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 ~2 Å 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.

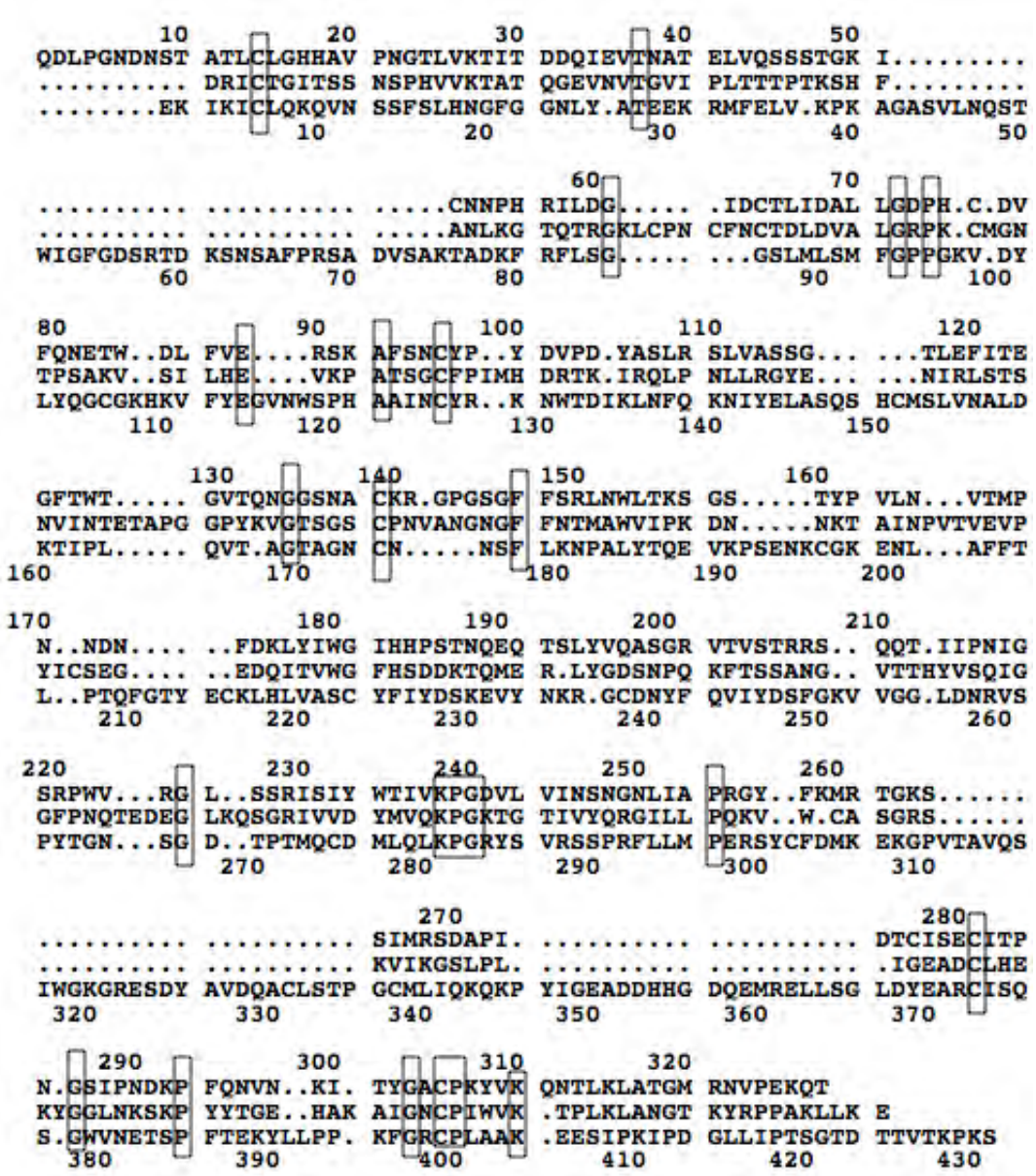
Hemagglutinin (HA)

- 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.

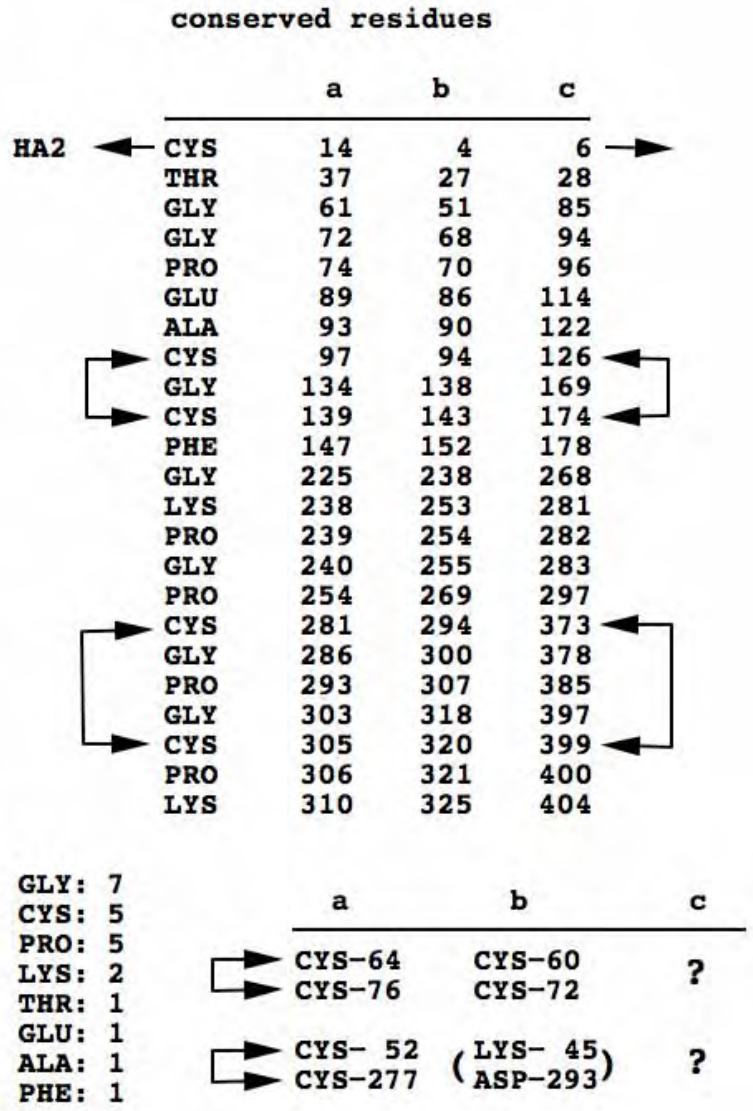
HA (continue)

- 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.

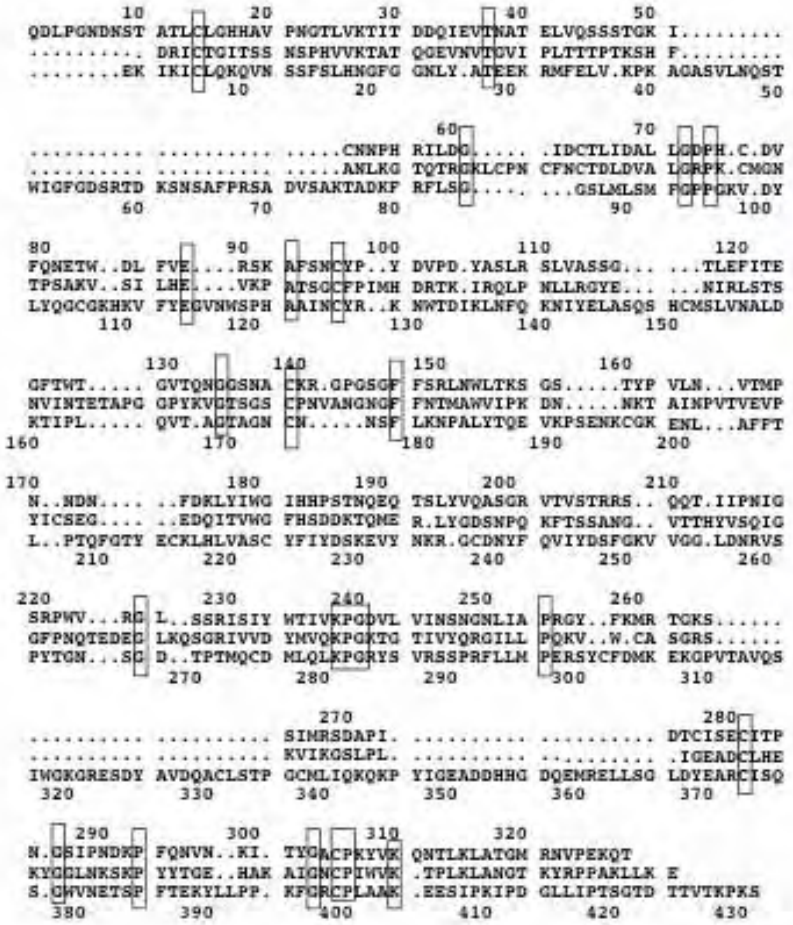
3-way alignment

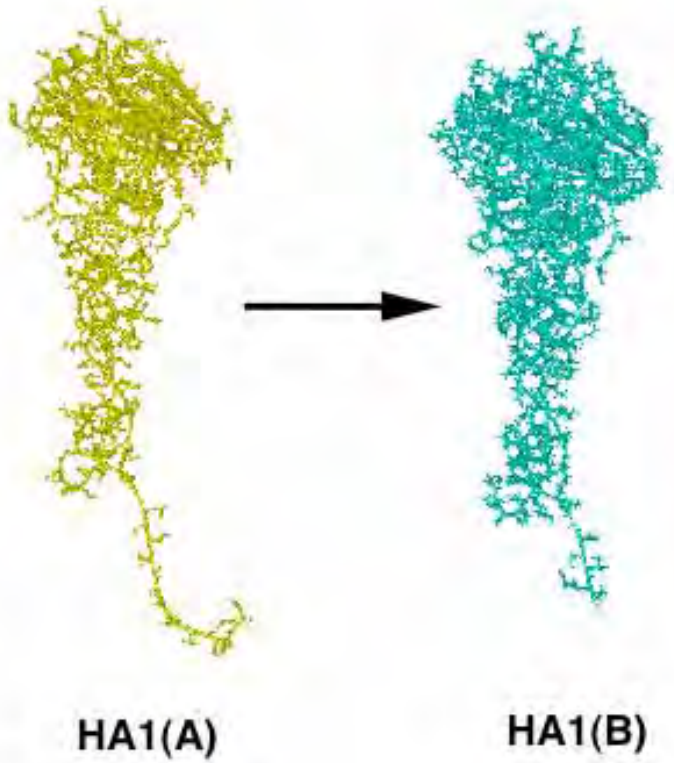


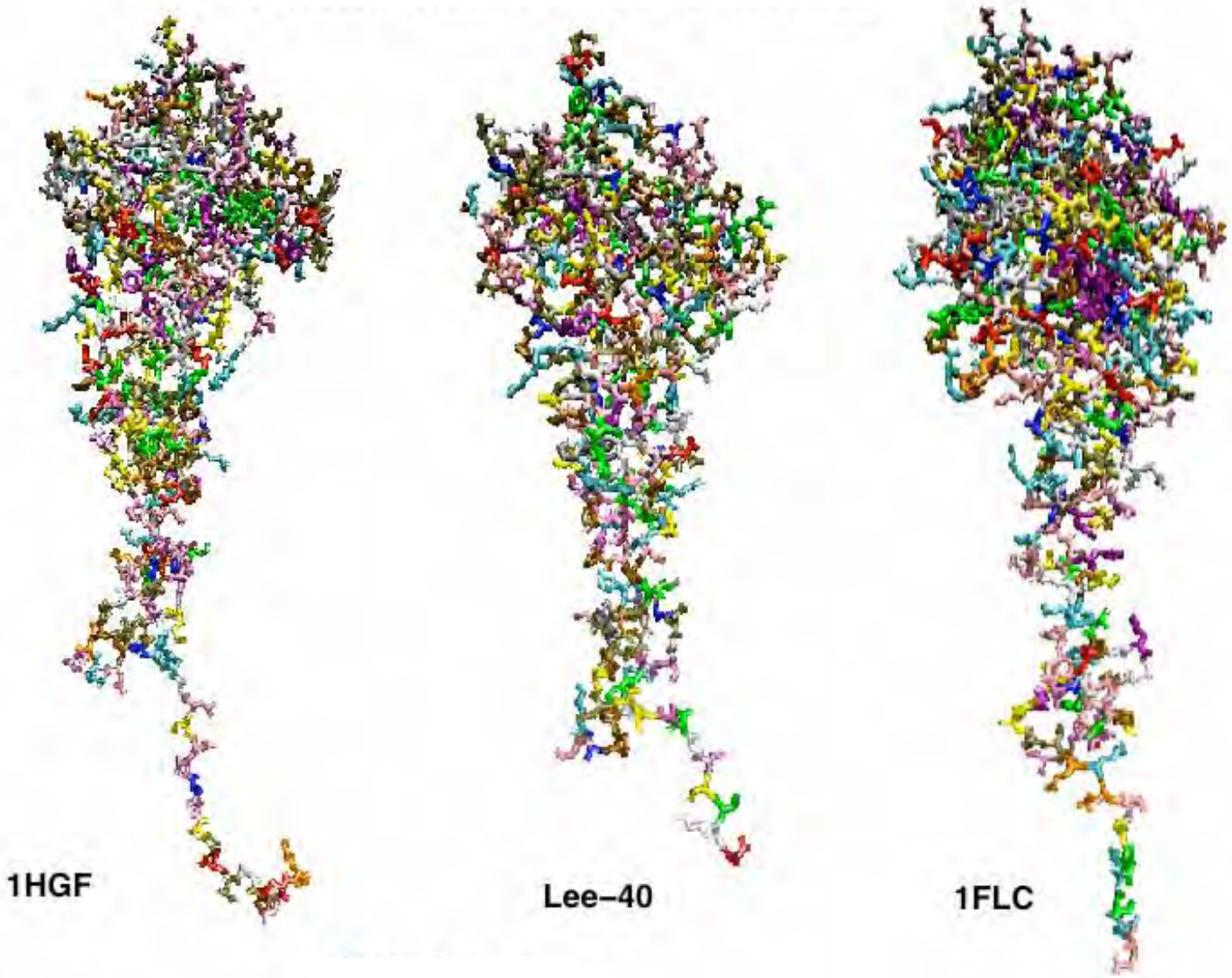
23 conserved residues



Model flu-b HA1 structure






1HGF
Lee–40
1FLC

Model validation

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

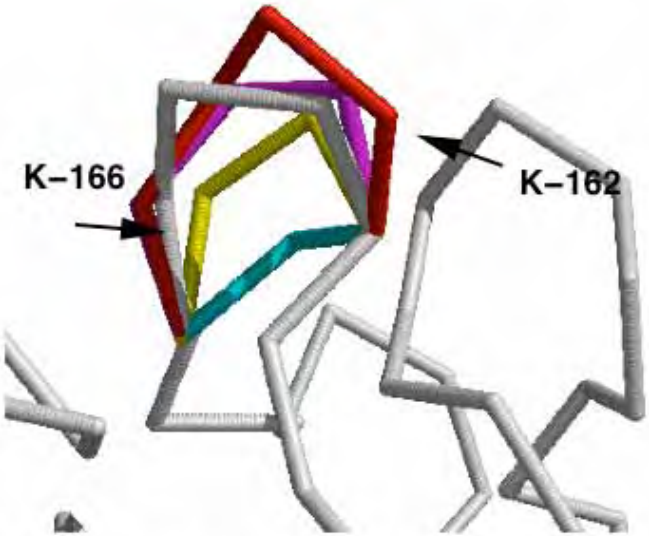
Quality of the model

- 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-DNNKIA
B/Bonn/43           VPK---NKIA
B/Osaka/70          VPK-N-NKIA
B/Finland/85        VPKNNNNKIA
```



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).

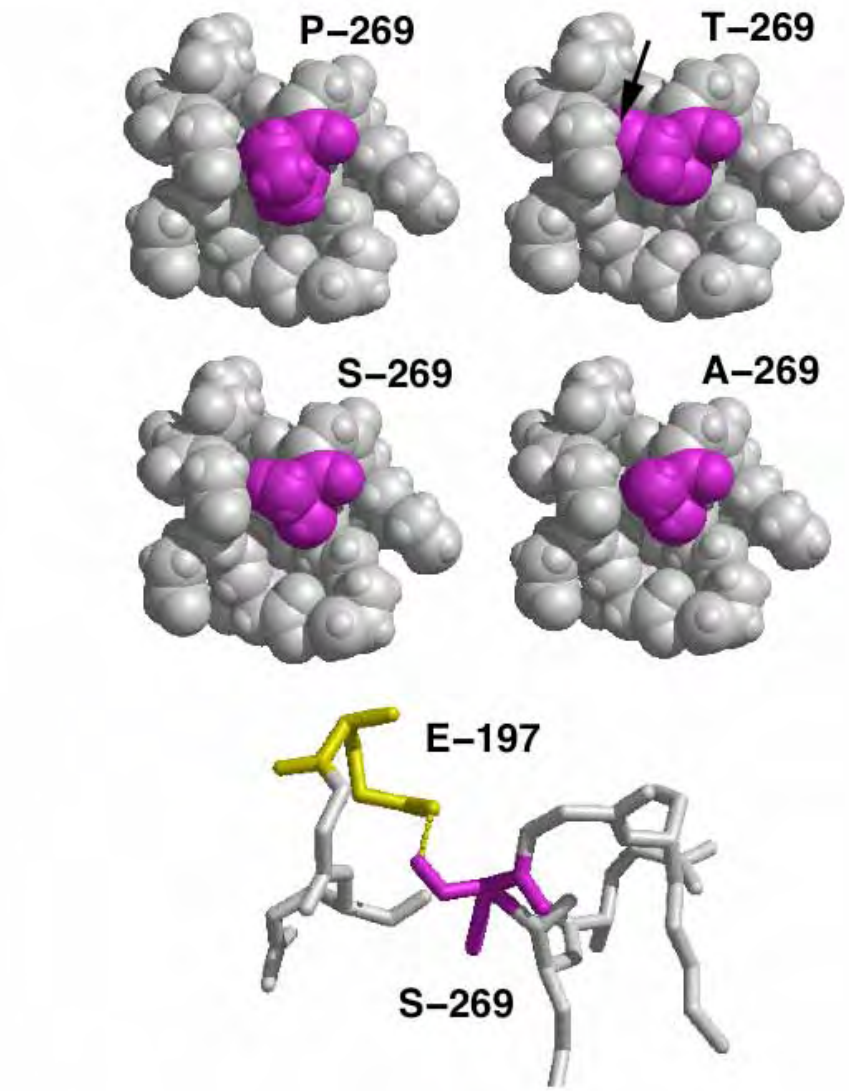
1 nucleotide change

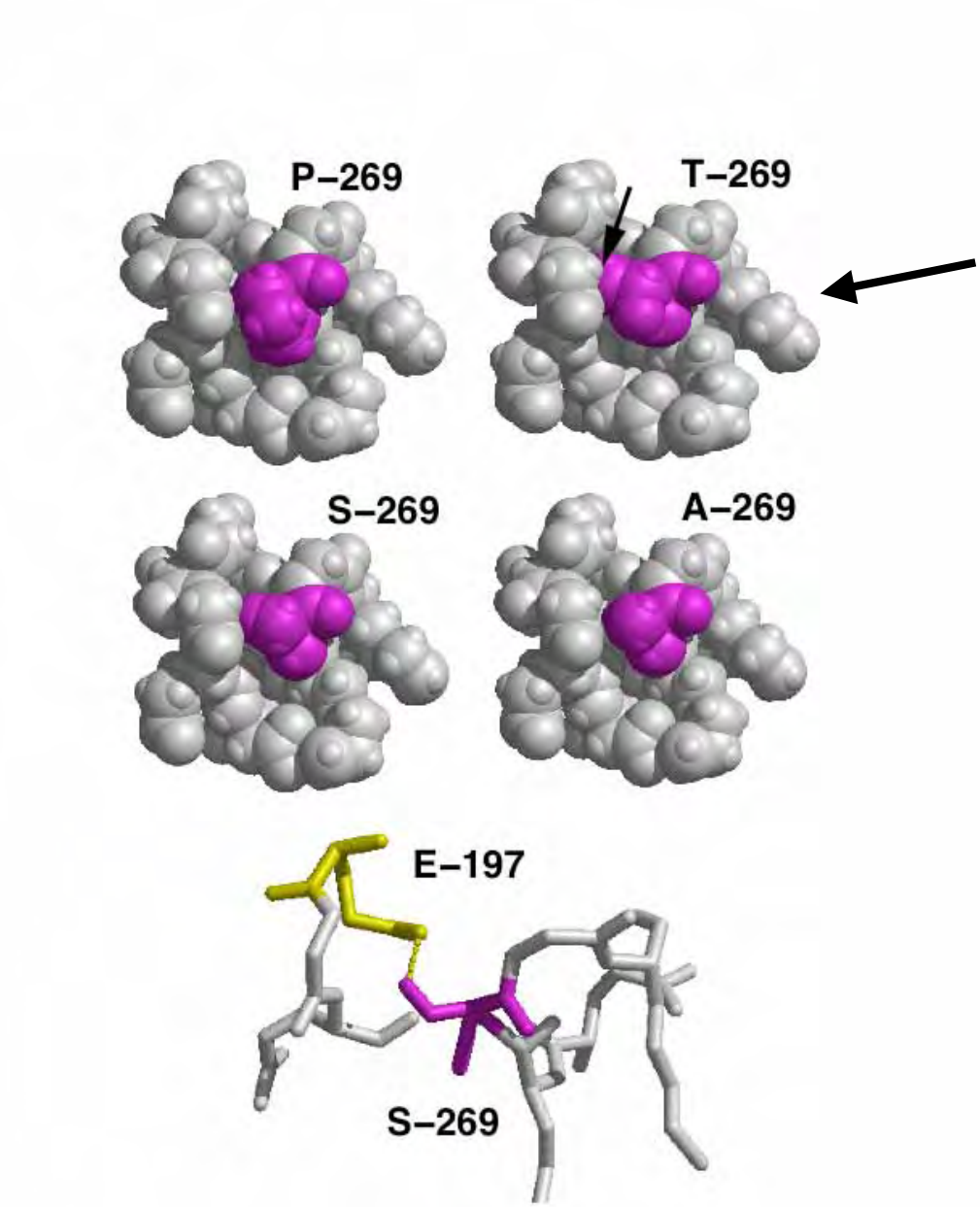
PRO: CCU, CCC, CCA, CCG

SER: Ucu, Ucc, Uca, Ucg
THR: Acu, Acc, Aca, Acg
ALA: Gcu, Gcc, Gca, Gcg

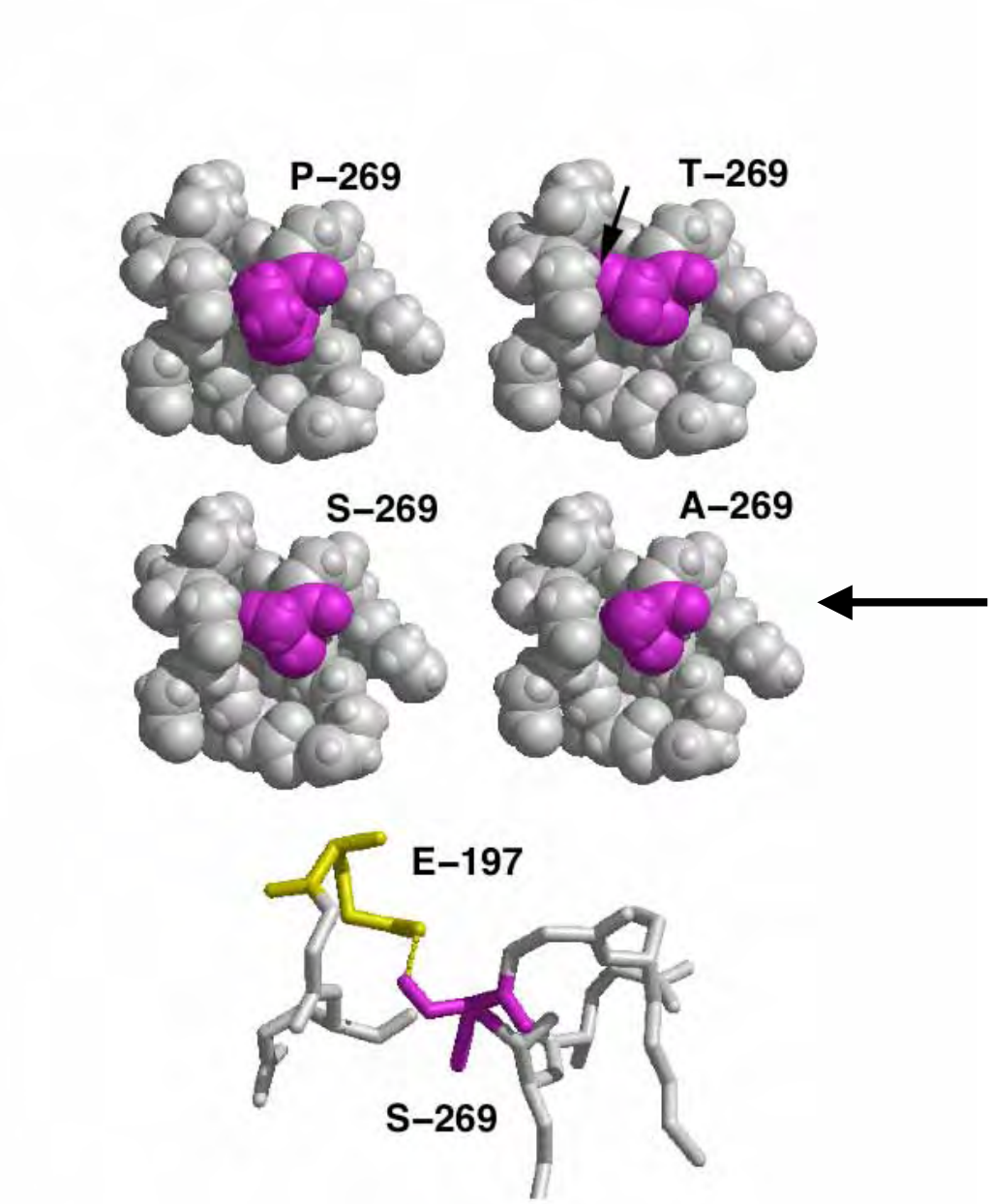
LEU: cUu, cUc, cUa, cUg
HIS: cAu, cAc
GLN: cAa, cAg
ARG: cGu, cGc, cGa, cGg

Different amino acid types at 269

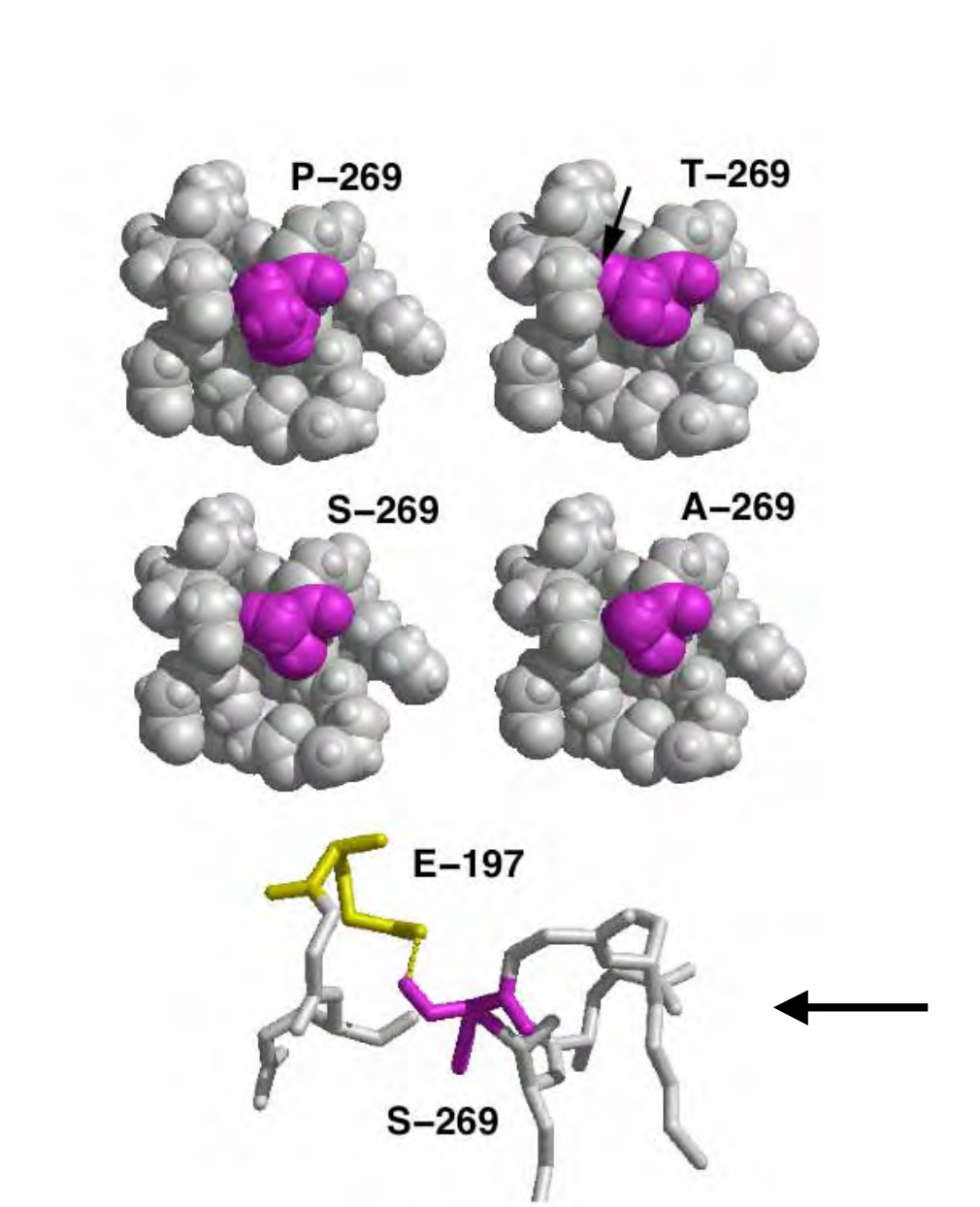




T-269 interfere with G-198 and E-199. Therefore, Thr, Leu, His, Gln or Arg are all unfavorable at this position.



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

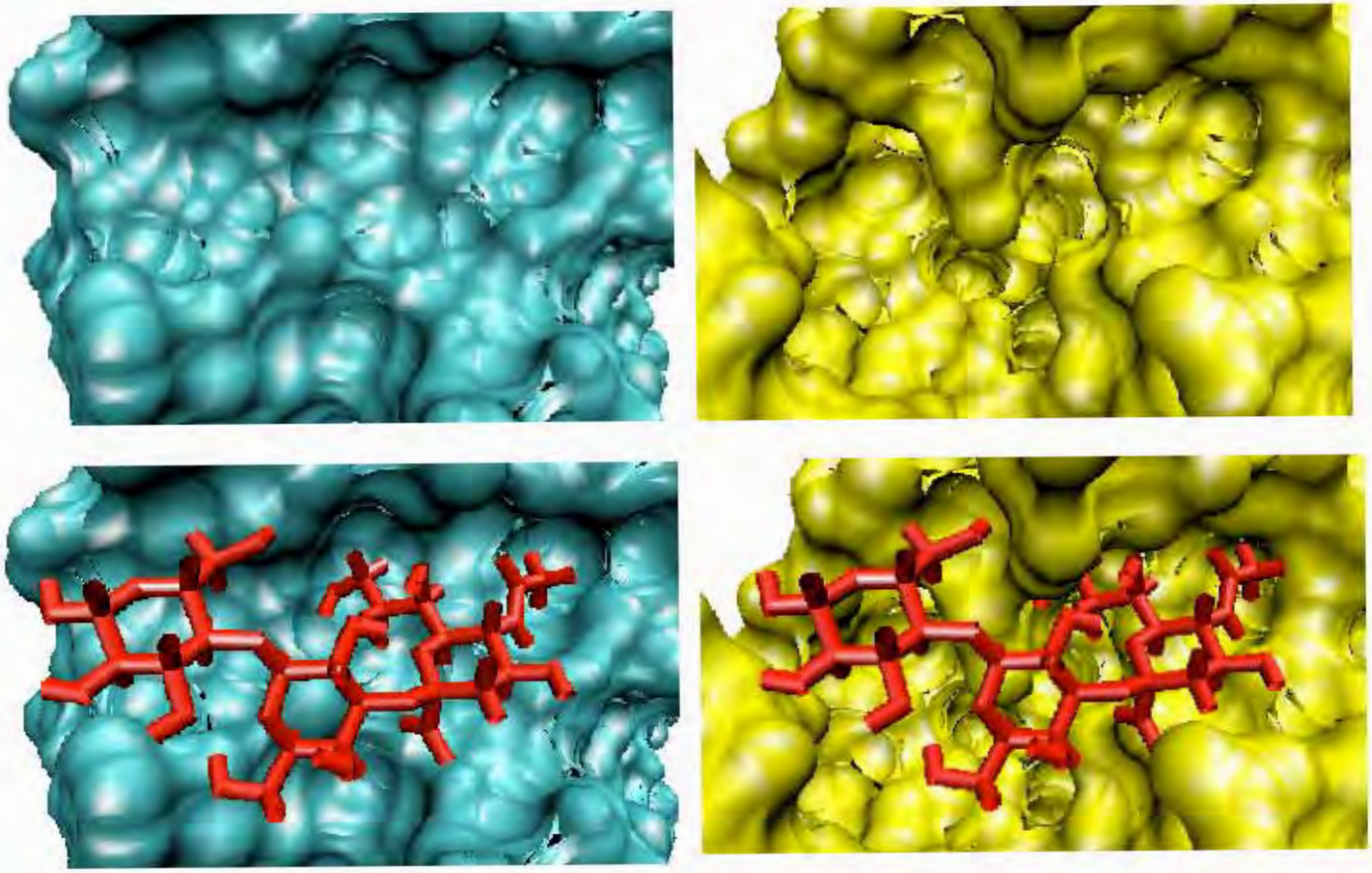


A H-bond between S-269 and E-197 stabilizes the structure

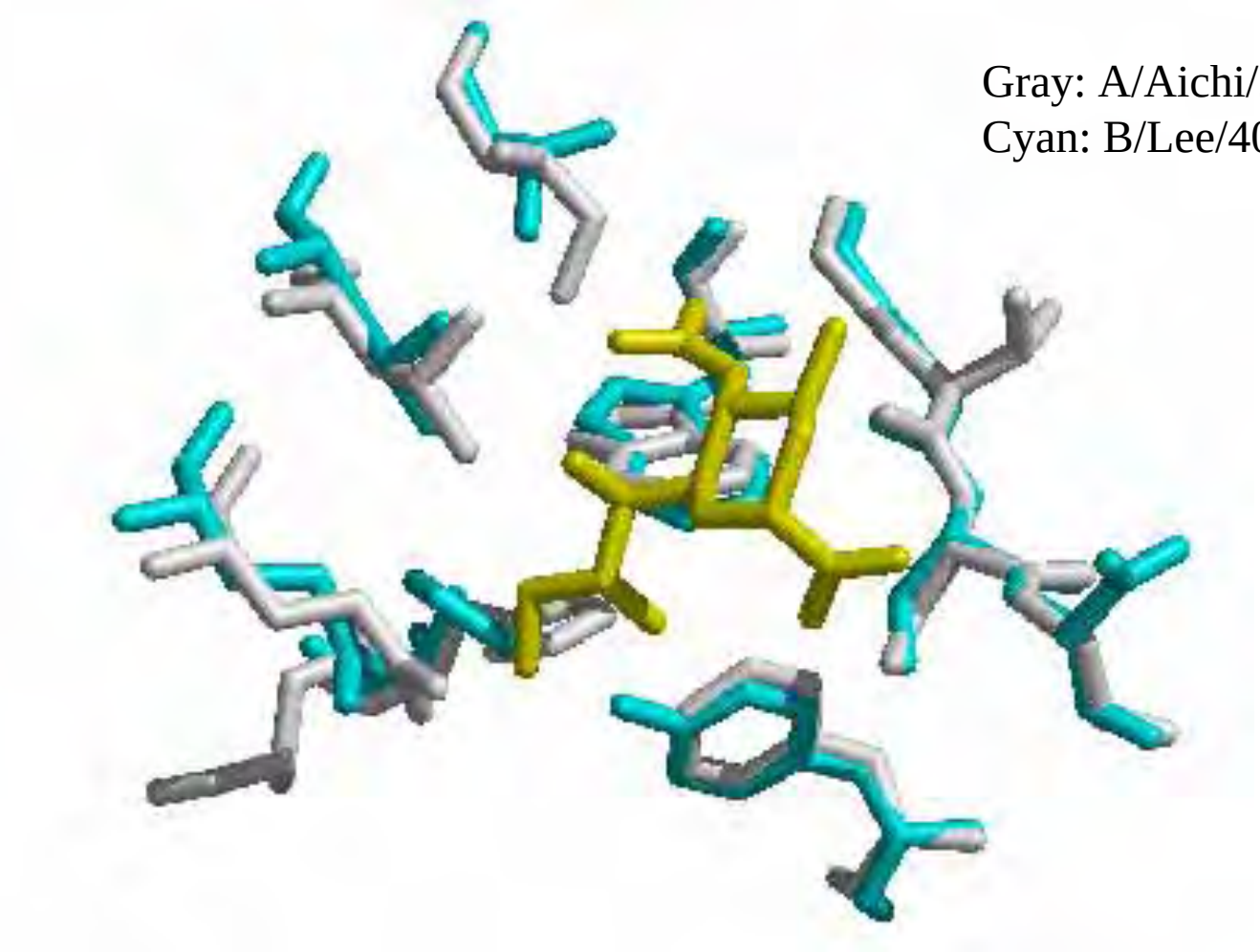
Receptor binding

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.

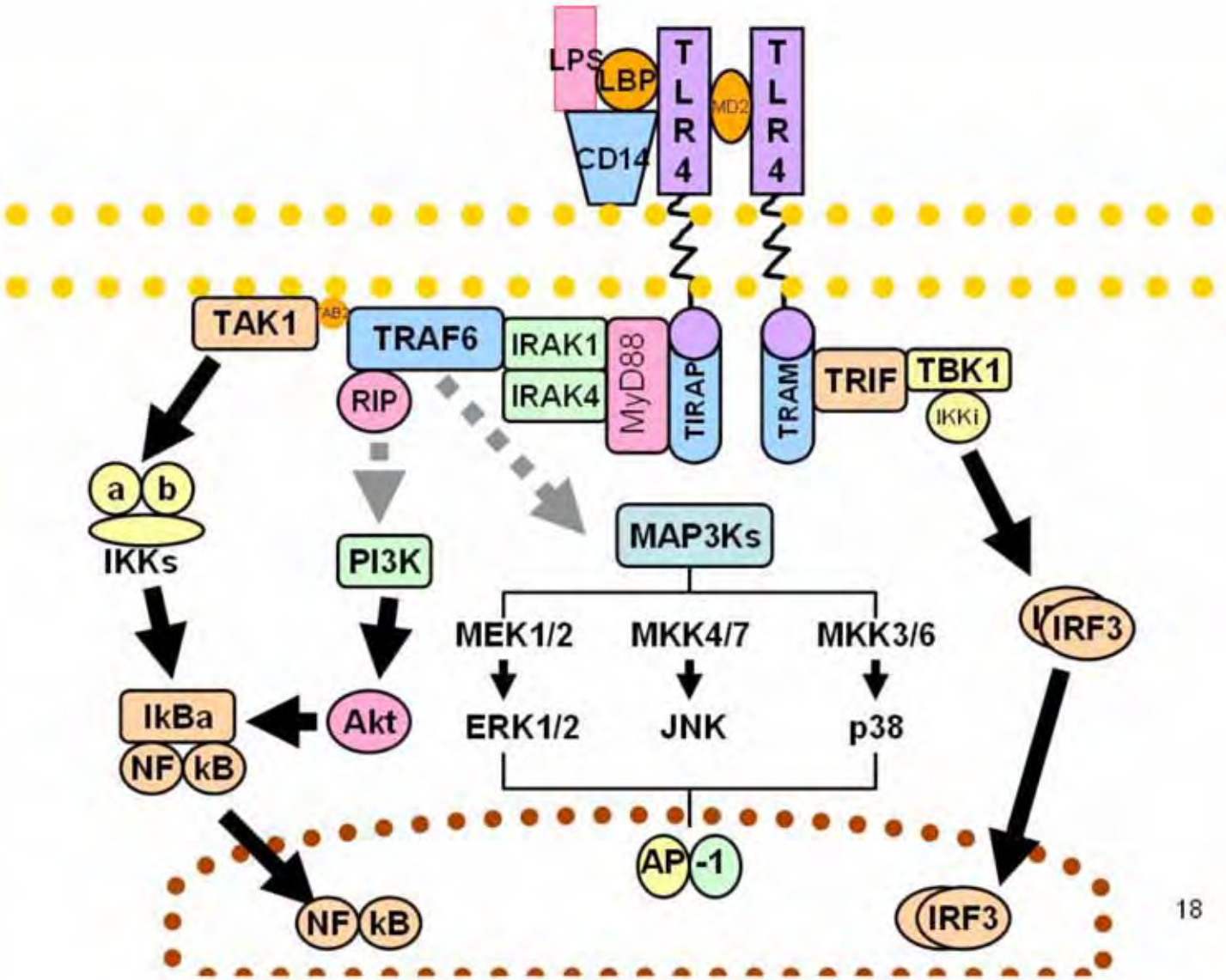
- Interface surface area:

  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.

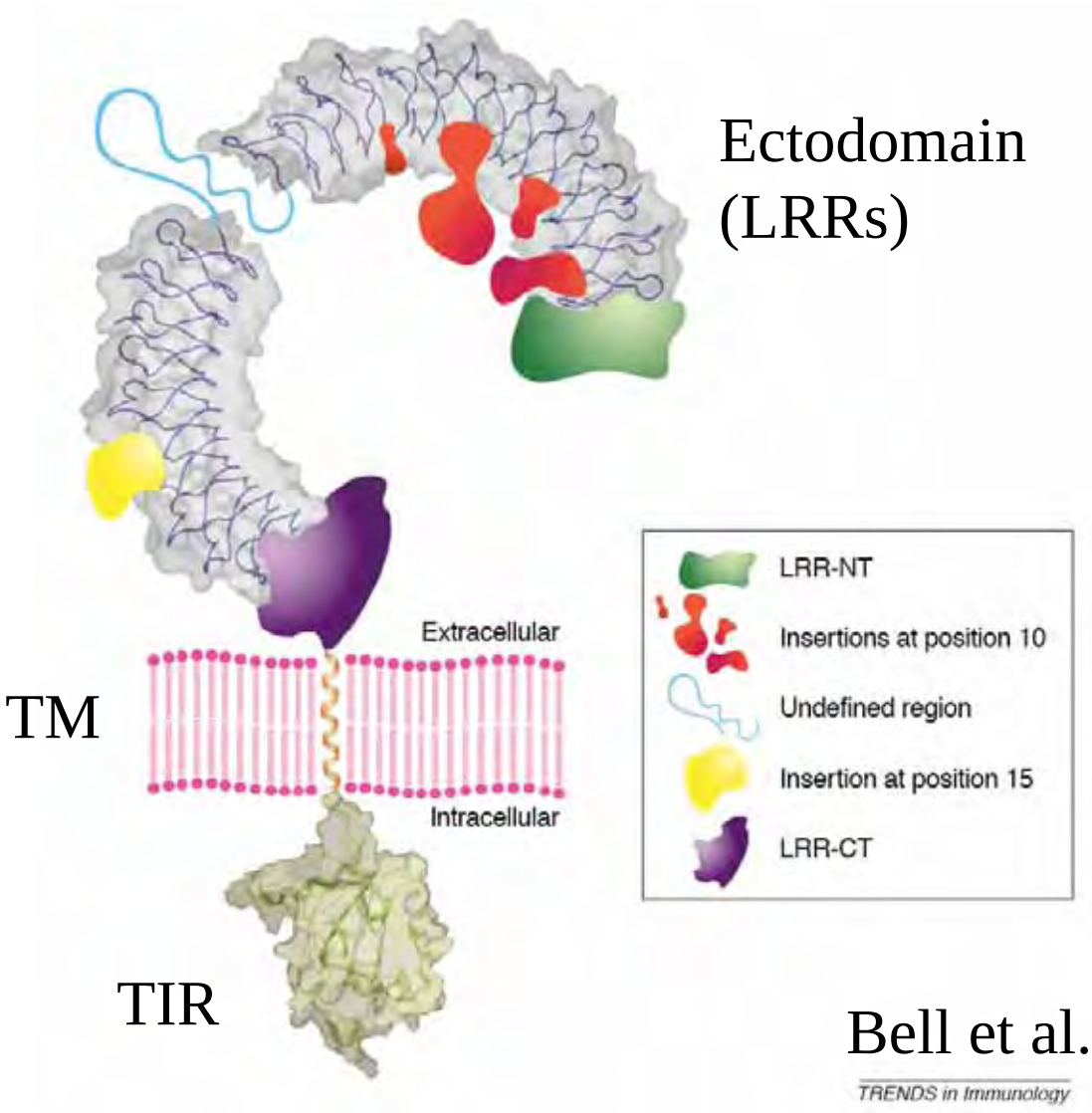
Signaling pathway of toll-like receptor



Toll like receptor

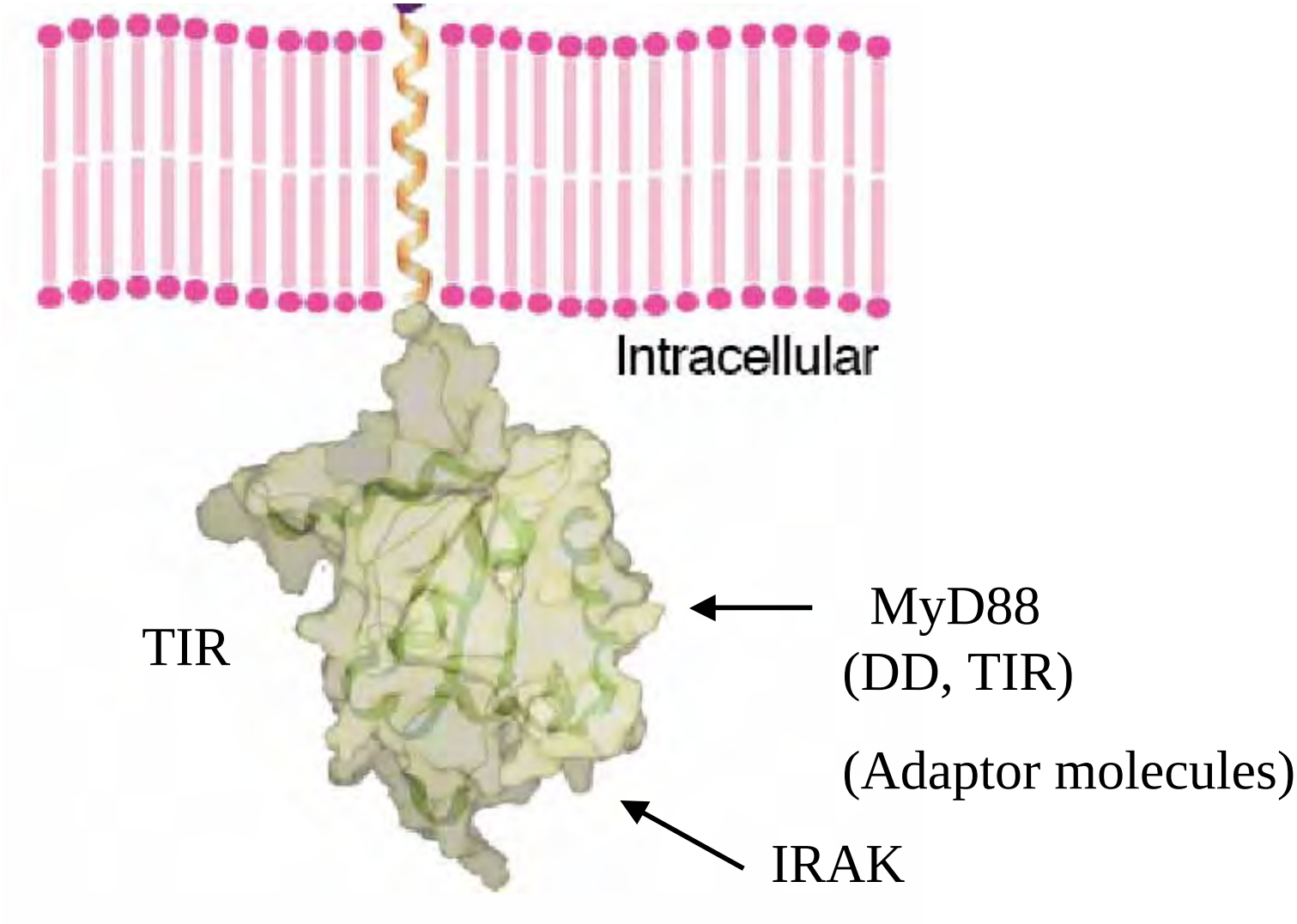
- Dunne and O'Neill www.stke.org/cgi/content/full/sigtrnas;2003/171/re3.
- Takeda, et el., 2003 Annu Rev Immmunol 21: 335-376.
- http://en.wikipedia.org/wiki/Toll-like_receptor

Structure of TLR

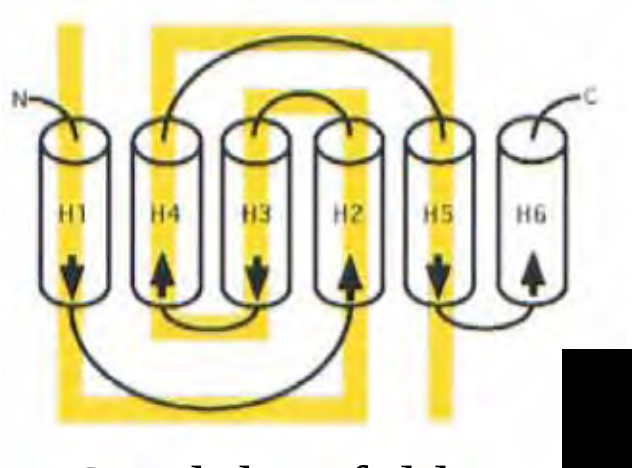


Bell et al., 2003

TRENDS in Immunology

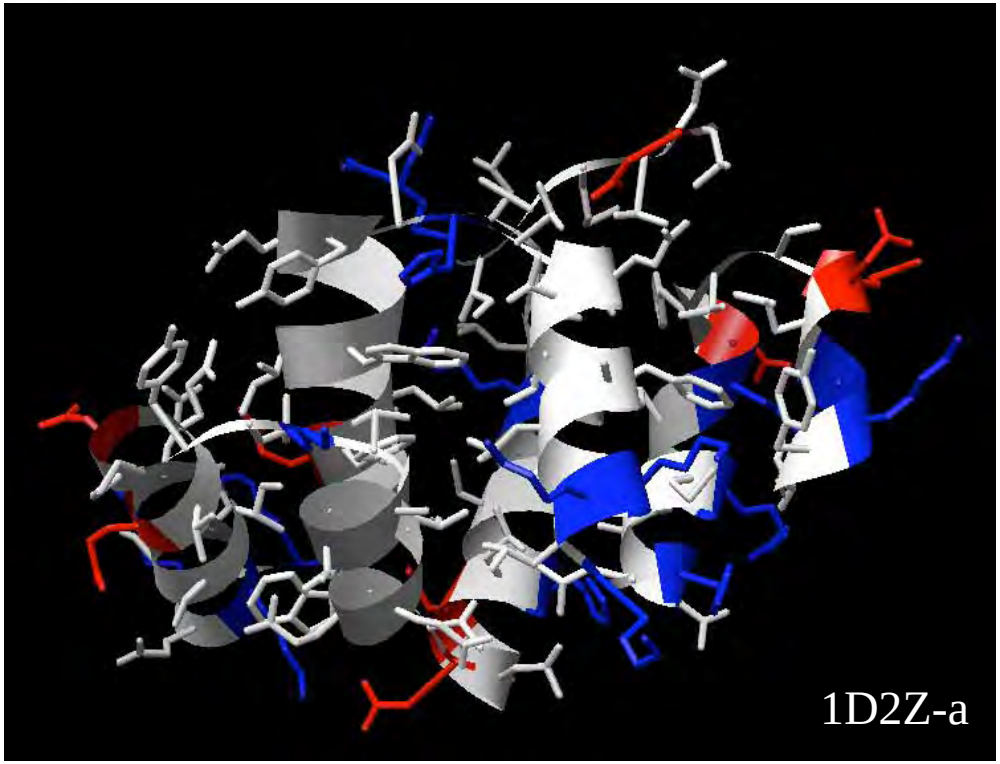

Intracellular
TIR
MyD88
(DD, TIR)
(Adaptor molecules)
IRAK

Death domain (DD)



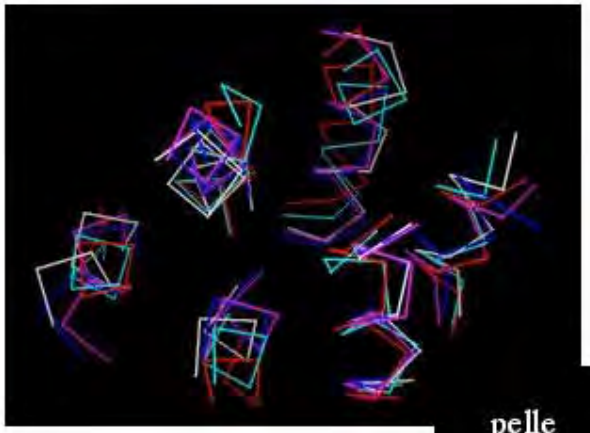
Greek key fold

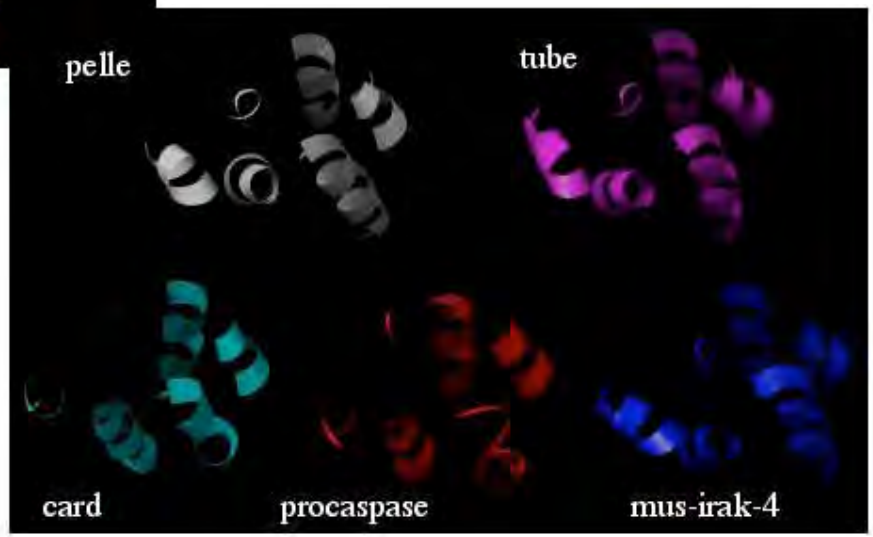
Pelle DD



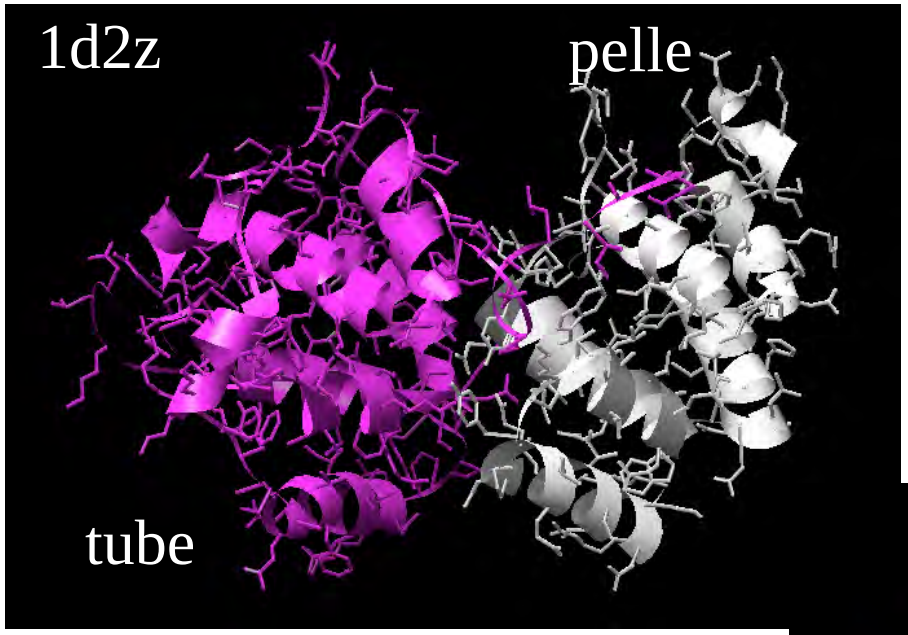
DD is a structural motif

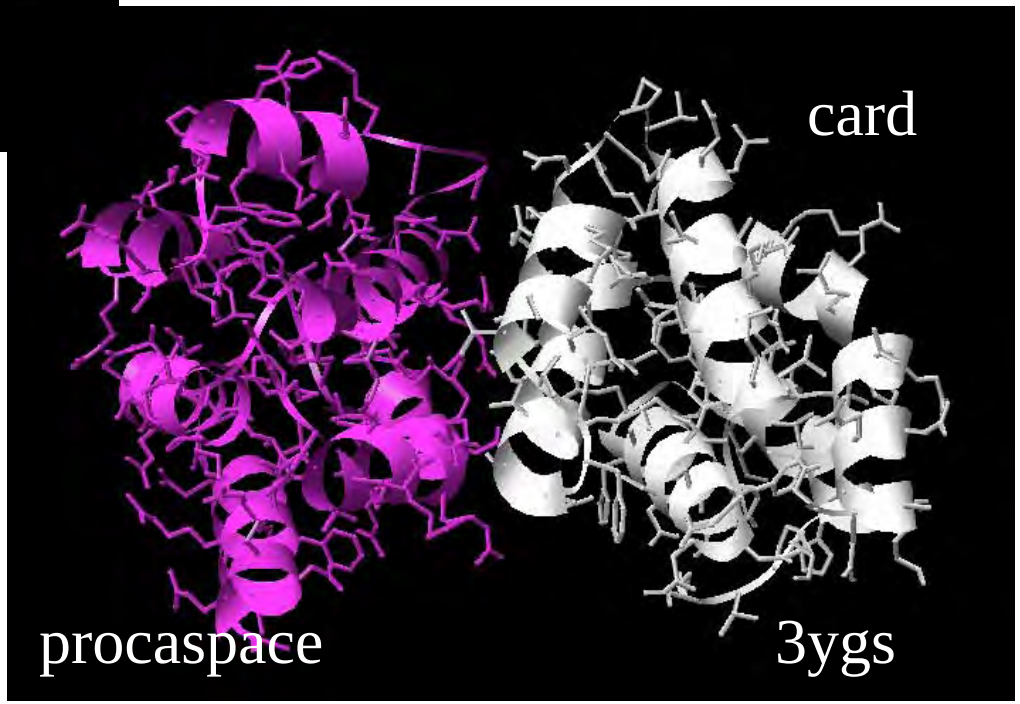
	pelle	tube	card	procaspace	mus-irak-4
pelle	0.00	1.24	2.06	2.12	1.21
tube	17.3% (5.3%)	0.00	1.98	1.93	0.90
card	14.6% (7.1%)	17.8% (6.9%)	0.00	1.11	2.20
procaspace	20.2% (8.9%)	18.5% (7.3%)	19.4% (18.4%)	0.00	2.19
mus-irak-4					0.00



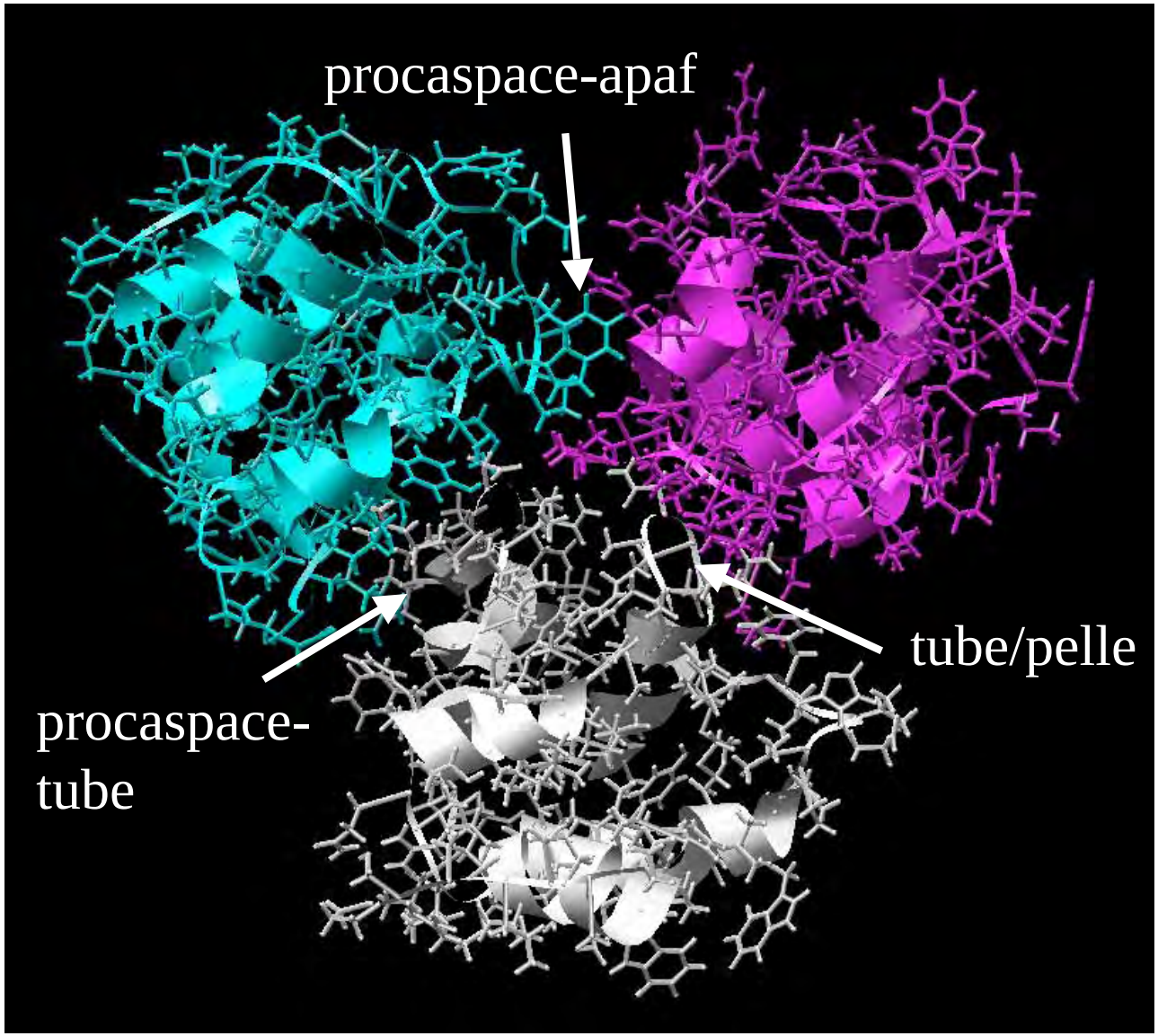


crystal contacts



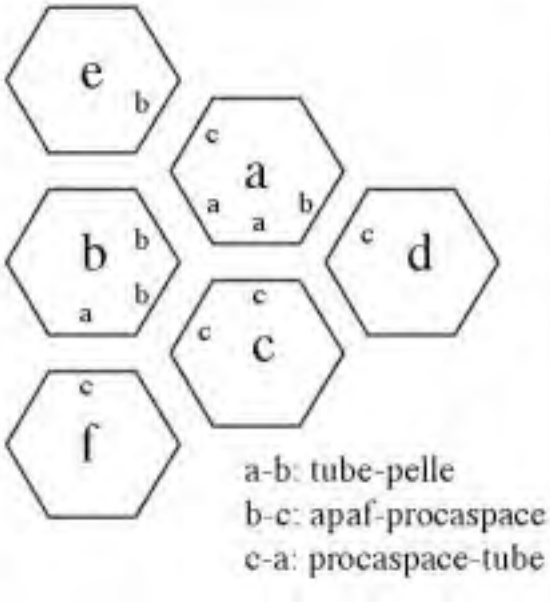


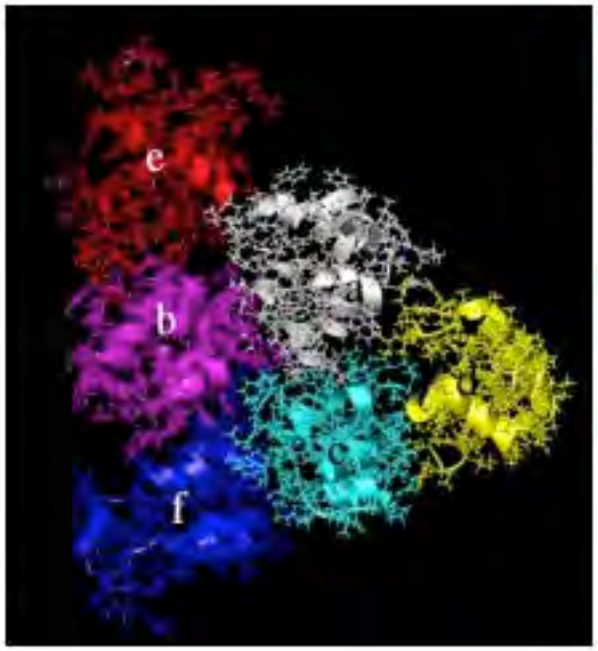
3-mer model

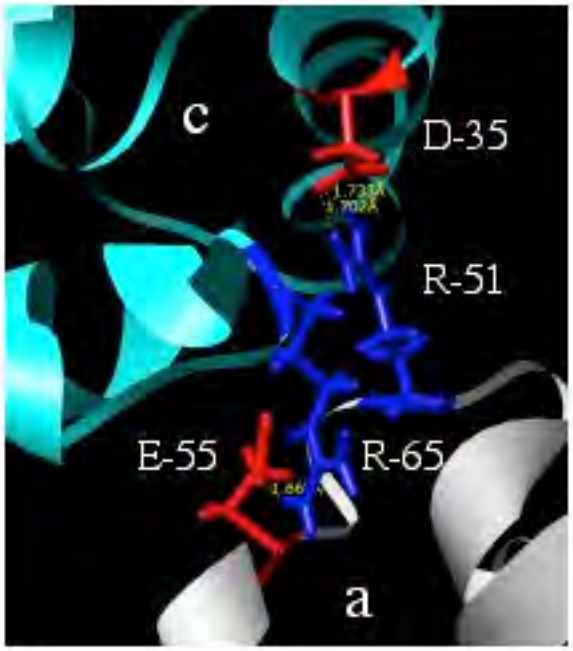


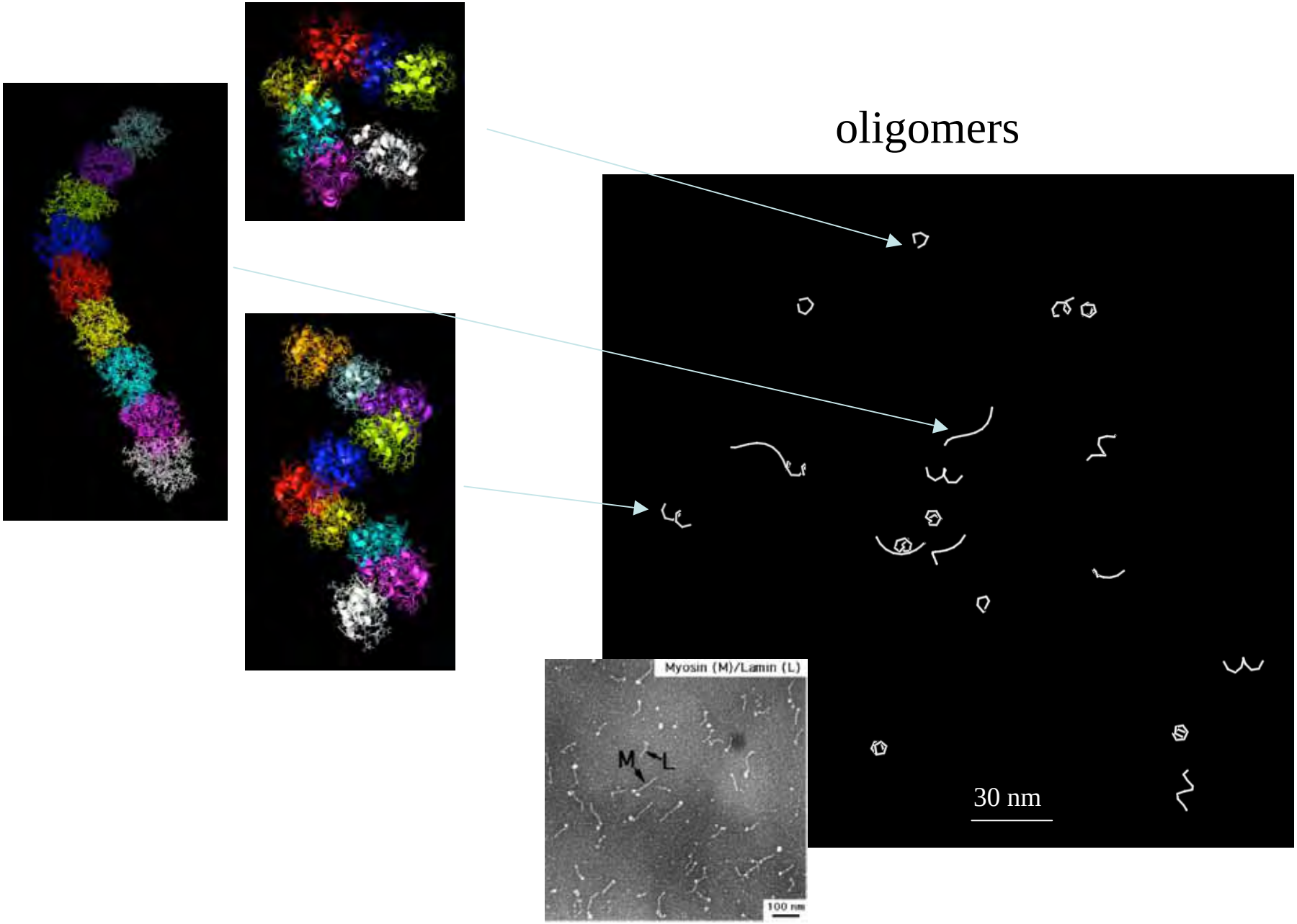
6-mer model

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



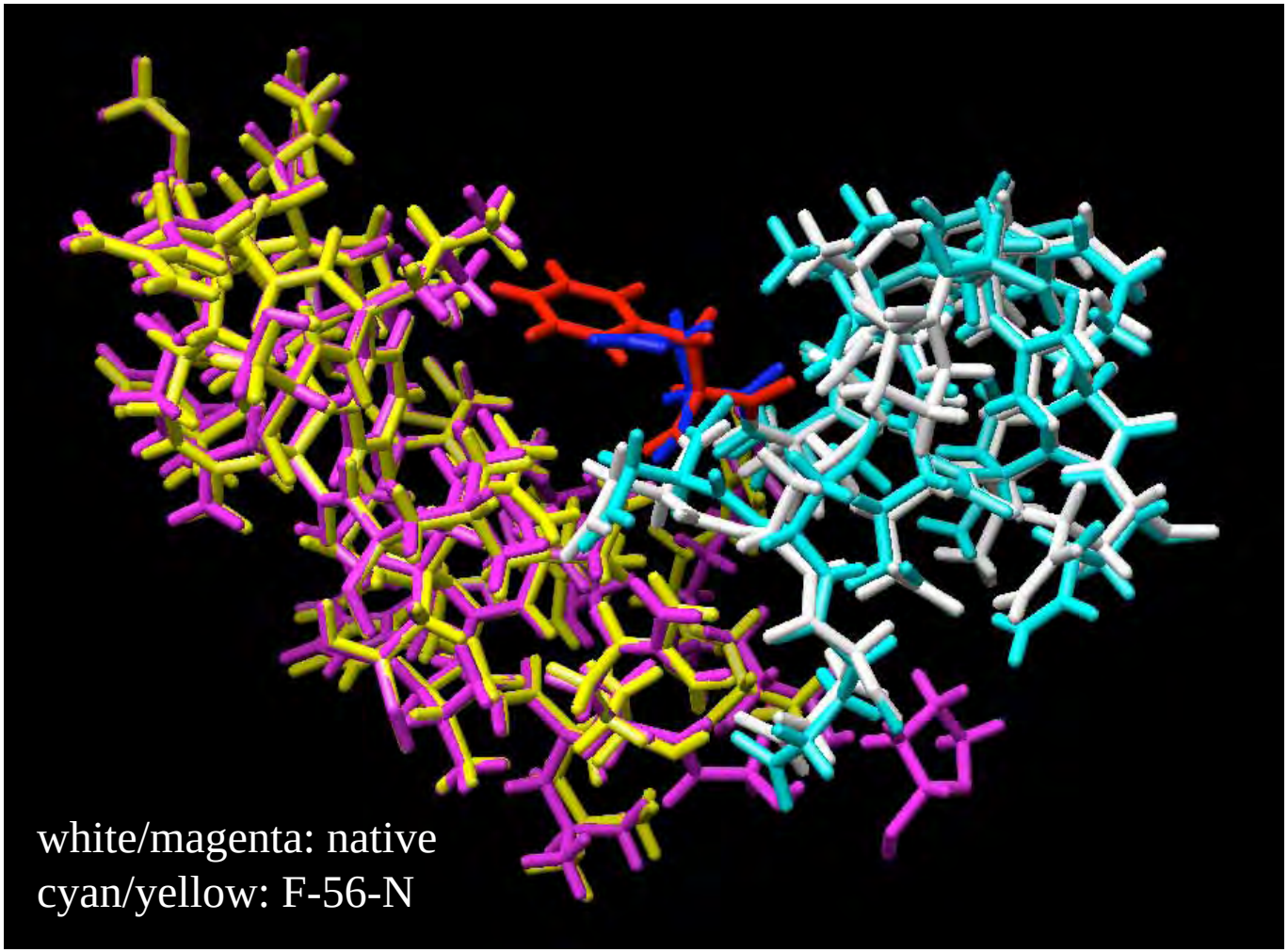





oligomers
Myosin (M)/Lamin (L)
M
L
100 nm
30 nm

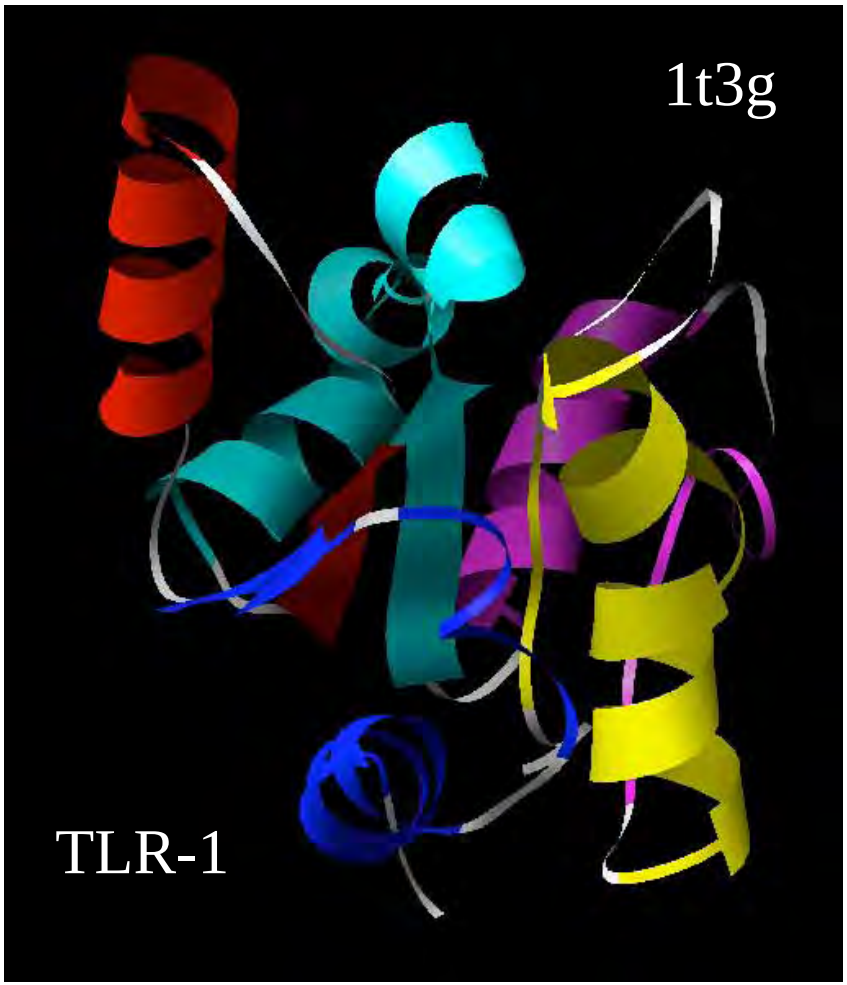
F-56-N mutation prevents dimerization of MyD88 DD. (Burns et al., 1998)

F-56-N mutation



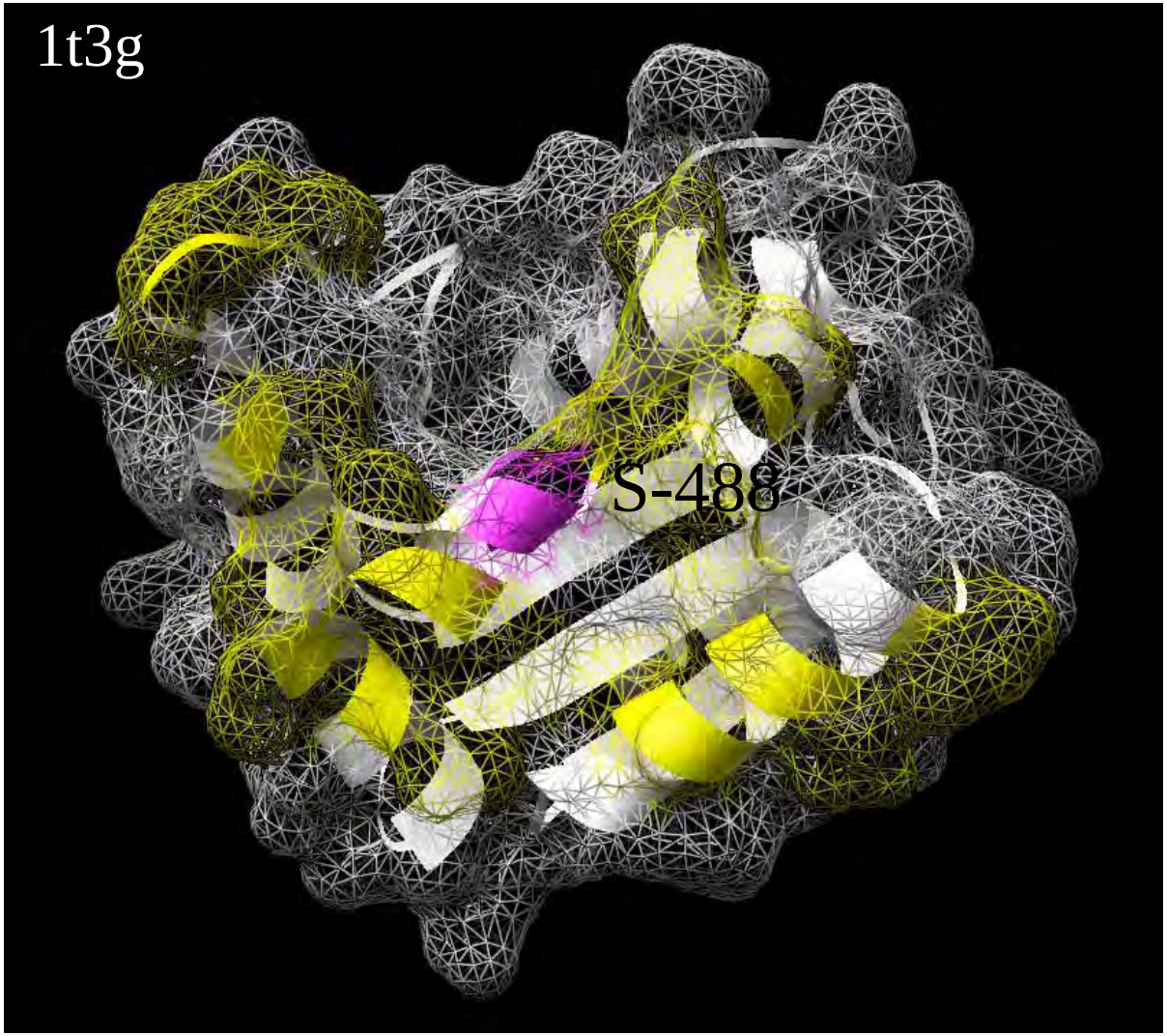
Loss 125 Å^2 of interface area due to mutation.

TIR: Toll/Interleukin-1 receptor domain

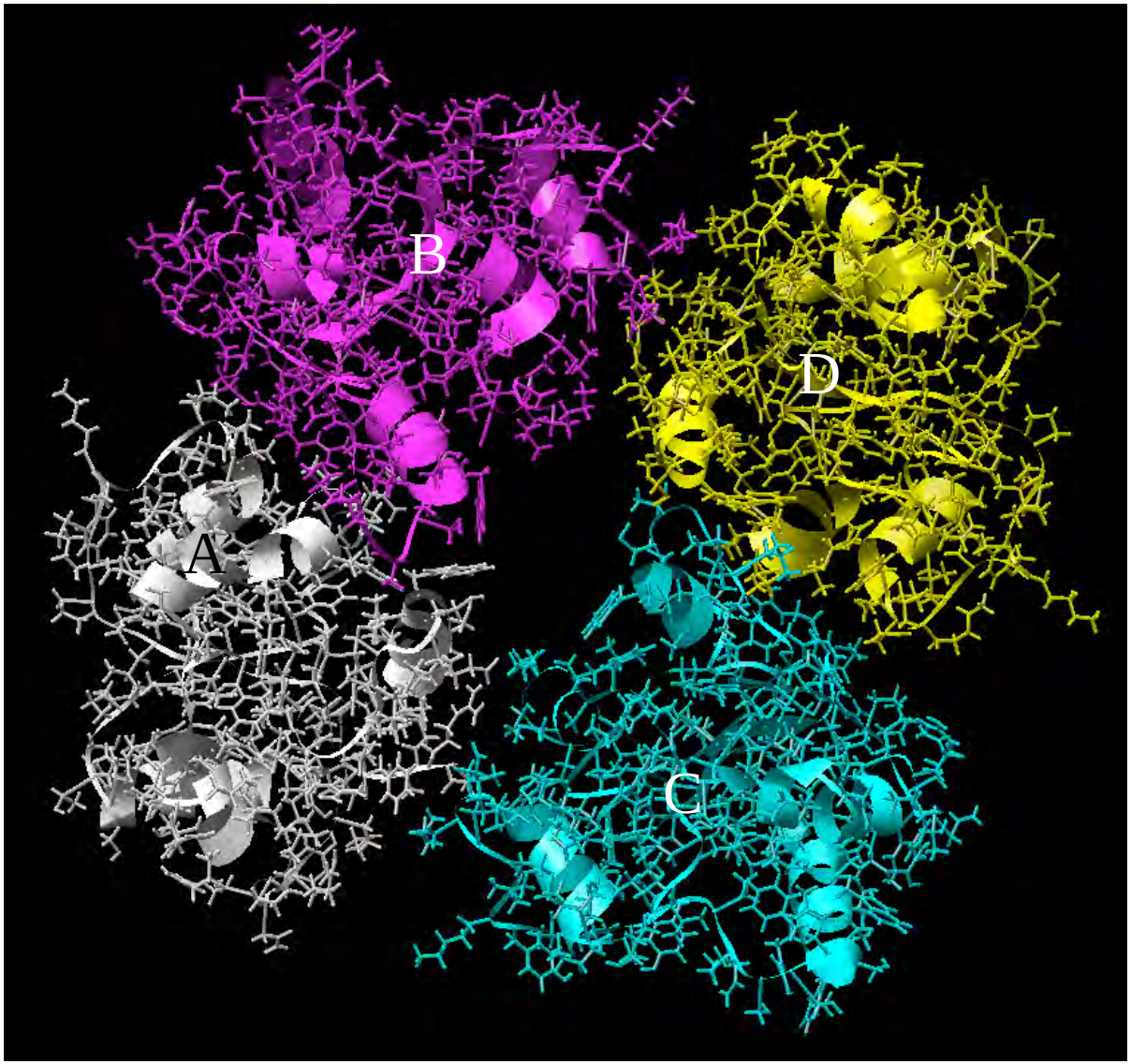


β–α folds

Interface area



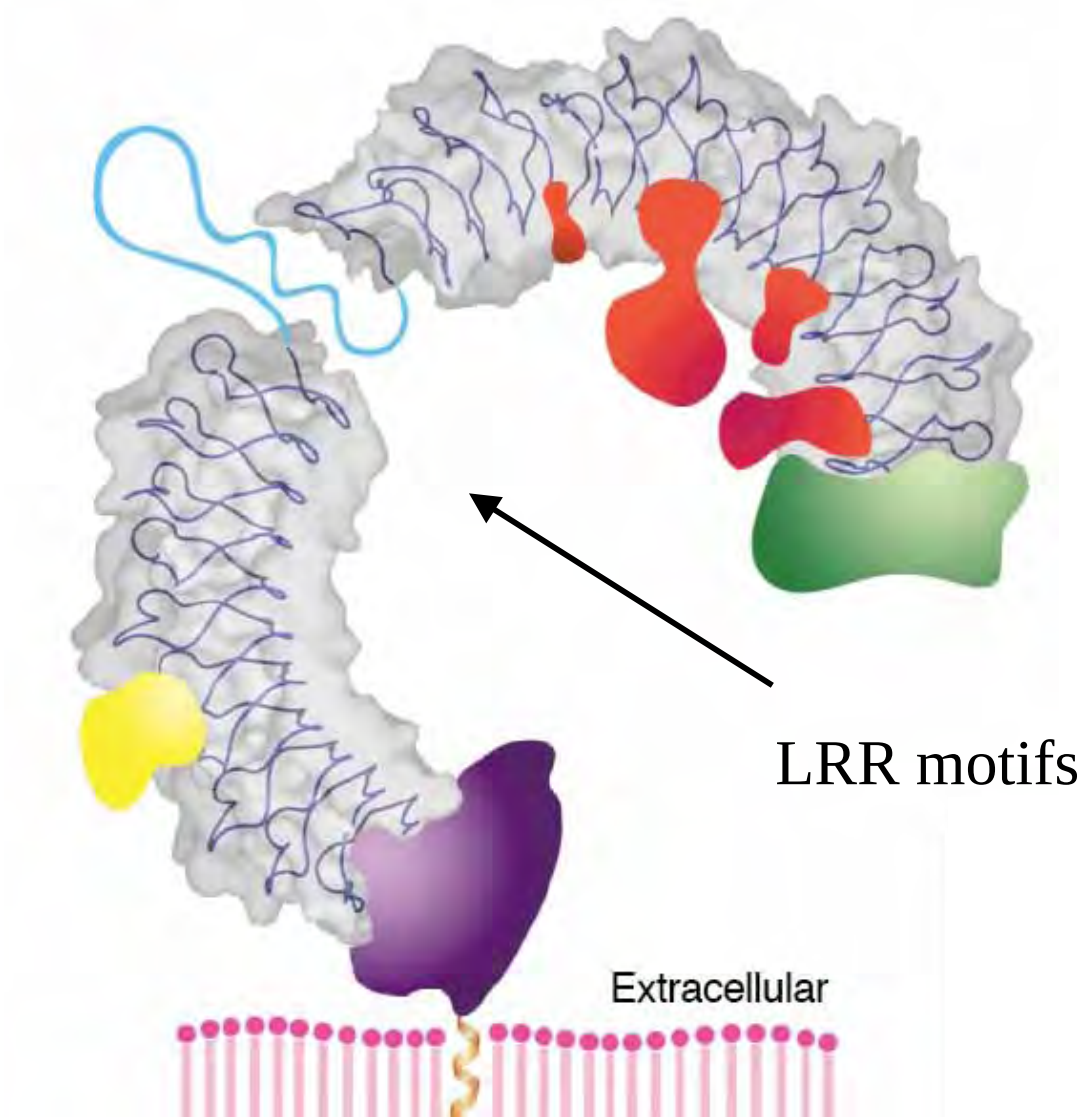
MyD88 TIR 4-mer



Interface areas

A-B: 1345 Å^2
A-C: 540 Å^2
B-D: 638 Å^2
C-D: 1116 Å^2

Ectodomain of TLR



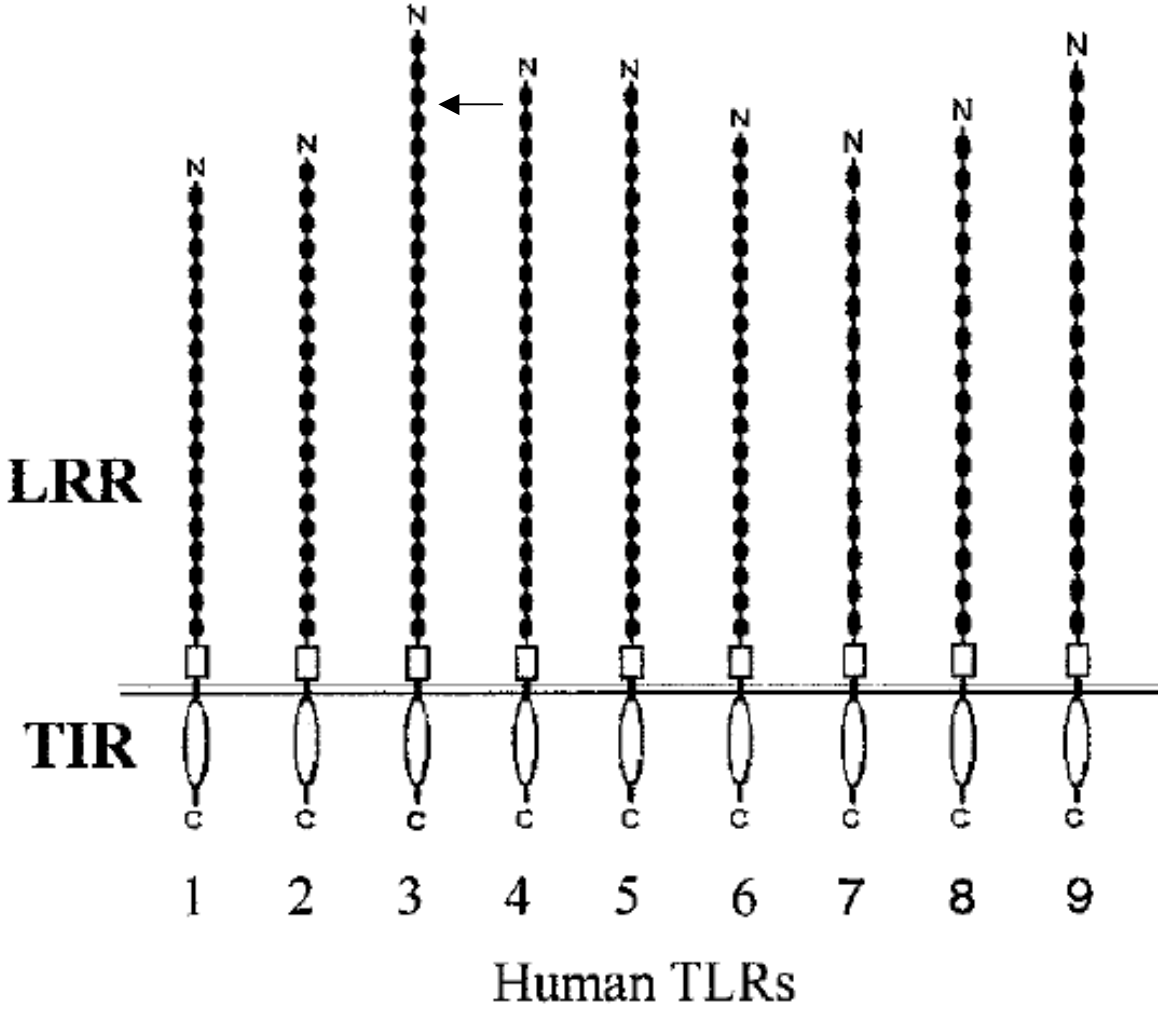
LRR motif (24 residues)

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

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

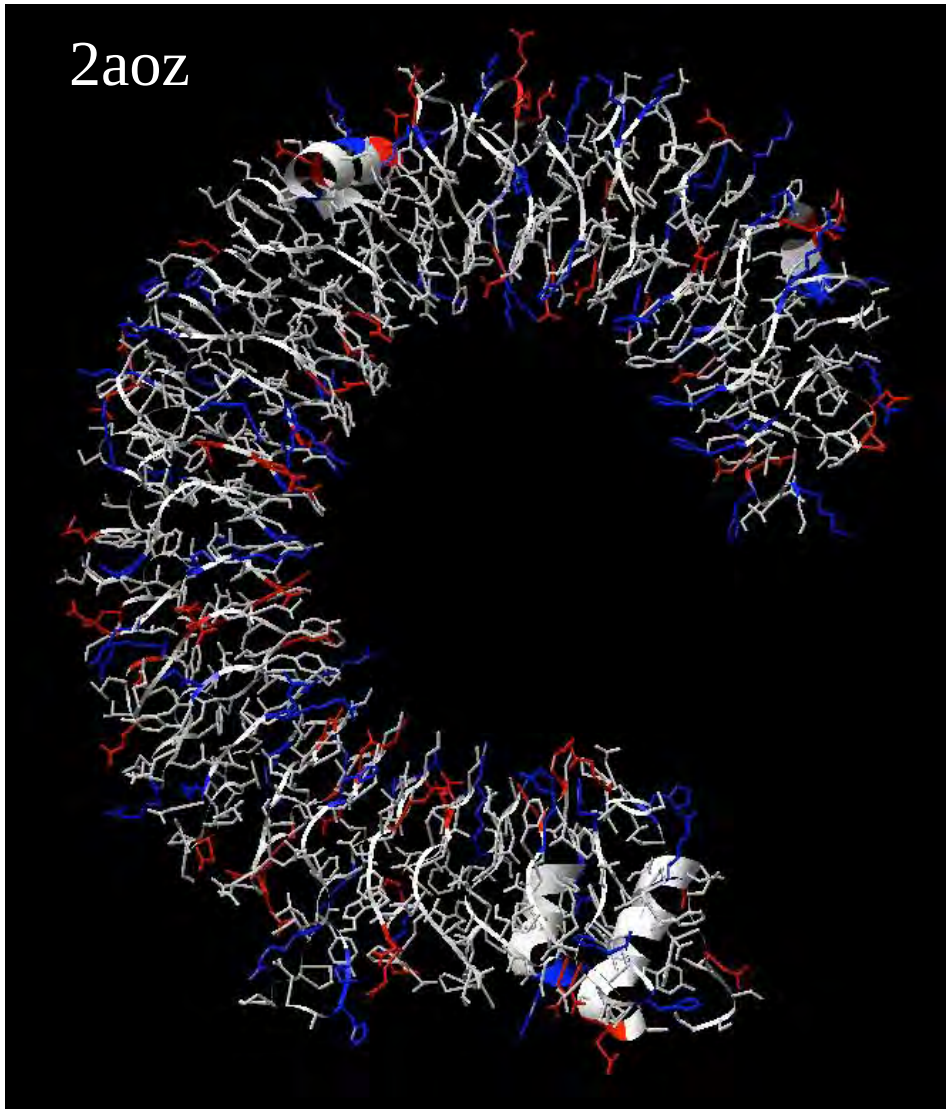
F is a conserved phenylalanine;

N is a conserved asparagine



19-25 tandem copies of LRRs in human TLRs

Ectodomain of TLR-3



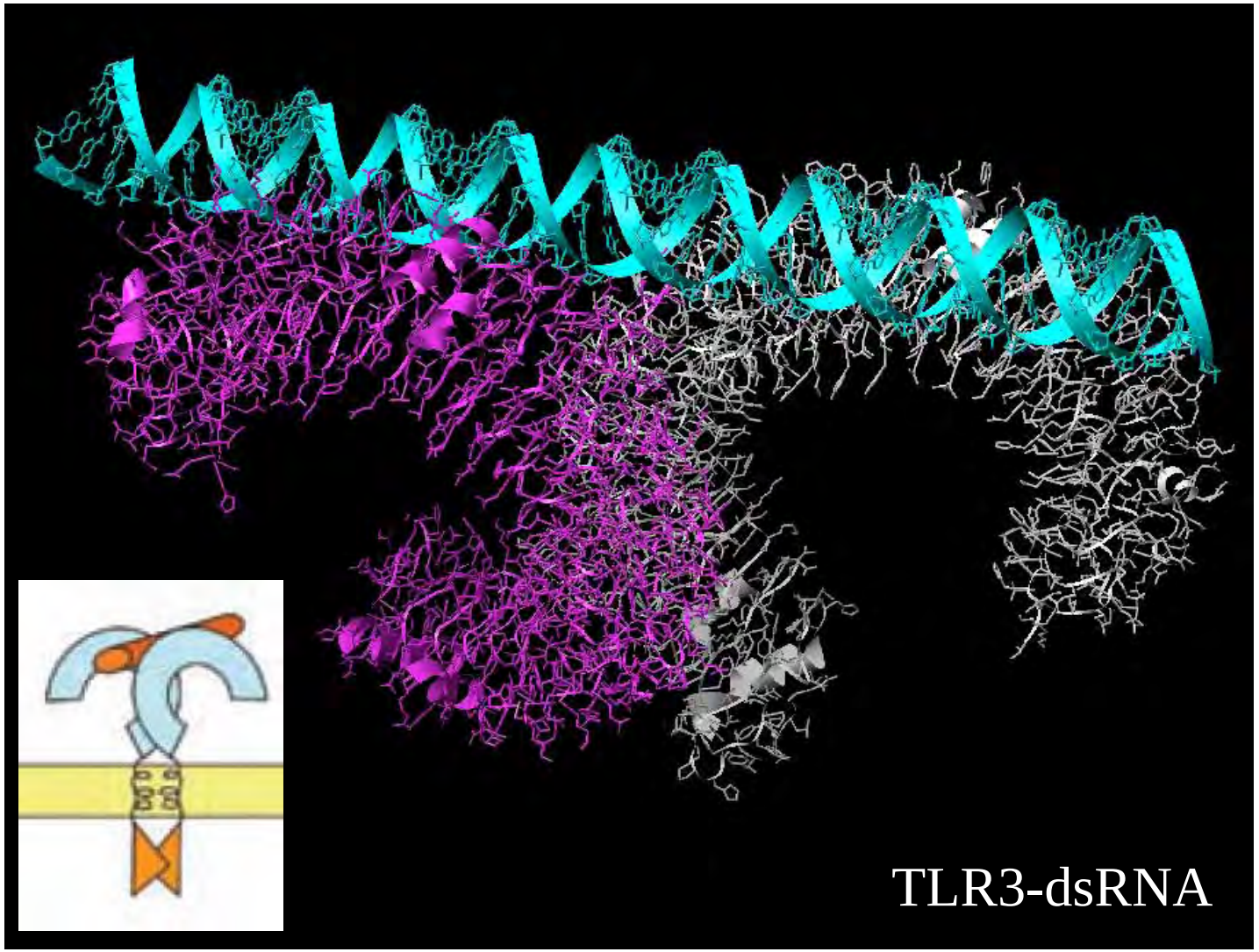
Choe et al., 2005
Bell et al., 2005

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 (~600 $Å^2$).
- Ligand binding increase the stability of the receptor dimer?

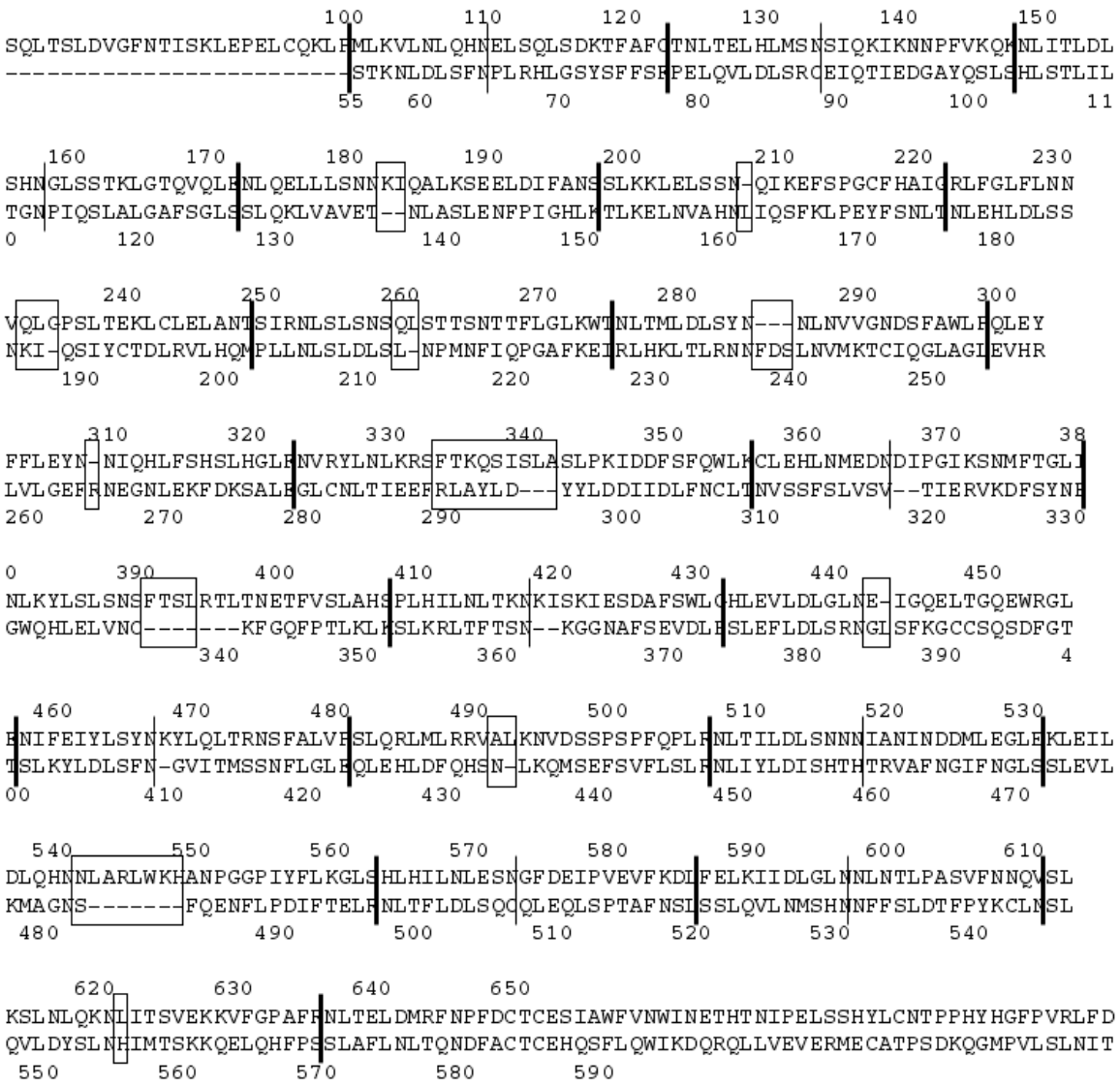
TLR3 ectodomain dimer + dsRNA

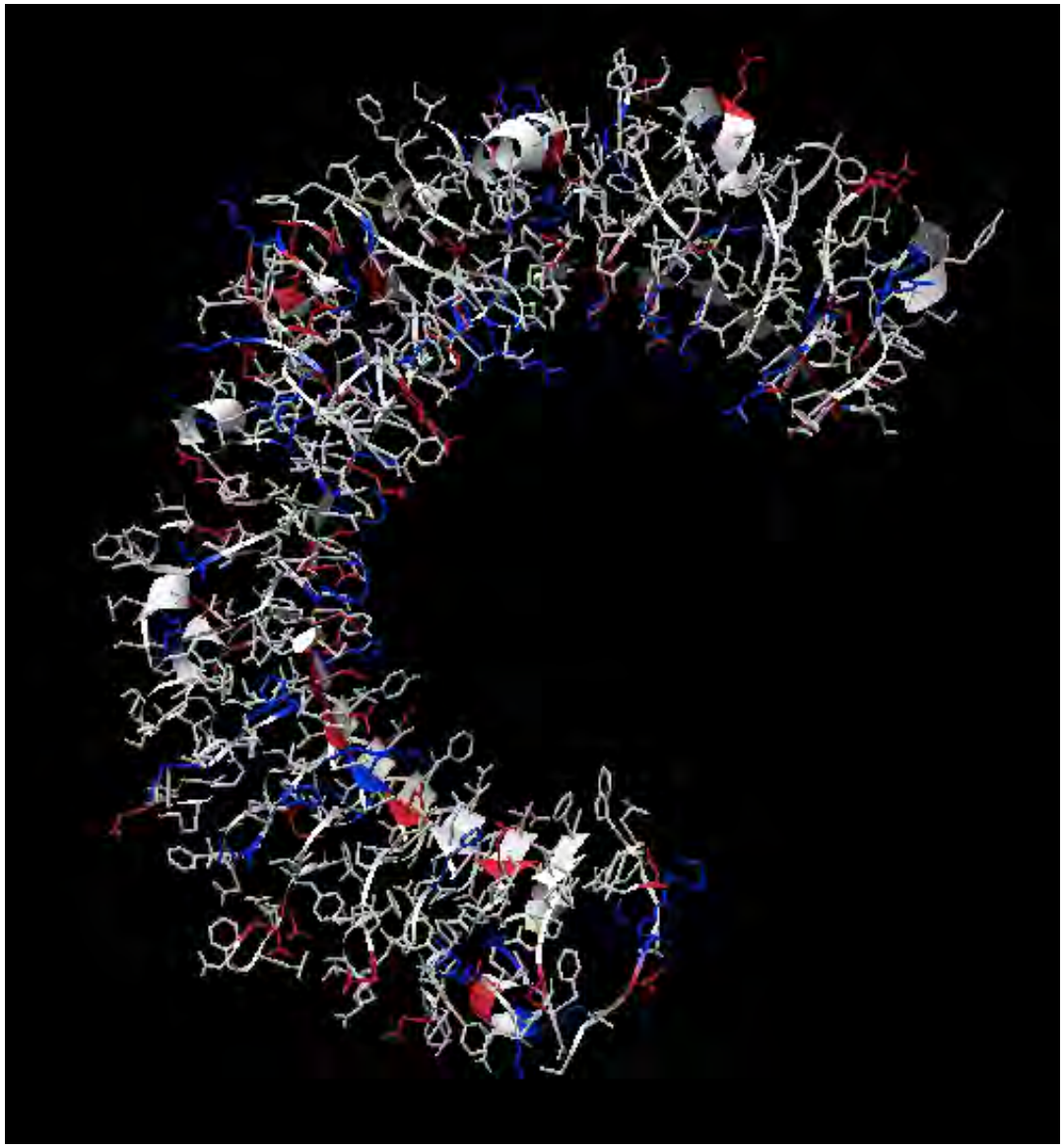


Modeling TLR4 ectodomain

- 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





TLR-4
ectodomain

(Bell et al., 2003)