

In silico Structural Biology of Signaling Proteins

Chang-Shung Tung

Group Leader

Theoretical Biology & Biophysics, LANL

ct@lanl.gov

505-665-2597

Outline

- Methods for structural determination/modeling.
 - computational approaches.
 - homology modeling.
 - an example (what, why, how, things learned).
- Applied to signaling proteins.
 - toll-like receptors.
 - what can structural modeling do/help?

Ways to determine protein structures

- X-ray crystallography
- NMR
- Cryo EM
- Low-resolution methods (SAXS, Neutron)
- Computational (*in silico*) prediction

Protein folding prediction

- From sequence to 3-D structure
- *ab initio*
 - conformational sampling
 - target function (potential energy, etc)
- Knowledge based
 - homology modeling
 - threading

Protein folding prediction (continue)

ab initio - classical:

- 30 years of works (Scheraga, Karplus, Levitt, etc)
- many software developed: CHARMM, AMBER, GROMOS, NAMD, etc)
- capable of predicting structures of small proteins.

Protein folding prediction (continue)

knowledge-based, homology modeling

- proteins with similar sequence/function fold into similar folds.
- the accuracy is approaching mid-resolution crystal.
- independent of protein size.
- easy and straight-forward.
- many software available (modeller, swiss-prot, etc).

Homology modeling

(Comparative modeling)

- The target protein needs to have >30% sequence identity with template protein(s) of known structure(s).
- Accurate sequence alignment is crucial for the success of the model structure.
- Structural comparison using root-mean-square-deviation (RMSD) metric as a measure between two structures.
- a typical model has $\sim 2 \text{ \AA}$ agreement between the matched C_{α} atoms at 70% sequence identity.
- More info in:

http://en.wikipedia.org/wiki/Homology_modeling

Can homology modeling works with low sequence identity?

- Tramontano (1998), Methods: A companion to Methods in Enzymology 14: 293-300.
- Tung, et al., (2004) J Gen Virol 85: 3249-3259.

Hemagglutinin (HA)

- Surface glycoprotein (aka membrane fusion protein, envelope protein), has two components (HA1, HA2) linked by disulfide bond.
- The functional unit is a trimer.
- HA binds to receptor of the host cell and initiates membrane fusion.
- Structurally, influenza HA is best studied and served as a model system for understanding membrane fusion between virus and host cell.

HA (continue)

- Crystal structure of influenza-a HA was solved in the 70s.
- Crystal structure of influenza-c HA/NA fusion protein was solved in 1998.
- Structure of influenza-b HA is not known.
- To model the structure of the influenza-b HA using a knowledge-based approach.
- Pair-wise sequence identities between HAs from flu-a, flu-b, flu-c are all under 20%.
- Using structural alignment of HA from flu-a and flu-c, added sequence of HA from flu-b to produce a 3-way alignment.

3-way alignment

```

      10      20      30      40      50
QDLPGNDNST ATLC LGHAV PNGTLVKTIT DDQIEVTNAT ELVQSSSTGK I.....
..... DRICTGITSS NSPHVVKTAT QGEVNVITGVI PLTTTPTKSH F.....
.....EK IKICLQKQVN SSFSLHNGFG GNLY.ATEEK RMFELV.KPK AGASVLNQST
      10      20      30      40      50
.....
..... CNNPH RILDG..... .IDCTLIDAL LGDPH.C.DV
..... ANLKG TQTRGKLCPN CFNCTDL DVA LGRPK.CMGN
WIGFGDSRTD KSNSAFP RSA DVSAKTADKF RFLSG..... ...GSLMLSM FGPPGKV.DY
      60      70      80      90     100
      80      90     100     110     120
FQNETW..DL FVE....RSK AFSNCYP..Y DVPD.YASLR SLVASSG... ..TLEFITE
TPSAKV..SI LHE....VKP ATSGCFPIMH DRTK.IRQLP NLLRGYE... ..NIRLSTS
LYQGCCKHKV FYEGVNWSPH AAINCYR..K NWTDIKLNFO KNIYELASQS HCMSLVNALD
      110     120     130     140     150
      130     140     150     160
GFTWT..... GVTQNGGSNA CKR.GPGSGF FSRLNWLTKS GS....TYP VLN...VTMP
NVINTETAPG GPYKVGTS GS CPNVANGNGF FNTMAWVIPK DN....NKT AINPVTVEVP
KTIPL..... QVT.AGTAGN CN....NSF LKNPALYTQE VKPSENKCGK ENL...AFPT
160      170     180     190     200
      170     180     190     200     210
N..NDN.... ..FDKLYIWG IHHPSTMQEQ TSLYVQASGR VTVSTRRS.. QQT.IIPNIG
YICSEG.... ..EDQITVWG FHSDDKTQME R.LYGDSNPQ KFTSSANG.. VTHYVSQIG
L..PTQFGTY ECKLHLVASC YFIYDSKEVY NKR.GCDNYF QVIYDSFGKV VGG.LDNRSV
      210     220     230     240     250     260
      220     230     240     250     260
SRPWV...RC L..SSRISY WTIVKPGDVL VINSNGNLIA PRGY..FKMR TGKS.....
GFPNQTEDEG LKQSGRIVVD YMVQKPGKGT TIVYQRGILL PQKV..W.CA SGRS.....
PYTGN...SG D..TPTMQCD MLQLKPGRYS VRSSPRFLLM PERSYCFDMK EKGPVTAVQS
      270     280     290     300     310
      270     280     290     300     310
..... SIMRSDAPI..... DTCISECITP
..... KVIKGSPLPL..... .IGEADCLHE
IWGKGRES DY AVDQA CLSTP GCMLIQKQKP YIGEADDDHG DQEMRELLSG LDYEARCISQ
      320     330     340     350     360     370
      290     300     310     320
N.GSIPNDKP FQNVN..KI. TYGACPKYVK QNTLKLATGM RNVPEKQT
KYGGLNKS KP YYTGE..HAK AIGNCPIWVK .TPLKLANGT KYRPPAKLLK E
S.GWVNETSP FTEKYL LPP. KFGRCPLAAK .EESIPKIPD GLLIPTSGTD TTVTKPKS
      380     390     400     410     420     430

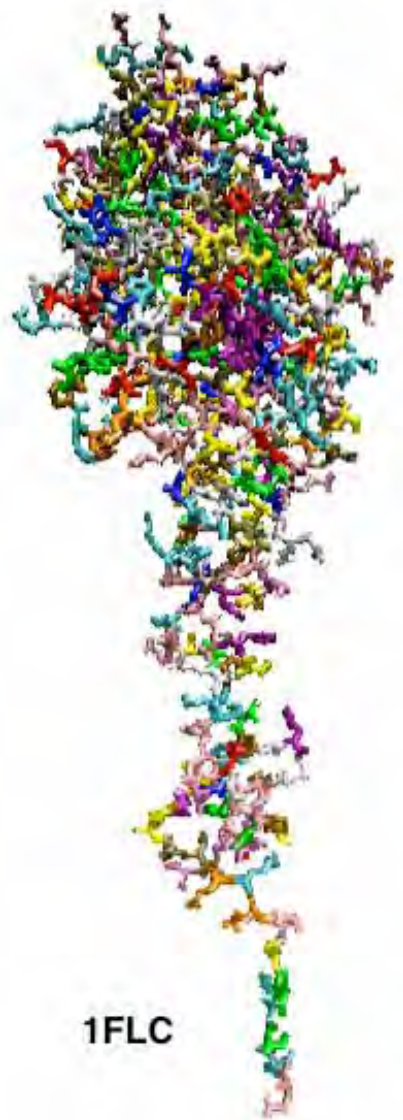
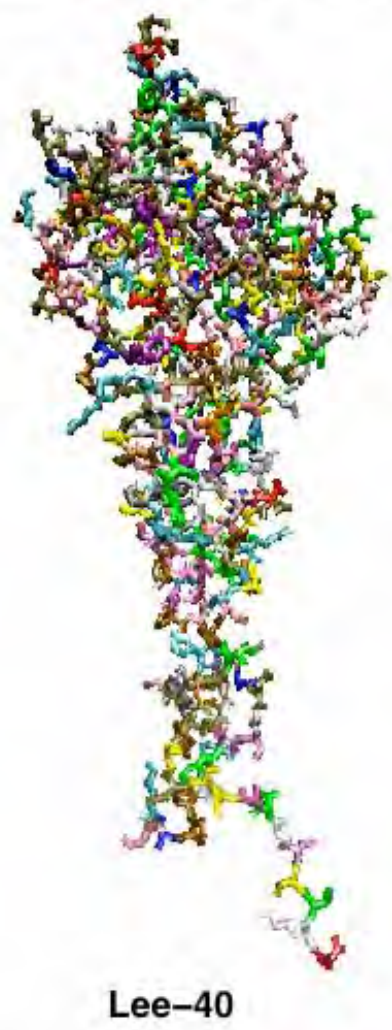
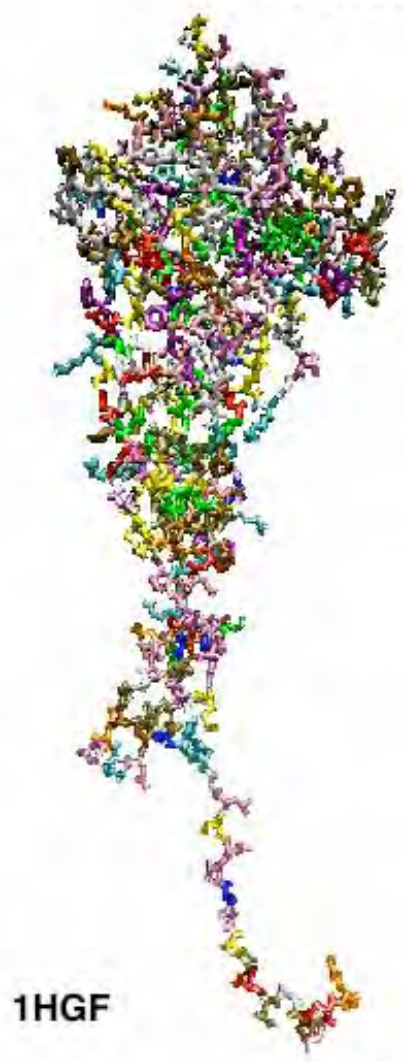
```

23 conserved residues

conserved residues

		a	b	c	
HA2	← CYS	14	4	6	→
	THR	37	27	28	
	GLY	61	51	85	
	GLY	72	68	94	
	PRO	74	70	96	
	GLU	89	86	114	
	ALA	93	90	122	
	→ CYS	97	94	126	←
	→ GLY	134	138	169	←
	→ CYS	139	143	174	←
	PHE	147	152	178	
	GLY	225	238	268	
	LYS	238	253	281	
	PRO	239	254	282	
	GLY	240	255	283	
	PRO	254	269	297	
	→ CYS	281	294	373	←
	→ GLY	286	300	378	←
	PRO	293	307	385	
	GLY	303	318	397	
	→ CYS	305	320	399	←
	PRO	306	321	400	
	LYS	310	325	404	

GLY: 7			
CYS: 5			
PRO: 5			
LYS: 2	→ CYS-64	CYS-60	?
THR: 1	→ CYS-76	CYS-72	?
GLU: 1			
ALA: 1	→ CYS- 52	(LYS- 45)	?
PHE: 1	→ CYS-277	(ASP-293)	?



Model validation

- Good stereochemistry (procheck) ?
- Functionality of HA1 -- binding of the sialic acid?
- Can the model accommodate naturally occurring deletions/insertions?
- Supporting observed mutations?

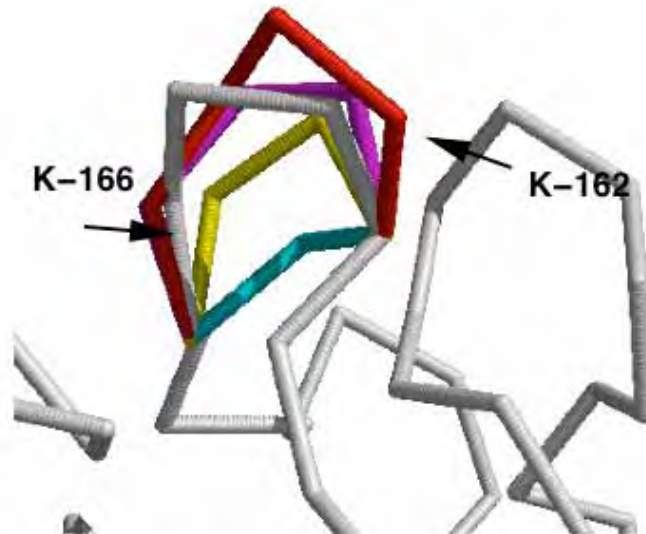
Quality of the model

- Bonds.
- Van der waal contacts.
- Main-chain torsional angles (ϕ, ψ).

98% in the combined core and allowed regions,
none in the disallowed region.

naturally occurring deletions/insertions

	162	166
B/Lee/40	IPK-DNNKTA	
B/Bonn/43	VPK---NKTA	
B/Osaka/70	VPK-N-NKTA	
B/Finland/85	VPKNNNNKTA	



white: homology modeled
magenta: loop modeled; yellow: 1 aa deletion
cyan: 2 aa deletion; red: 1 aa insertion

Residue 269

- A signature of the sublineages
- “Pro” in Yamagata sublineage.
- “Ser” in Victoria sublineage.
- “Pro” to “Ser” is a non-conservative change.

changes involves both charge (neutral to polar) and size (“Ser” is larger).

1 nucleotide change

PRO: CCU, CCC, CCA, CCG

SER: Ucu, Ucc, Uca, Ucg

THR: Acu, Acc, Aca, Acg

ALA: Gcu, Gcc, Gca, Gcg

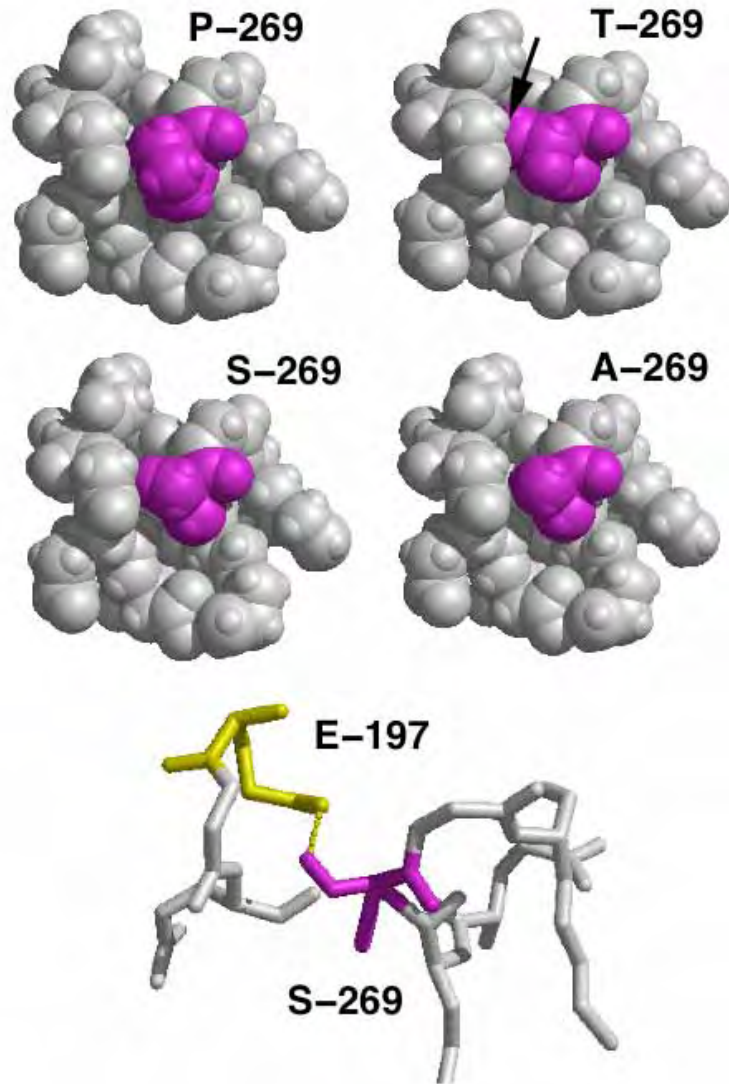
LEU: cUu, cUc, cUa, cUg

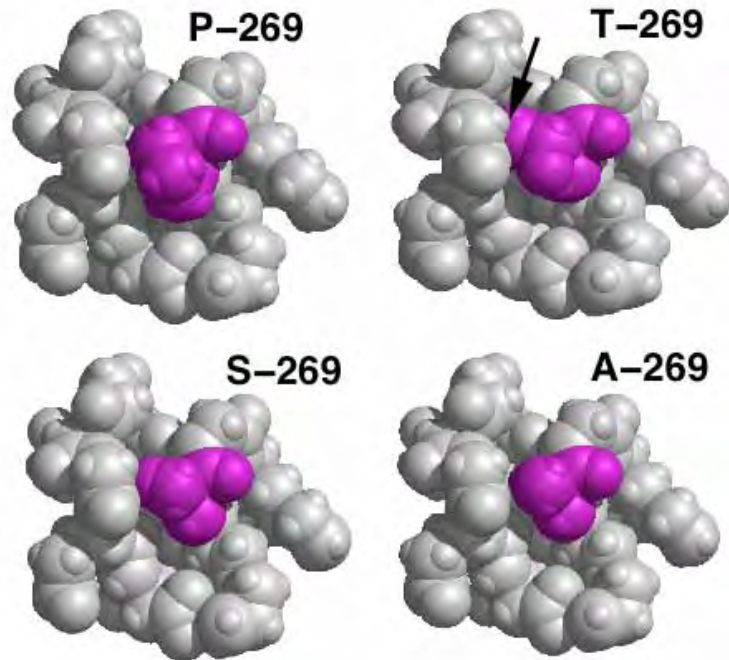
HIS: cAu, cAc

GLN: cAa, cAg

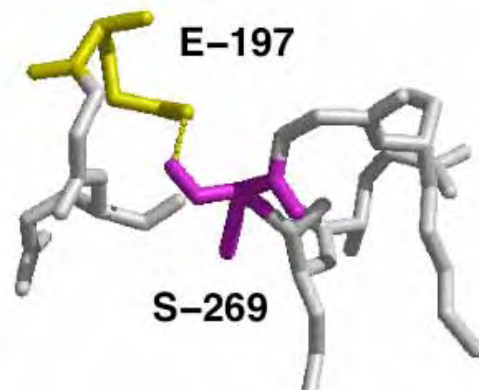
ARG: cGu, cGc, cGa, cGg

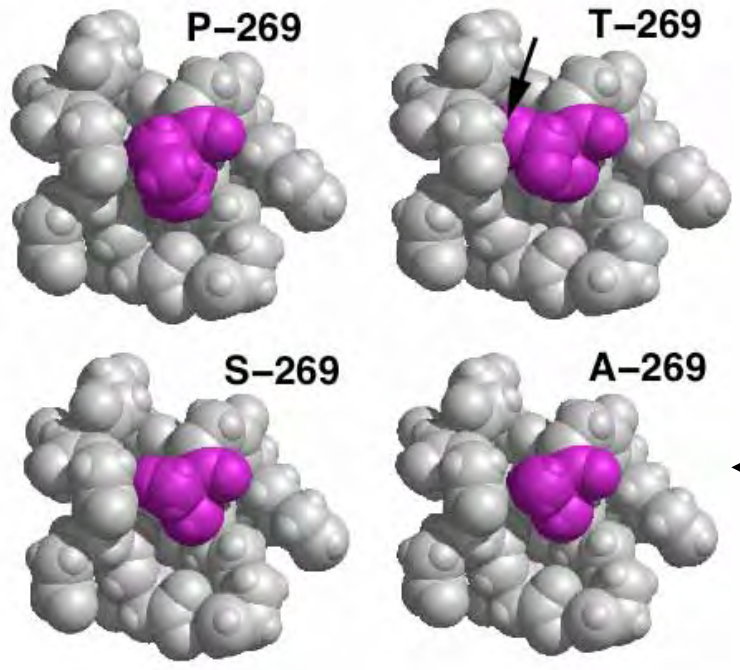
Different amino acid types at 269



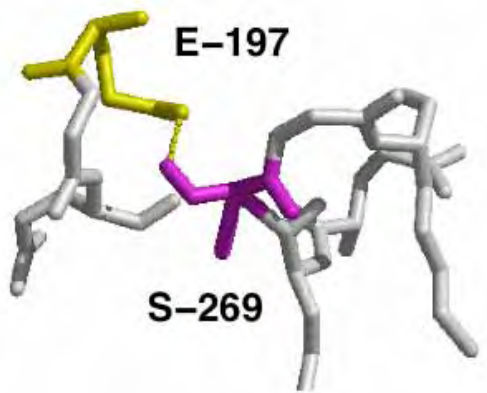


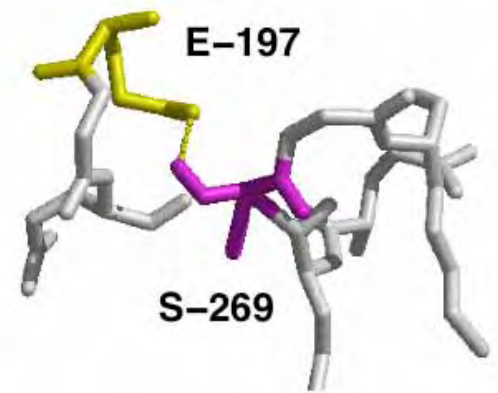
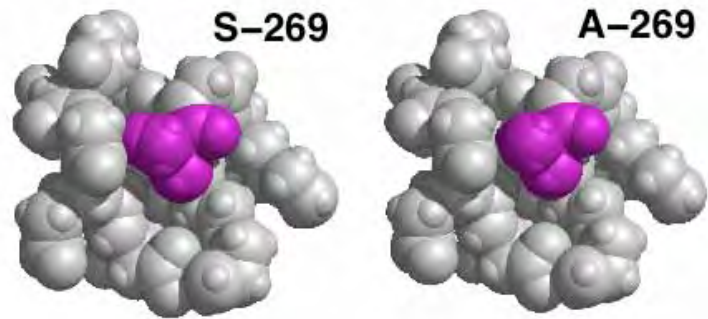
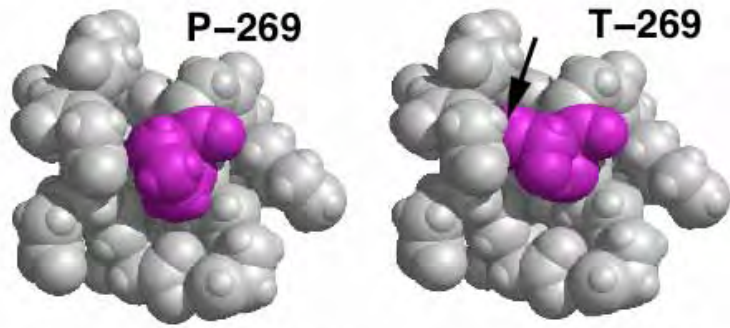
T-269 interfere with G-198 and E-199. Therefore, Thr, Leu, His, Gln or Arg are all unfavorable at this position.





Both Ser and Ala
are smaller than
Pro, lost some
favorable contacts

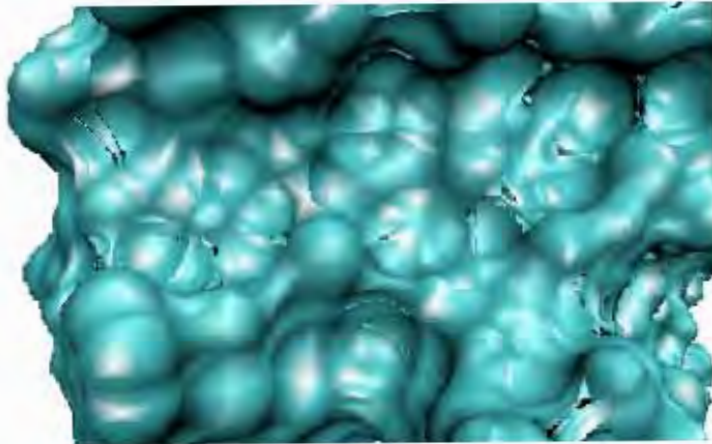




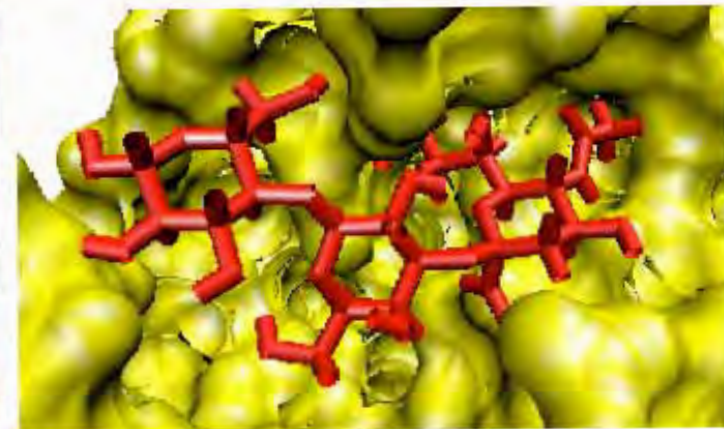
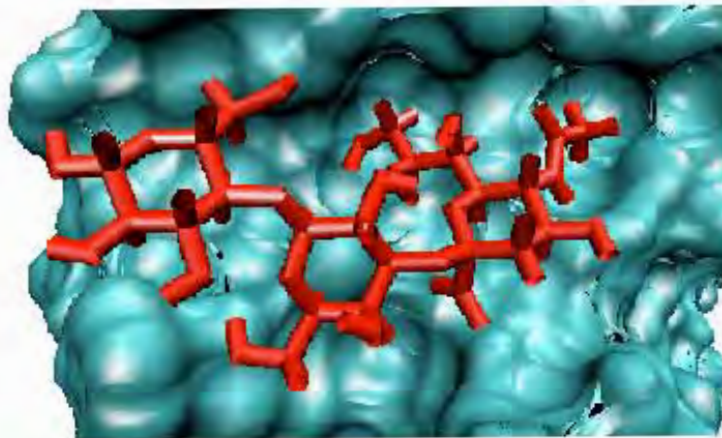
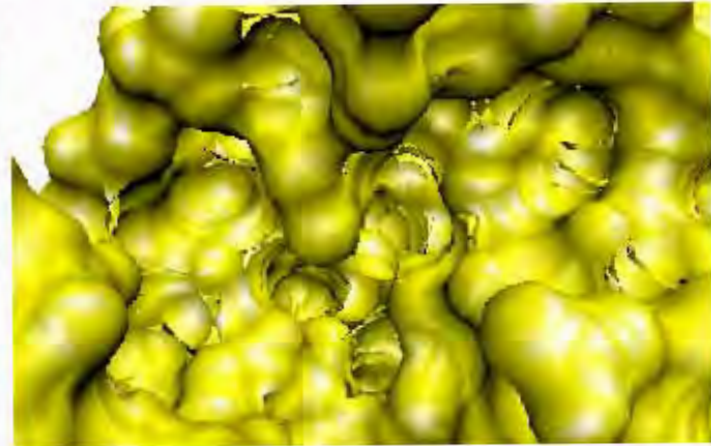
A H-bond between S-269 and E-197 stabilizes the structure

Receptor binding

A/Aichi/2/68



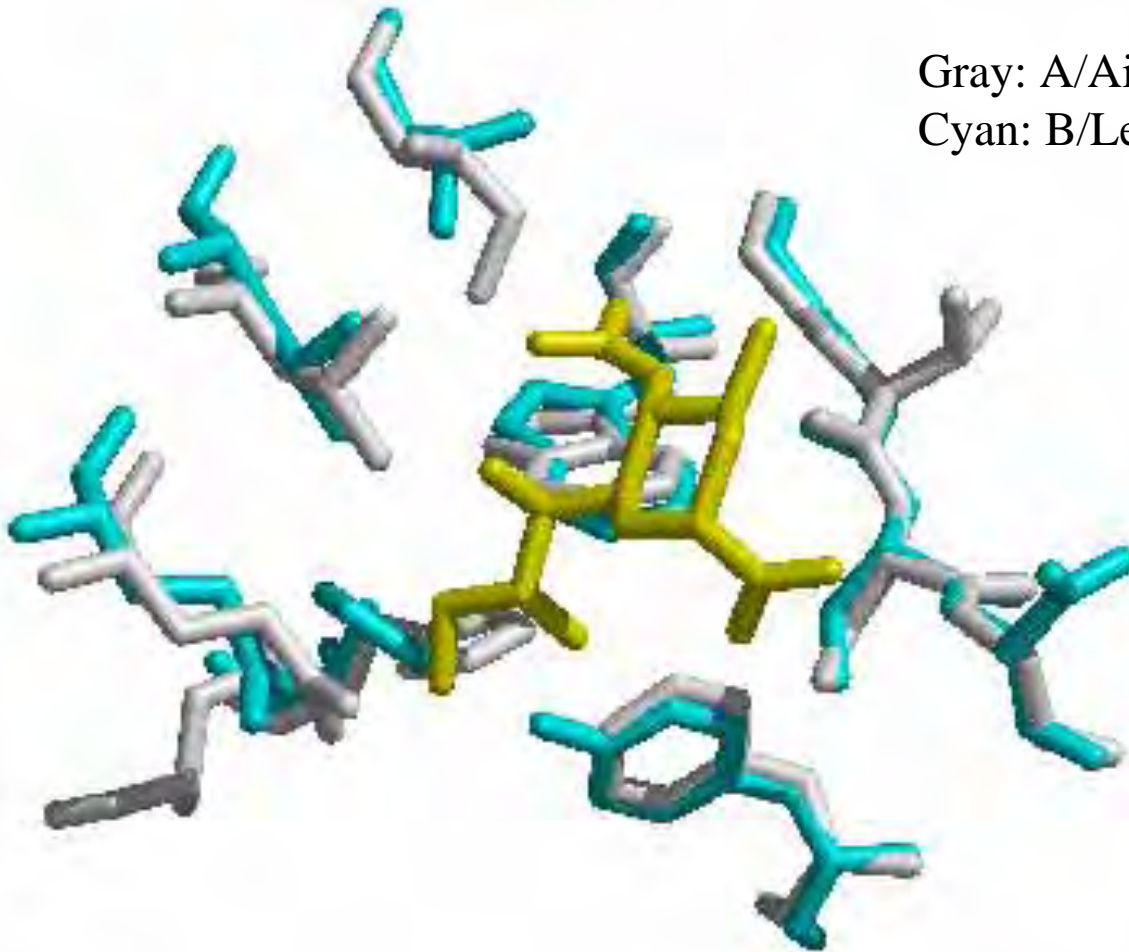
B/Lee/40



Receptor binding (continue)

Gray: A/Aichi/2/68

Cyan: B/Lee/40



Background:

- Structural motifs are functionally relevant.
- Folds are preserved, binding interfaces are shared among proteins in the same family.
- Structures of interacting molecules can be modeled computationally with reasonable accuracy.
- Predictions can be tested experimentally.
- Experimental results can be used to refine structural models.

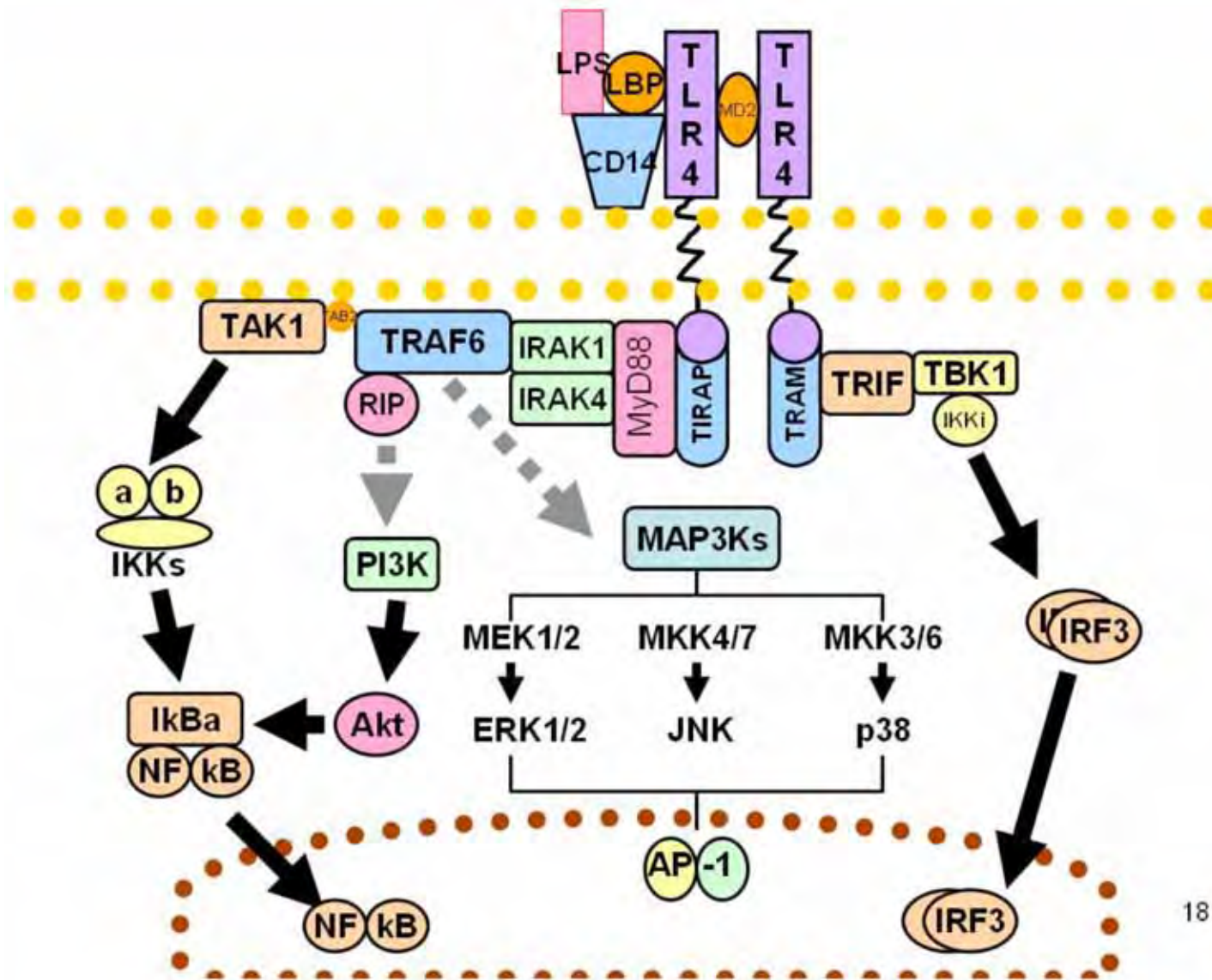
Things to look for

- Type of binding surface:
dimers, trimers, tetramers, etc.
- Specifics in binding:
H-bonds, ion pairs, hydrophobic interactions, shape, etc.
- Interface surface area:
correlates to binding strength.

Toll-like receptor (TLR)

- Part of our innate immune system.
- Pattern recognition receptors that recognize molecules that are broadly shared by pathogens.
- Presents in vertebrates and invertebrates.
- 13 mammalian toll-like receptor families.
- First human toll-like receptor was described by Nomura et al., in 1994.

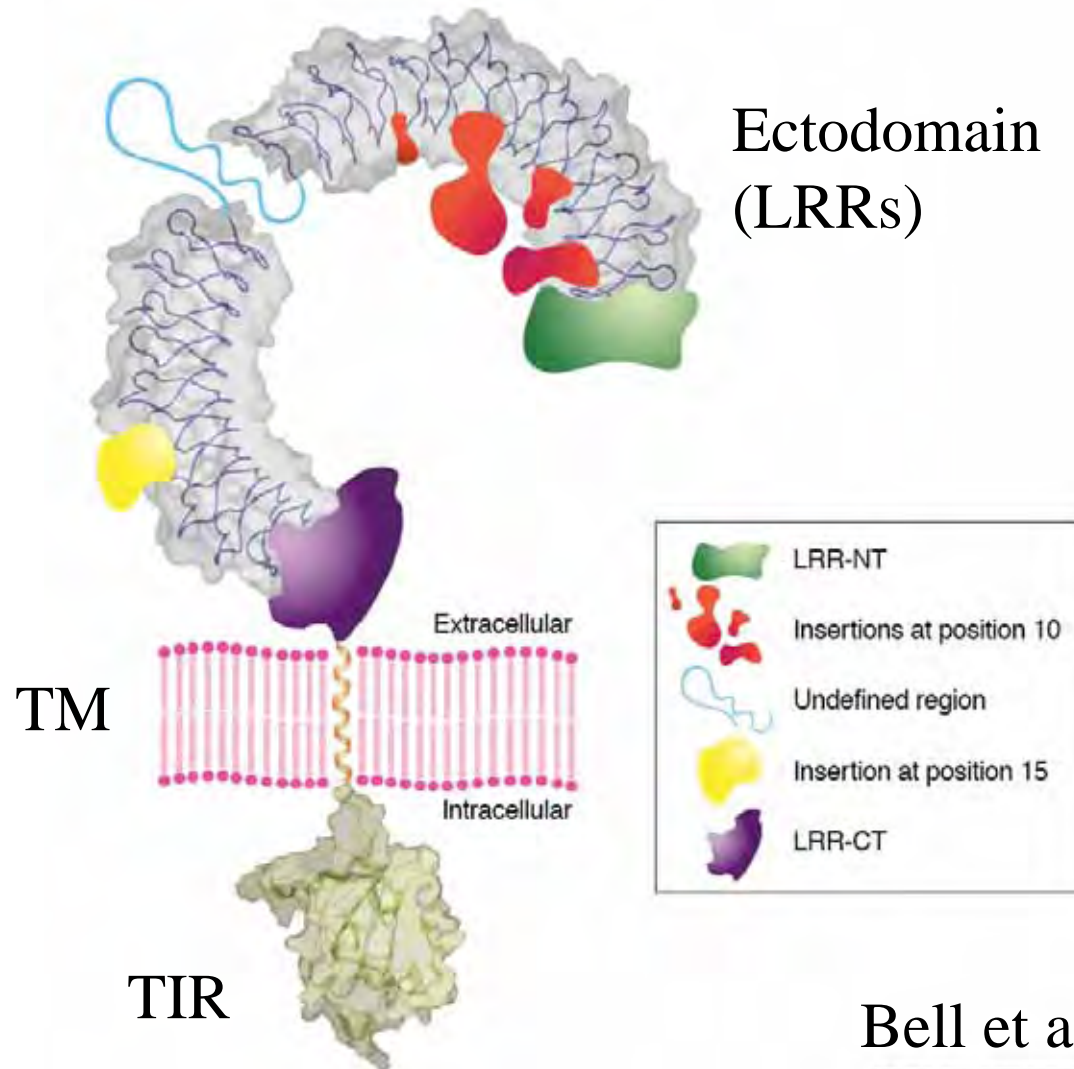
Signaling pathway of toll-like receptor



Toll like receptor

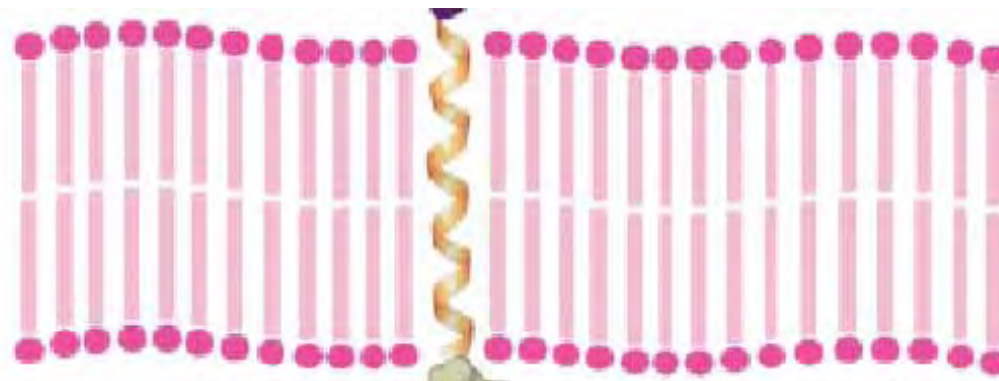
- Dunne and O'Neill
www.stke.org/cgi/content/full/sigtrnas;2003/171/re3.
- Takeda, et al., 2003 Annu Rev Immunol 21: 335-376.
- http://en.wikipedia.org/wiki/Toll-like_receptor

Structure of TLR



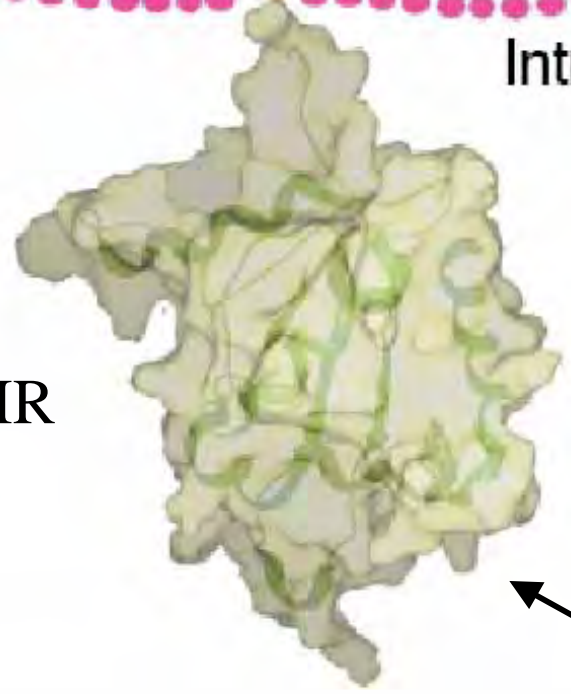
Bell et al., 2003

TRENDS in Immunology



Intracellular

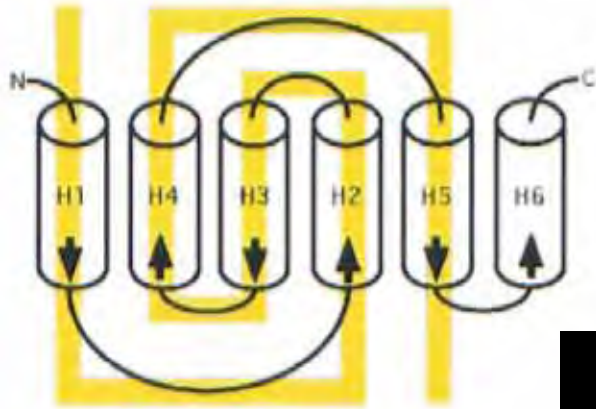
TIR



MyD88
(DD, TIR)
(Adaptor molecules)

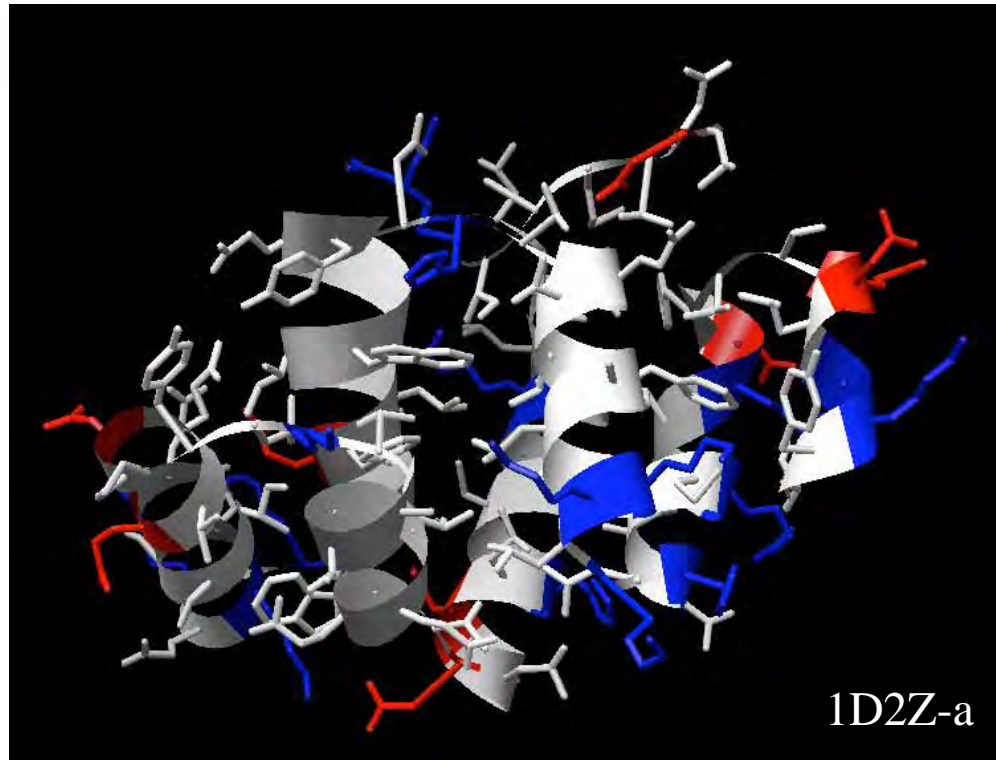
IRAK

Death domain (DD)



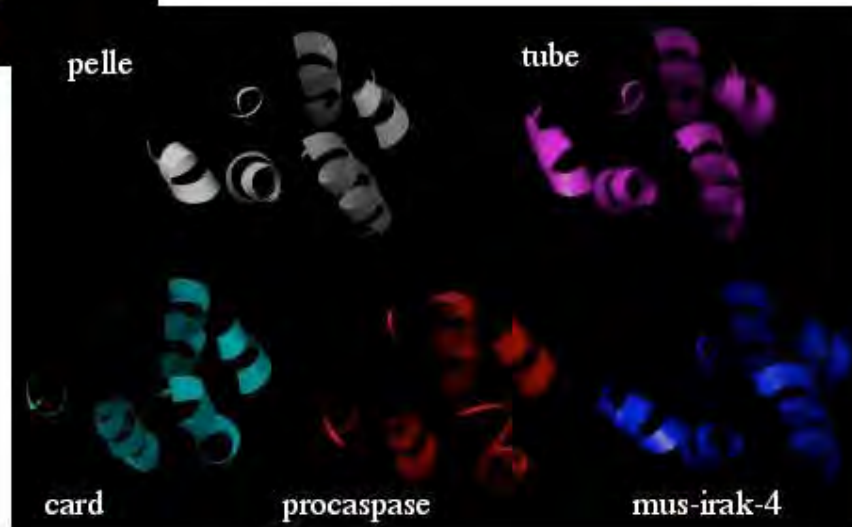
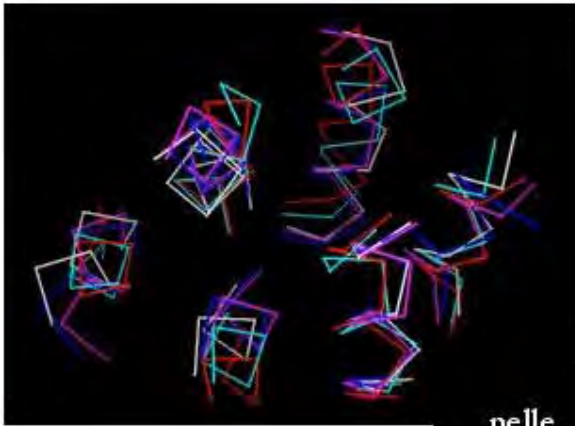
Greek key fold

Pelle DD

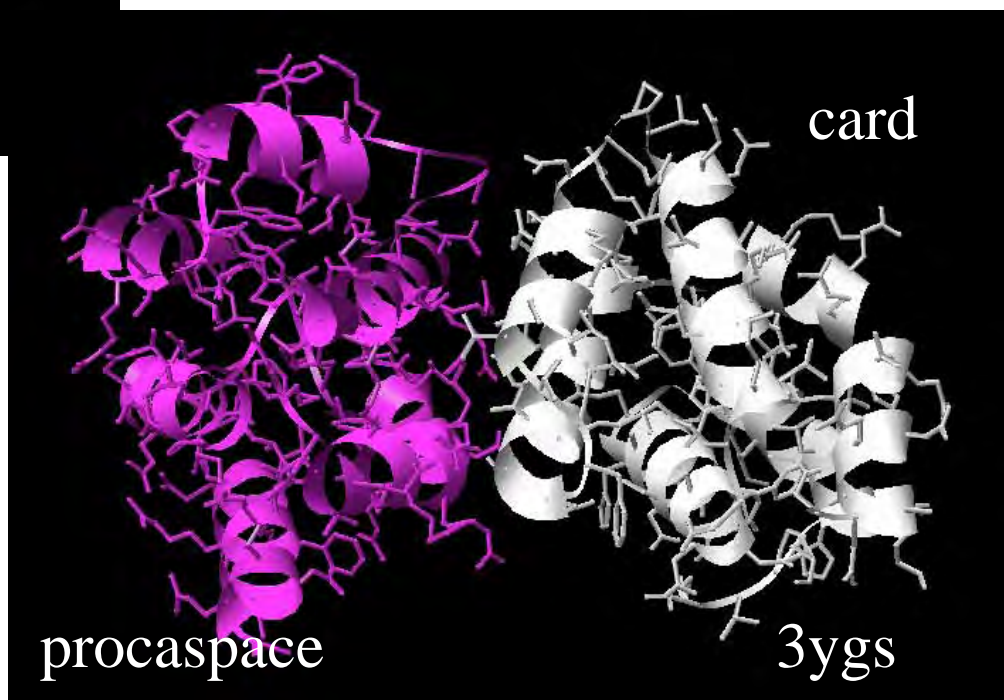
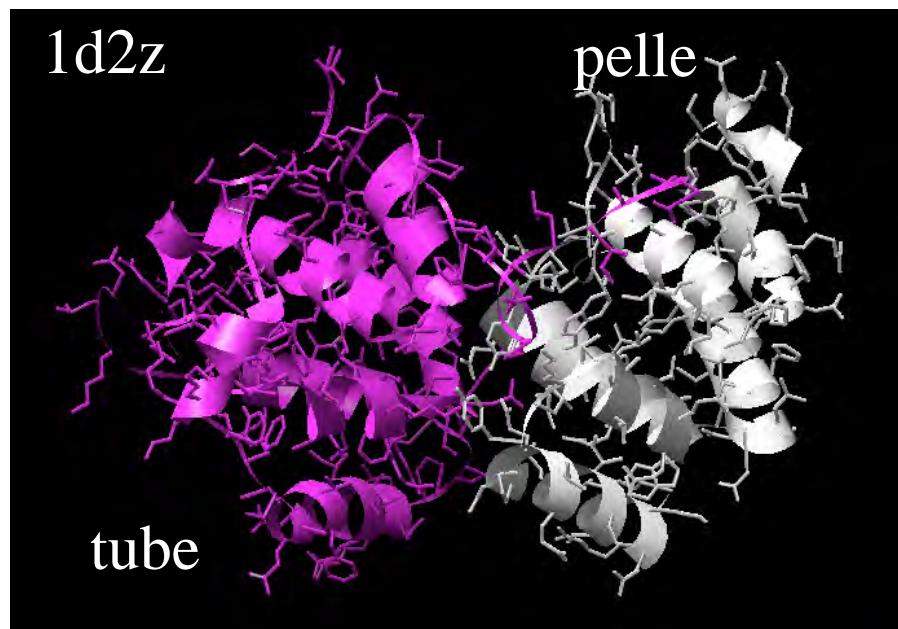


DD is a structural motif

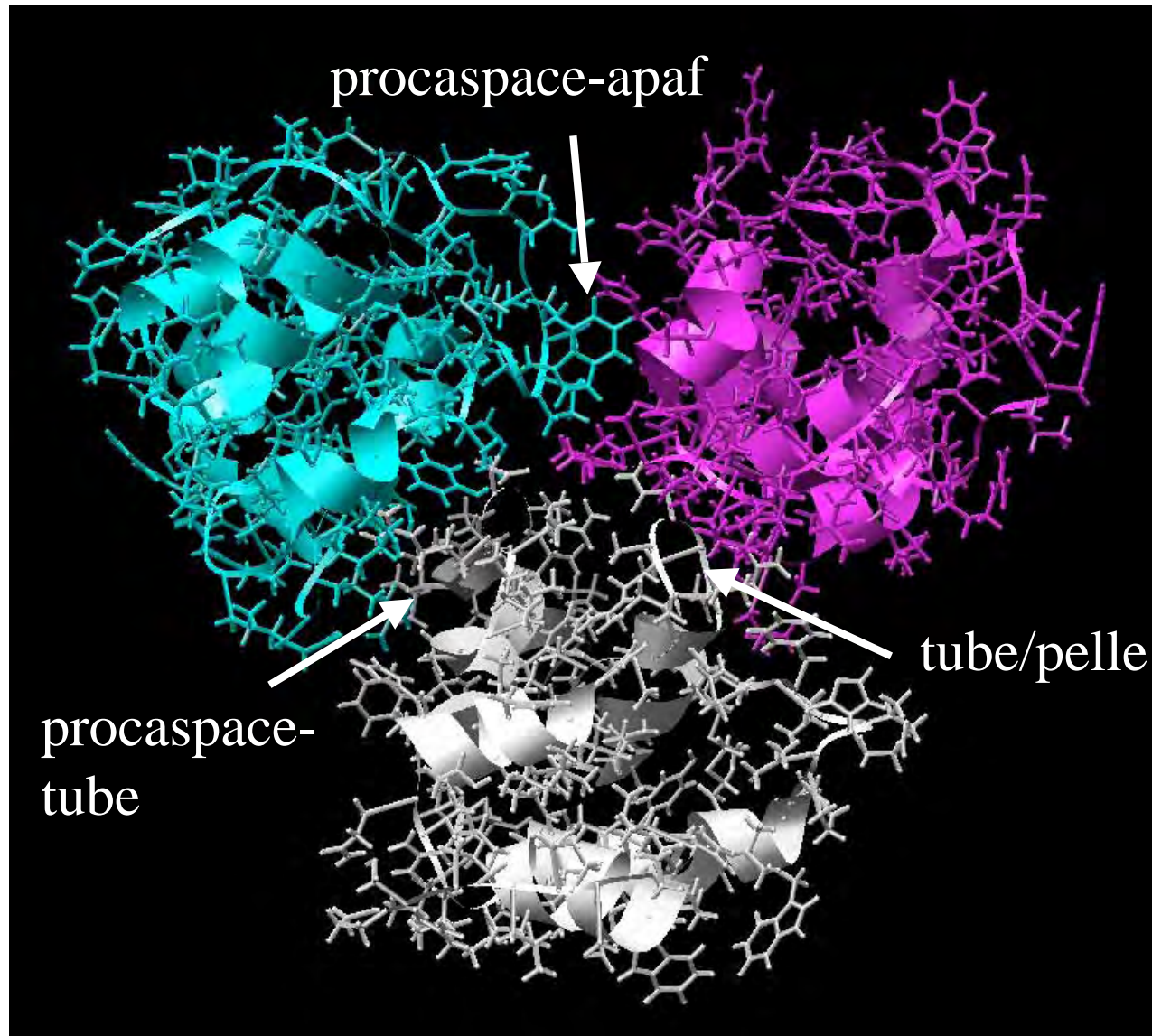
	pelle	tube	card	procaspase	mus-irak-4
pelle	0.00	1.24	2.06	2.12	1.21
tube	17.3% (5.3%)	0.00	1.98	1.93	0.90
card	14.6% (7.1%)	17.8% (6.9%)	0.00	1.11	2.20
procaspase	20.2% (8.9%)	18.5% (7.3%)	19.4% (18.4%)	0.00	2.19
mus-irak-4					0.00



crystal contacts

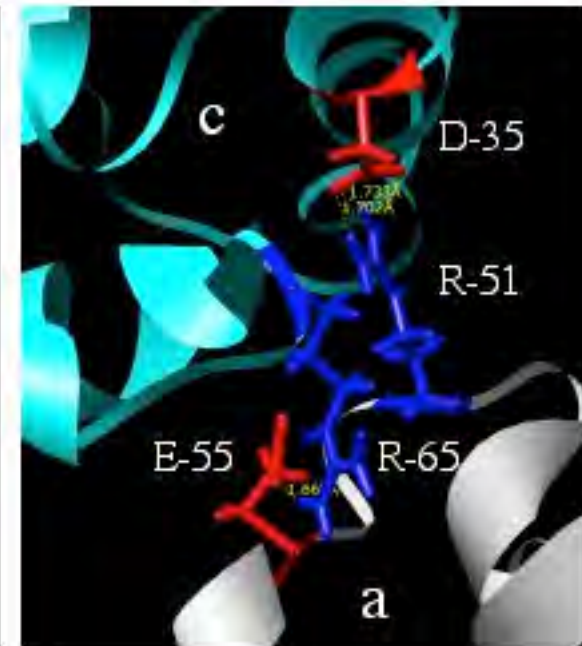
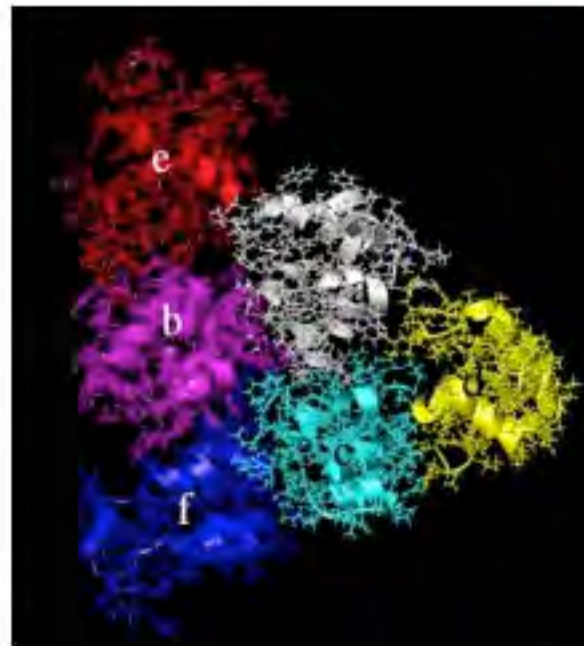
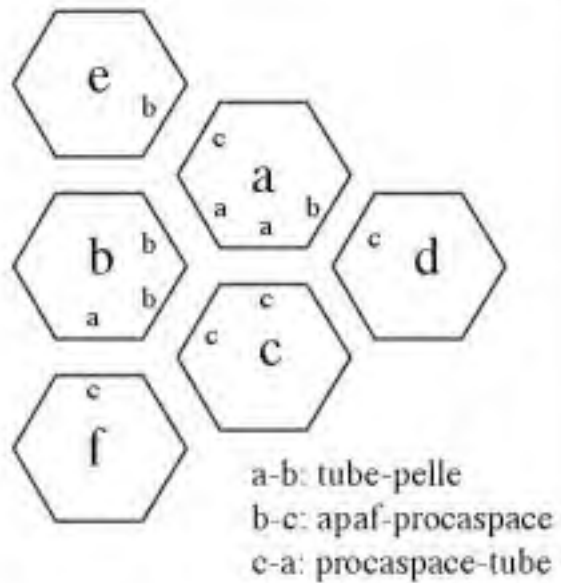


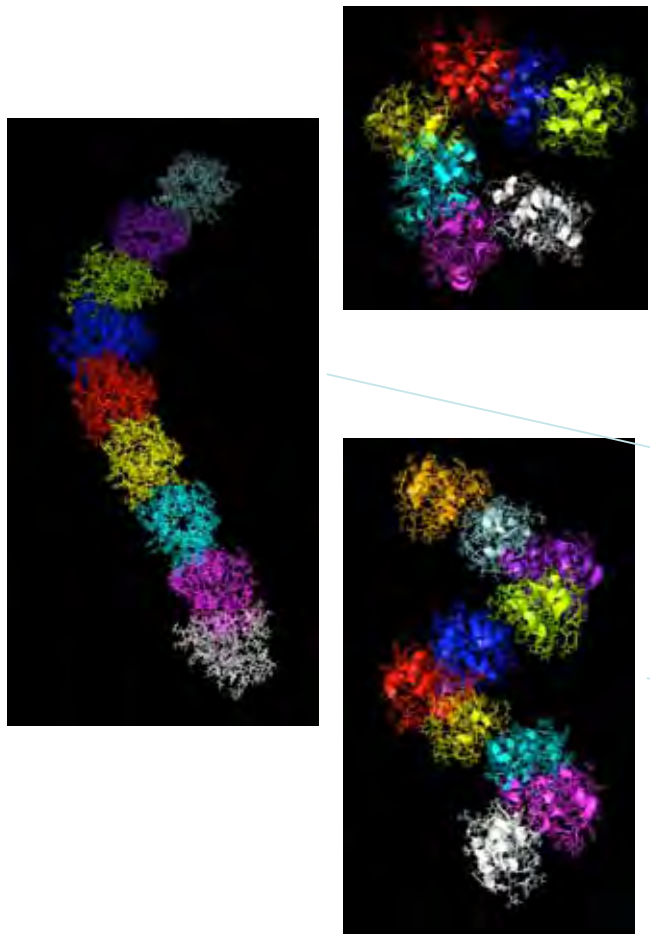
3-mer model



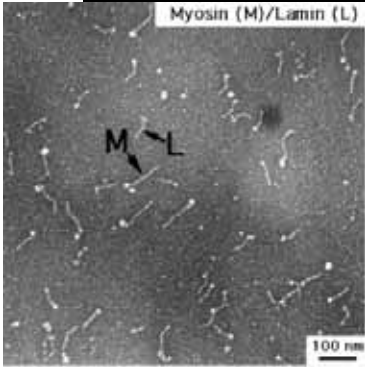
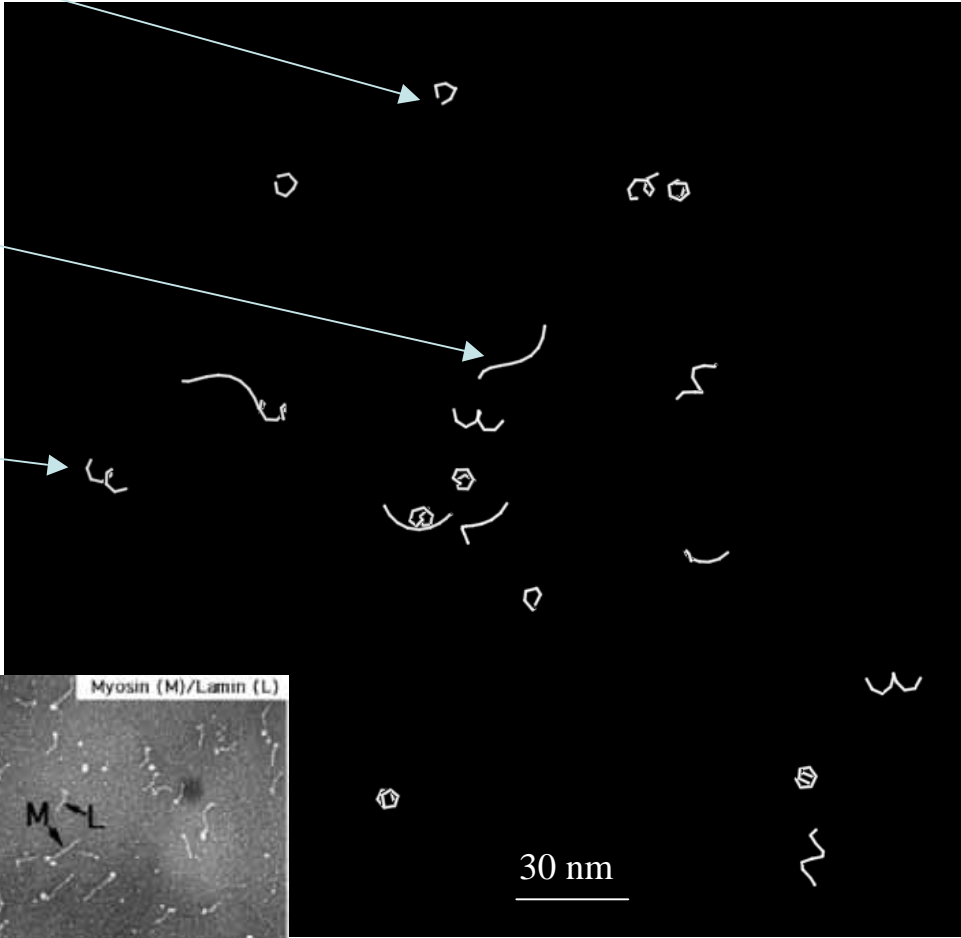
6-mer model

a,b,c being IRAK-1 DD; d,e,f being MyD88 DD



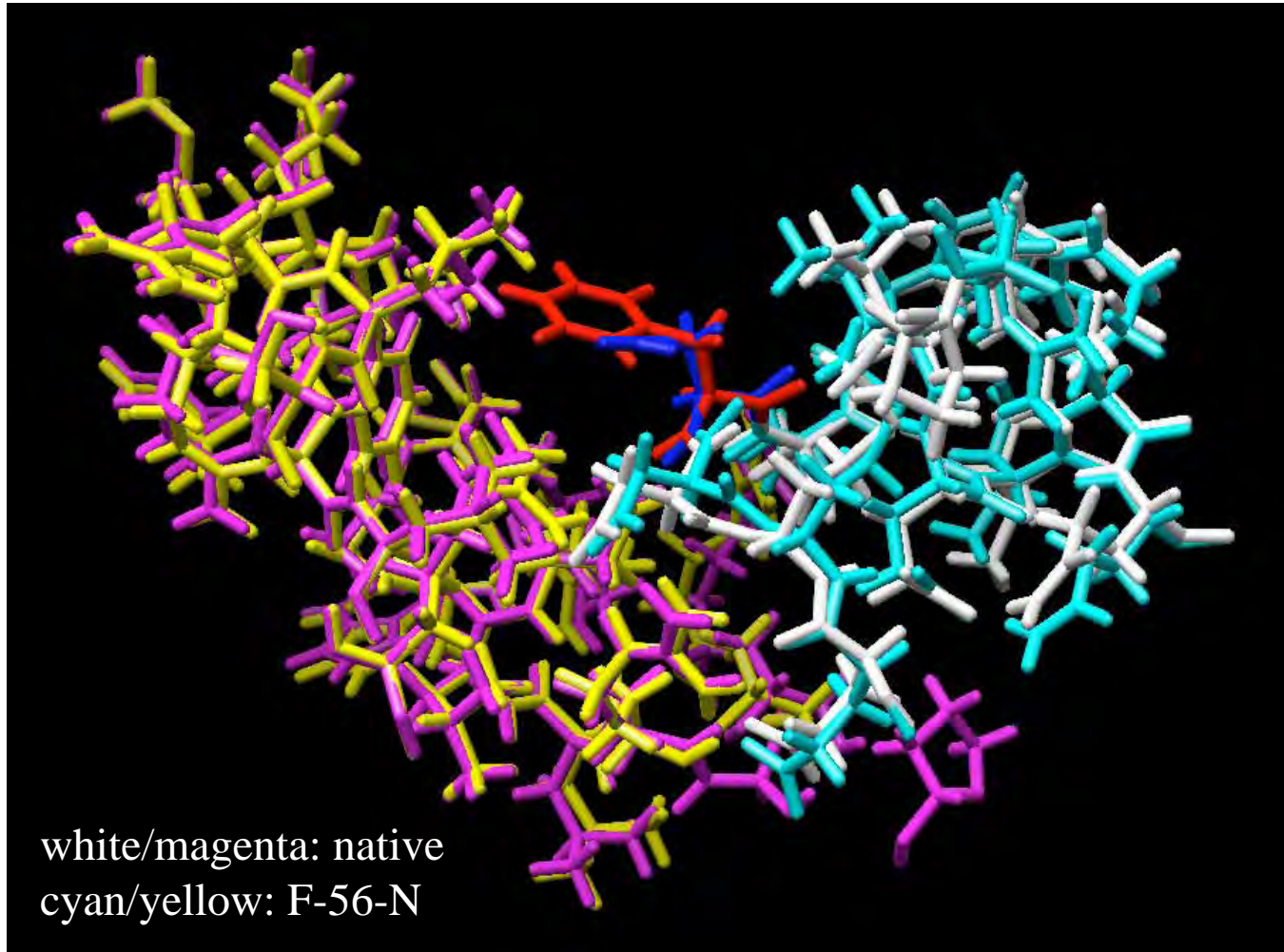


oligomers



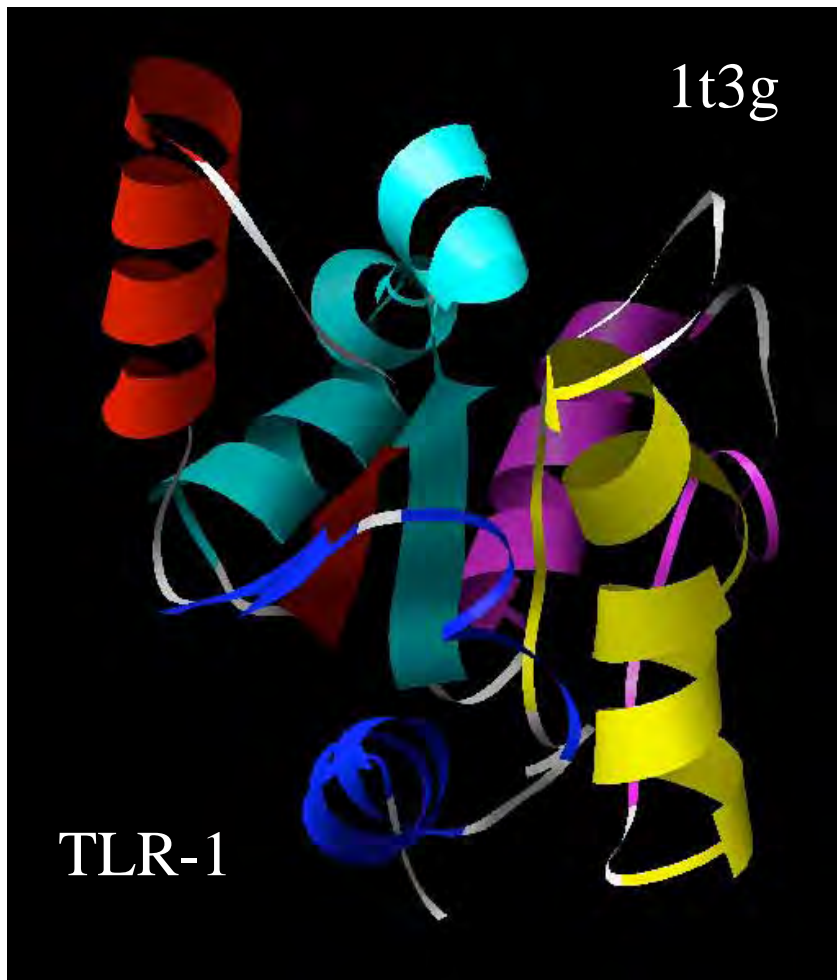
F-56-N mutation prevents dimerization
of MyD88 DD. (Burns et al., 1998)

F-56-N mutation



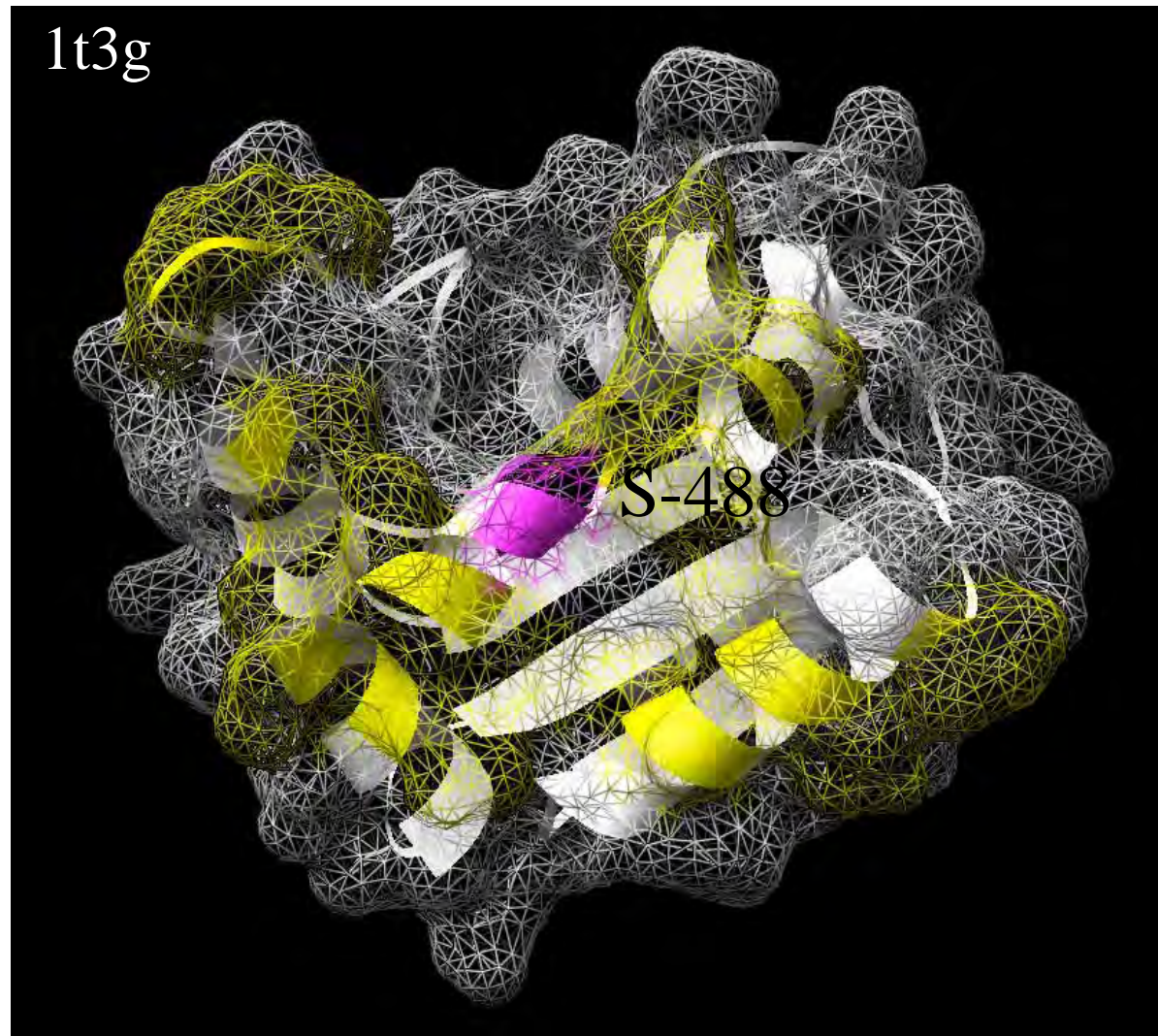
Loss 125 Å² of interface area due to mutation.

TIR: Toll/Interleukin-1 receptor domain

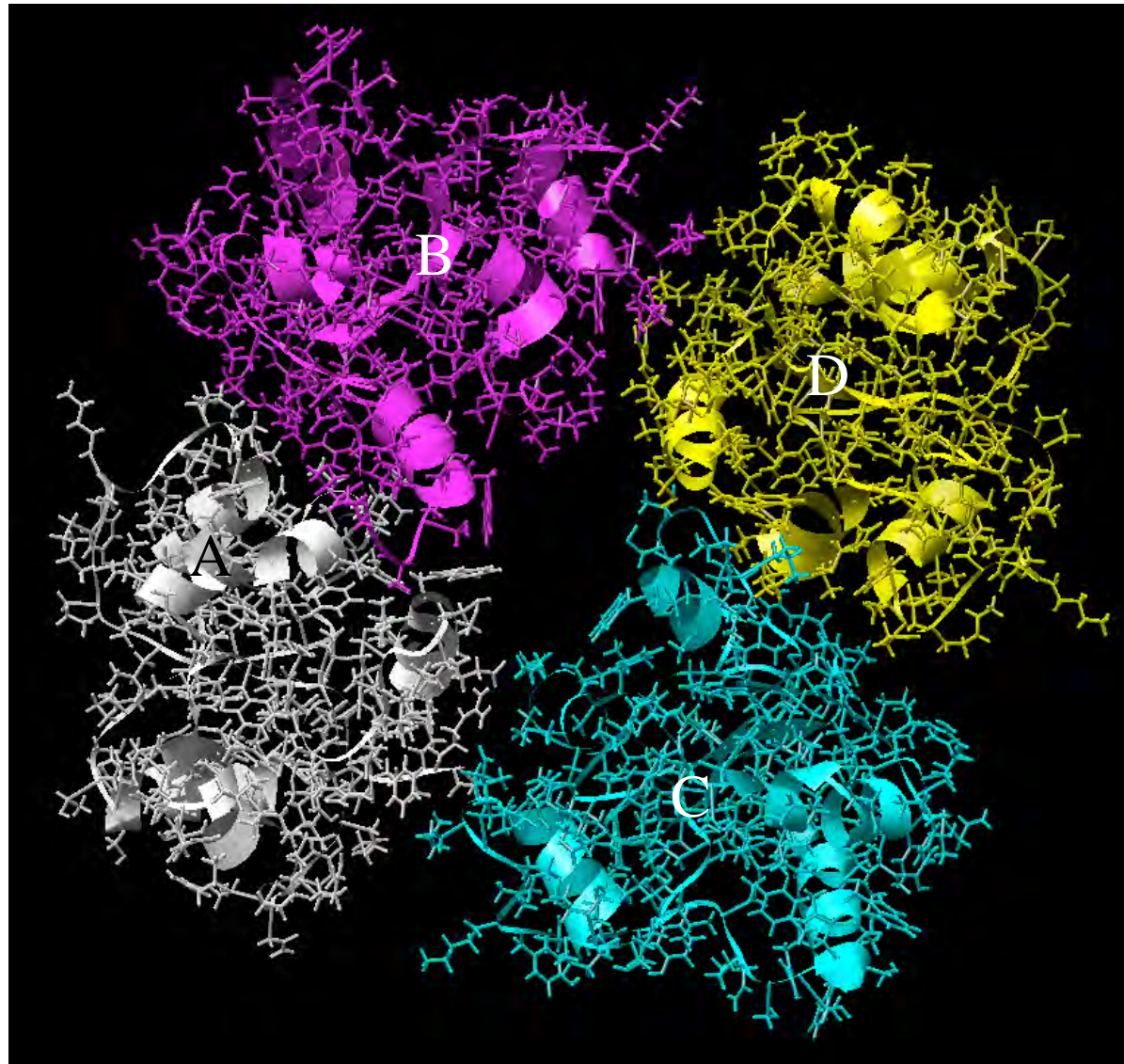


β - α folds

Interface area



MyD88 TIR 4-mer



Interface areas

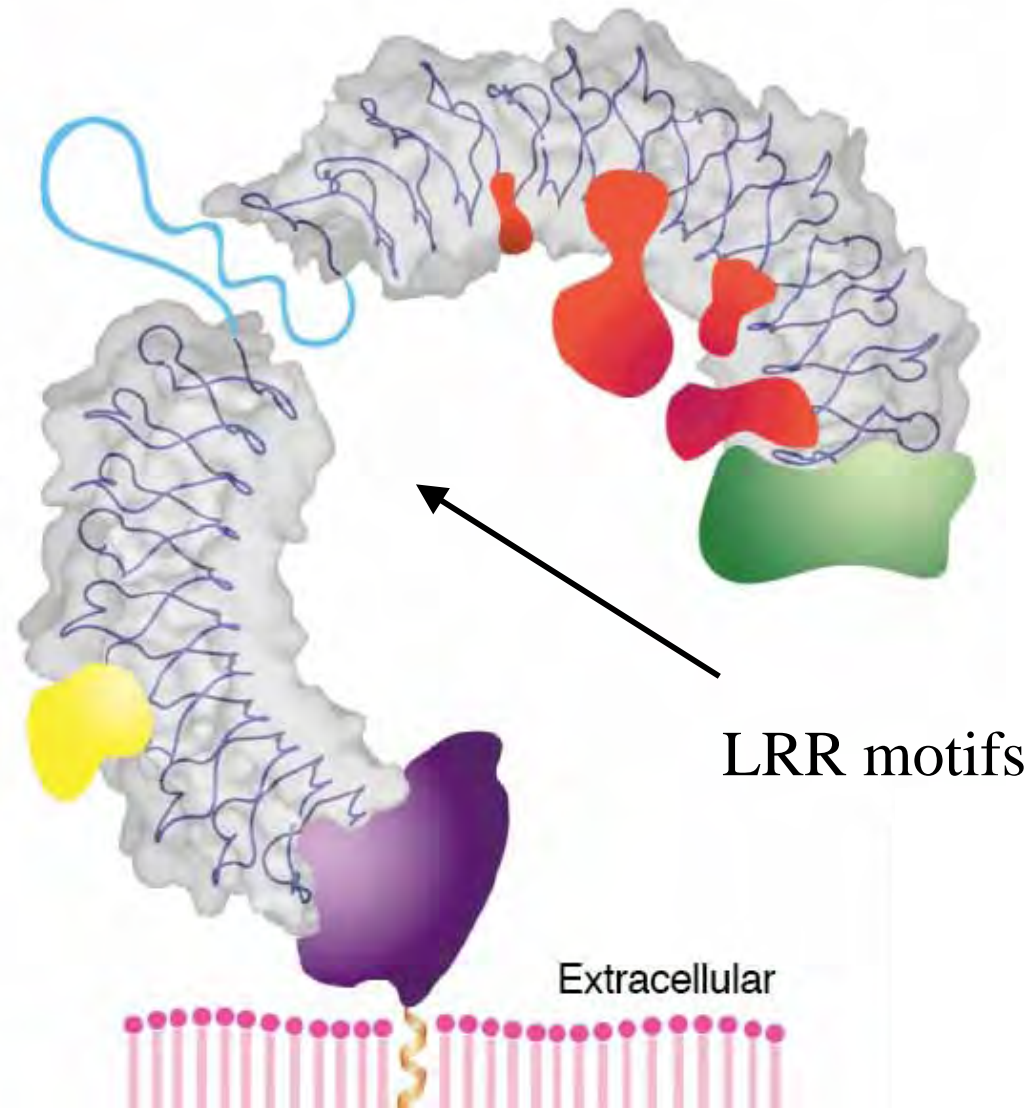
A-B: 1345 Å²

A-C: 540 Å²

B-D: 638 Å²

C-D: 1116 Å²

Ectodomain of TLR



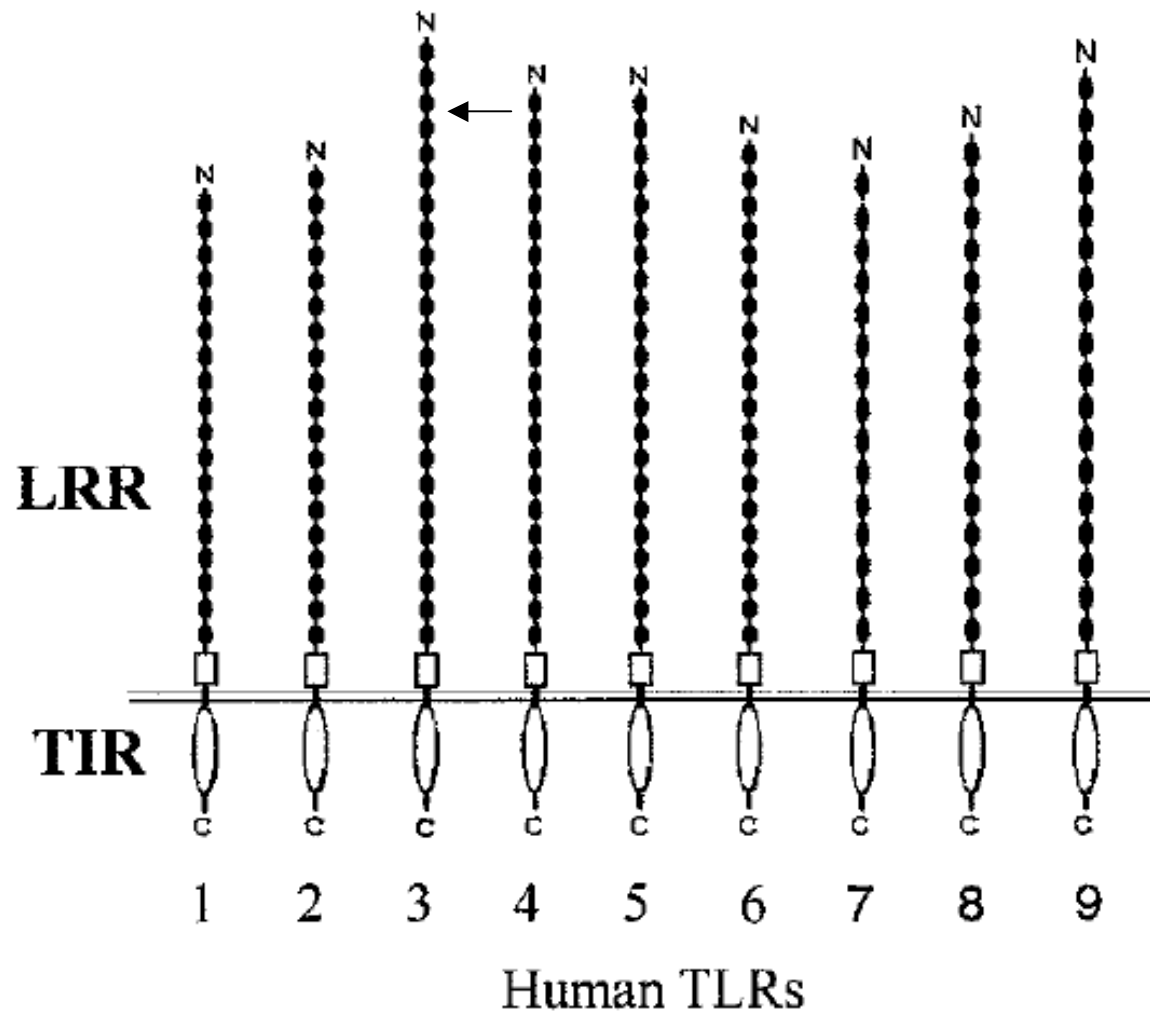
LRR motif (24 residues)

$xL^2xxL^5xL^7xxN^{10}xxL^{15}xxxxF^{20}xxL^{23}x$

L represents obligate hydrophobic residues including:
isoleucine, valine, methionine, and phenylalanine;

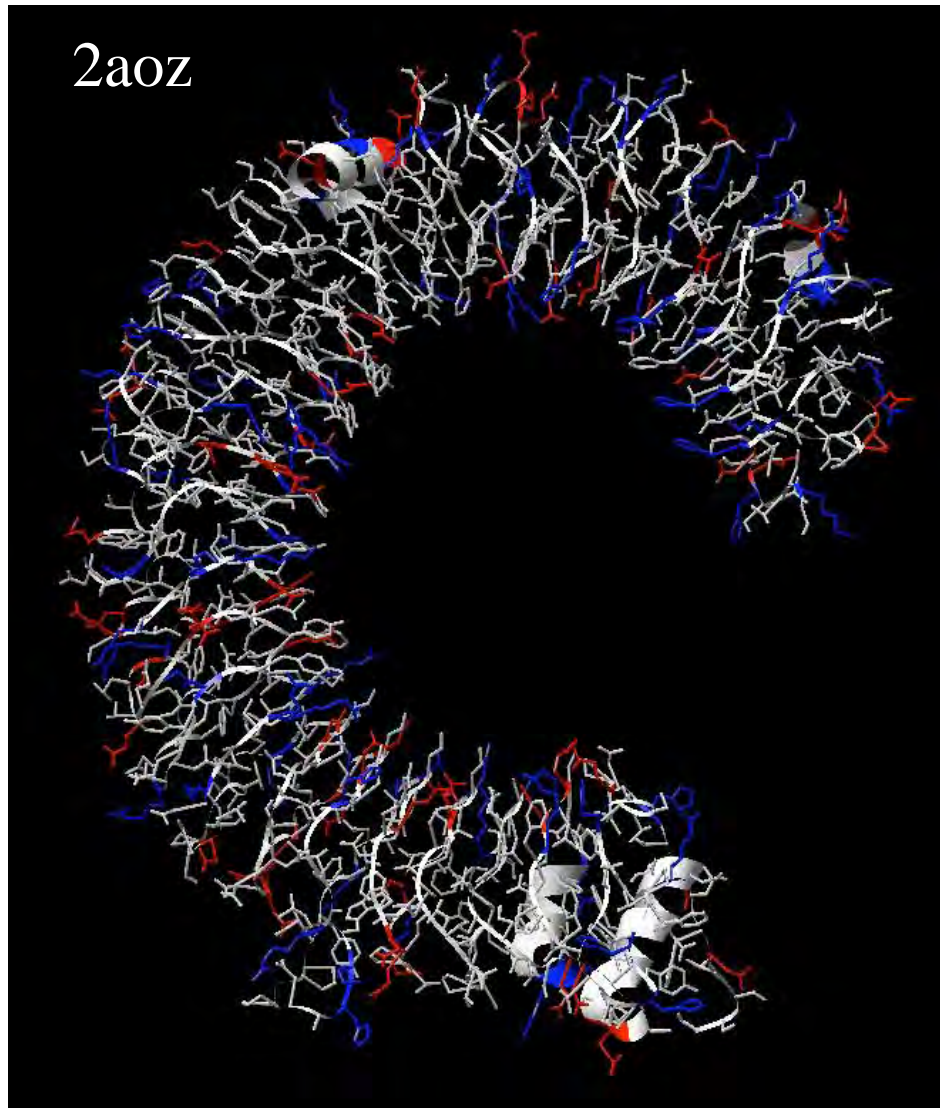
F is a conserved phenylalanine;

N is a conserved asparagine



19-25 tandem copies of LRRs in human TLRs

Ectodomain of TLR-3



Choe et al., 2005

Bell et al., 2005

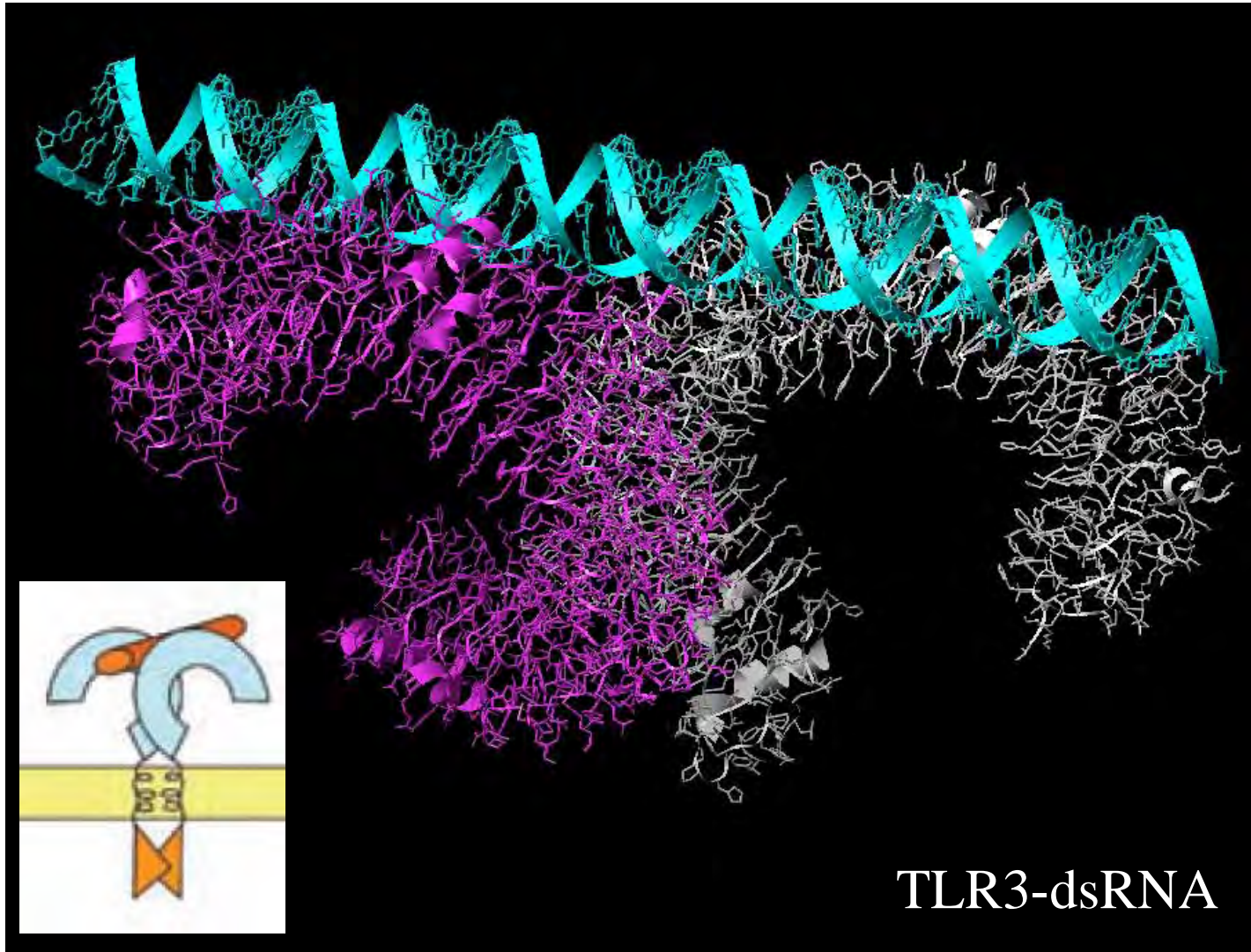
23 LRRs

Horseshoe shaped

Receptor-ligand interactions

- Using a multiscale docking procedure to develop TLR3 ectodomains/ds RNA structural complex.
- Interface surface area for TLR3 ectodomain dimer is small ($\sim 600 \text{ \AA}^2$).
- Ligand binding increase the stability of the receptor dimer?

TLR3 ectodomain dimer + dsRNA



Modeling TLR4 ectodomain

- Structure of TLR3 ectodomain is known.
- Sequence identity between TLR3 and TLR4 ectodomains is low (26%).
- Due to LRR motifs, a structure-based alignment can be used to align the two sequences.

Structure-based alignment

```
      100      110      120      130      140      150
SQTSLDVG FNTISKLEPEL CQKLEMLKVLNLQHNELSQLSDKTF AFQTNLTELHLMSNSIQKIKNNPFVKQHNLTITLDDL
-----STKNL DLSFNPLRHLGYSYFFSPELQVLDLSRCEIQTI EDGAYQSLSHLSTLIL
      55      60      70      80      90     100     110

      160      170      180      190      200      210      220      230
SHNGLSSTKLG TQVQLENLQELLSSNNKIQALKSEELDI FANSLK KLELSSNQIKEFSPGCFHAICRLFGFLF LNN
TGNPIQSLALGAFSGLS SLQKLVAVET--NLASLENFPIGHLH TLKELNVAHMLIQSFKLPEYFSNLTNLEHLDLSS
0      120      130      140      150      160      170      180

      240      250      260      270      280      290      300
VQLGPSL TEKLCLELANTSIRNLSLSNSQISTTSNTTFLGLKWTNLTMLDLSYN---NLNVVGND SFAWLHOLEY
NKI-QSIYCTDLRVLHQMP LLNLSLDLSL-NPMNFIQPGAFKEIRLHKLTLRNNFDSLNV MKTCIQGLAGIEVHR
      190      200      210      220      230      240      250

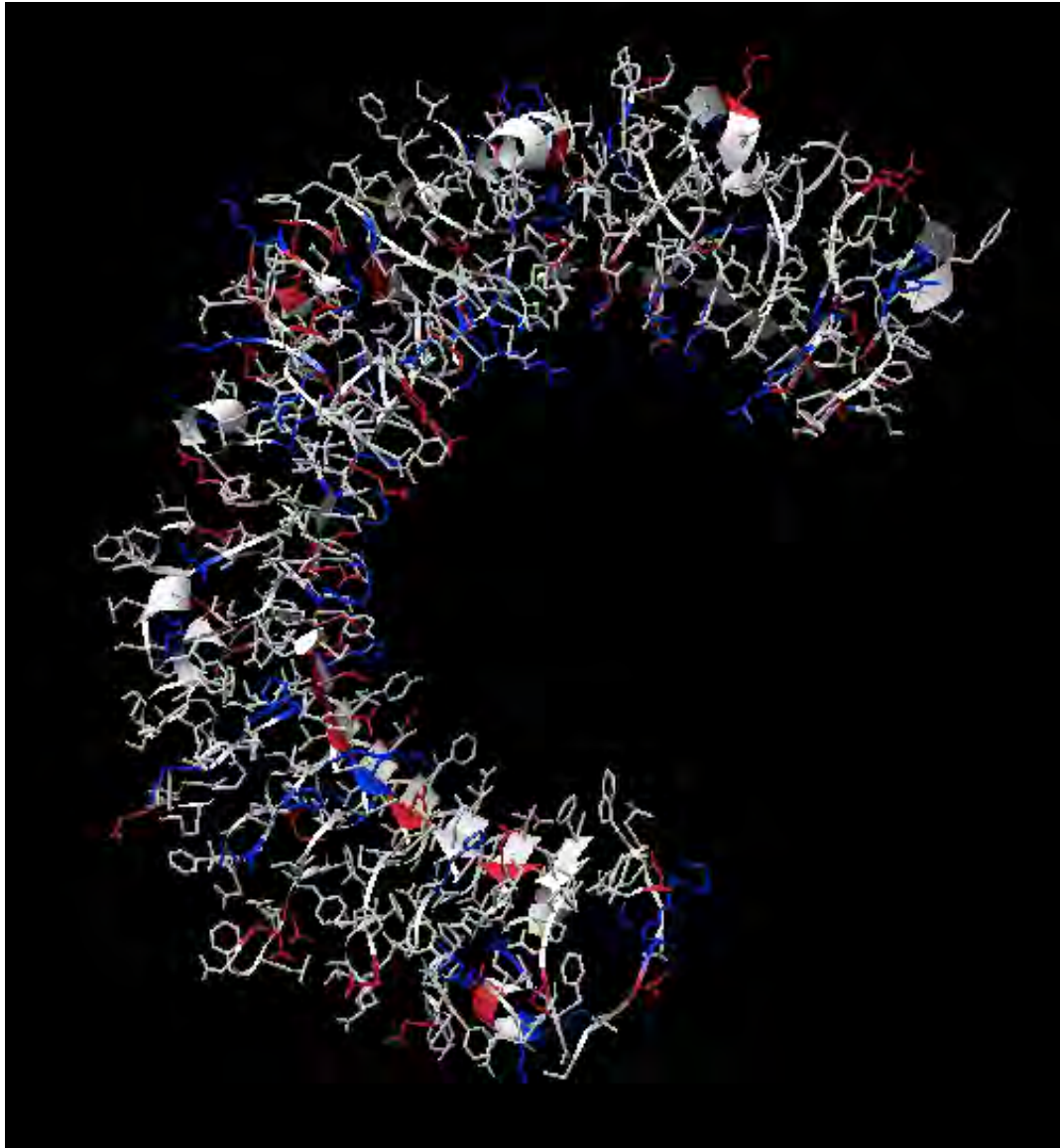
      310      320      330      340      350      360      370      380
FFLEYN-NIQHLFSHSLHGLE NVRYLNLKRSFTKQSI SLASLPKIDDFSFQWLKCLEHLN MEDNDIPGIKSNMFTGLI
LVLGEFRNEGNLEKFDKSALEGLCNLTIEEFRLAYLD---YLLDDIIDLFNCLTNVSSFSLSVSV--TIERVKDFS YNE
260      270      280      290      300      310      320      330

0      390      400      410      420      430      440      450
NLKYL SLSNSFTSLR TLTNET FVSLAHSPLHILNLTKNKISKIESDAFSWLCHLEVL DLGLNE-IGQELTGQEW RGL
GWQHLELVN C-----KFQGFPTLKLPSLKR LTF TSN--KGNAFSEVDLH SLEFLDLSRNGLSFKGCCSQSDFGT
      340      350      360      370      380      390      400

      460      470      480      490      500      510      520      530
ENIFEIYLSYNKYLQLTRNSFALVHSLQRLMLRRVALKNVDSSSPFPQPLNLTILDL SNNNIANINDDMLEGLEKLEIL
TSLKYLDLSFN-GVITMSSNFLGLRQLEHLDFQHSN-LKQ MSEFSVFLSLNLIYLDI SHTRVAFNGIFNGLS SLEVL
00      410      420      430      440      450      460      470

      540      550      560      570      580      590      600      610
DLQHNNLARLWKHANPGGPIYFLKGLSHLHILNLESNGFDEIPVEVFKDIFELKIIDLGLN NLNLTLPASVFNNQVSL
KMA GNS-----FQENFLPDIFTELPNLTFLLDLSQCQLEQLSPTAFNSLSSLQVLM SHNFFSLDTFPYKCLNSL
      480      490      500      510      520      530      540

      620      630      640      650
KSLNLQKNLITSVEKKVFGPAFNLTELDMRFNPFDCESIAWFVNWINETH TNIPELSSH YLCNTPPHYHGFPVRLFD
QVLDYSLNHIMTSKKQELQHFPSSLAFLNLTQNDFACTCEHQSFQWIKDQRQLLVEVERMECATPSDKQGPVLSLNI T
      550      560      570      580      590
```



TLR-4
ectodomain

(Bell et al., 2003)

