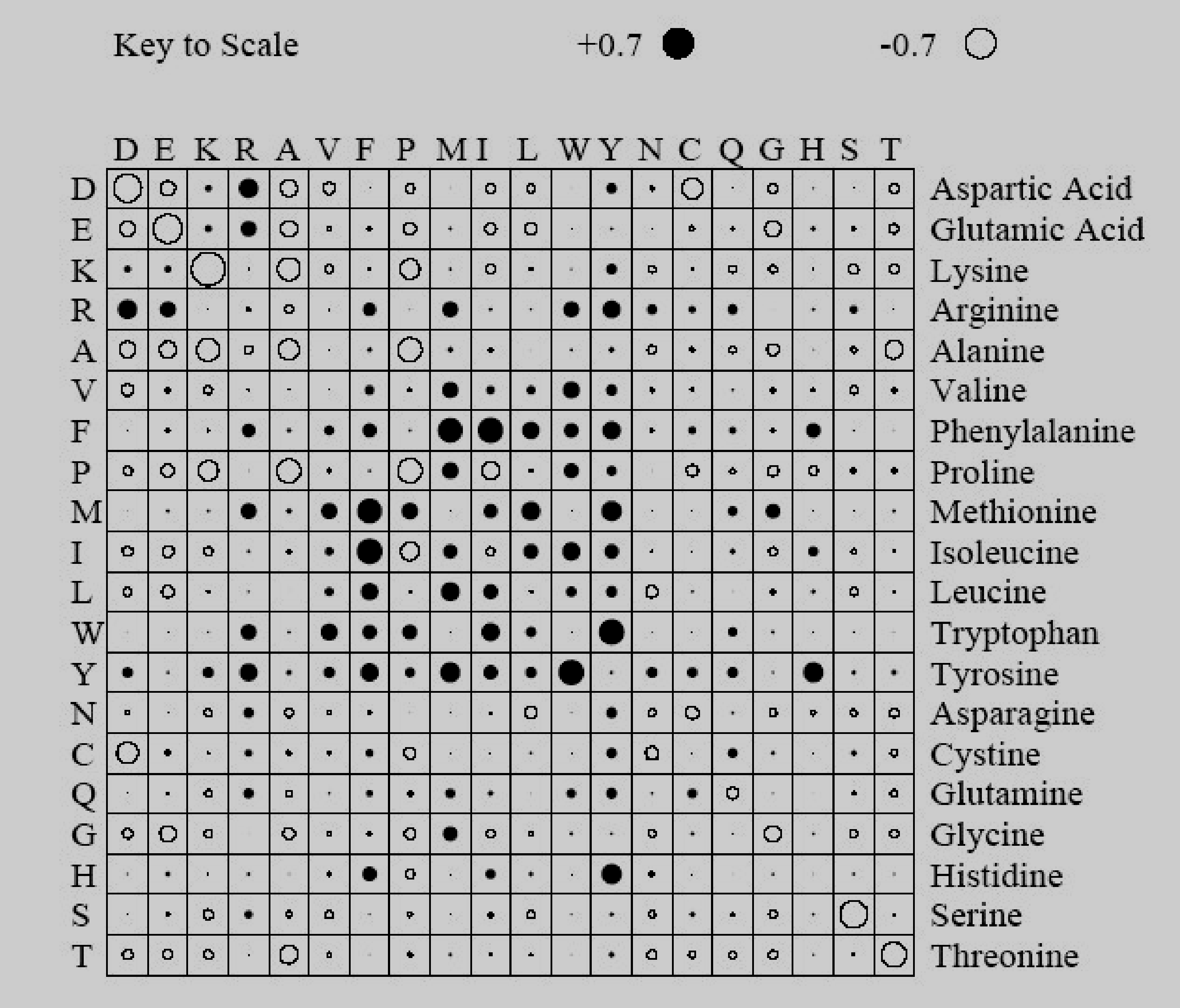
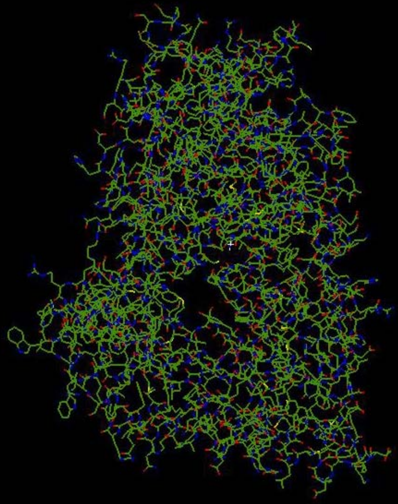
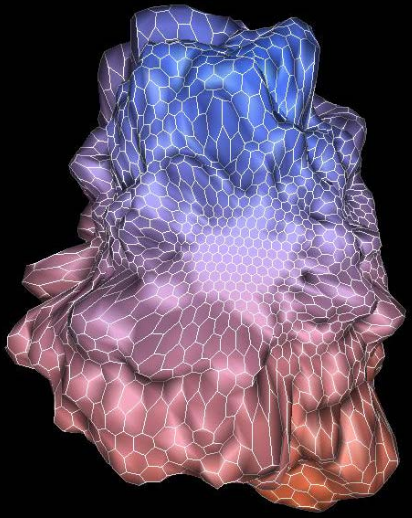
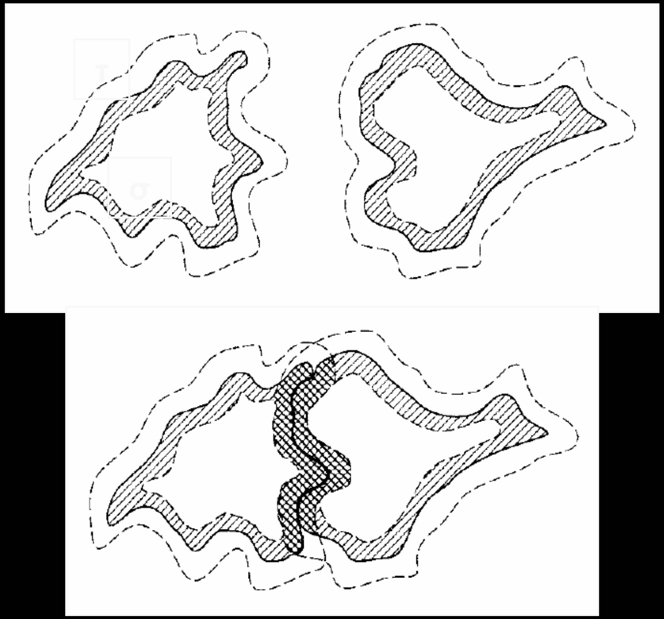


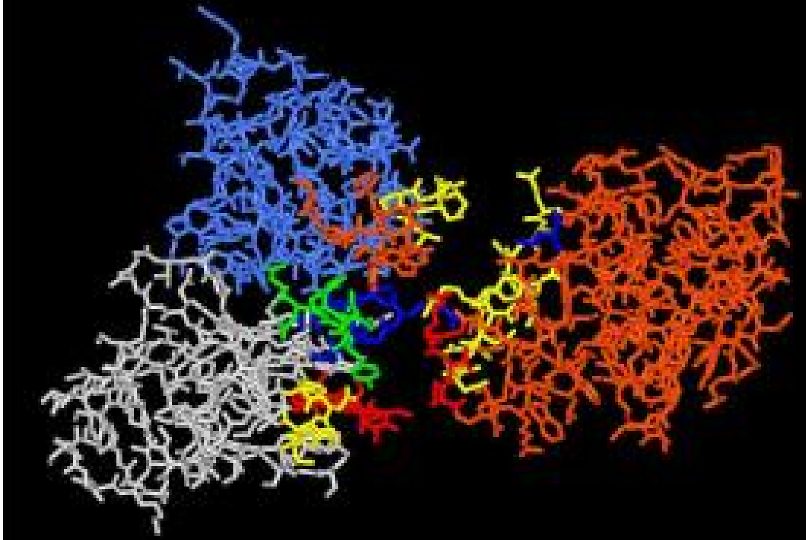
Figure 2: Grid discretisation of molecules and calculation of surface complementarity

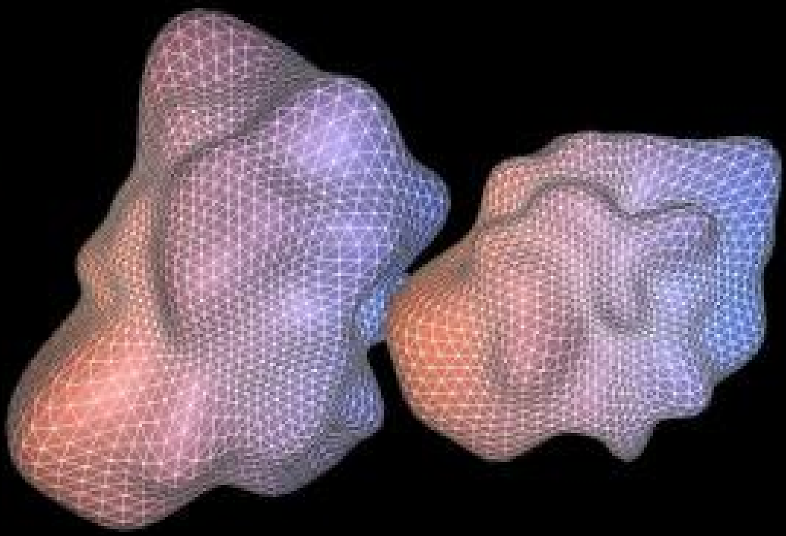

Key to Scale
+0.7 ●
-0.7 ○
D E K R A V F P M I L W Y N C Q G H S T
D Aspartic Acid
E Glutamic Acid
K Lysine
R Arginine
A Alanine
V Valine
F Phenylalanine
P Proline
M Methionine
I Isoleucine
L Leucine
W Tryptophan
Y Tyrosine
N Asparagine
C Cystine
Q Glutamine
G Glycine
H Histidine
S Serine
T Threonine



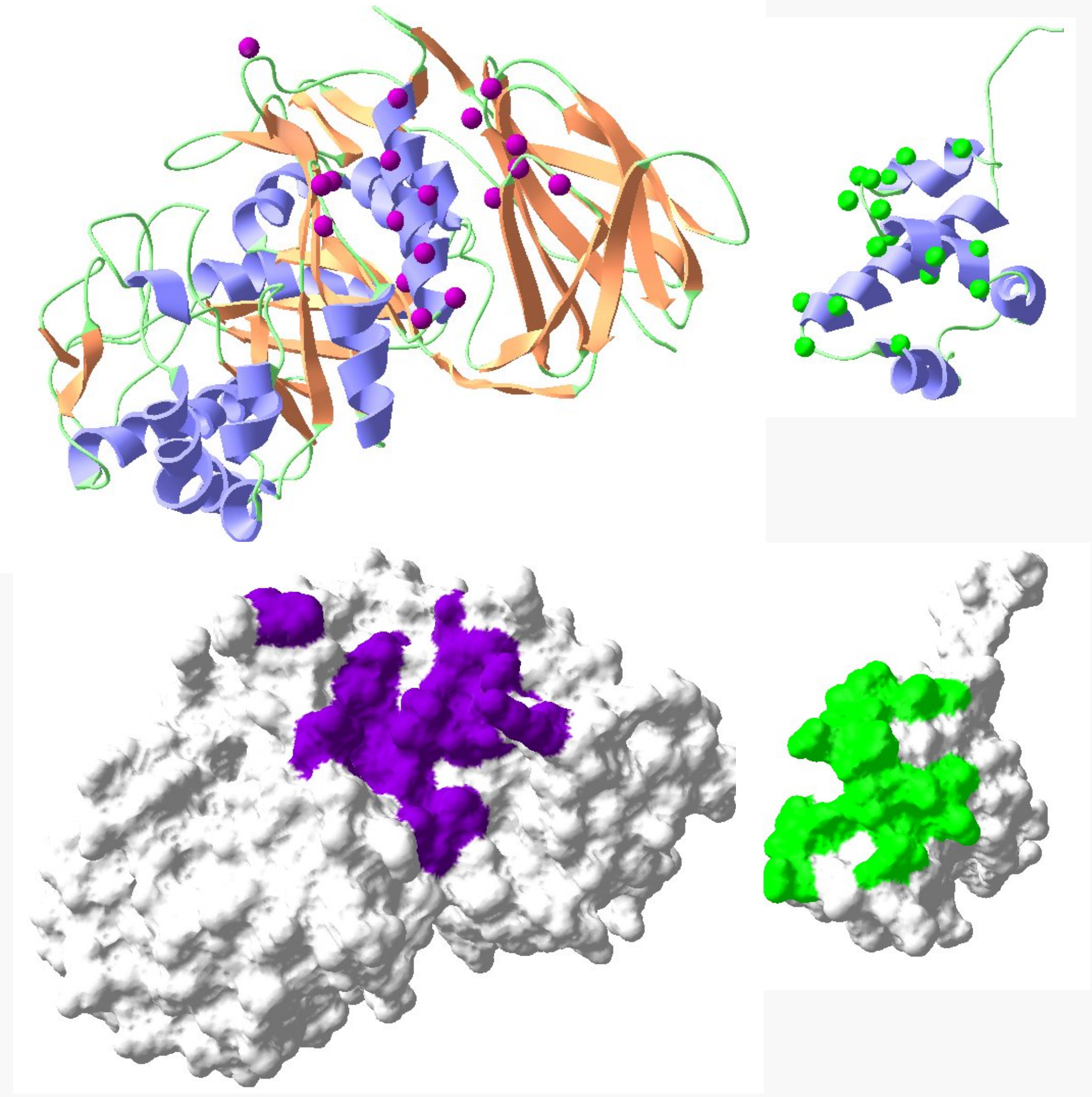


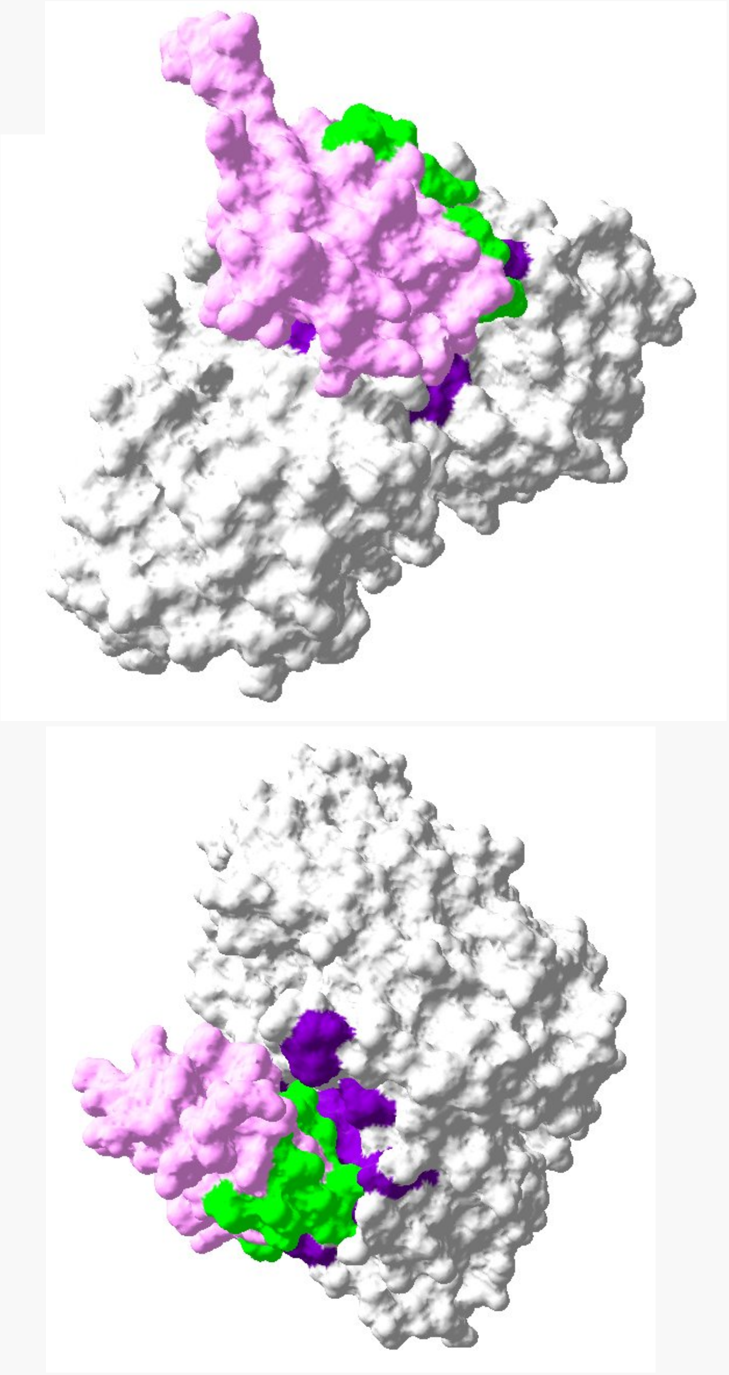






Integration of Evolutive Information: Correlated Mutations

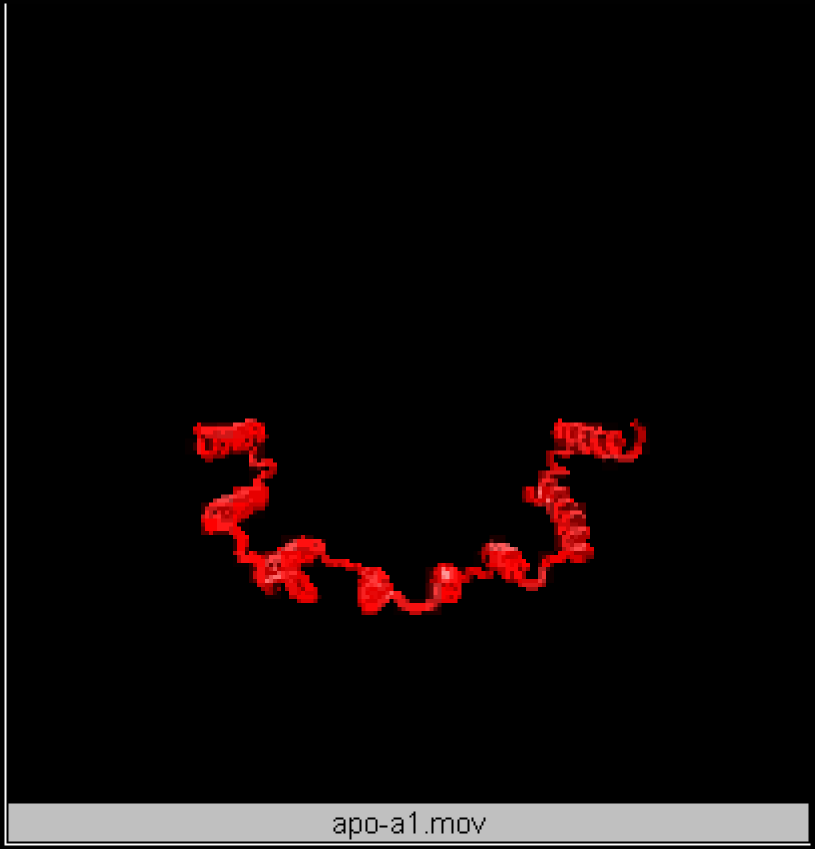




S. Atrian

structure-function

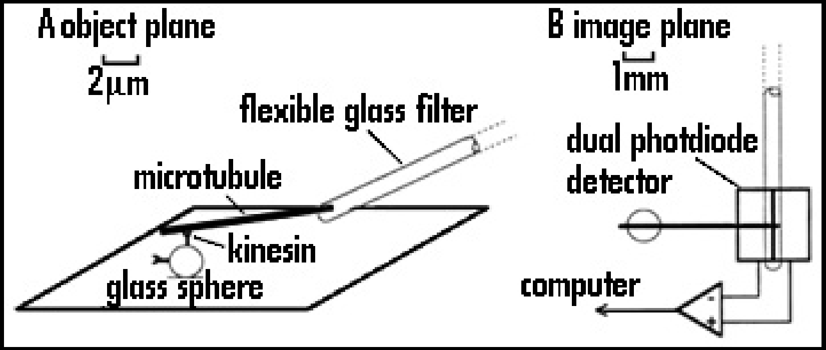
(a pair of examples)

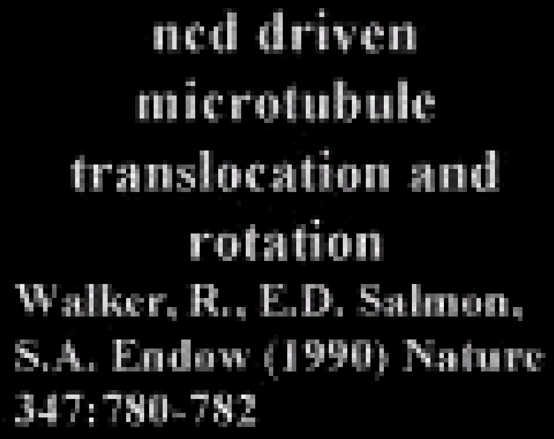


High Density Lipoprotein Particle. The model surrounded a circular disk of 160 lipids, forming a bilayer, with two apoA-I proteins and 6000 water molecules, altogether a 46,000 atom system.

Kinesin Motility

Kinesin is a force-generating enzyme, or motor protein, which converts the free energy of the gamma phosphate bond of ATP into mechanical work. This work is used to power the transport of intracellular organelles along microtubules.

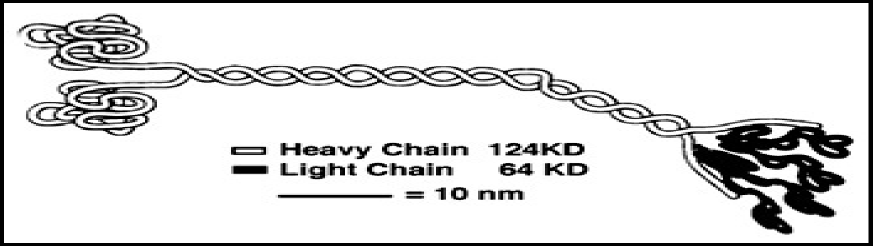



ncd driven
microtubule
translocation and
rotation
Walker, R., E.D. Salmon,
S.A. Endow (1990) Nature
347:780-782

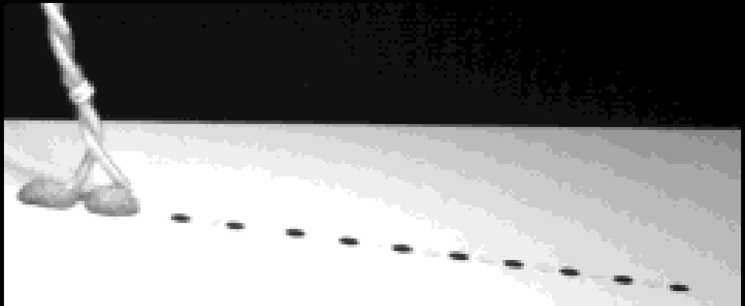
http://www.proweb.org/~kinesin/ Liz Greene, Steve Henikoff & Sharyn Endow

Kinesin Motility

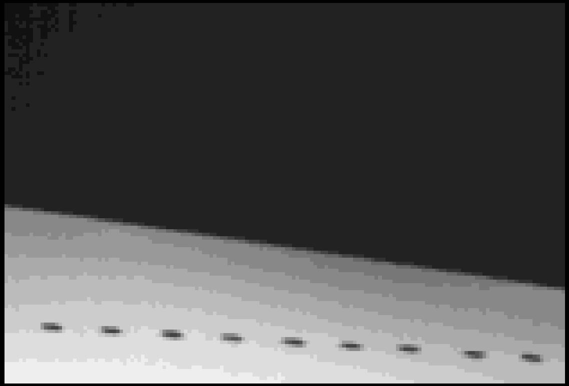
Kinesin itself is an alpha2-beta2 heterotetramer of two heavy and two light chains



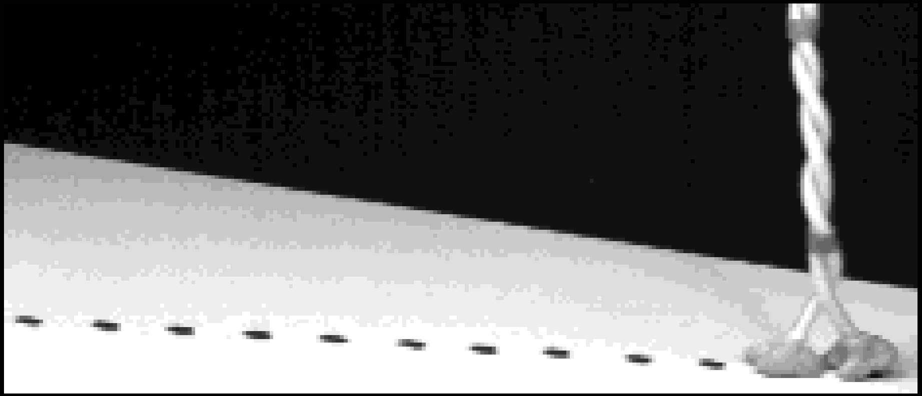
How does kinesin walk along a MT protofilament?



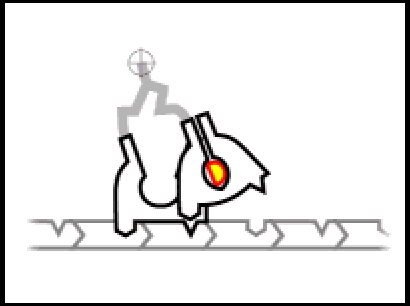
Floppy logic model



Twisting model

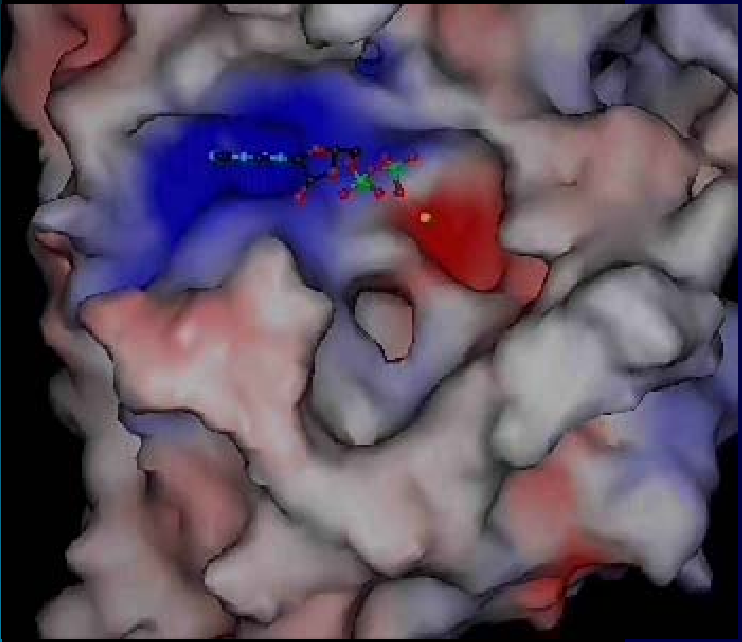


Alternate-sides model

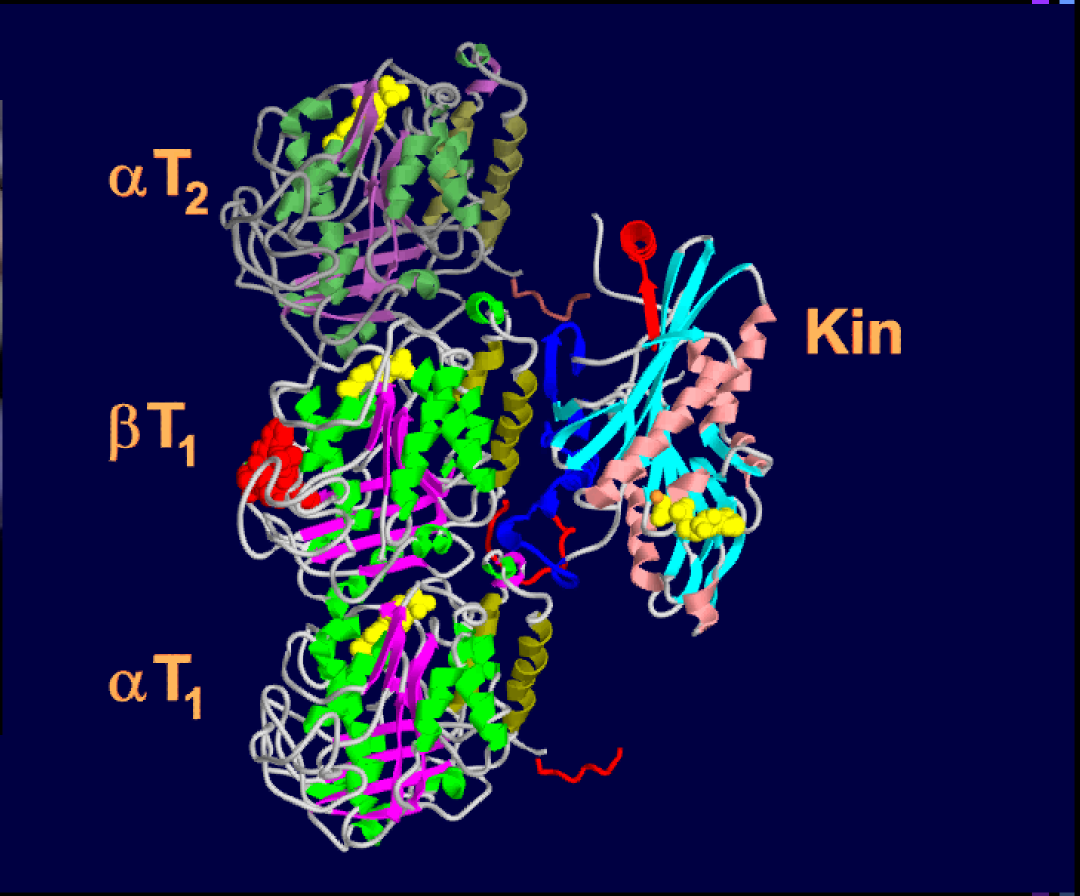


http://mc11.mcri.ac.uk/wrongtrousers.html Rob Cross & Minia Alonso

Kinesin Motility



Neurospora crassa kinesin



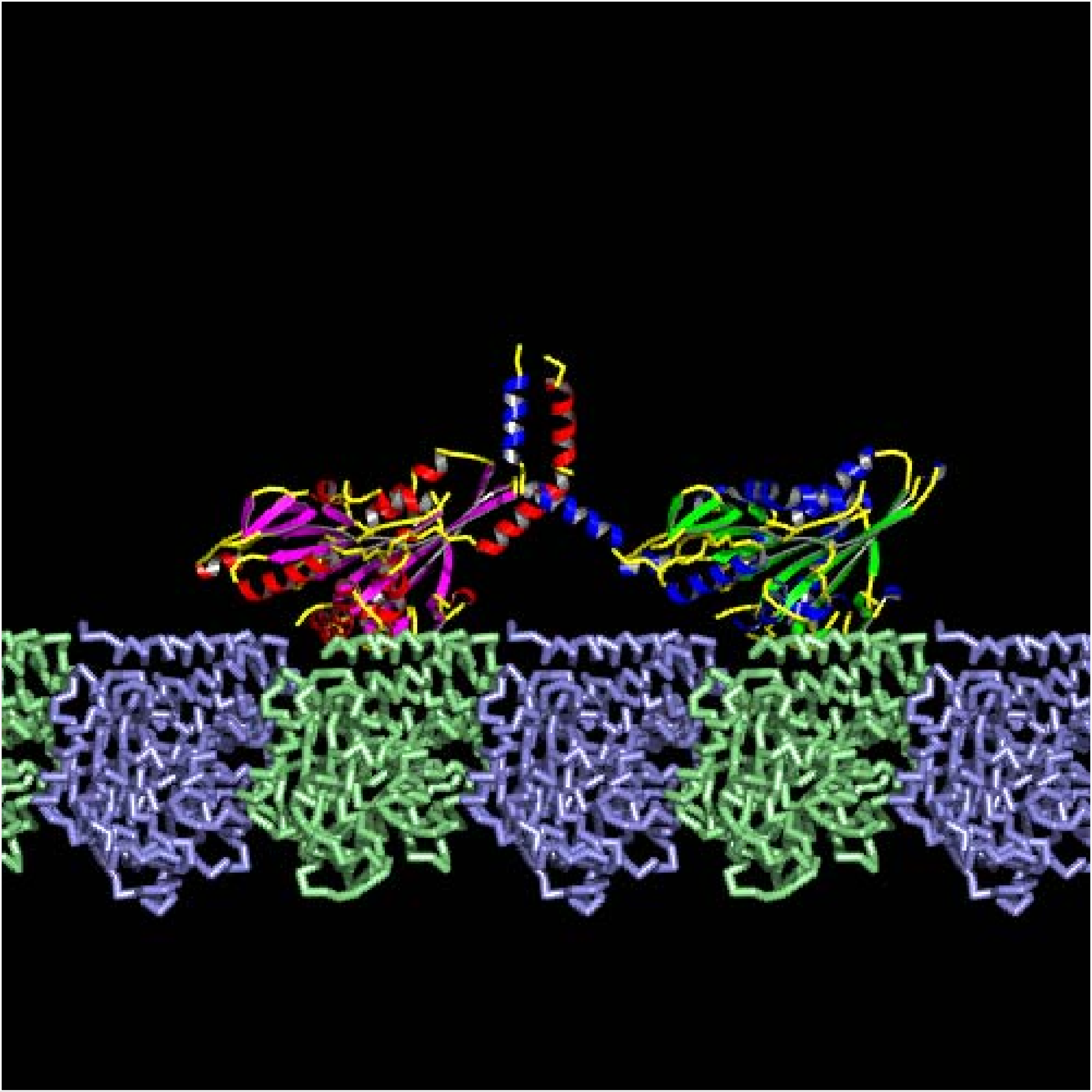
Song et al. (2001)

http://www.mpasmb-hamburg.mpg.de/ktdock/

Mandelkow lab

Kinesin Motility

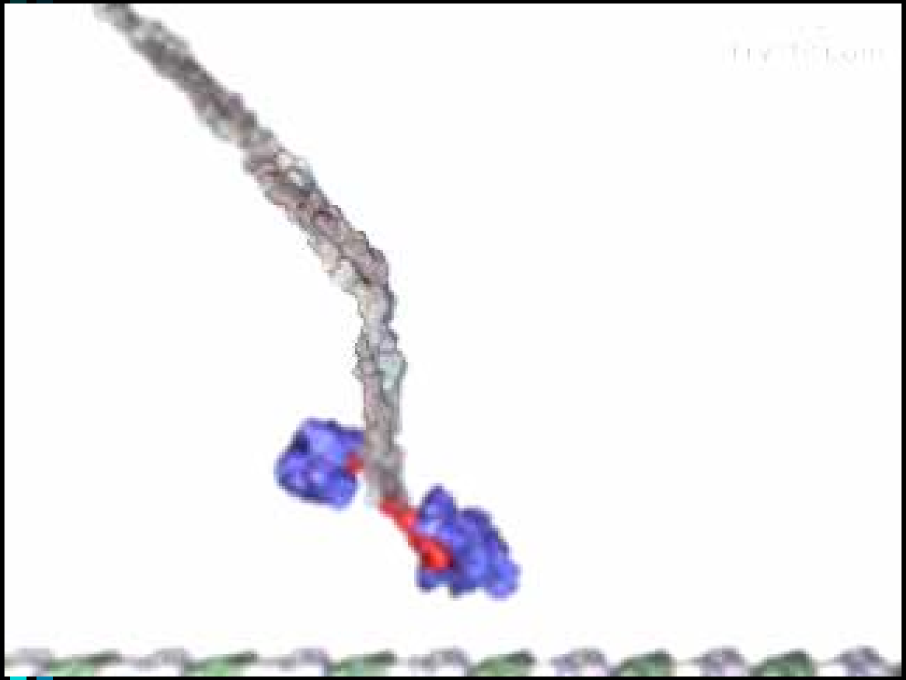
Model of kinesin dimer walking along a microtubule protofilament. (Hoenger et al.2000).



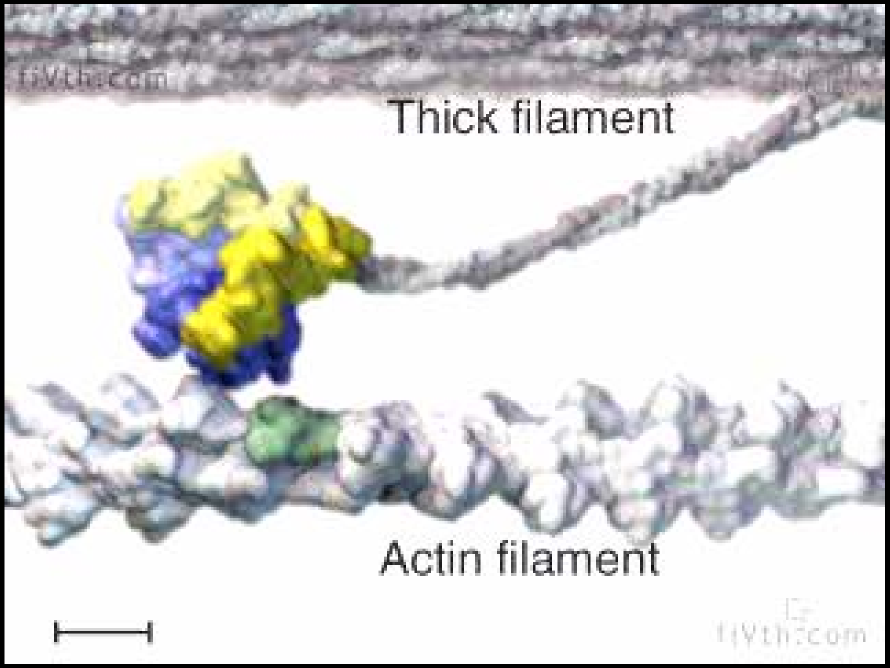
http://www.mpasmb-hamburg.mpg.de/ktdock/

Mandelkow lab

Kinesin Motility



Animated Model for
Processive Motility by Kinesin



Animated Model for Muscle
Myosin-Based Motility

http://www.scripps.edu/milligan/projects.html **Ron Milligan Lab**

Preguntas, por favor...

<pagomez@cbm.uam.es>

http://www.cbm.uam.es/pagomez