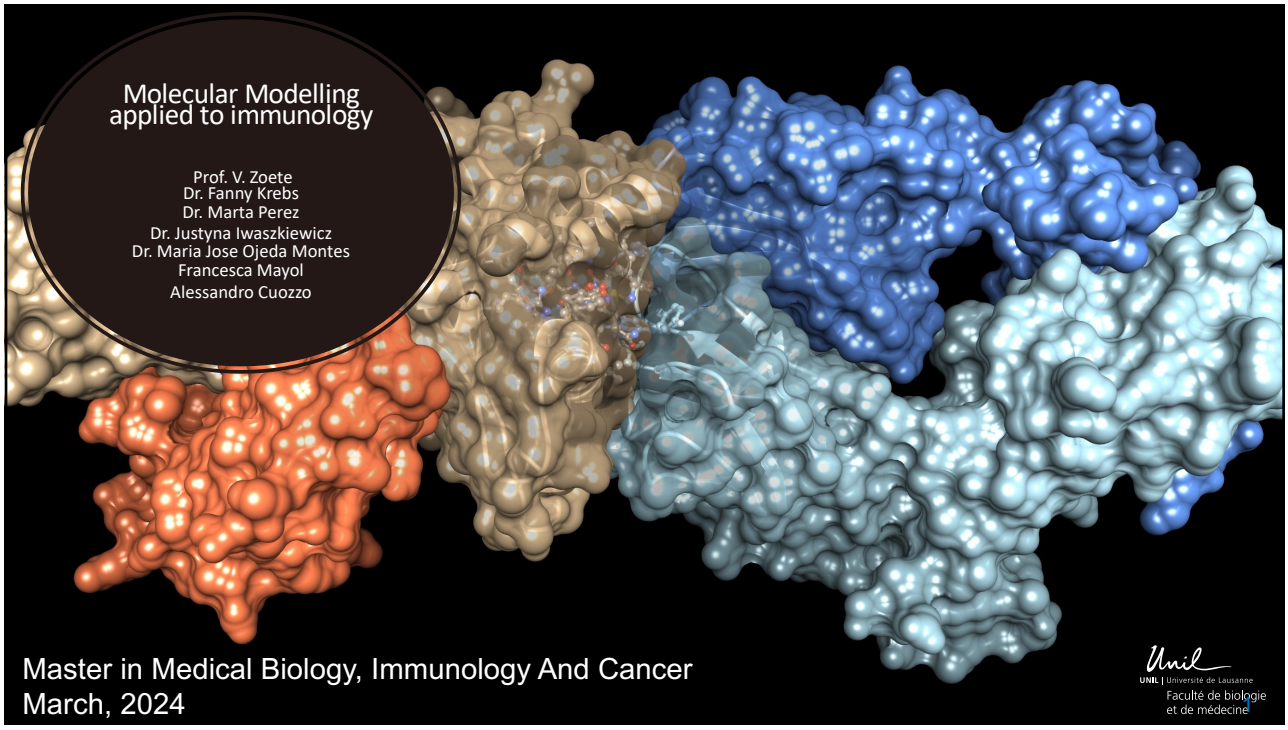


1

Lecture and Practice Proceedings & Objectives

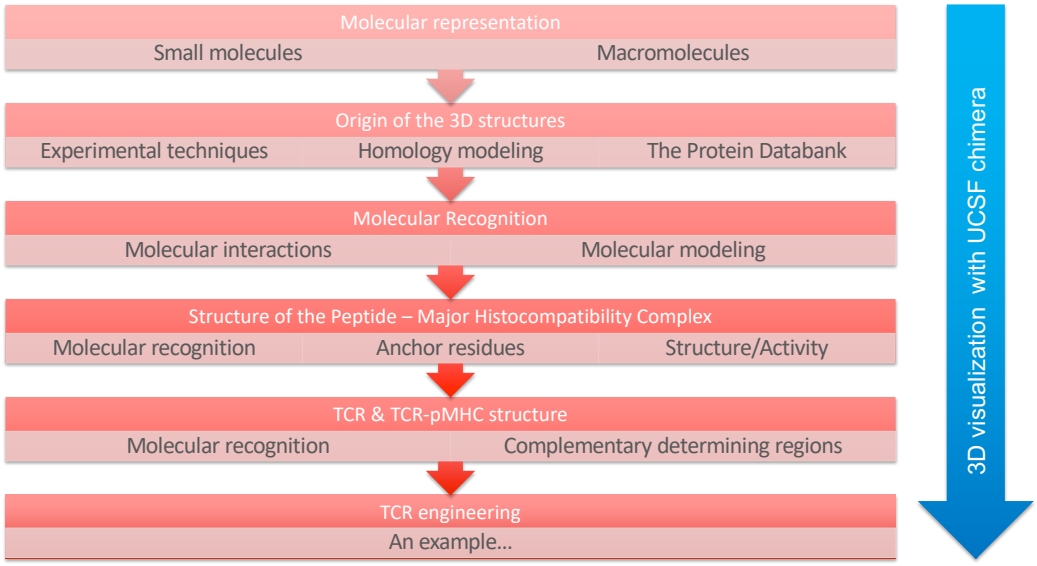
- Tutors are present to help you. Contact them!!!

- Get a flavor of molecular structure and modeling
- Acquire some basic theoretical background
- Practice the molecular graphics techniques
- Use them to uncover structure-activity relationship
- Get a detailed knowledge about the 3D structures of pMHC and TCR/pMHC

➔ You should be able to perform simple tasks of molecular graphics and analysis

2

Lectures & Practices Agenda



3

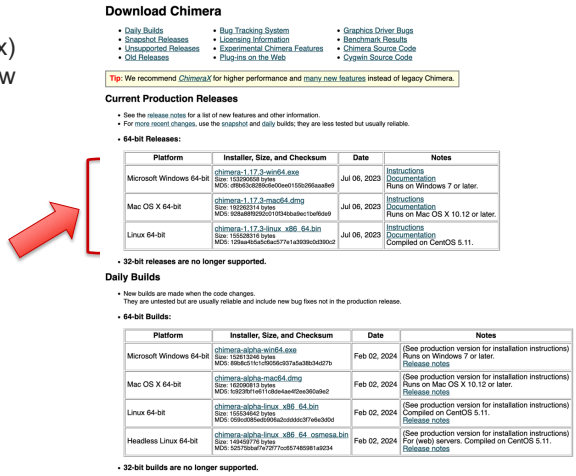
Installing UCSF Chimera

In this lecture/practical you will use the software UCSF Chimera for 3D structure visualisation and analysis.

This software is:
- free for teaching or academic research
- available for the most current platforms (Windows, Mac, Linux)
- open source (you can modify it for your needs if you know how to code in python. This is out of the scope of this lecture).

You can download the latest production release here:
https://www.cgl.ucsf.edu/chimera/download.html

Please, install this software on your machine.
It will be mandatory for the practicals, but also useful for the theoretical lectures



Download Chimera

- Daily Builds
- Snapshot Releases
- Unsupported Releases
- Old Releases
- Bug Tracking System
- Licensing Information
- Experimental Chimera Features
- Plug-ins on the Web
- Graphics Driver Bugs
- Benchmark Results
- Chimera Source Code
- Cygwin Source Code

Tip: We recommend *ChimeraX* for higher performance and many new features instead of legacy Chimera.

Current Production Releases

- See the release notes for a list of new features and other information.
- For more recent changes, use the snapshot and daily builds; they are less tested but usually reliable.
- **64-bit Releases:**

Platform	Installer, Size, and Checksum	Date	Notes
Microsoft Windows 64-bit	chimera-1.17.3-win64.exe Size: 153290656 bytes MD5: d8bb63c8286c6e00ee0155b266aaa8e9	Jul 06, 2023	Instructions Documentation Runs on Windows 7 or later.
Mac OS X 64-bit	chimera-1.17.3-mac64.dmg Size: 192262314 bytes MD5: 928a8891f250c01f354bba9ec1bef6de9	Jul 06, 2023	Instructions Documentation Runs on Mac OS X 10.12 or later.
Linux 64-bit	chimera-1.17.3-linux_x86_64.bin Size: 155528316 bytes MD5: 129aa4c6a5c6ac577e1a3939c0d390c2	Jul 06, 2023	Instructions Documentation Compiled on CentOS 5.11.

- **32-bit releases are no longer supported.**

Daily Builds

- New builds are made when the code changes.
  They are untested but are usually reliable and include new bug fixes not in the production release.
- **64-bit Builds:**

Platform	Installer, Size, and Checksum	Date	Notes
Microsoft Windows 64-bit	chimera-alpha-win64.exe Size: 152813246 bytes MD5: 86b8c511c1cf9056c937a3a3ffb34d27b	Feb 02, 2024	(See production version for installation instructions) Runs on Windows 7 or later. Release notes
Mac OS X 64-bit	chimera-alpha-mac64.dmg Size: 162306813 bytes MD5: fc923fbf1e611c8de4ae4f2ee360a9e2	Feb 02, 2024	(See production version for installation instructions) Runs on Mac OS X 10.12 or later. Release notes
Linux 64-bit	chimera-alpha-linux_x86_64.bin Size: 155534662 bytes MD5: 059cd085ed0908a2cdddd3f7e6e3d0d	Feb 02, 2024	(See production version for installation instructions) Compiled on CentOS 5.11. Release notes
Headless Linux 64-bit	chimera-alpha-linux_x86_64_osmesa.bin Size: 149493776 bytes MD5: 52573bbaf7e7277ccd55748598fa9234	Feb 02, 2024	(See production version for installation instructions) For (web) servers. Compiled on CentOS 5.11. Release notes

- **32-bit builds are no longer supported.**

4

The dedicated web site

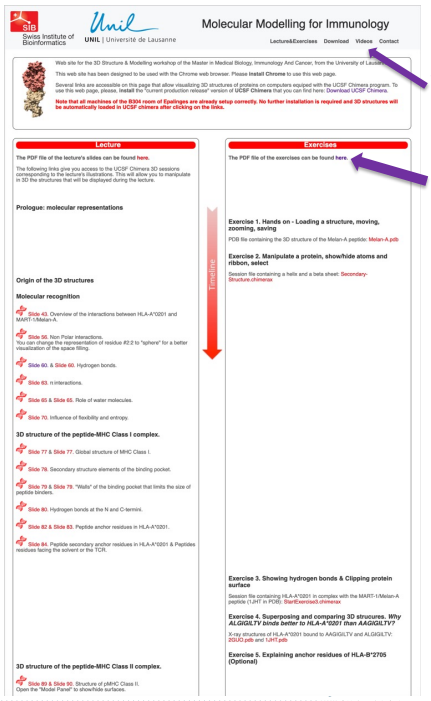
This teaching has been conceived to alternate theoretical lectures and practicals, so that you will:

- experiment yourself the visualisation and analysis of protein 3D structures
- get a 3D view of the systems mentioned in the lecture (to prevent being limited by the 2D views in the slides)

To facilitate the process, a web site has been especially conceived for this teaching. You can find it here:

http://www.immunology-and-modelling.ch

1. This web site will indicate you when to switch between lecture and practicals. For instance, you will be able to make exercices 1 and 2 just after the prologue regarding molecular representation, while exercises 3 to 5 will be made after the lecture section dedicated to MHC-I and before that of MHC-II
2. Videos on how to execute exercices 1 to 5 have been made for your help. There are without sound, but all instructions are detailed in the booklet
3. The booklet of the practicals and the PDF of the lecture can be downloaded from the web site too



5

The dedicated web site – Seeing the lecture in 3D

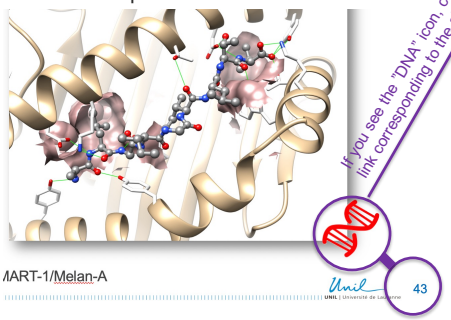
4. Once you will be trained in using UCSF Chimera, you will be able to use this tool to see in 3D, on your computer, the systems that are displayed in 2D in the lecture.

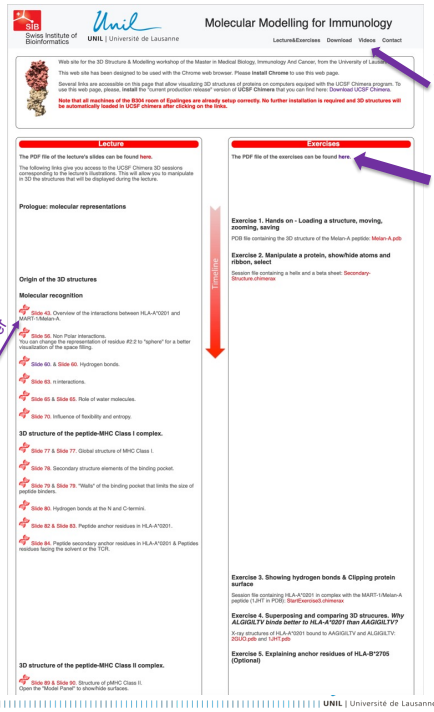
So, when you see this icon in the lecture slide, you can check the slide number, and click on the corresponding link on the web site.

Upon clicking, the molecular system should be automatically displayed in UCSF Chimera, if your browser is managing it correctly.

Otherwise, you can download the linked file, with the .chimerax extension, on your machine. Then, double clicking on the file, or drag&droping it on UCSF chimera will open it in this software.

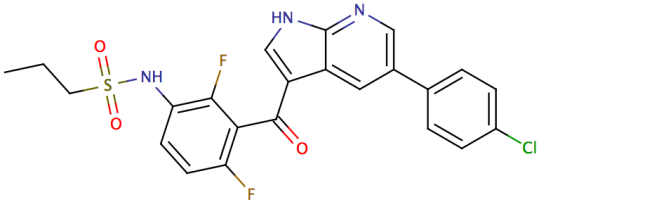
Although you are encouraged to do it, to beneficiate from a full 3D "immersion", this is totally optional.
This possibility is provided to you to go beyond 2D images in the slides. But if it does not work for you, or if you don't want to use it, it will not prevent you from understanding the lecture.

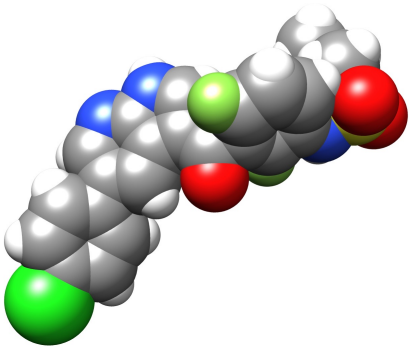


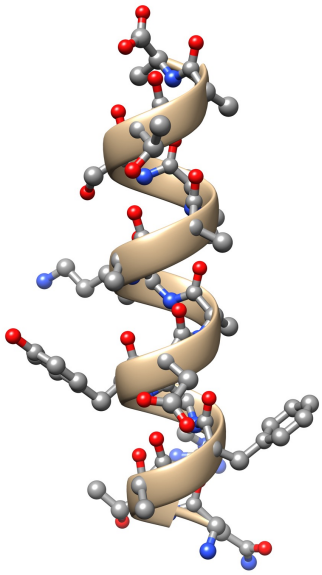


6

Prologue: molecular representations







7

Molecular representations – “small” molecules

Organic molecules of less than ~ 100 atoms are often referred to as “small” molecules, as opposed to biological macromolecules (i.e. proteins, DNA, etc.)

Small molecules can be represented in 1D, 2D or 3D:

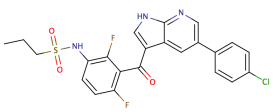
Example of Vemurafenib (BRAF V600E inhibitor)

1D

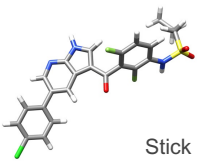
SMILES: CCCS(=O)(=O)NC1=C(F)C(C(=O)C2=CNC3=NC=C(C=C23)C2=CC=C(Cl)C=C2)=C(F)C=C1

InChI: 1S/C23H18ClF2N3O3S/c1-2-9-33(31,32)29-19-8-7-18(25)20(21(19)26)22(30)17-12-28-23-16(17)10-14(11-27-23)13-3-5-15(24)6-4-13/h3-8,10-12,29H,2,9H2,1H3,(H,27,28)

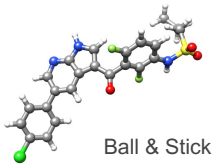
2D



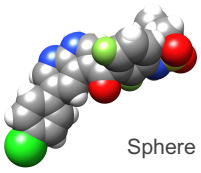
3D



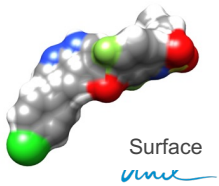
Stick



Ball & Stick



Sphere



Surface

8

Molecular representations – biological macromolecules

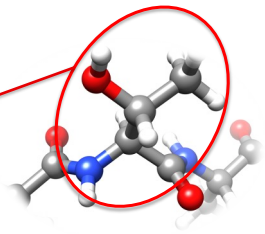
Biological macromolecules can also be represented in 1D, 2D or 3D:

Example of proteins

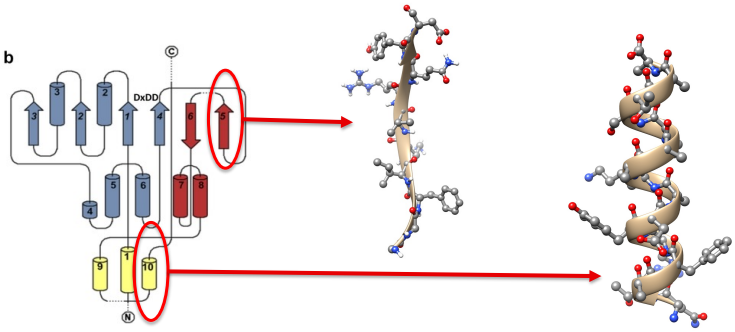
1D

Sequence if amino-acids:

```
MGCTLSAEDKAAVERSKMIDRNLREDGEKAAREVKLLLLGAGESGKSTIVKQMKIIHEAG
YSEEECKQYKAVVYSNTIQSIIAIIRAMGRLKIDFGDSARADDARQLFVLAGAAEEGFMT
AELAGVIKRLWKDSGVQACFNRSREYQLNDSAAYYLNDLDRIAQPNYIPTQQDVLRTRVK
TTGIVETHFTFKDLHFKMFDVGGQRSERKKWIHCFEGVTAIIFCVALSDYDLVLAEDEEM
NRMHESMKLFDSICNNKWFTDTSIILFLNKKDLFEEKIKKSPLTICYPEYAGSNTYEEAA
AYIQCQFEDLNKRKDTKEIYTHFTCATDTKNVQFVFDAVTDVIIKNNLKDCGLF
```



2D



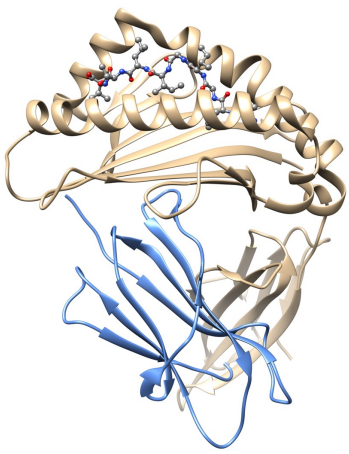
9

Molecular representations – biological macromolecules

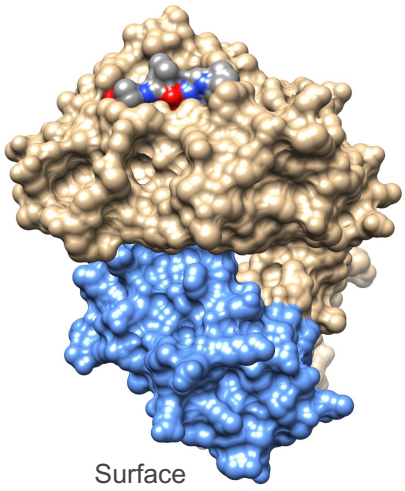
Biological macromolecules can also be represented in 1D, 2D or 3D:

Example of proteins (Crystal structure of HLA-A2*0201 in complex with MART-1/Melan-A)

3D



Ribbon / cartoon



Surface

10

Molecular representations – biological macromolecules

Biological macromolecules can also be represented in 1D, 2D or 3D:

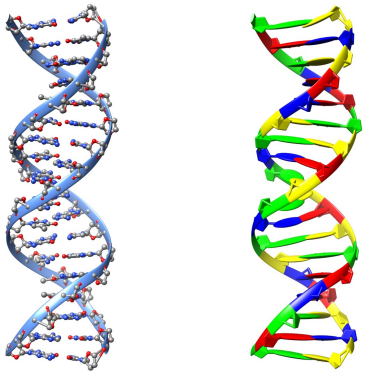
Example of DNA

1D

Sequence if nucleotides:

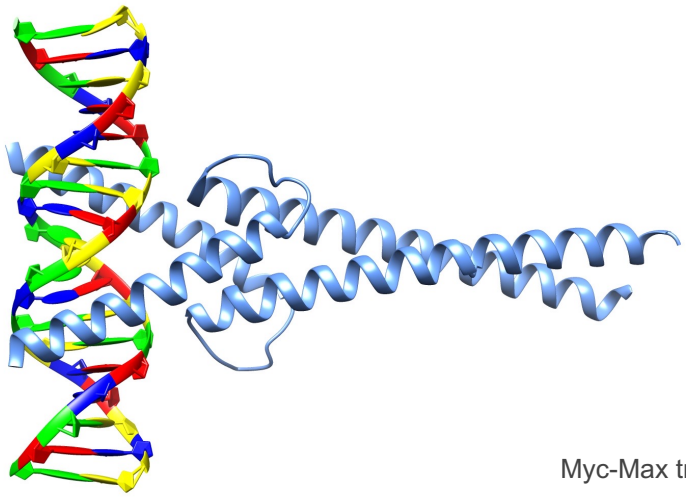
GAGTAGCACGTGCTACTC

3D



11

Molecular representations – biological macromolecules



Myc-Max transcription factor

12

Origin of the 3D structures

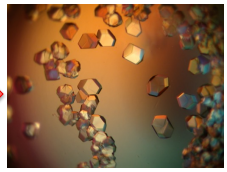
13

Experimental methods – Xray crystallography

Purified protein

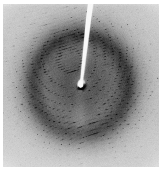
Crystallization

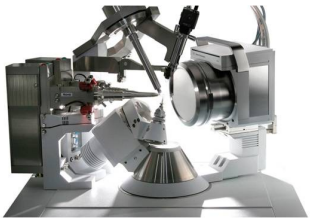
Protein crystals



X-ray

Diffraction spectrum

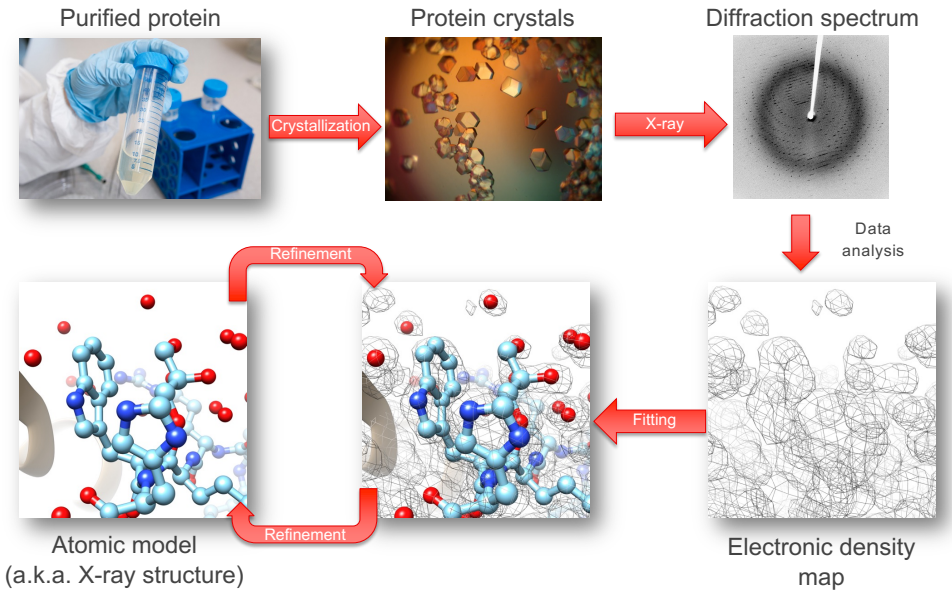




Xray diffraction

14

Experimental methods – Xray crystallography

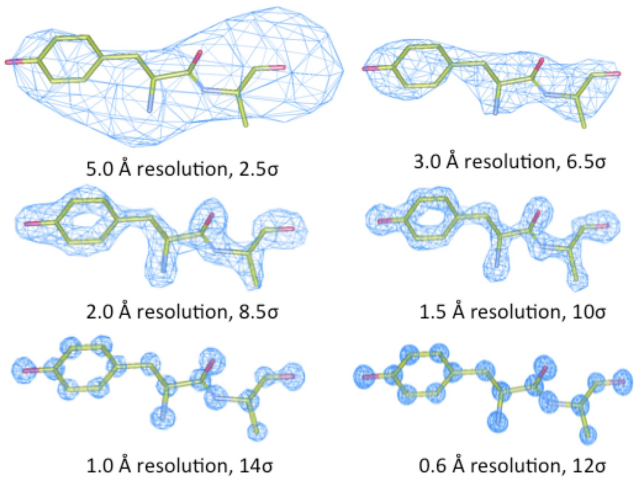


15

Experimental methods – Xray crystallography

Important measures of accuracy:

- **Resolution** (in Å): measures the amount of detail that may be seen in the experimental data. The lower the better (typically around 2 Å)



Source: PLoS One. 2015 Apr 20;10(4):e0123146.

16

Experimental methods – Xray crystallography

3 important measures of accuracy:

- **Resolution** (in Å): measures the amount of detail that may be seen in the experimental data. The lower the better (typically around 2 Å)
- **R-value**: measures how well the atomic model is supported by the experimental data found in the structure factor file (Perfect fit R-value = 0.0; Random fit R-value = 0.63; Typical R-value ~ 0.20) The atomic model is used to simulate a diffraction spectrum, which is compared to the experimental one.
- **R-free value**: idem than R-value, but calculated for a set of experimental data that have not been used to create the model (~10% of the data are removed before refinement, in order to be used in this test). Generally, R-free value > R-value; Typically R-free value ~ 0.26 for a good quality structure.

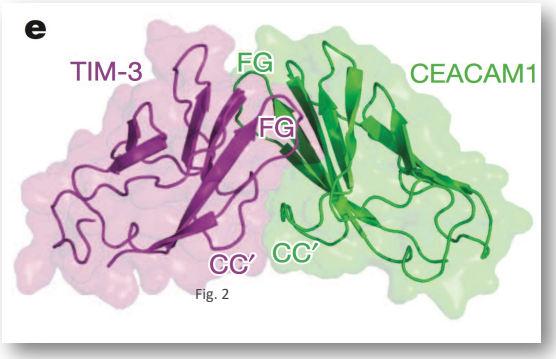
Typical limitations:

- Hydrogen atoms are generally not visible
- Some regions are not defined (e.g. flexible loops or flexible side chains)
- X-ray structures are models. They can be totally wrong!!

Experimental methods – Xray crystallography

X-ray structures are models. They can be totally wrong!!

Huang, Y.-H., et al. CEACAM1 regulates TIM-3-mediated tolerance and exhaustion. *Nature,* 2015, *517*(7534), 386–390.



Xray structure of the complex
CEACAM1/TIM3
PDB ID: 4QYC
Resolution: 3.4Å
R-value: 0.232

Correction

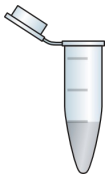
5DZL

Crystal structure of the protein human CEACAM1

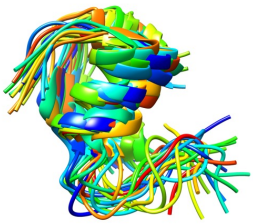
DOI: 10.2210/pdb5dzl/pdb Entry 5DZL **supersedes** 4QYC

It was a homodimer of CEACAM1…!

Experimental methods – NMR spectroscopy



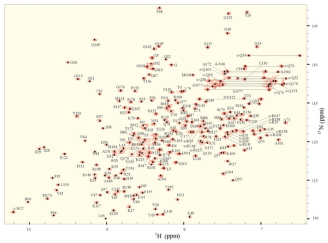
Purification
Concentration



Ensemble of structures
Example: insulin

Distance constraints

Modeling



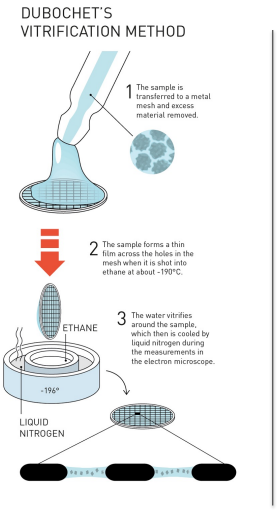
2D-Spectra

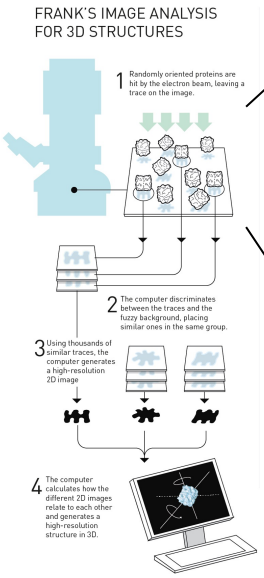
Pros : Structure in solution

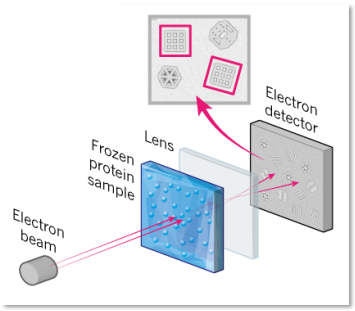
Cons:
- Limited to small proteins
- Low resolution
- Highly flexible regions

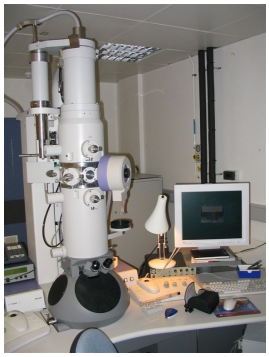
19

Experimental methods – CryoEM





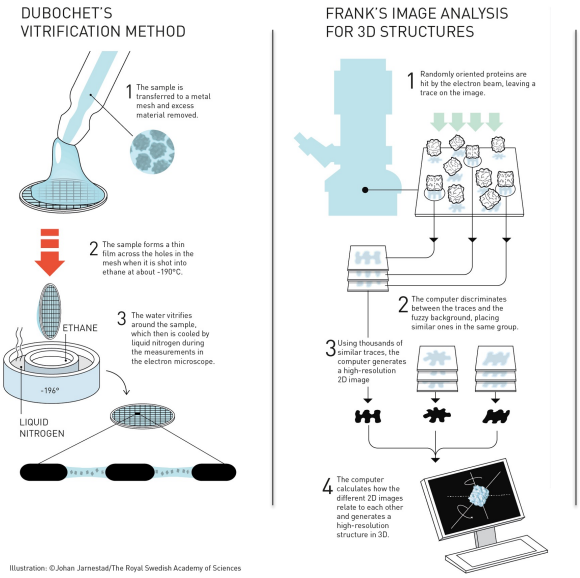




- Very power electronic beam
- Better resolution than light (smaller wave length)
- In vacuo in the microscope
- Frozen sample (77 K or 4 K)
- Vitrified water

20

Experimental methods – CryoEM



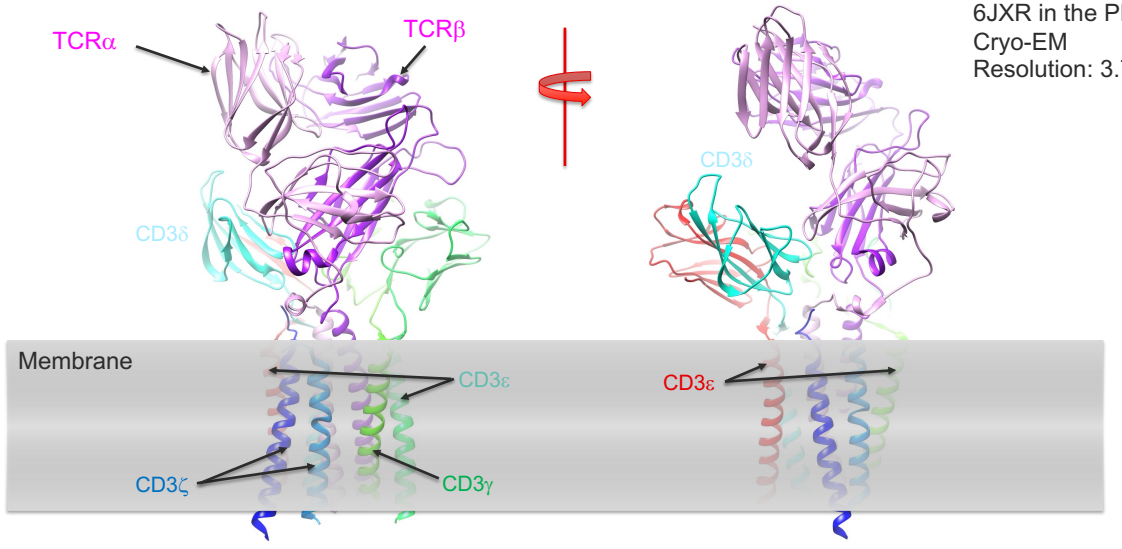
Until recently:

- Only low resolution structures. Need to be used together with Xray crystallography or NMR (for example, insertion of Xray structures into the Cryo-EM density map)
- Limited to large-size systems (which can actually be seen as a pros or a cons)

Nowadays:

- Resolution close to that of Xray crystallography
- Applicable to smaller systems
- More Cryo-EM structures produced every year than NMR structures
- Capture structures in relevant states (isolated molecules, in solution, at a given salt concentration and pH)

Experimental methods – CryoEM

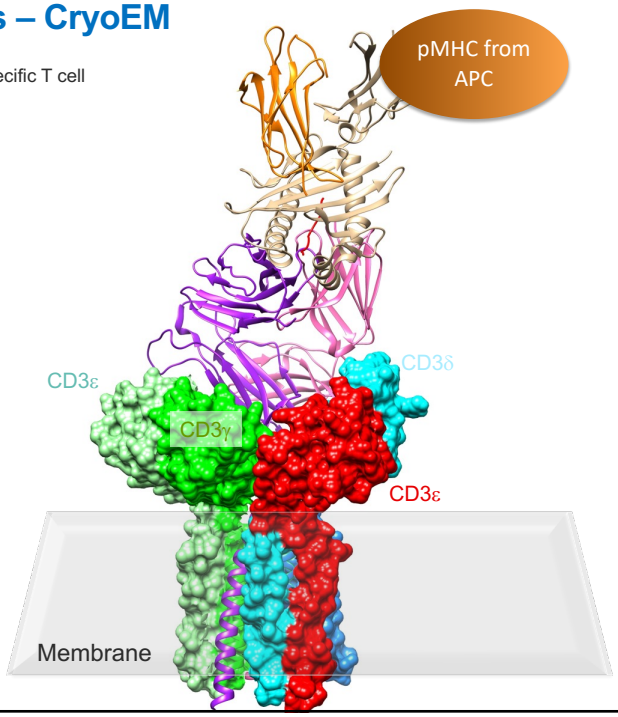


6JXR in the PDB
Cryo-EM
Resolution: 3.7 Å

Structural basis of assembly of the human T cell receptor-CD3 complex.
Zheng, L., Lin, J., Zhang, B., Zhu, Y., Li, N., Xie, S., Wang, Y., Gao, N., Huang, Z.
(2019) Nature 573: 546-552

Experimental methods – CryoEM

Structure of a fully assembled tumor-specific T cell receptor ligated by pMHC
Lukas Sušac et. al.
Cell, 2022, 185 (17), 3201-3213.e19



7PHR in the PDB
Cryo-EM
Resolution: 3.1 Å

Experimental methods – CryoEM

The Nobel Prize in Chemistry 2017

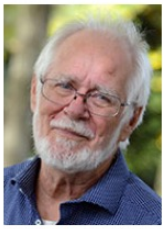


Photo: Félix Imhof © UNIL [CC BY-SA 4.0]
Jacques Dubochet
Prize share: 1/3

Photo: B. Winkowski © Columbia University Medical Center
Joachim Frank
Prize share: 1/3

Photo: MRC Laboratory of Molecular Biology
Richard Henderson
Prize share: 1/3

The Nobel Prize in Chemistry 2017 was awarded to Jacques Dubochet, Joachim Frank and Richard Henderson *"for developing cryo-electron microscopy for the high-resolution structure determination of biomolecules in solution".*

Experimental methods - Summary

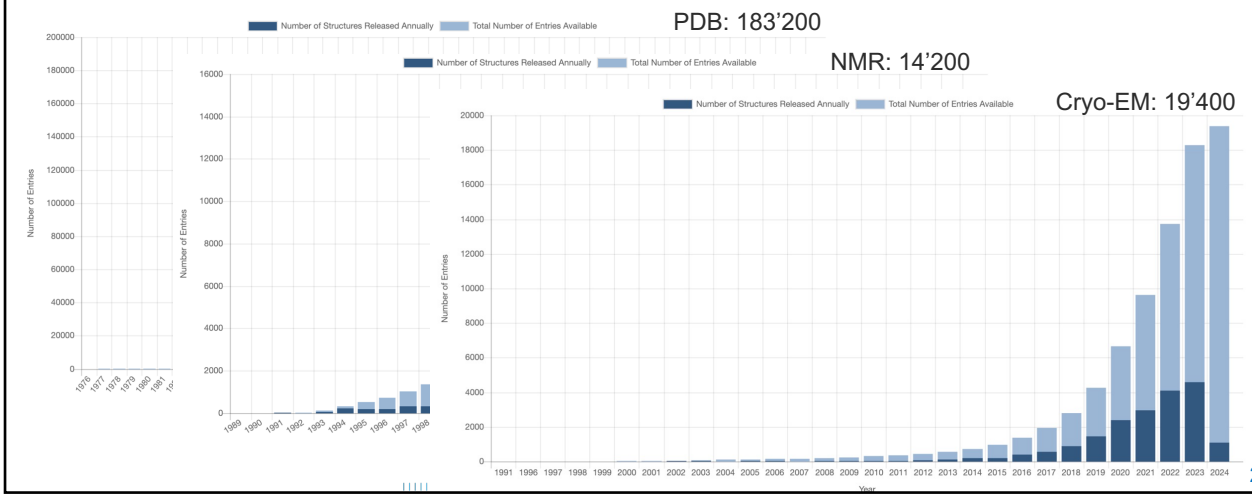
Technique	Advantages	Disadvantages
Xray crystallography	High resolution (1 to 3 Å)	Requires to crystallize the protein Does not allow studying transmembrane or very flexible proteins
NMR	Does not require protein crystallization ~ High resolution	Generally limited to small proteins
Cryo-EM	Does not necessitate to crystallize the protein: possible to study transmembrane proteins, and more flexible proteins than Xray. New techniques allow studying smaller proteins, and increasing resolution	Generally limited to large proteins Low resolution, 4 to 20 Å (a lot of progresses have been done recently)

Where to find experimental 3D structures? The protein databank

Experimental 3D structures are stored in the **Protein Data Bank (PDB)**

Worldwide Protein Data Bank (wwPDB)	https://www.wwpdb.org/
Protein Data Bank in US (PDB)	http://www.rcsb.org/
Protein Data Bank in Europe (PDBe)	http://www.ebi.ac.uk/pdbe/

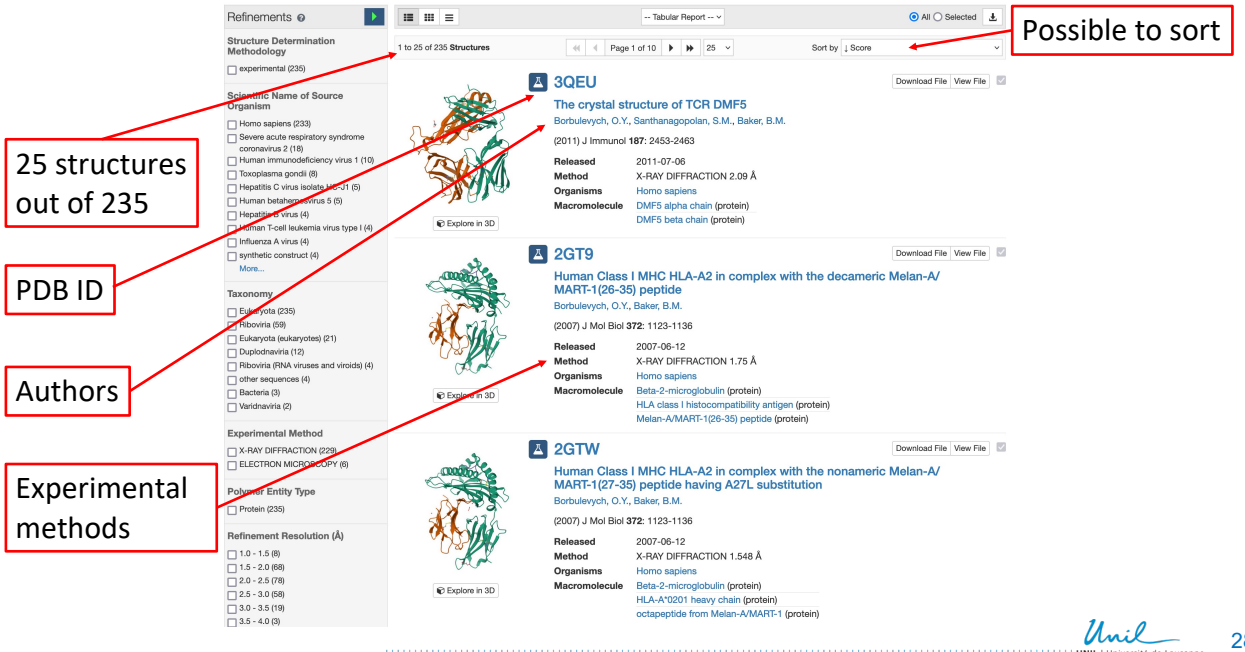
217'200 structures in March 2024



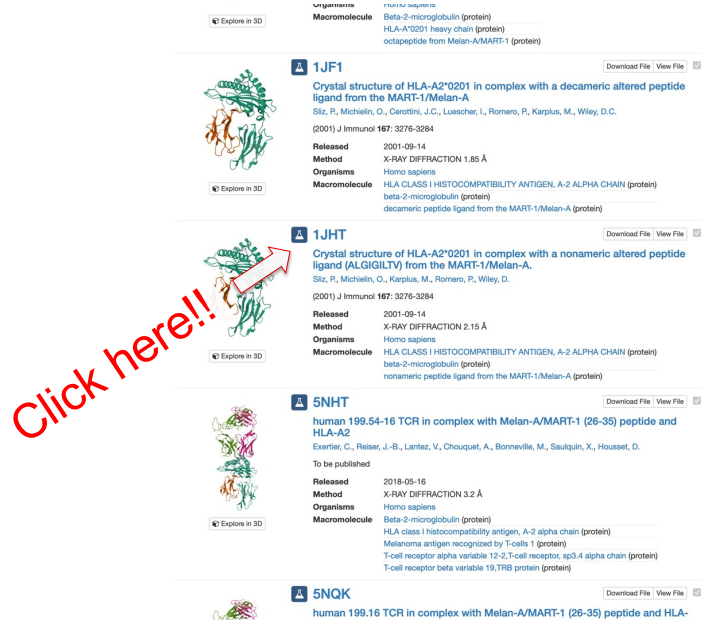
Where to find experimental 3D structures? The protein databank

http://www.rcsb.org/

Where to find experimental 3D structures? The protein databank

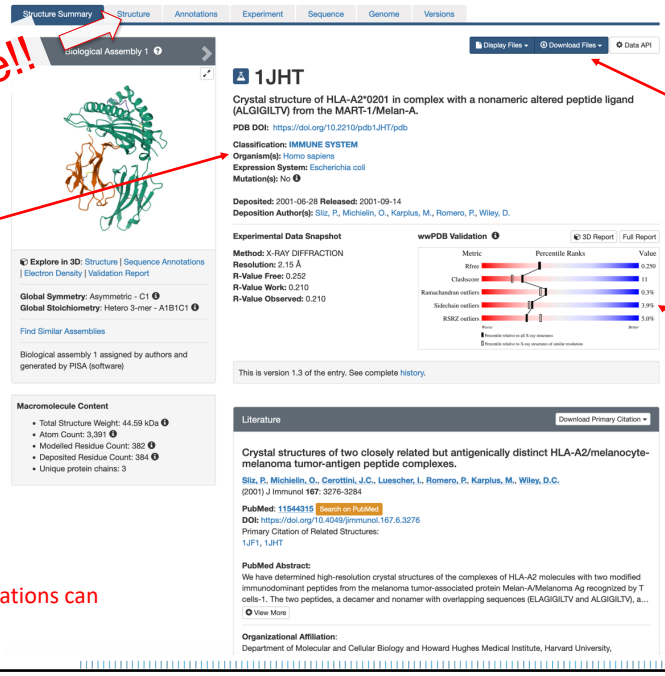


Where to find experimental 3D structures? The protein databank



29

Where to find experimental 3D structures? The protein databank



Organism (origin of the sequence) & expression system (synthesis et post-translational modifications)

Dowload or online visualization

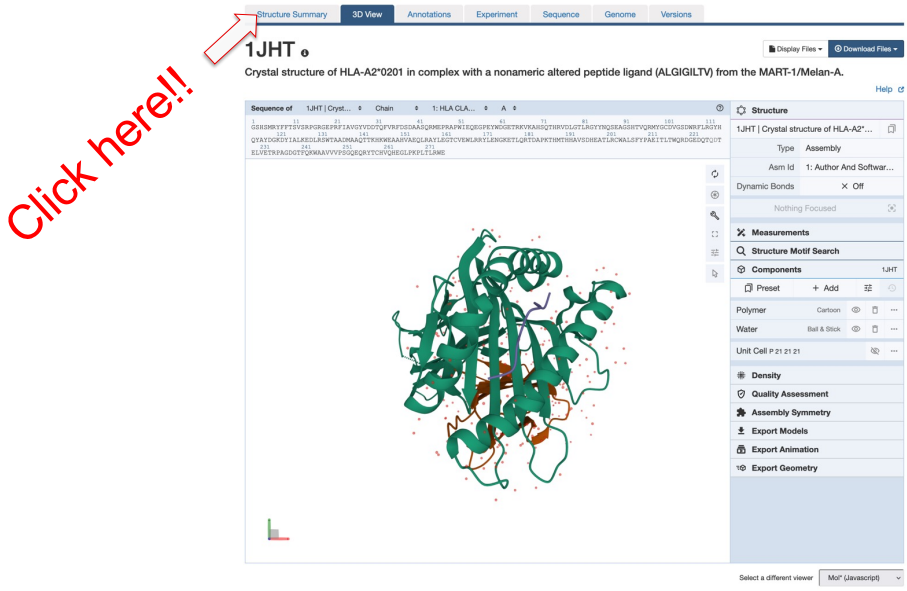
Experimental method and quality

Note: post-translational modifications can differ between organisms

30

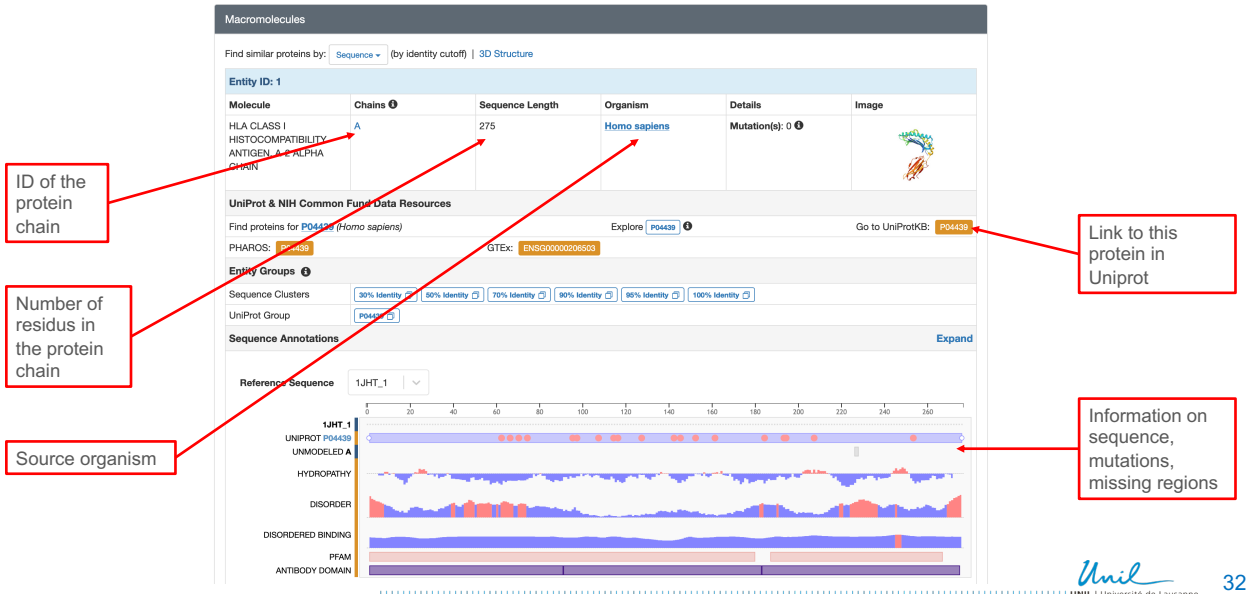
Where to find experimental 3D structures? The protein databank

Online visualization



Where to find experimental 3D structures? The protein databank

Information regarding the protein, and what is present in the experimental structure



Where to find experimental 3D structures? The protein databank

Download / display

Click here!

Display Files | Download Files

FASTA Sequence

1JHT

Crystal structure of HLA-A2*0201 in complex with a ... tide ligand (ALGIGILTV) ...

PDB DOI: 10.221...

Classification: IM...

Organism(s): Ho...

Expression Syst...

Mutation(s): No ...

Deposited: 2001...

Deposition Auth...

Experimental Da...

Method: X-RAY D...

Resolution: 2.15 ...

R-Value Free: 0.2...

R-Value Work: 0....

R-Value Observe...

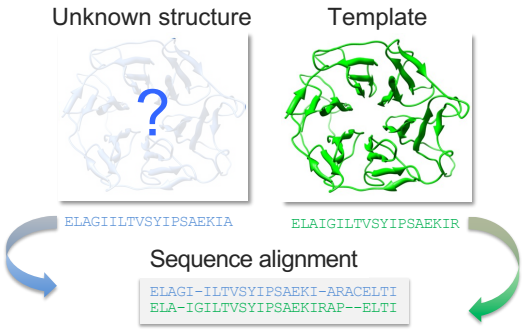
A header with information on the protein and experimental conditions

3D Cartesian coordinates of each visible atom

Need a Visualization software…

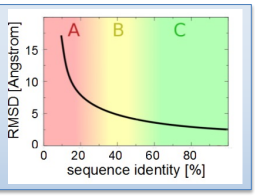
Ex.: Swiss PDB Viewer, UCSF Chimera, Pymol

And when there is no experimental structure? Homology modeling

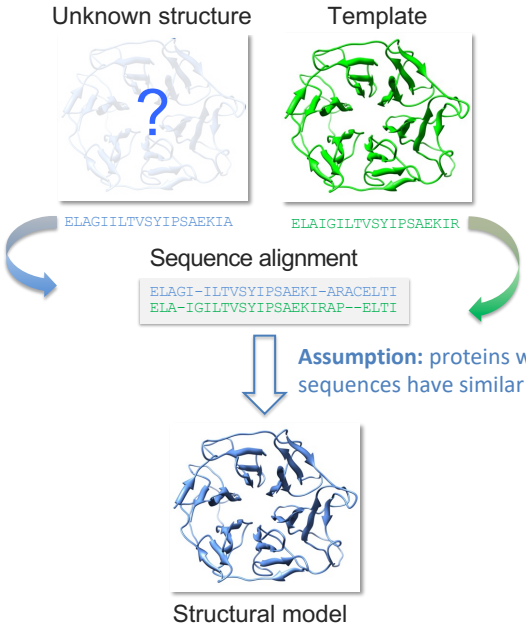


Homology modeling:
Building a theoretical model of the 3D structure of a protein based on experimentally known 3D structure of related proteins.

Quality/relevance of the prediction depends on the sequence identity.
Needs at least 40% seq. id.



And when there is no experimental structure? Homology modeling



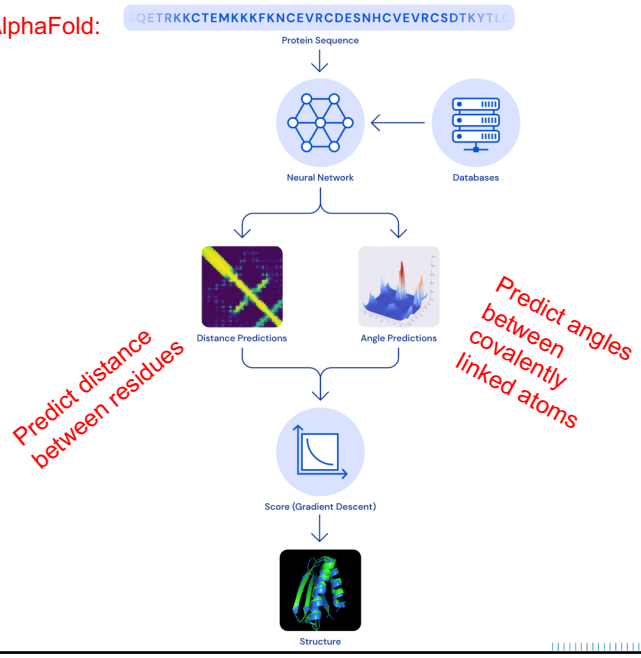
Homology modeling:
Building a theoretical model of the 3D structure of a protein based on experimentally known 3D structure of related proteins.

Programs et web servers:
- Modeller
- I-Tasser
- Robetta
- HHPred
- …

Databases of structural models:
- Swiss-model
- Modbase
- …

And when there is no experimental structure? Deep Learning

AlphaFold:



Predict distance between residues:

Sequence of the target protein

Sequence alignment of all apparented proteins

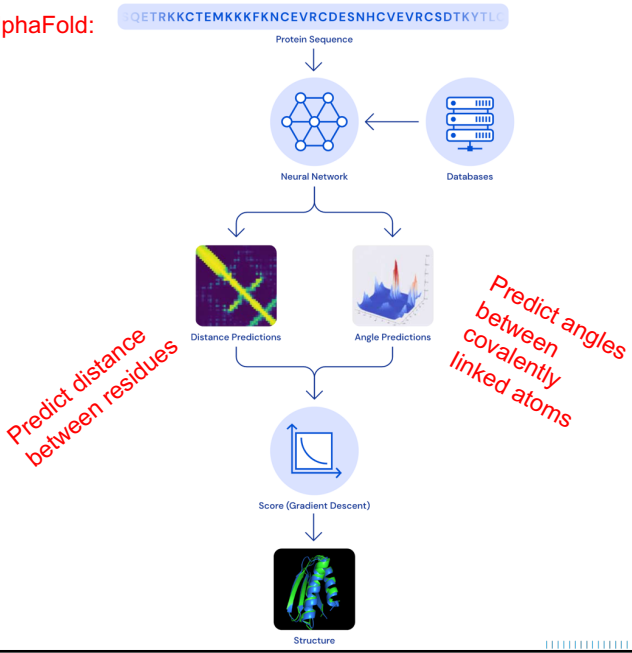
```
C T S Y P I K L M D F E R T S W Q A P R I M T G H K
C S S Y P I K L M D W E R T S W Q A P R I C T G Y K
C Q S Y P L K L M D F E R T S W Q V P R I P T G H K
C N S Y P L K L M D C E R T S W Q V P R I D T G C K
C S S Y P I K L M D F E R T S W Q A P R I F T G H K
C D S Y P V K L M D F E R T S W Q L P R I G T G H K
C C S Y P I K L M D K E R T S W Q A P R I M T G E K
C S S Y P A K L M D F E R T S W Q L P R I K T G H K
C T S Y P I K L M D D E R T S W Q A P R I L T G R K
```

Correlated mutations

Correlated mutations

And when there is no experimental structure? Deep Learning

AlphaFold:



Predict distance between residues:

Sequence of the target protein

↓

Sequence alignment of all apparented proteins

↓

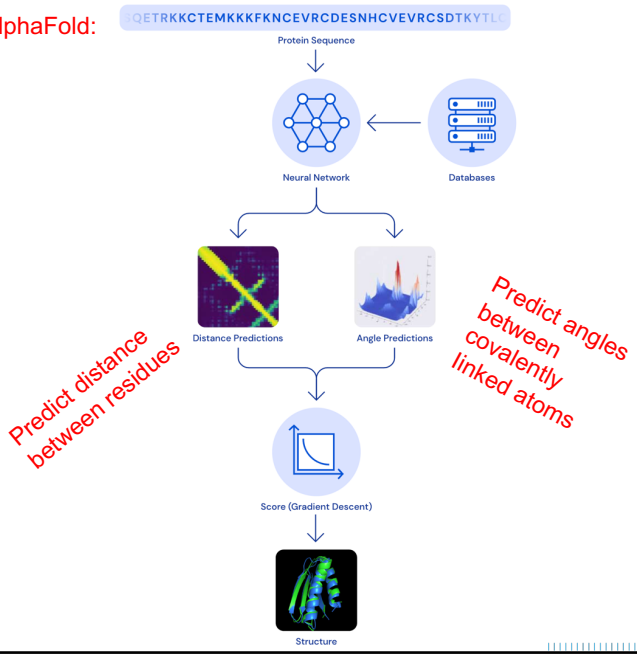
Correlations between residues in the sequence alignments indicate that these residues are close in 3D, even if they are distant in the sequence alignment

↓

Prediction of the distance between residues

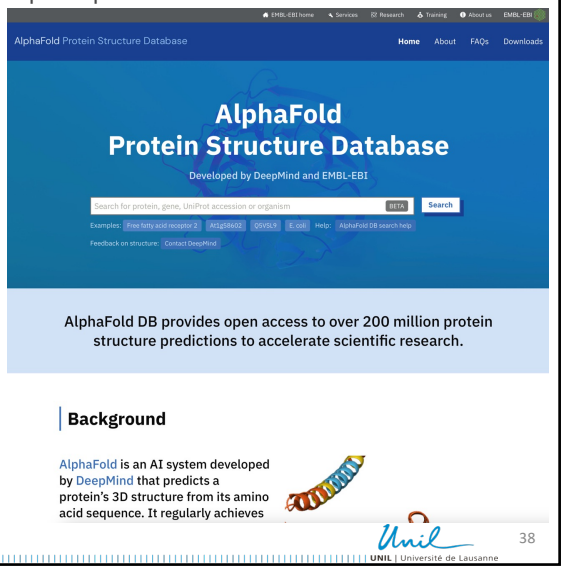
And when there is no experimental structure? Deep Learning

AlphaFold:



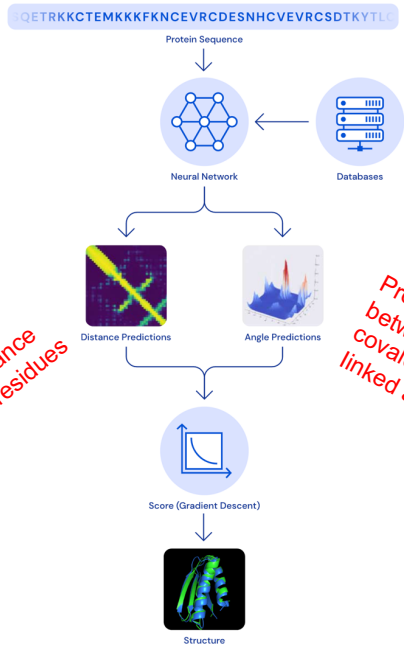
Free database of AlphaFold models:

https://alphafold.ebi.ac.uk



And when there is no experimental structure? Deep Learning

AlphaFold:



WARNINGS!!!!

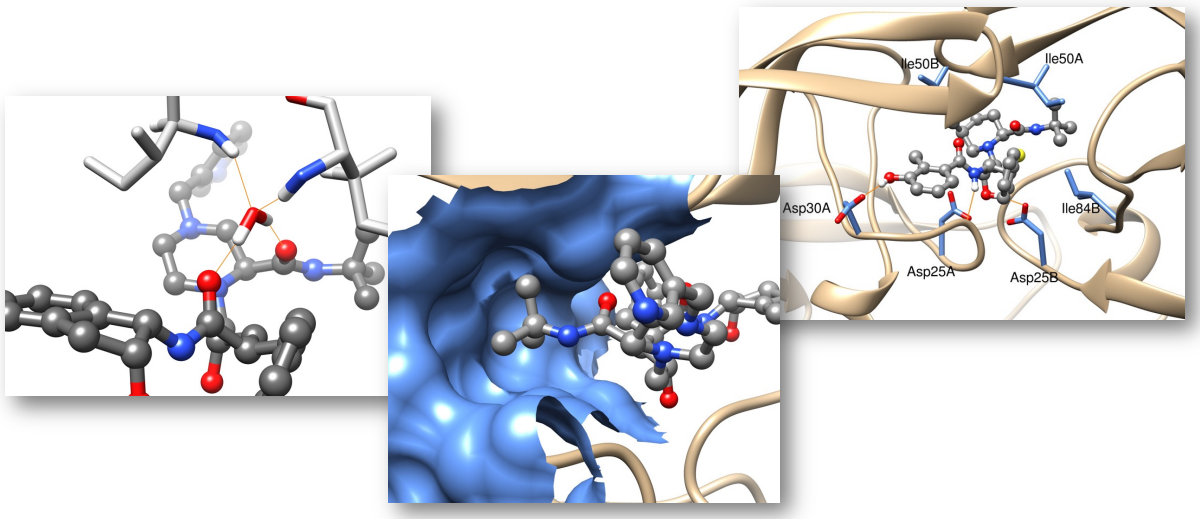
- AlphaFold models can contain errors and should be interpreted with care

- When existing, experimental structures should be prefered to AlphaFold models

- Although it can provide a good overall structure of a TCR or TCRpMHC, AlphaFold is not good at predicting the right conformation of the Complementary Determining Loops (see later)

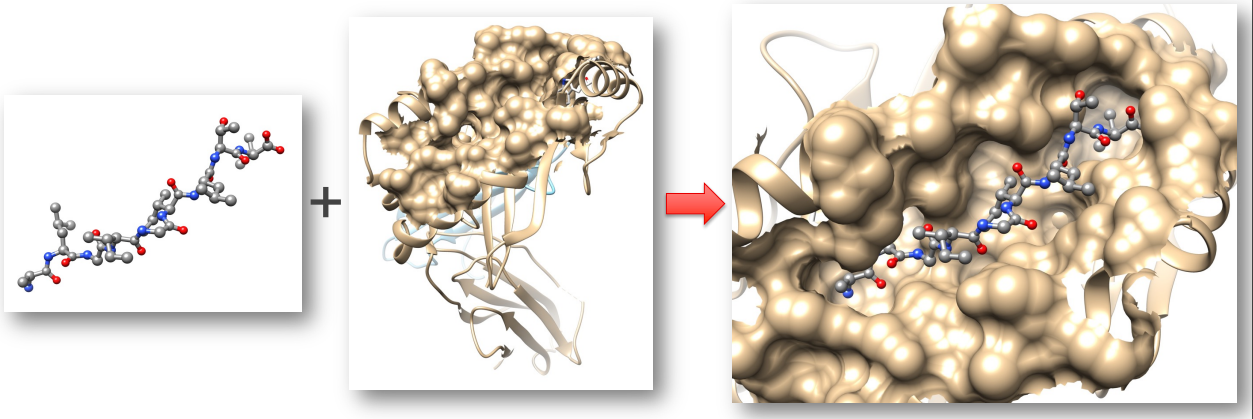
Possible to use dedicated tools, like
- TCRmodel2, https://tcrmodel.ibbr.umd.edu
- TCRdock, https://github.com/phbradley/TCRdock

Molecular Recognition



Molecular recognition

Molecular interactions → Molecular recognition ⇒ Biological response



41

Molecular recognition – Historical models

"Lock and key" model.
Emil Fischer in the 1890s.
The protein has a particular shape into which the ligand fits exactly.

Induced fit model
Daniel Koshland 1958.
The binding site of the macromolecule is flexible and its shape can be modified as the ligand interacts with it.

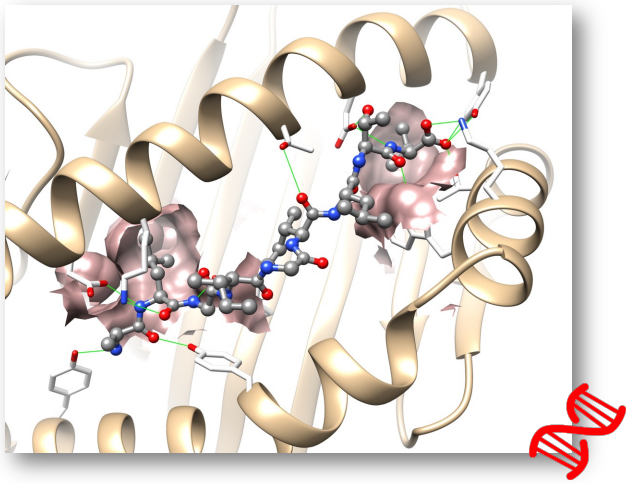
→ **Molecular recognition**:
Collection of **interactions** between molecules that govern their **binding**.

Qualitative **nature** of the interactions?
Quantitative **intensity** of the molecular recognition?

42

Molecular recognition - type of interactions

Non covalent interactions between atoms :
- non-polar interactions (shape recognition)
- electrostatic interactions (salt bridge and hydrogen bond)
- π interactions
- metal/ion interactions



Crystal structure of HLA-A2*0201 in complex with MART-1/Melan-A

Molecular recognition - type of interactions

Non Polar:

Ala, Val, Leu, Ile, Pro, Met, ~Cys

Polar:

Ser, Thr, Asn, Gln, Tyr, His, Trp, ~Cys

Aromatic:

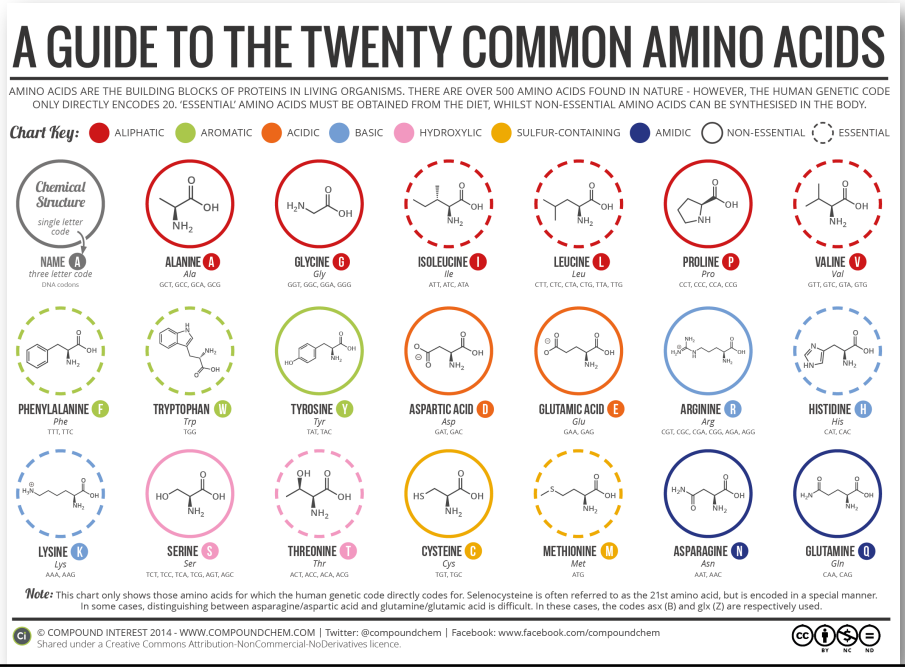
Phe, Tyr, Trp, His

Negatively charged:

Asp, Glu

Positively charged:

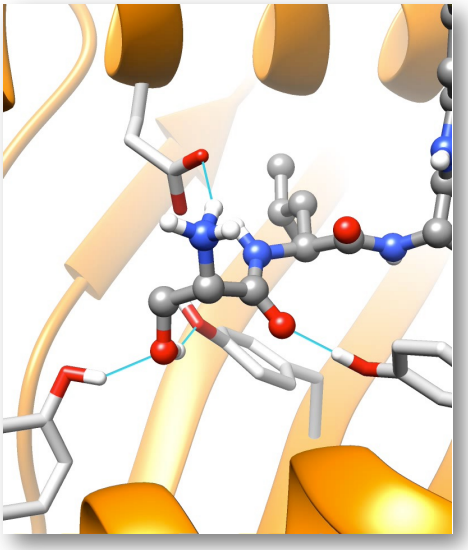
Arg, Lys, ~His



Molecular recognition – introduction to molecular mechanics

Structure determination

Biological events

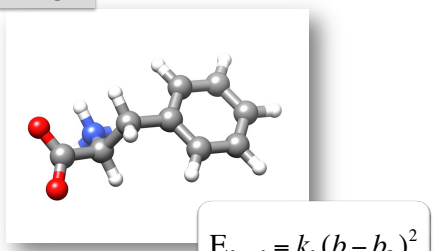


45

Molecular recognition – introduction to molecular mechanics

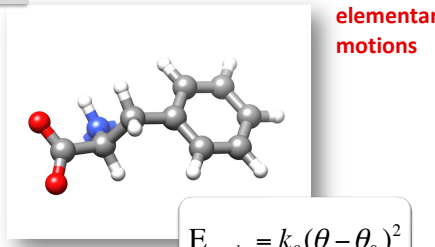
Molecular dynamics is decomposed into elementary motions

Bond length



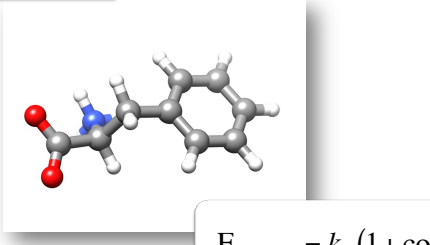
$$E_{bond} = k_b(b - b_0)^2$$

Bond angle



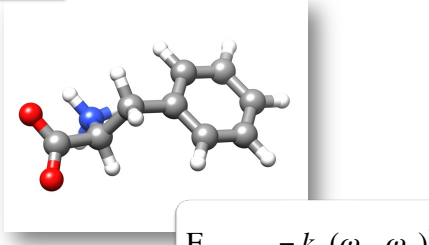
$$E_{angle} = k_\theta(\theta - \theta_0)^2$$

Dihedral angle



$$E_{dihedral} = k_\varphi\left(1 + \cos(n\varphi - \delta)\right)$$

Improper angle



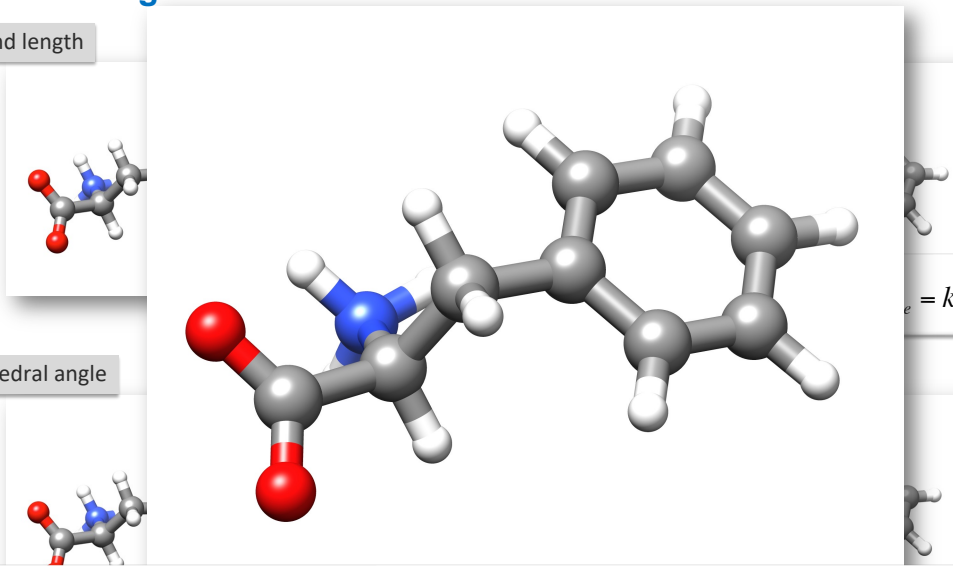
$$E_{improper} = k_\omega(\omega - \omega_0)^2$$

46

Molecular recognition – introduction to molecular mechanics

Molecular dynamics is decomposed into elementary motions

Bond length



$= k_\theta(\theta - \theta_0)^2$

Dihedral angle

$$E_{bonded} = \sum_{bonds} k_b(b - b_0)^2 + \sum_{angles} k_\theta(\theta - \theta_0)^2 + \sum_{dihedrals} k_\varphi(1 + \cos(n\varphi - \delta)) + \sum_{impropers} k_\omega(\omega - \omega_0)^2$$

$\omega_0)^2$

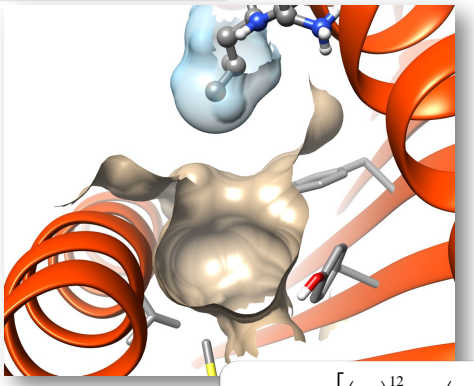
47

47

Molecular recognition – Molecular interactions

Molecular recognition is driven by non-polar and electrostatic interactions

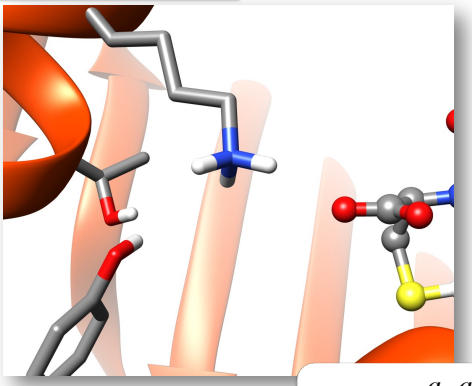
Non-polar interactions



$$E_{vdW} = \varepsilon\left[\left(\frac{r_m}{r_{ij}}\right)^{12} - 2\left(\frac{r_m}{r_{ij}}\right)^{6}\right]$$

Shape complementarity

Electrostatic interactions



$$E_{elec} = \frac{q_i\, q_j}{4\pi\varepsilon_0\varepsilon\, r_{ij}}$$

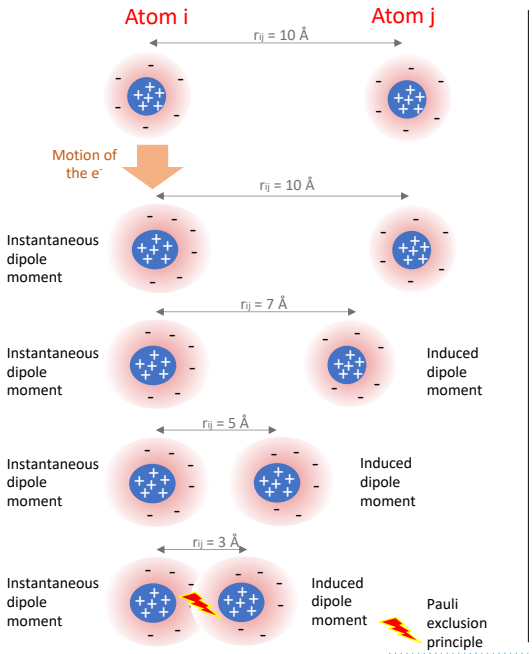
Specificity

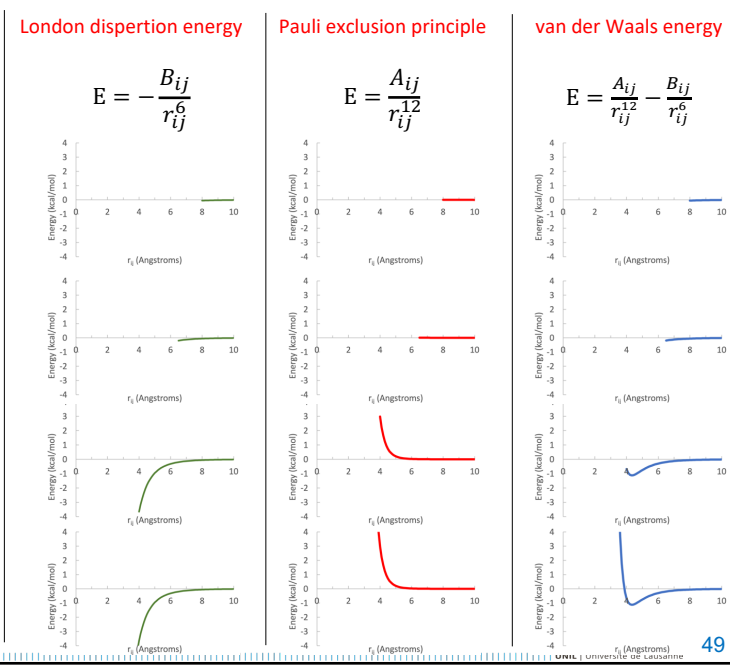
Unil UNIL | Université de Lausanne

48

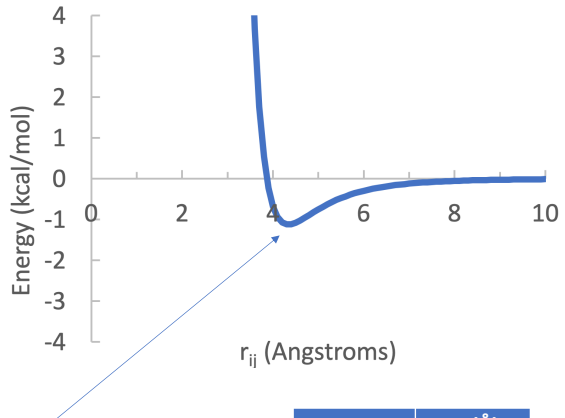
48

Molecular recognition – Van der Waals interactions





Molecular recognition – Van der Waals interactions



Optimum at $r_{ij} = R_{i,vdw} + R_{j,vdw}$

R_{vdw}: van der Waals radius

Optimal interaction when atoms are « touching » each other

Atom	R_{vdw} (Å)
Hydrogen	1.2
Carbon	1.7
Nitrogen	1.55
Oxygen	1.52
Sulfur	1.8

Described by the Lennard-Jones potential

$$E = \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6}$$

Interaction energy follows $1/r^6$ and $1/r^{12}$

Short range interaction
Typically 3.5 Å

The optimal energy is weak between a given pair of atoms (Typically 0.5 kcal/mol)

However it is **cumulative** over all atoms involved in molecular recognition

Molecular recognition – Van der Waals interactions

Do not require charges or partial charges on atoms

↳ van der Waals interactions are considered as **non-polar interactions**
… even though they are electrostatic by nature

Interactions particularly important for non-polar residues:
- Alanine, Valine, Leucine, Isoleucine, Proline
- Cysteine, Methionine
- Phenylalanine, Tyrosine, Tryptophan

Molecular recognition – Van der Waals interactions

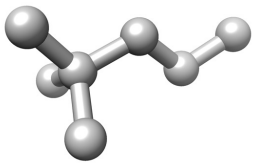
Each atom tries to be positioned at optimal distance from its neighbors

2 atoms

3 atoms

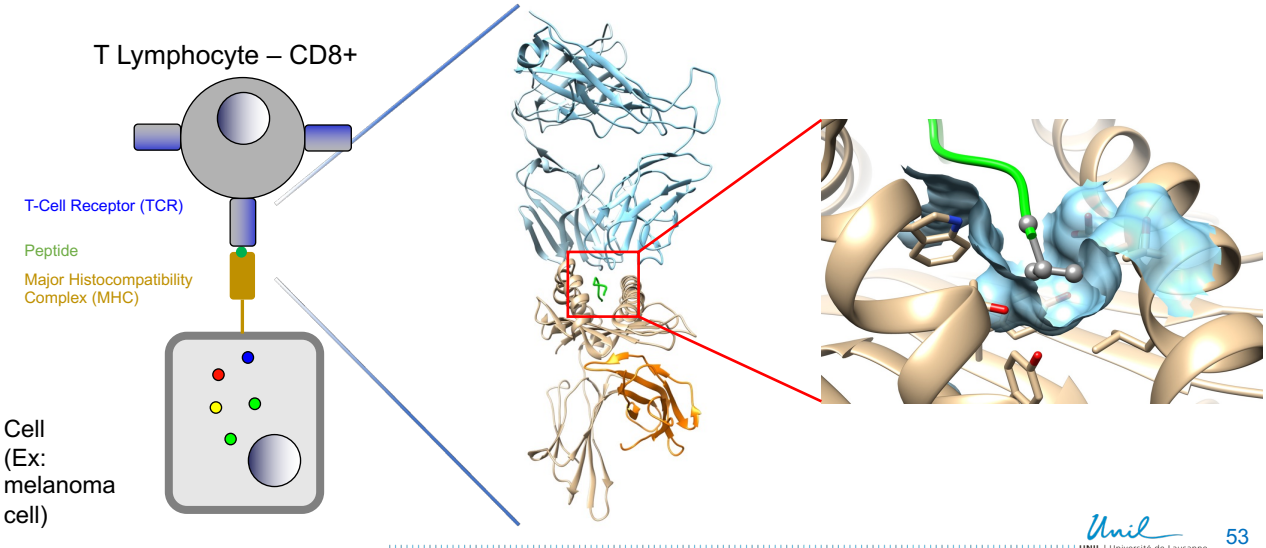
4 atoms

However, in molecules, atoms are also linked via covalent bonds, which force a geometry…



Molecular recognition – Van der Waals interactions

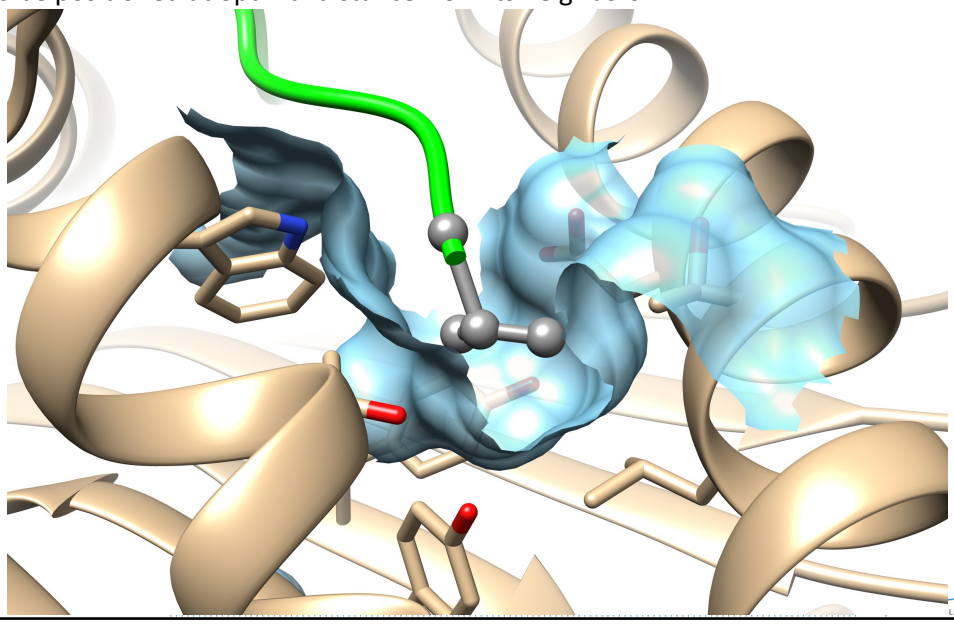
Each atom tries to be positioned at optimal distance from its neighbors



53

Molecular recognition – Van der Waals interactions

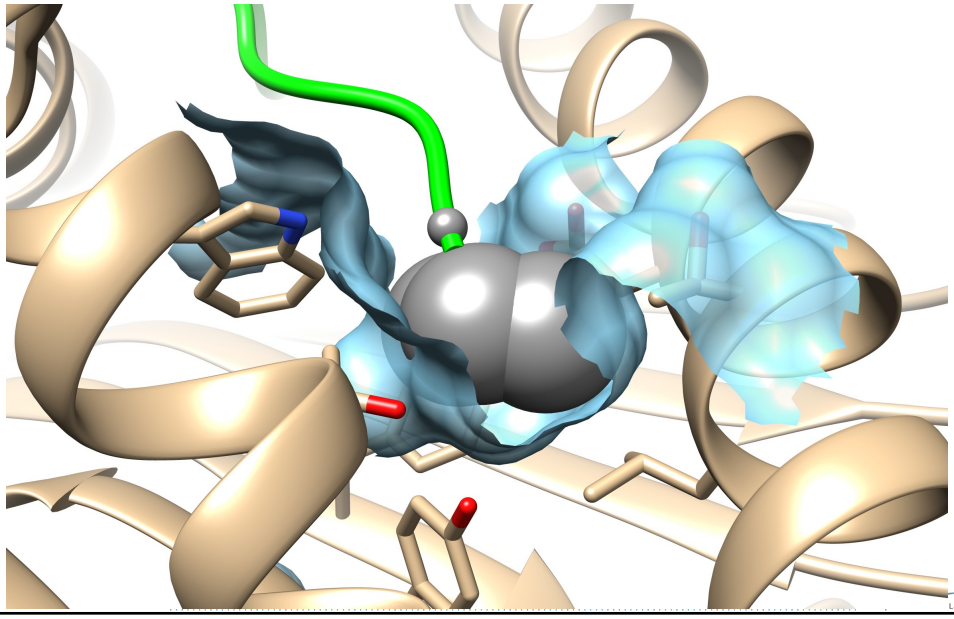
Each atom tries to be positioned at optimal distance from its neighbors



54

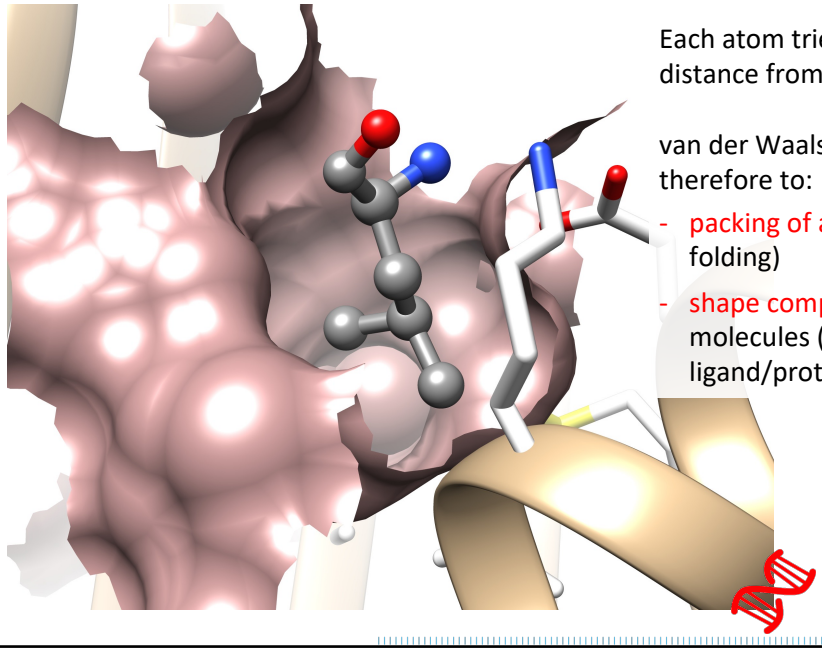
Molecular recognition – Van der Waals interactions

Each atom tries to be positioned at optimal distance from its neighbors



55

Molecular recognition – Van der Waals interactions



Each atom tries to be positioned at optimal distance from its neighbors

van der Waals interactions contribute therefore to:

- packing of atoms (and macromolecule folding)
- shape complementarity between binding molecules (example: protein/protein or ligand/protéine complexes)

56

Molecular recognition – Electrostatic interactions

The interaction between two point charges in a uniform medium is described by the **Coulomb law**

Coulomb energy

$$E_{Coul} = \frac{1}{4\pi\varepsilon_o\varepsilon}\frac{q_i q_j}{r_{ij}}$$

ε_o : dielectric constant of vacuo

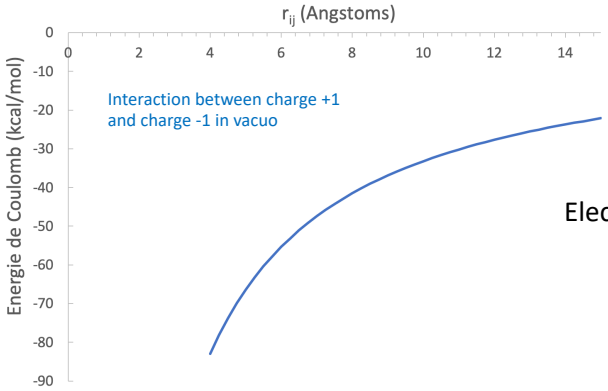
$\frac{1}{4\pi\varepsilon_o}$ =332 (kcal/mol) Å/q_e^2

ε: dielectric constant of medium

ex: $\varepsilon_{(vacuo)} = 1$; $\varepsilon_{(water)} = 80$

Interaction between charges +1 et -1 at 5 Å :

- -66 kcal/mol in vacuo
- -0.8 kcal/mol in water



Electrostatic interaction energy follows a 1/r expression

↳ **Long range interaction**

57

Molecular recognition – Electrostatic interactions

Electrostatic interactions can involve:

- Integer charge – integer charge

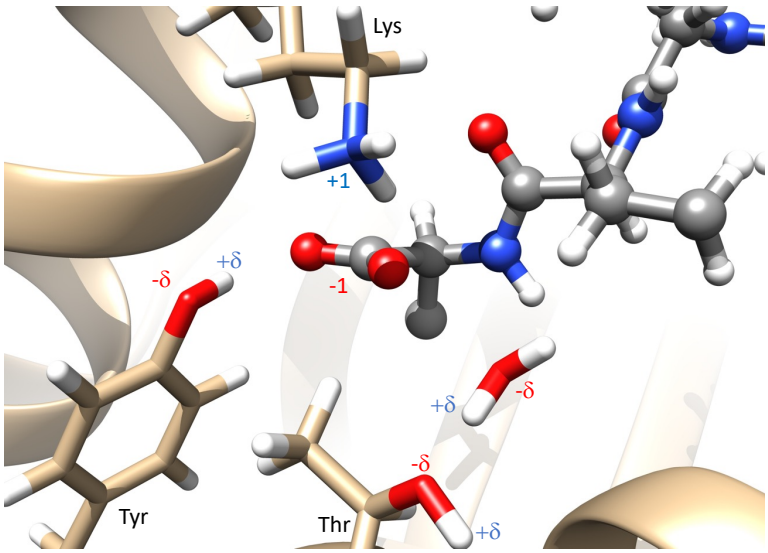
  Called ionic interactions.
  At short distance (~ 4/5 Å), ionic interactions are called salt bridges.

- Integer charge – permanent dipole

  Ex: charged assisted hydrogen bond

- Permanent dipole – permanent dipole

  Ex: hydrogen bond

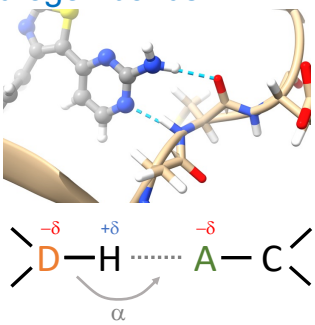


58

Molecular recognition – Electrostatic interactions – Hydrogen bonds

Typically between two dipoles:

- D-H where D is the hydrogen bond **donor**
- A-C where A is the hydrogen bond **acceptor** and C a carbon atom



Extremely frequent in proteins and nucleic acids

Important factor of the architecture of bio-macromolecules

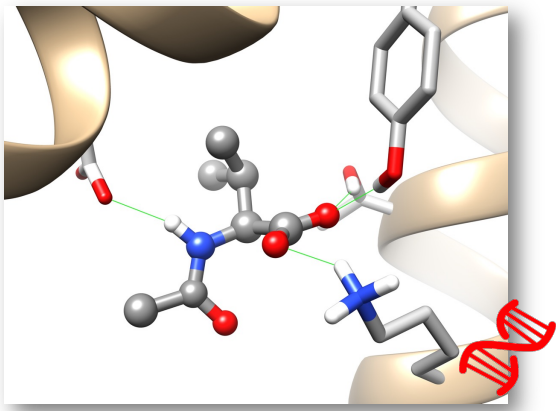
Typical distances in hydrogen bonds:

- Between H and A : ~ 1.95 Å
- Between A and D : O – O : 2.50 – 2.70 Å
  O – N : 2.75 – 2.85 Å
  N – N : 2.70 – 3.00 Å

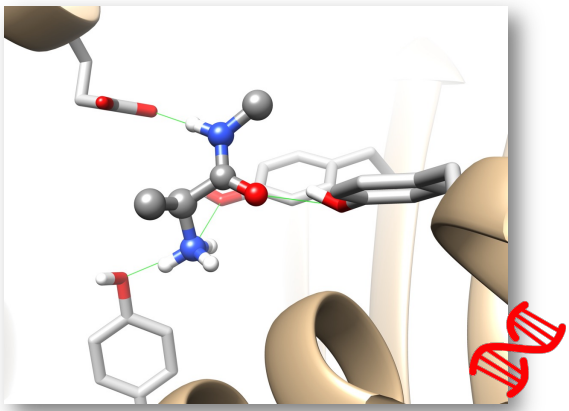
Angle α depends on atom types and atom hybridization

59

Molecular recognition – Electrostatic interactions – Hydrogen bonds



Ex: H-bonds between residue Val9 of MART-1/Melan-A and pocket F of HLA-A2*0201



Ex: H-bonds between residue Ala1 of MART-1/Melan-A and pocket A of HLA-A2*0201

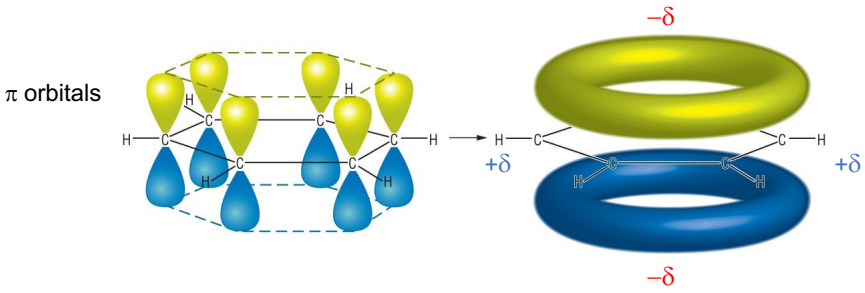
Electrostatic interactions are **local and directional** (H-bonds even more than salt bridges)

→ **Directionality / locality of interactions**
Specificity of molecular recognition

60

Molecular recognition – π interactions

Electronic structure of benzene:

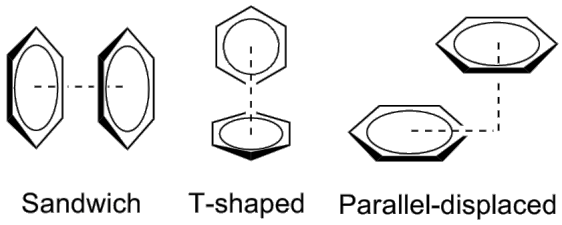


Aromatic cycles (Phenyl, Tyrosine, Tryptophan & Histidine) can interact with:

- Other aromatic cycles (stacking)
- Metals
- Polar groups
- Hydrogen bond donors

61

Molecular recognition – π interactions

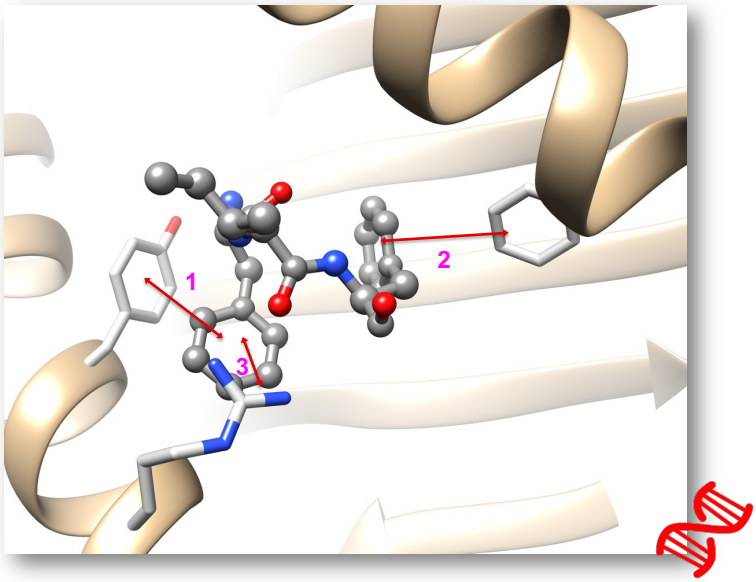


(source: Wikipedia)

T-shaped and parallel-displaced π-π interactions are the most frequent

62

Molecular recognition – π interactions



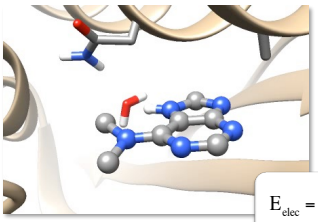
1. π - stacking
2. T – staking (the two aromatic cycles are orthogonal)
3. Cation - π interaction

Ex: π interactions between murine coronavirus epitope RCFIFANI and H-2Kb (4PV8 in PDB)

Molecular recognition – Other factors

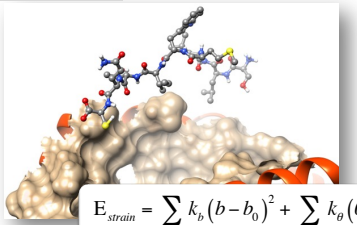
Many other factors impact the molecular recognition and binding affinity

Water bridges



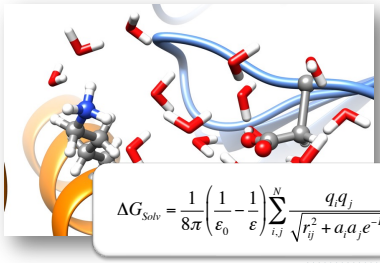
$$E_{elec} = \frac{q_i\, q_j}{4\pi\varepsilon_0 \varepsilon\, r_{ij}}$$

Conformational changes



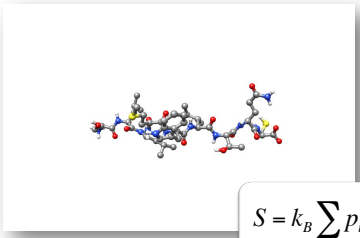
$$E_{strain} = \sum_{bonds} k_b (b - b_0)^2 + \sum_{angles} k_\theta (\theta - \theta_0)^2 + \ldots$$

Desolvation and elec. shielding



$$\Delta G_{Solv} = \frac{1}{8\pi}\left(\frac{1}{\varepsilon_0} - \frac{1}{\varepsilon}\right)\sum_{i,j}^{N} \frac{q_i q_j}{\sqrt{r_{ij}^2 + a_i a_j e^{-D}}} \quad , \quad D = \left(\frac{r_{ij}}{2\sqrt{a_i a_j}}\right)^2$$

Entropy changes

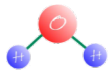


$$S = k_B \sum p_i \ln(p_i)$$

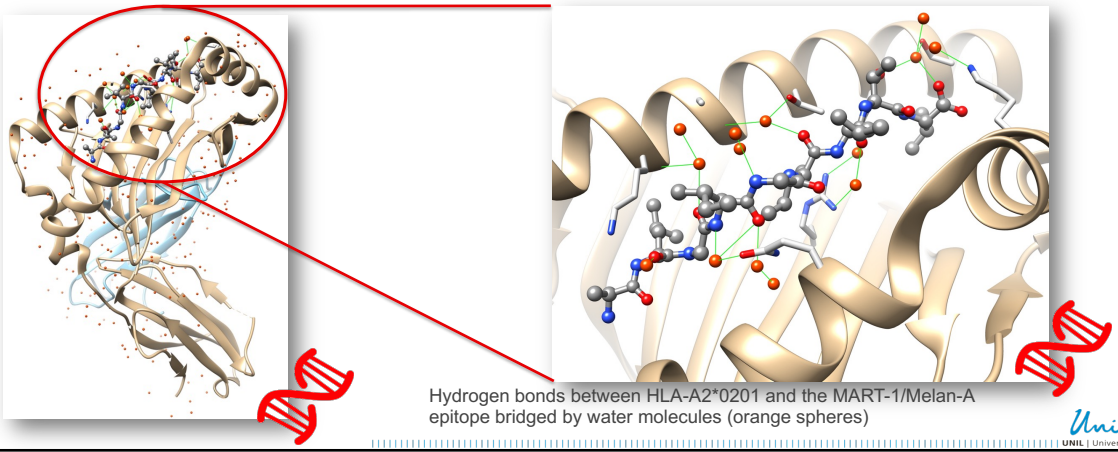
Molecular recognition – Other factors – Water

Molecular recognition between small molecule and protein takes place in an **aqueous environment.**

Discrete water molecules

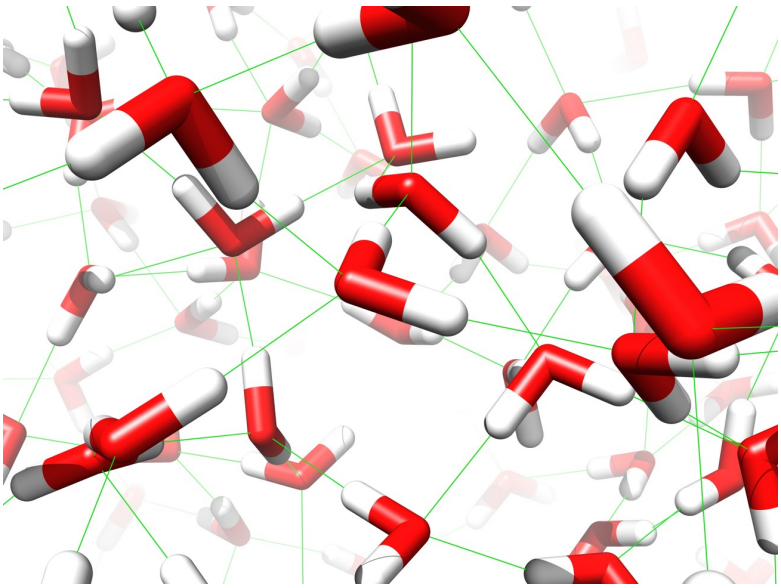


- Bridge interactions through H-bonds or OH... π ➔ favorable to binding.
- Displacement from the protein cavity ➔ favorable to binding.



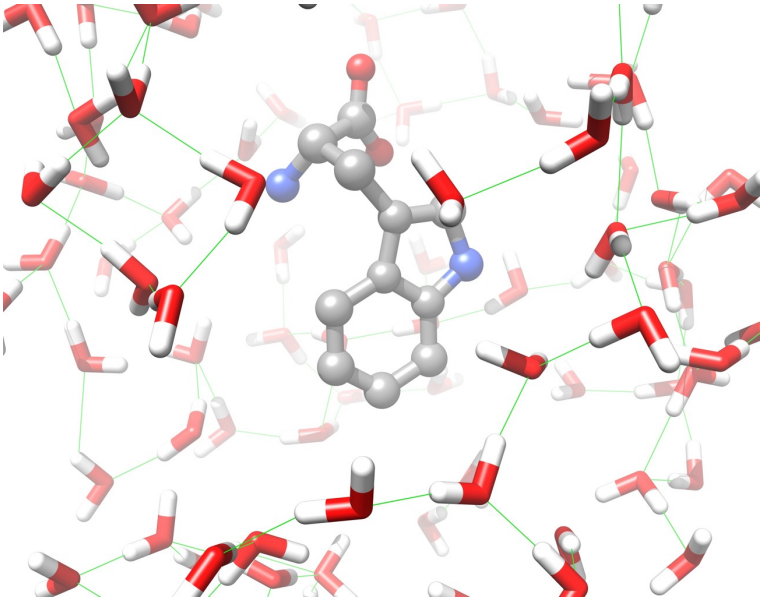
Hydrogen bonds between HLA-A2*0201 and the MART-1/Melan-A epitope bridged by water molecules (orange spheres)

Molecular recognition – Other factors – Water – Hydrophobic effect



Water structure is stabilized by hydrogen bonds and dipole interactions

Molecular recognition – Other factors – Water – Hydrophobic effect

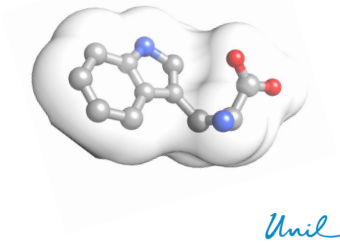


The presence of a solute decreases water-water interactions

Non-polar solvation energy is proportional to the solvent accessible surface area (SASA) for large molecules:

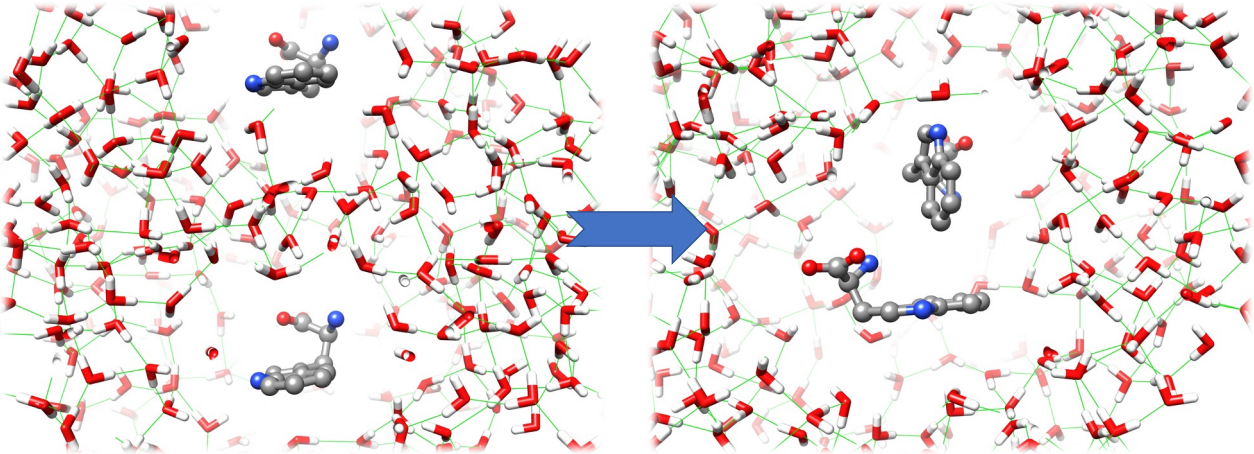
$$E = \sigma \times SASA$$

σ = 0.025 kcal/Å^2



Molecular recognition – Other factors – Water – Hydrophobic effect

Solutes aggregate to limit their deleterious on water structure

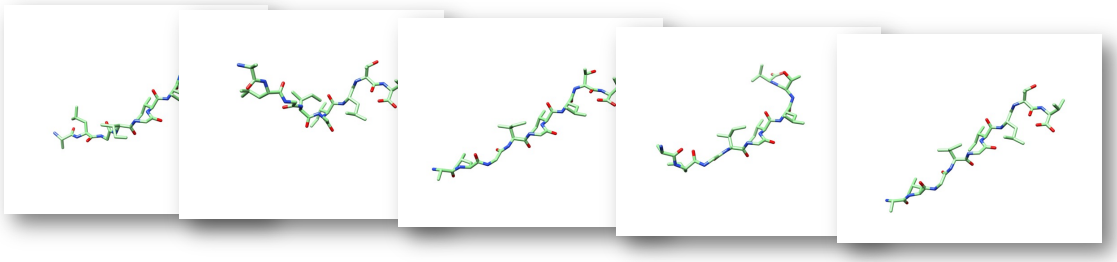


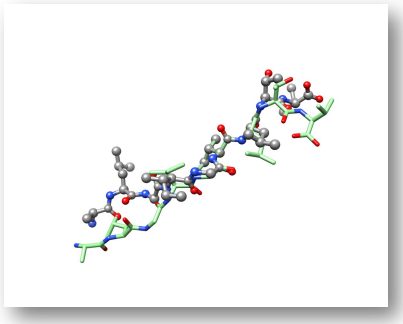
Energy of non-polar desolvation: $\Delta G_{np} = \sigma \times \Delta SASA$

The solvent-accessible surface area of aggregated solutes is lower than the sum of those of the separated solutes (ΔSASA<0). ΔG_{np} is therefore favorable to aggregation (binding of solutes)

Molecular recognition – Other factors – Conformational changes

Molecules have many conformations (conformers)





Peptide **bioactive** conformation **(geometry as bound to the protein)**

does **NOT** correspond to

Lowest energy conformation **(most stable geometry in solution)**

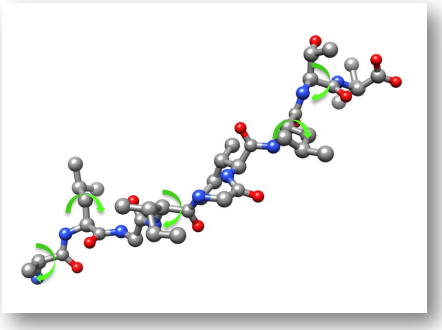
BUT is a low energy conformation (within 3 to 5 kcal/mol)

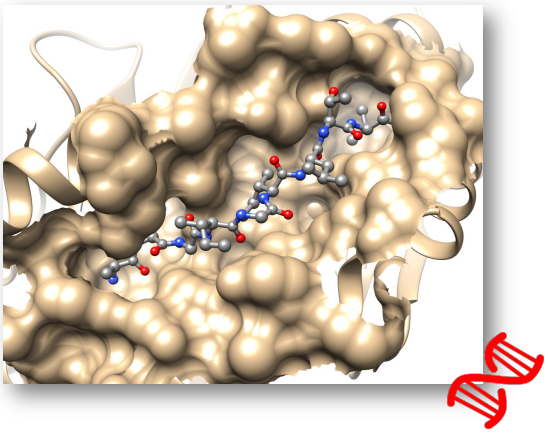
Bioactive conformation (in protein)
Lowest energy conformation (in solution)

69

Molecular recognition – Other factors – Entropy changes

Entropy is a measure of disorder. Nature likes disorder!
Loss of entropic energy when entropy (disorder) decreases.
Gain of entropic energy when entropy (disorder) increases.





Two main events **upon ligand binding** to protein:

- **Conformational** degrees of **freedom** (rotatable bonds) are **blocked**: **unfavorable!**
- **Water** molecules are **kicked-out** from the protein binding site **to bulk**: **favorable!**

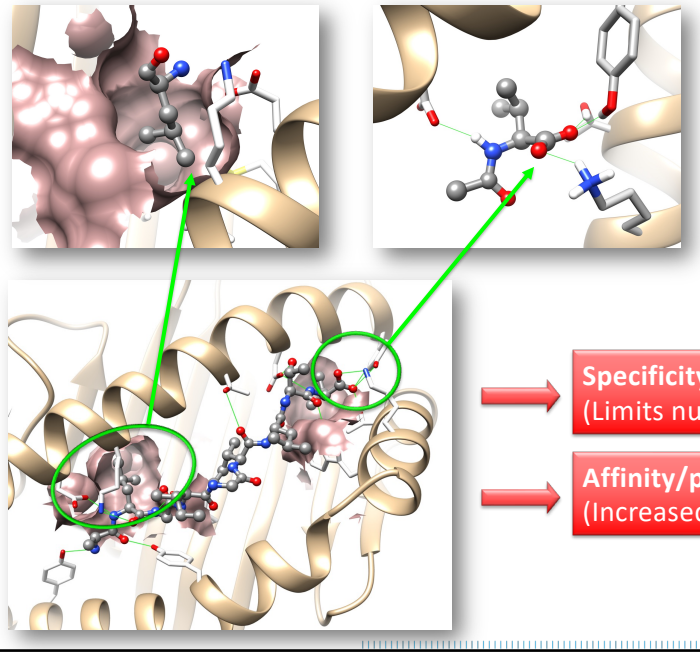
70

Molecular recognition – Summary

Category	Interaction	Distance	Residues involved	Remarks
Electrostatic	Ionic (charge-charge)	Long range	Arg, Lys, Asp, Glu His (if charged)	Called salt bridge at short distance
	Hydrogen bond	Short range	Arg, Lys, Asp, Glu His, Tyr Ser, Thr, Asn, Gln Cys	Directionality / locality of interactions Specificity of molecular recognition
	π interaction	Short range	Phe, Tyr, Trp, His	
Electrostatic/Non-polar	Van der Waals	Short range	Ala, Val, Ile, Leu, Pro, Cys, Met Phe, Tyr, Trp, His	Packing of atoms Shape complementarity
Non-polar	Hydrophobic effect	-	All	Solute aggregation

71

Molecular recognition – Potency and specificity



Various and numerous ligand-protein interactions:

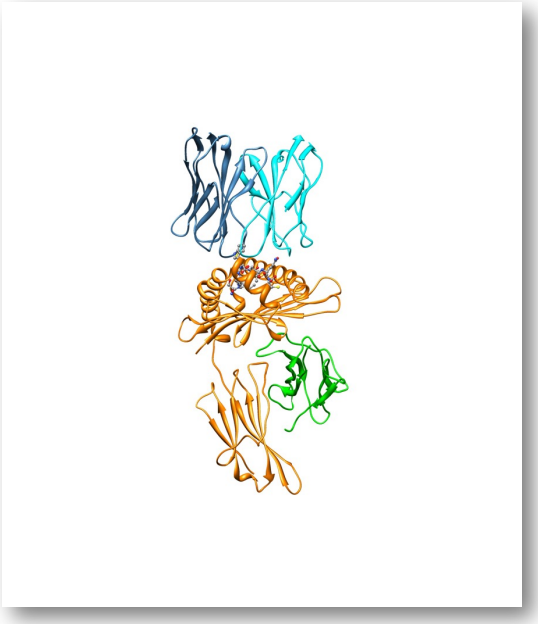
- local and directional interactions
- shape complementarity

→ **Specificity**
(Limits number/nature of possible epitopes)

→ **Affinity/potency**
(Increased epitope recognition)

72

Molecular recognition - Molecular Mechanics - Molecular Dynamics



- Adding explicit droplet of water:

System solvated with explicit water molecules (TIP3P model):
- **~ 29,500** water molecules
- **~ 100,000** atoms in total

- Molecular Dynamics (**MD**)

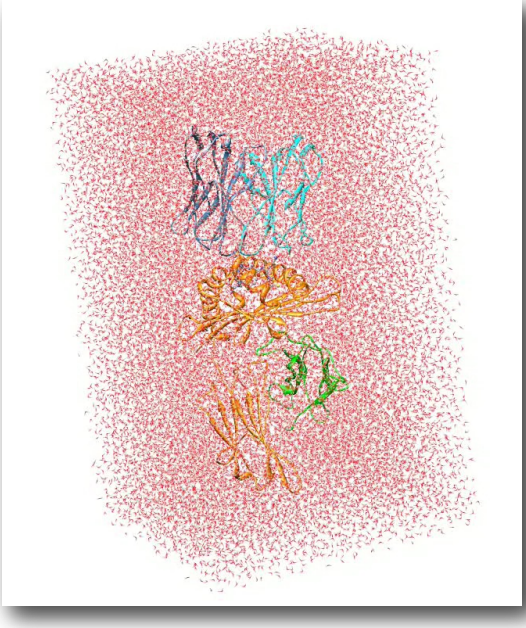
Atom motions are calculated to follow Newton's equation of motion,
at **300 K** and **1 atm**.

Typical simulation times: from **0.5 ns to ~ 100 ns** (10^{-9} s).

➔ **Simulation closer to physiological reality, but more computationally intensive**

73

Molecular recognition - Molecular Mechanics - Molecular Dynamics



- Adding explicit droplet of water:

System solvated with explicit water molecules (TIP3P model):
- **~ 29,500** water molecules
- **~ 100,000** atoms in total

- Molecular Dynamics (**MD**)

Atom motions are calculated to follow Newton's equation of motion,
at **300 K** and **1 atm**.

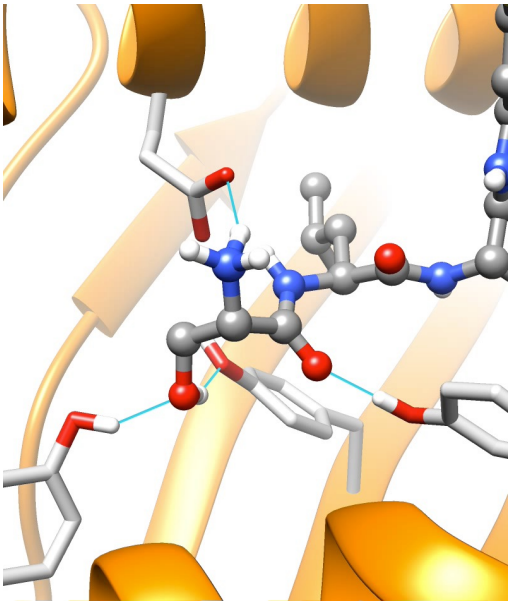
Typical simulation times: from **0.5 ns to ~ 100 ns** (10^{-9} s).

➔ **Simulation closer to physiological reality, but more computationally intensive**

74

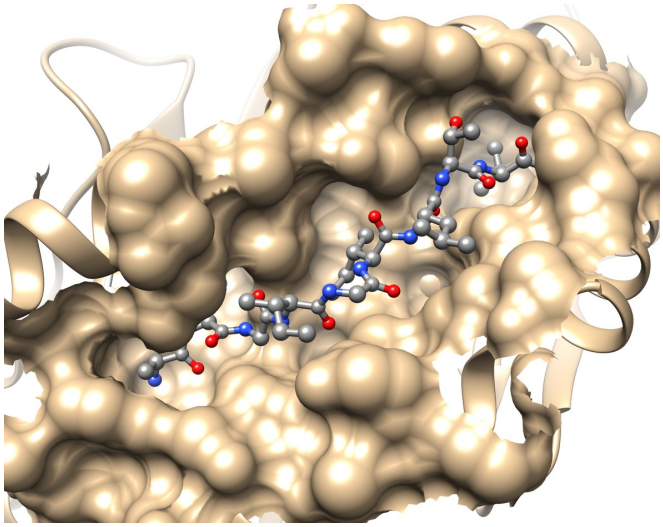
Molecular recognition - Molecular Mechanics - Molecular Dynamics

Typical motions in a peptide/MHC complex at room temperature:



75

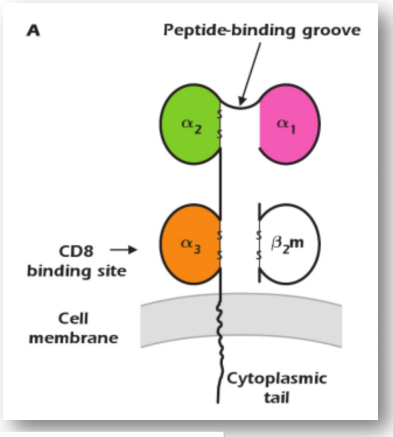
3D structure of peptides bound to the Major Histocompatibility Complex Class I (pMHC)

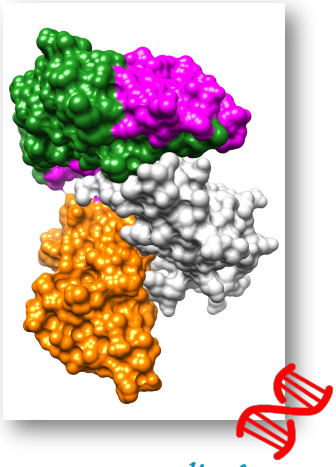


76

Structure of pMHC Class I

- Transmembrane glycoprotein (43 kDa)
- 3 domains α_1, α_2 and α_3
- Deep groove to bind a peptide (8-10 residues)
- Expressed at cell surface
- Non-covalent association with β_2 microglobulin (β_2m)





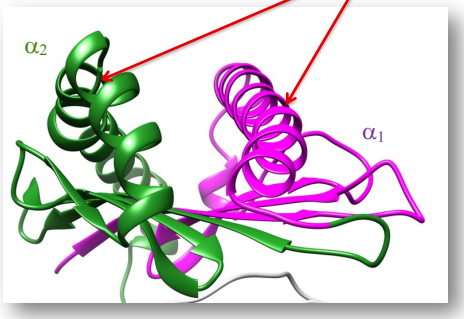
From "Immunology: A Short Course" by Richard Coico, Geoffrey Sunshine

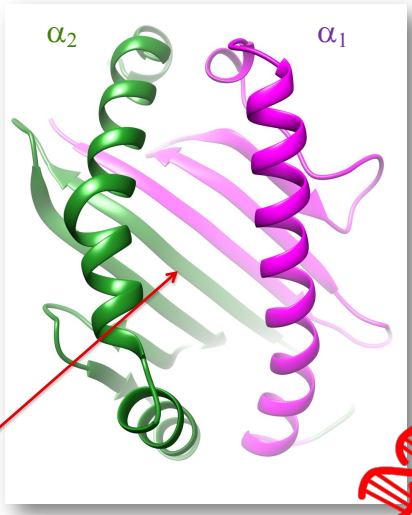
77

77

Structure of pMHC Class I - The binding pocket

Two α helices



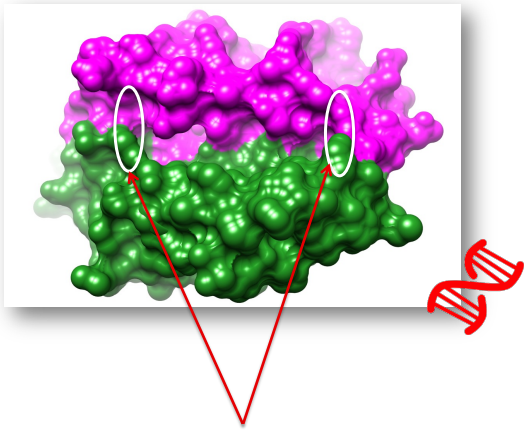


β sheet formed of 8 β strands

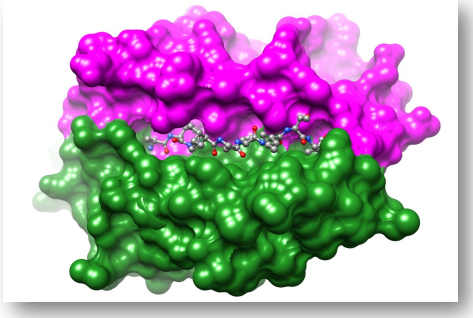
78

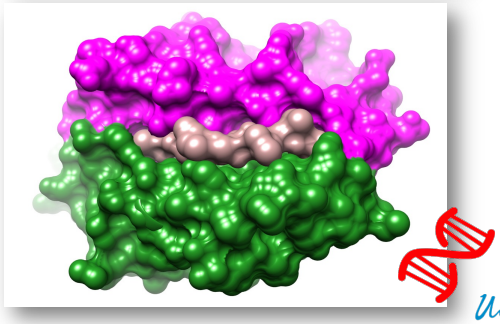
78

Structure of pMHC Class I - The binding pocket



Walls defining the width of the pocket



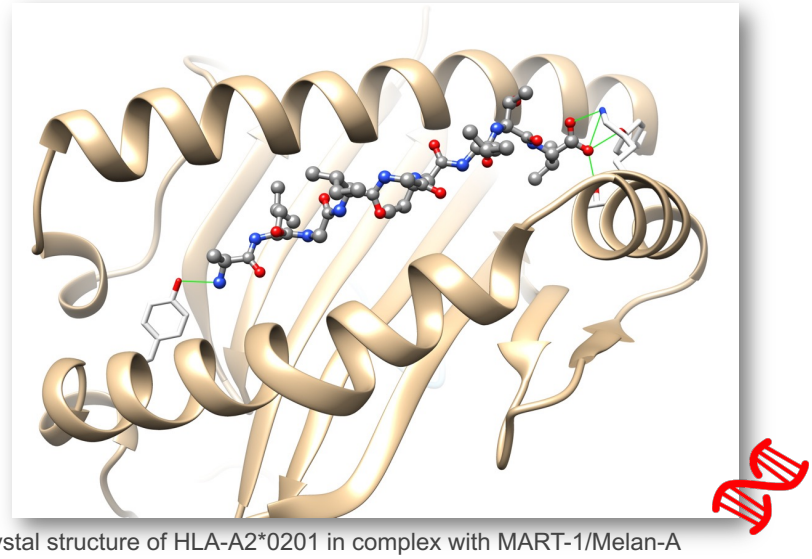


Binding peptides limited to 8 – 10 residues

79

Structure of pMHC Class I - Peptide binding to MHC groove

N and C-termini of the peptide are stabilized by H-bonds networks



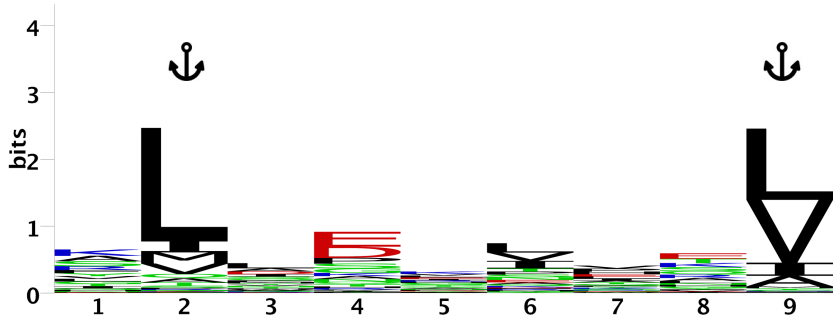
Crystal structure of HLA-A2*0201 in complex with MART-1/Melan-A

80

Structure of pMHC Class I - Peptide binding to MHC groove

- The specificity of the peptide binding is provided by the anchor residues, i.e. invariant or closely related residues in the peptide sequence.
- A peptide binding to MHC Class I has typically 2 main anchor residues and 2/3 secondary anchor residues.
- Other positions are variable.

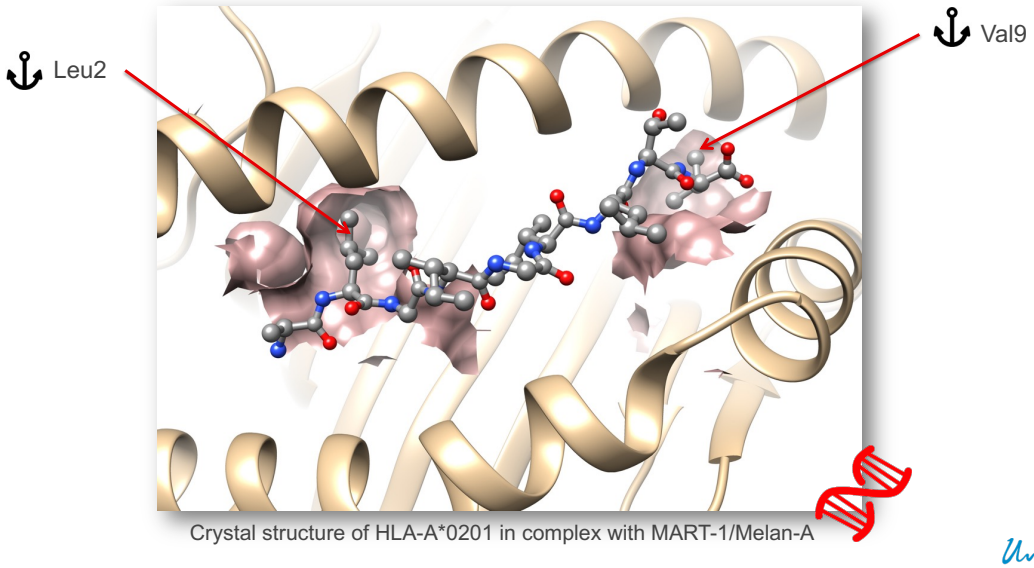
Sequence logo for **HLA-A2*02:01**



81

Structure of pMHC Class I - Peptide binding to MHC groove

Anchor residues make strong interactions with the MHC binding pockets

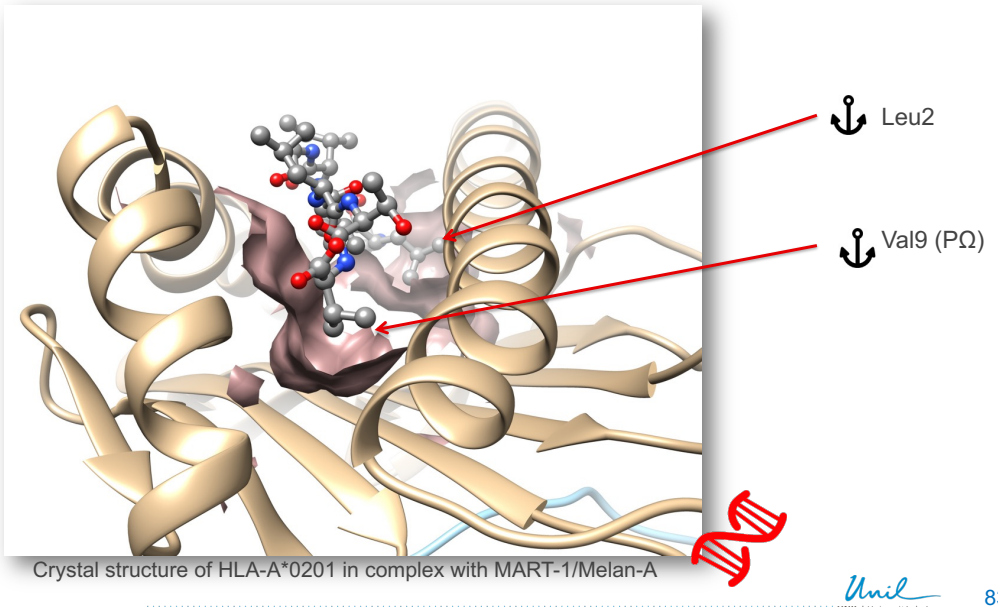


Crystal structure of HLA-A*0201 in complex with MART-1/Melan-A

82

Structure of pMHC Class I - Peptide binding to MHC groove

Anchor residues make strong interactions with the MHC binding pockets

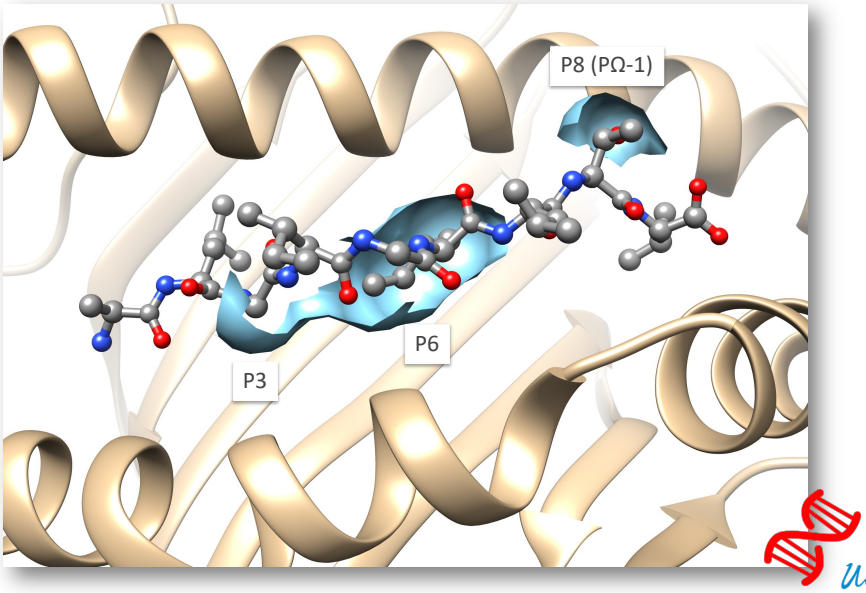


Crystal structure of HLA-A*0201 in complex with MART-1/Melan-A

83

Structure of pMHC Class I - Peptide binding to MHC groove

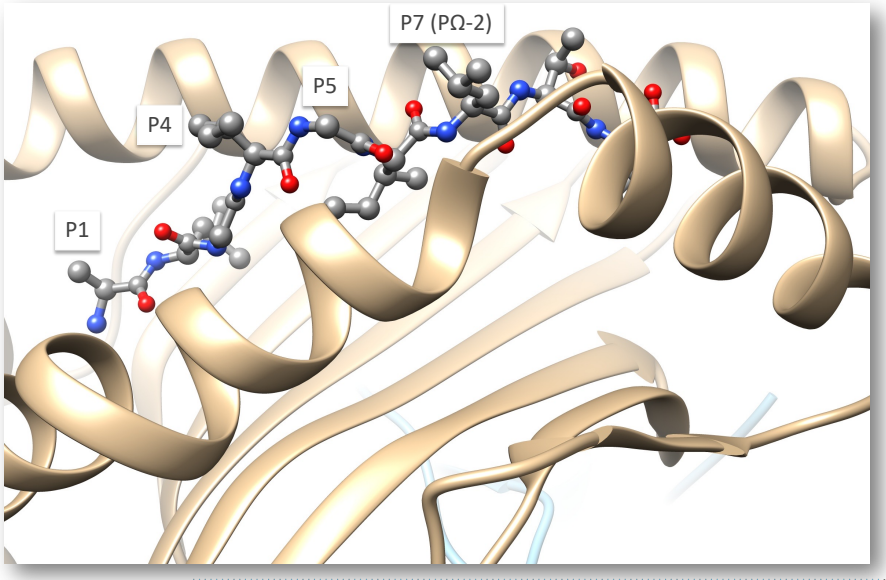
Secondary anchor residues also make interactions with the MHC binding pockets, but are more variable



84

Structure of pMHC Class I - Peptide binding to MHC groove

Other positions are facing the solvent (or the TCR) and are variable



85

Structure of pMHC Class I - Peptide binding to MHC groove

Exercise 4 of the booklet

- Can you explain the difference in binding affinities of these two peptides for HLA-A*0201?

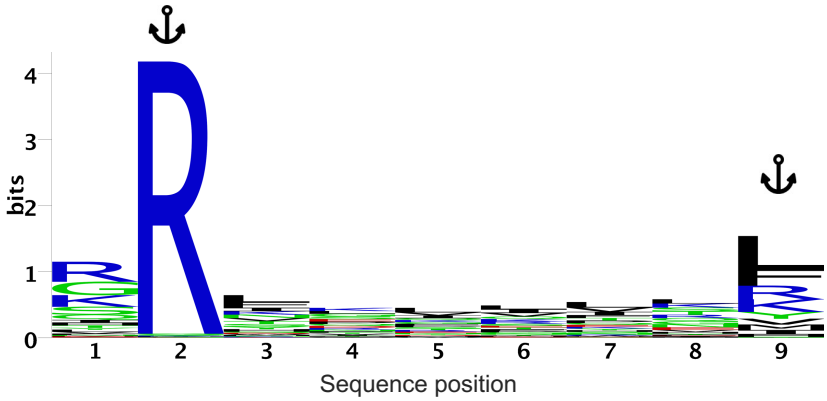
  - AAGIGILTV : 60 µM. PDB ID.: 2GUO
  - ALGIGILTV : 1.5 µM. PDB ID.: 1JHT

86

Structure of pMHC Class I - Peptide binding to MHC groove

Exercise 5 of the booklet

Sequence logo for **HLA-B*2705**:

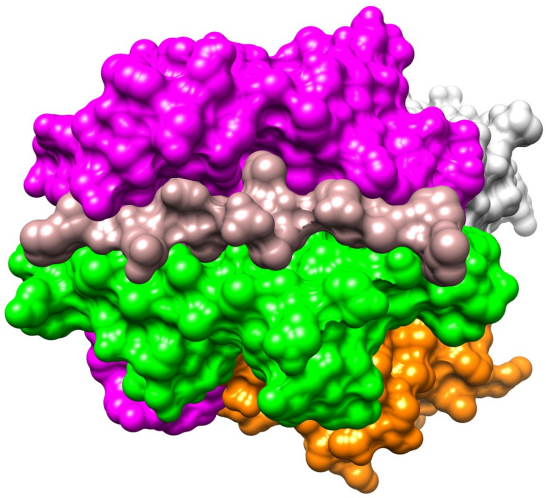


Bassani-Sternberg, […] Gfeller. (2017). *PLoS Computational Biology*, *13*(8), e1005725.

- What are the preferred amino-acids for the two anchor residues?
- Is it different from that of HLA-A*0201?
- Using the PDB and UCSF Chimera, can you explain this?

87

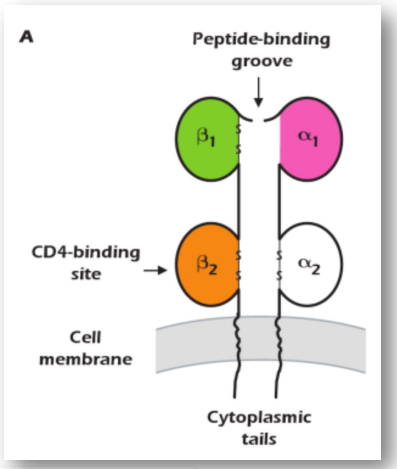
3D structure of peptides bound to the Major Histocompatibility Complex Class II

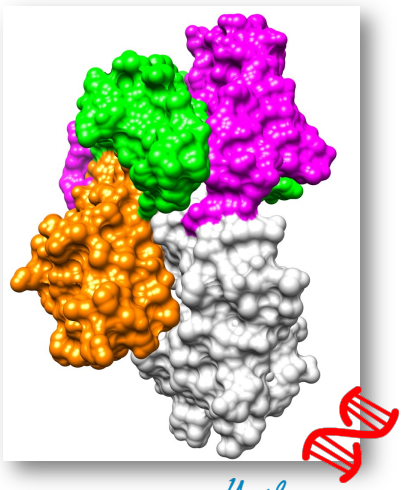


88

Structure of pMHC Class II

- Transmembrane glycoprotein
- 2 chains α (35 kDa) and β (28 kDa), forming 4 domains, α_1, α_2, β_1 and β_2.
- Deep groove to bind a peptide (12-18 residues) between α_1 and β_1
- Expressed at cell surface

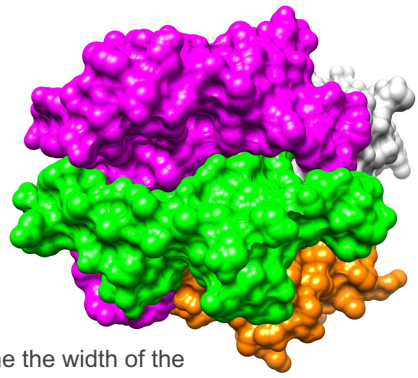




From "Immunology: A Short Course" by Richard Coico, Geoffrey Sunshine

89

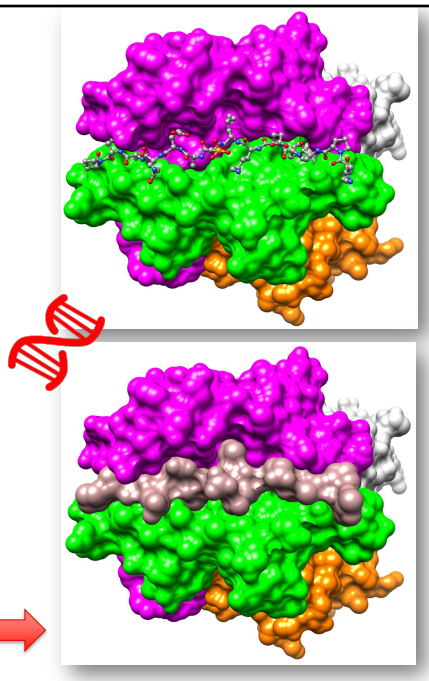
Structure of pMHC Class II – The binding groove



No walls to define the width of the pocket

Can adapt very large peptides, 12 – 18 residues

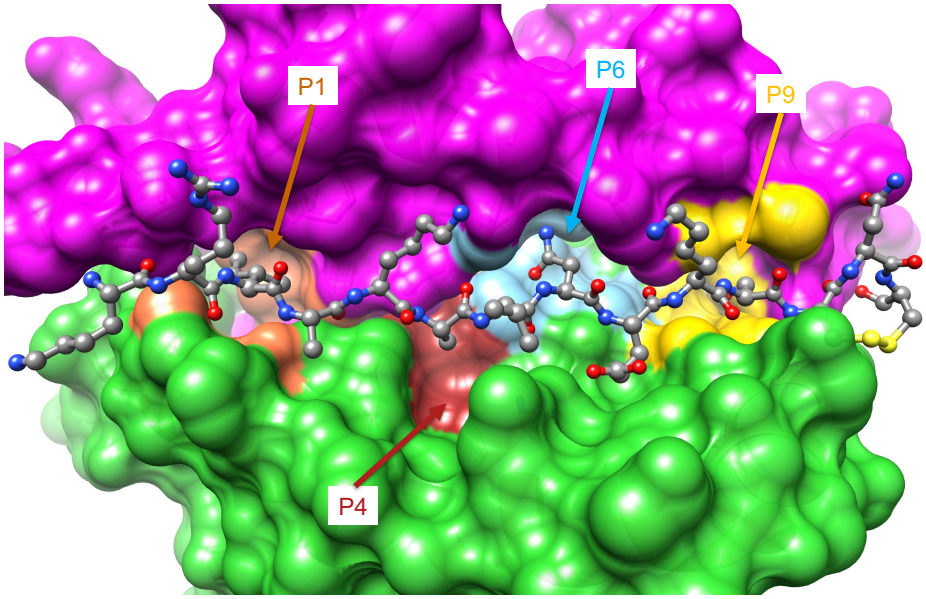
Structure of MHC class II molecule HLA-DR1 in complex with phosphopeptide MART-1 (15 residues)



90

Structure of pMHC Class II – The binding pockets

3 to 4 major anchoring pockets to accommodate the primary peptide anchor residues



91

Structure of pMHC Class II – The binding pocket

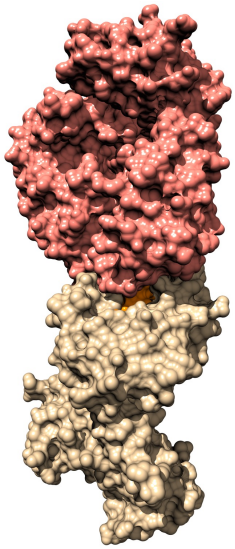
A large number of peptide residues interacting with MHC-II

Exercise 6 of the booklet

- Load structures 3L6F, 1FYT and 1AQD in UCSF Chimera
- Only keep chains A, B and C
- What is the common HLA protein?
- Show chains A and B as ribbons (hiding residues) and chains C as ball and stick (hiding the ribbon)
- The peptides are composed of how many residues ?
- What can we say about the position of the backbone of the peptides?
- Use the “Tools/Structure Comparison/Match->Align” tool to align the peptide sequences (chains C, only). Use the “Headers” menu to remove “RMSD” from the header, and add “Conservation”, “Consensus” and “Charge variation”.
- What are the most conserved positions?
- Are they facing HLA or the solvent?
- What are the residues defining the pocket in which they are bound?
- What type of interactions are taking place?

92

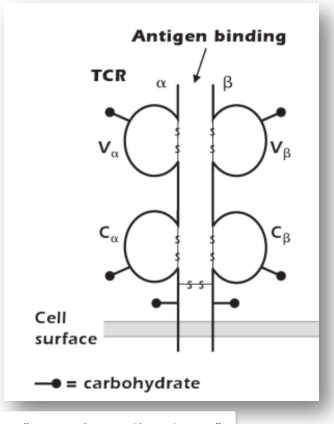
3D structure of the complex between the T-cell receptor (TCR) and pMHC



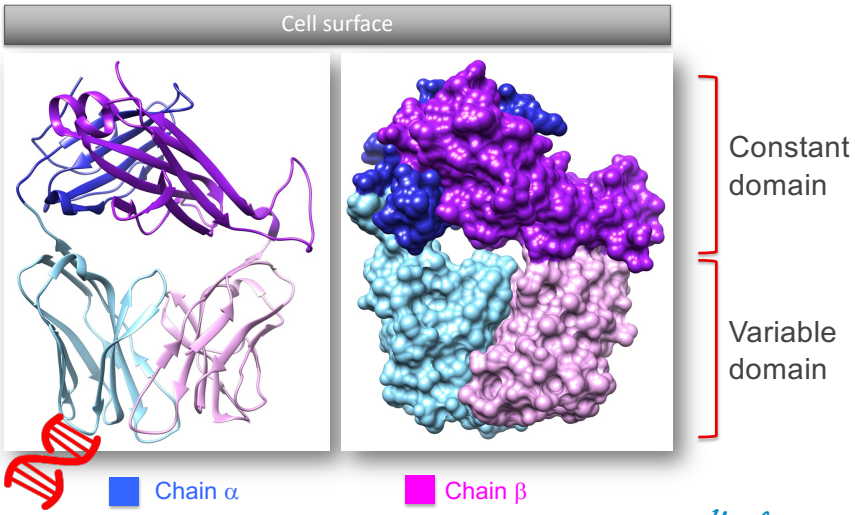
93

TCR 3D structure

- Membrane-anchored heterodimeric protein
- 2 chains α and β
- Each chain is composed of 2 extracellular domains: a variable domain V, and a constant domain C



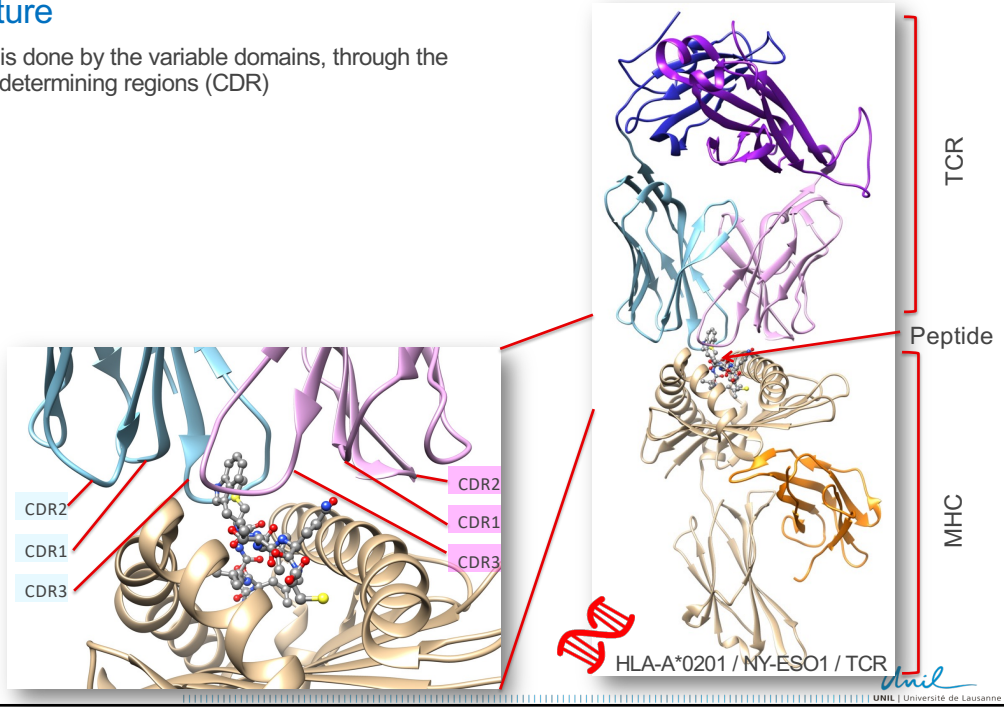
From "Immunology: A Short Course" by Richard Coico, Geoffrey Sunshine



94

TCR 3D structure

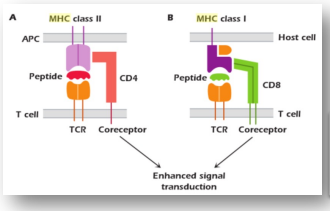
- Antigen binding is done by the variable domains, through the complementary determining regions (CDR)



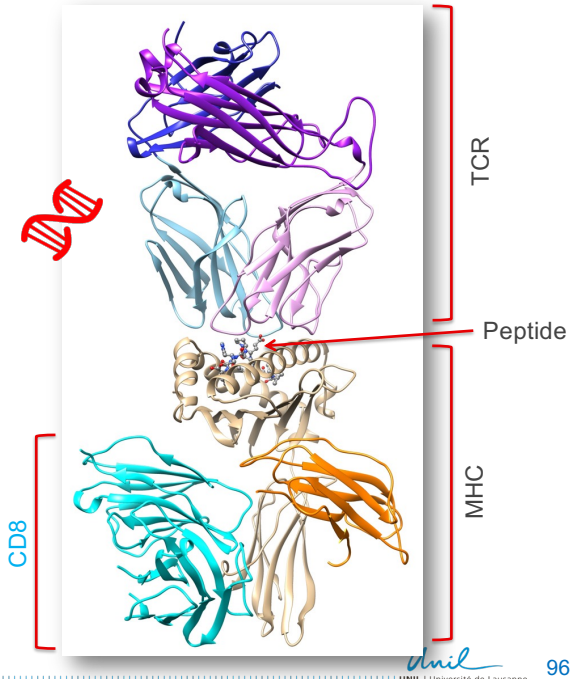
95

TCR 3D structure

- Antigen binding is done by the variable domains, through the complementary determining regions (CDR)

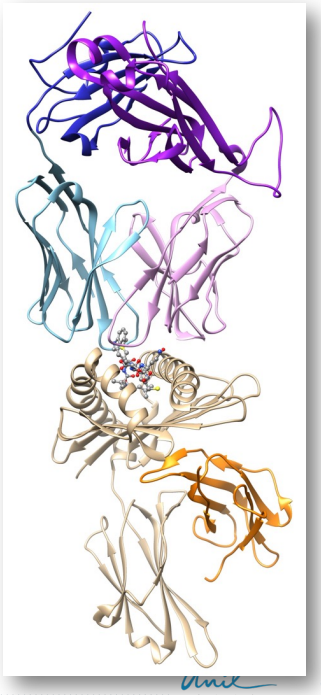


From "Immunology: A Short Course" by Richard Coico, Geoffrey Sunshine



96

TCR 3D structure
Exercise 7 of the booklet



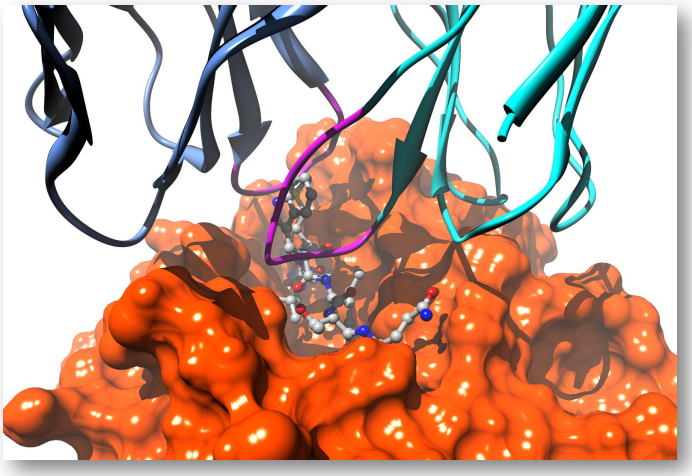
- Load the 2BNR PDB structure into UCSF chimera
- Produce a molecular representation similar to the one you see on the right

Residues in the CDR for TCRα:
- CDR1: 28 - 32
- CDR2: 51 - 55
- CDR3: 94 - 101

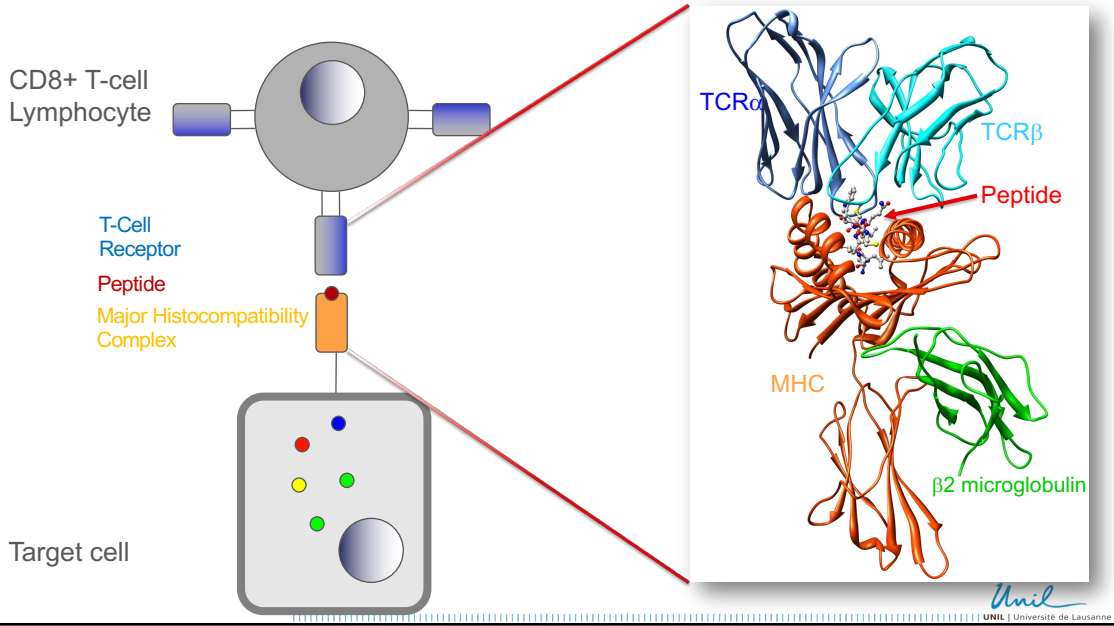
Residues in the CDR for TCRβ:
- CDR1: 25 - 29
- CDR2: 49 - 53
- CDR3: 94 - 100

- Display the atoms of the CDRs for TCRα and TCRβ
- What are the CDRs that make most of the contacts with the peptide epitope? And with the MHC?
- What are the TCR residues that make contact with peptide Trp5? What types of interactions are taking place?
- What MHC residues are close to TCRβ Ala51?
- If you wanted to change residue 51 of TCRβ, to increase the affinity of TCR for pMHC, what mutation would you introduce? Why?

TCR engineering – an example



TCR engineering – an example

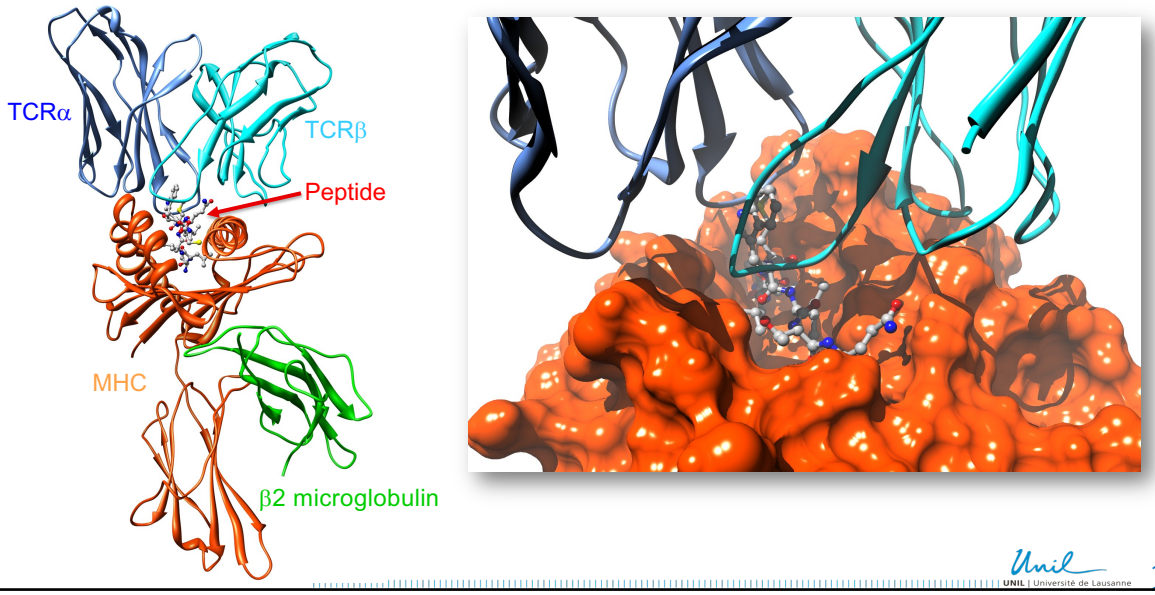


99

99

TCR engineering – 3D structure

TCR recognizes pMHC through complementary determining regions (CDR) loops

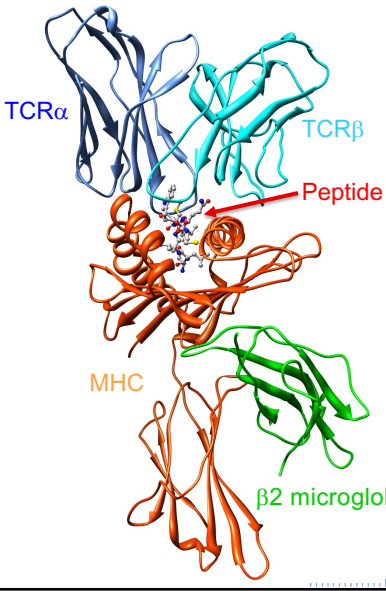


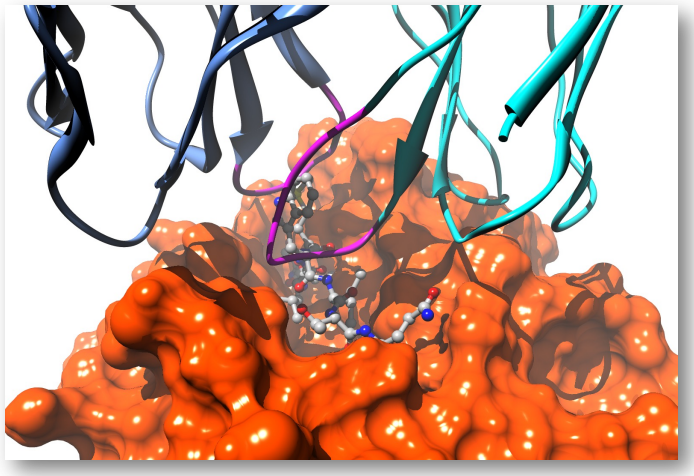
10

100

TCR engineering – 3D structure

TCR recognizes pMHC through complementary determining regions (CDR) loops



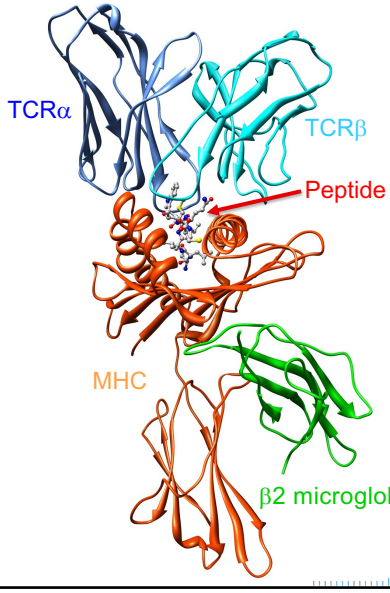


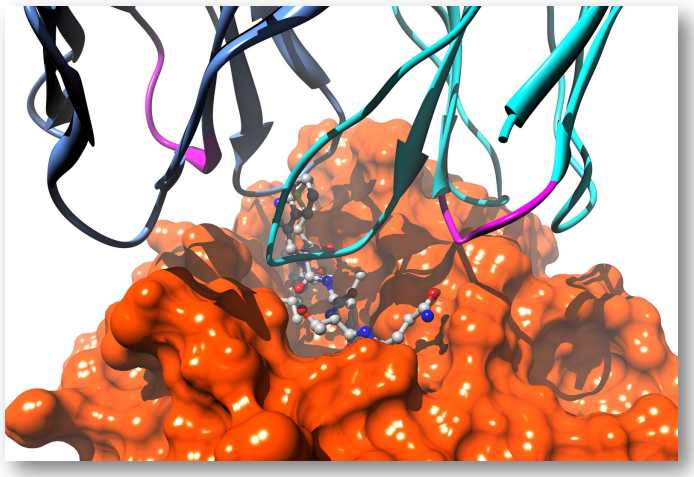
CDR3 in contact with MHC and peptide

101

TCR engineering – 3D structure

TCR recognizes pMHC through complementary determining regions (CDR) loops



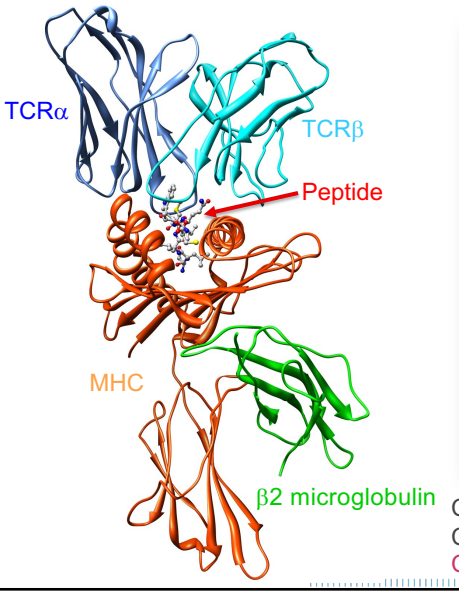


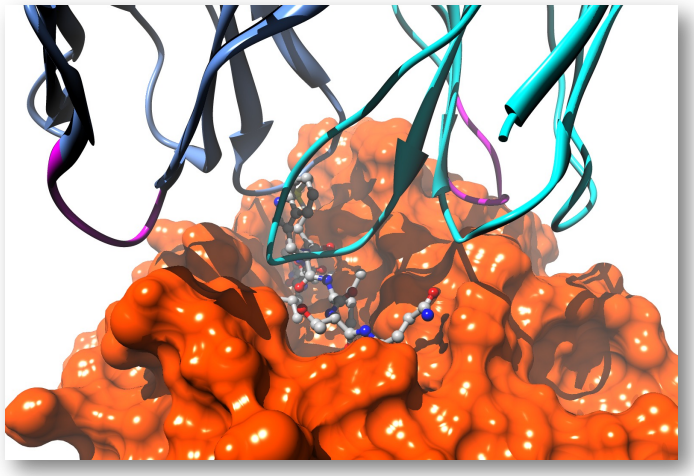
CDR3 in contact with MHC and peptide
CDR1 mainly in contact with MHC

102

TCR engineering – 3D structure

TCR recognizes pMHC through complementary determining regions (CDR) loops

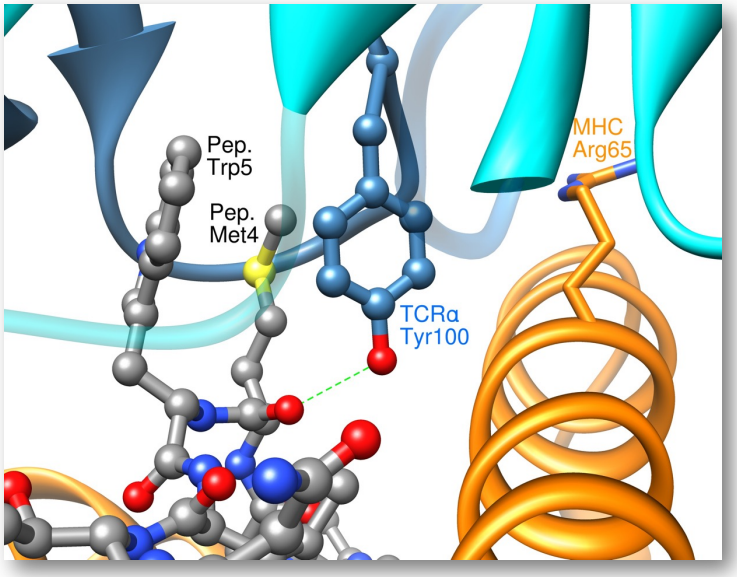




CDR3 in contact with MHC and peptide
CDR1 mainly in contact with MHC
CDR2 mainly in contact with MHC

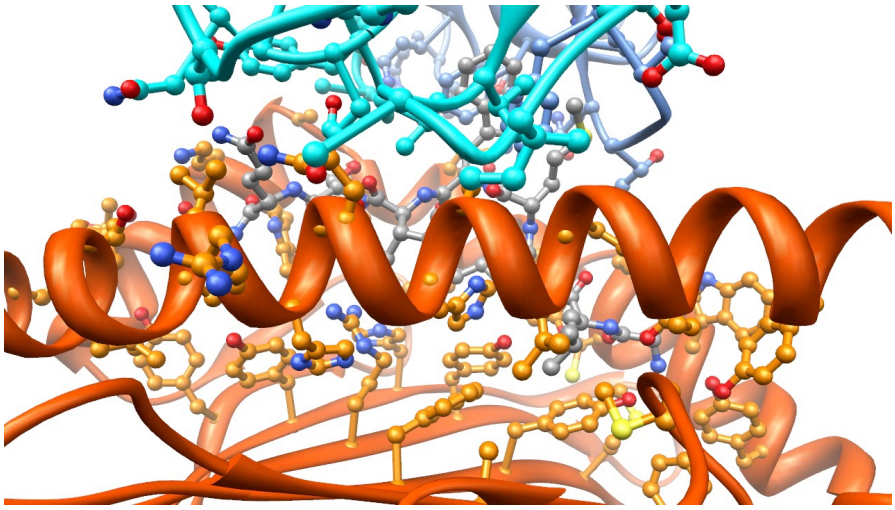
103

TCR engineering – Binding free energy



104

TCR engineering – Binding free energy



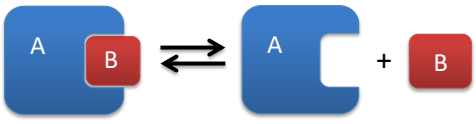
Visually: “important” interactions everywhere

Need for a physics-based method to estimate quantitatively the importance of each residue/interaction

105

TCR engineering – Binding free energy, ΔG_{bind}

Link between experiment and modeling

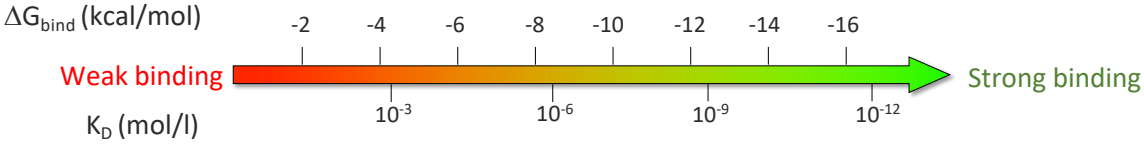


$$K_D = \frac{[A][B]}{[AB]}$$

K_D : dissociation constant

Accessible by computer-aided methods

$$\Delta G_{bind} = RT \ln(K_D) = \Delta H - T\Delta S$$



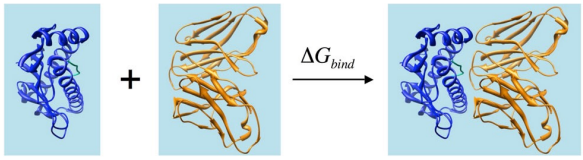
106

TCR engineering – Binding free energy, ΔG_{bind}

In silico methods to estimate ΔG_{bind}

MM-GBSA: $\langle \Delta G_{bind} \rangle = \langle \Delta H^{gas}_{bind} \rangle + \langle \Delta G_{desolv} \rangle - T\langle \Delta S \rangle$

Provides contribution of **each atom** to the association strength



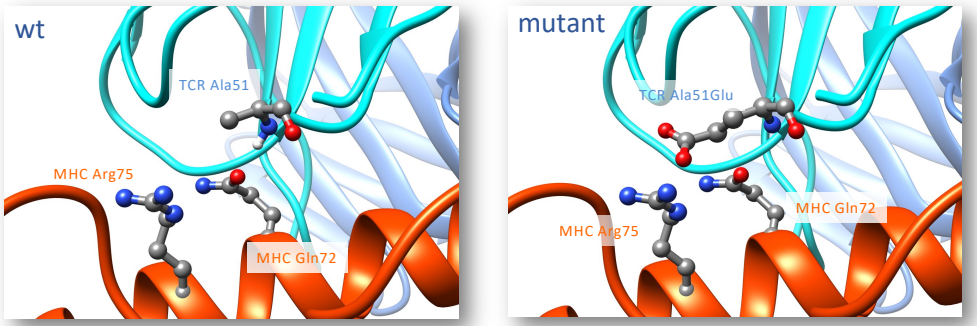
Zoete, V., Meuwly, M., & Karplus, M. *Proteins*, **2005**, *61*, 79–93.

Zoete, V.*, Meuwly, M.* *J. Comput. Chem.*, **2006**, *27*, 1843–1857.

MM-GBSA: Molecular Mechanics Generalized Born Surface Area

TCR engineering – Binding free energy, ΔG_{bind}

Using ΔG_{bind} to select mutations for experimental assay



ΔG^{wt}_{bind} ΔG^{mutant}_{bind}

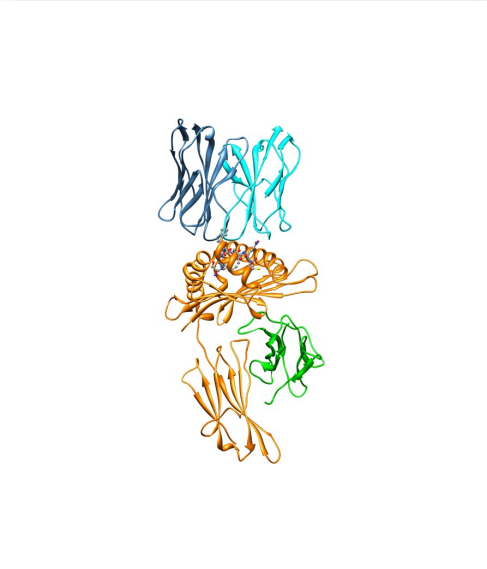
$\Delta\Delta G_{bind} = \Delta G^{mutant}_{bind} - \Delta G^{wt}_{bind}$

< 0 : selected for experiment

> 0 : rejected

Zoete, V., Irving, M. B., & Michielin, O. MM-GBSA binding free energy decomposition and T cell receptor engineering. *J. Molec. Rec.*, **2010**, *23*, 142–152.

TCR engineering – The computational pipeline



Molecular Dynamics (MD) simulation:

System solvated with explicit water molecules (TIP3P model):
- ~ 29,500 water molecules
- ~ 100,000 atoms in total

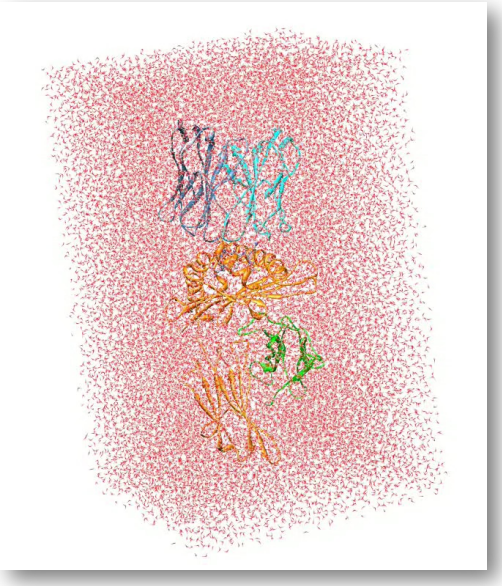
Atom motions are calculated to follow Newton's equation of motion, at 300 K and 1 atm.

Typical simulation times: from 0.5 ns to ~ 100 ns

Energy terms averaged over 200 to 500 frames extracted from the MD simulation

109

TCR engineering – The computational pipeline



Molecular Dynamics (MD) simulation:

System solvated with explicit water molecules (TIP3P model):
- ~ 29,500 water molecules
- ~ 100,000 atoms in total

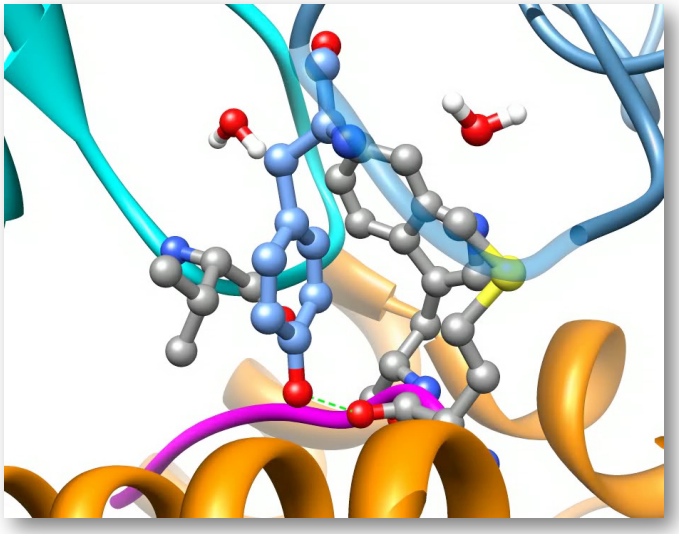
Atom motions are calculated to follow Newton's equation of motion, at 300 K and 1 atm.

Typical simulation times: from 0.5 ns to ~ 100 ns

Energy terms averaged over 200 to 500 frames extracted from the MD simulation

110

TCR engineering – The computational pipeline



Molecular Dynamics (MD) simulation:

System solvated with explicit water molecules (TIP3P model):

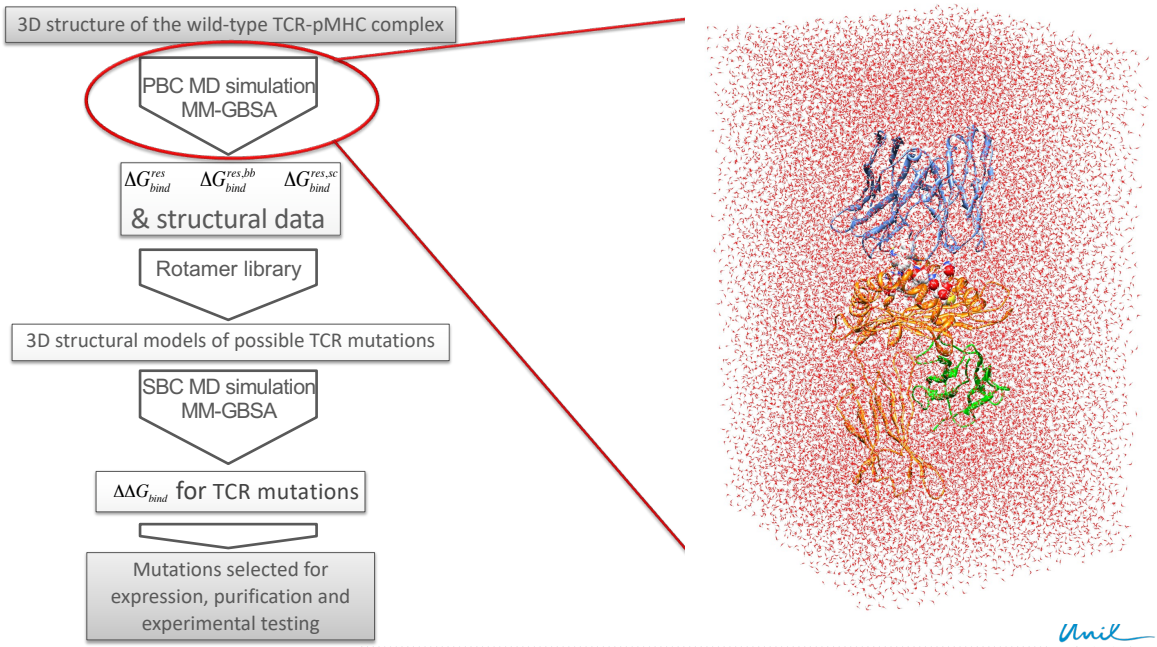
- ~ 29,500 water molecules
- ~ 100,000 atoms in total

Atom motions are calculated to follow Newton's equation of motion, at 300 K and 1 atm.

Typical simulation times: from 0.5 ns to ~ 100 ns

Energy terms averaged over 200 to 500 frames extracted from the MD simulation

TCR engineering – The computational pipeline



TCR engineering – The computational pipeline

3D structure of the wild-type TCR-pMHC complex

PBC MD simulation
MM-GBSA

ΔG_{bind}^{res} $\Delta G_{bind}^{res,bb}$ $\Delta G_{bind}^{res,sc}$
& structural data

Rotamer library

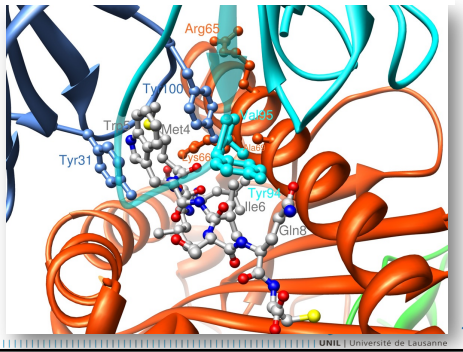
3D structural models of possible TCR mutations

SBC MD simulation
MM-GBSA

$\Delta\Delta G_{bind}$ for TCR mutations

Mutations selected for expression, purification and experimental testing

Residue	E_{vdW}	E_{elec}	$\Delta G_{desolv,elec}$	$\Delta G_{desolv,np}$	ΔG_{bind}^{res}
Glu29	-2.12	-58.92	71.25	-0.50	9.72
Asp55	-0.34	-56.52	59.24	-0.19	2.18
Arg93	-1.25	34.05	-30.60	-0.07	2.13
Gln95	-2.56	-4.37	8.26	-0.52	0.81
…					
Gly96	-2.30	-0.25	0.76	-0.20	-2.00
Ile53	-1.94	1.83	-1.77	-0.51	-2.39
Ser53	-1.22	-5.01	4.28	-0.47	-2.43
Gly98	-2.29	-5.26	5.28	-0.42	-2.69
Gln51	-2.09	-3.04	2.66	-0.29	-2.77
Tyr94	-1.84	1.02	-2.00	-0.18	-3.01
Val95	-3.18	-3.09	2.86	-0.39	-3.81
Tyr31	-5.25	-0.52	1.52	-0.54	-4.80
Tyr100	-5.07	-4.74	5.29	-0.72	-5.24



113

TCR engineering – The computational pipeline

3D structure of the wild-type TCR-pMHC complex

PBC MD simulation
MM-GBSA

ΔG_{bind}^{res} $\Delta G_{bind}^{res,bb}$ $\Delta G_{bind}^{res,sc}$
& structural data

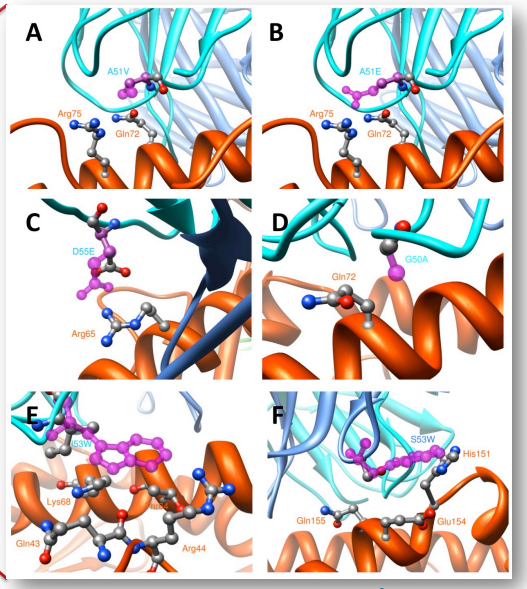
Rotamer library

3D structural models of possible TCR mutations

SBC MD simulation
MM-GBSA

$\Delta\Delta G_{bind}$ for TCR mutations

Mutations selected for expression, purification and experimental testing

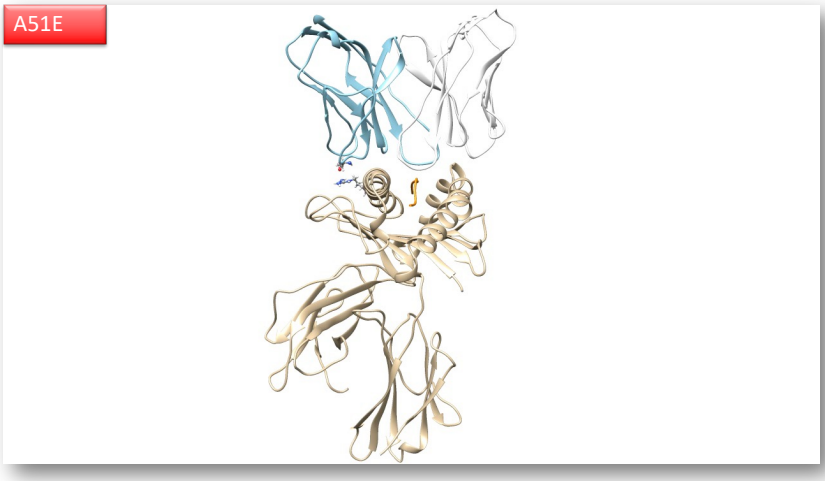


114

TCR engineering – The computational pipeline

Increasing affinity

A51E



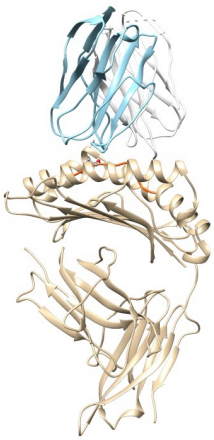
Gain in binding free energy:
-7.3 kcal/mol

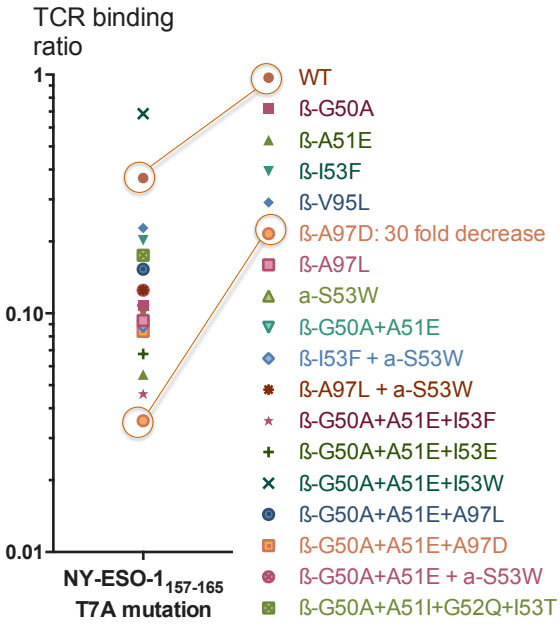
115

TCR engineering – The computational pipeline

Increasing selectivity

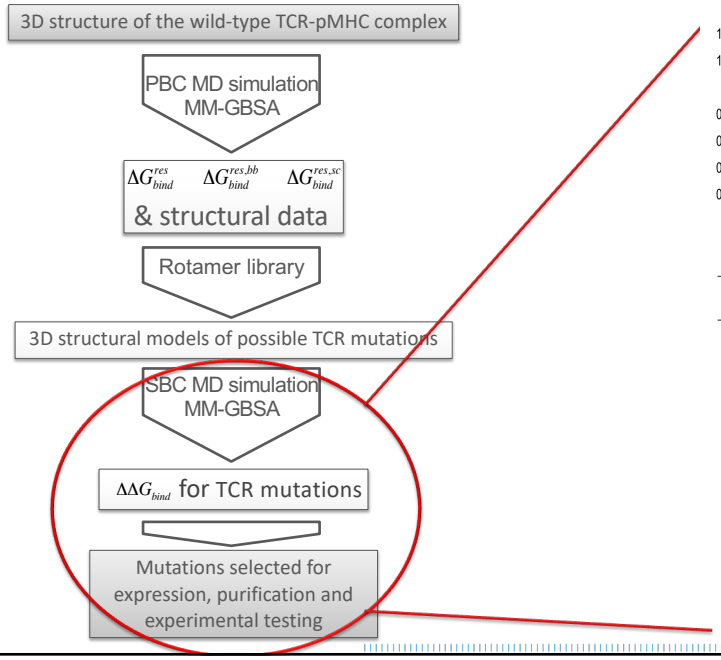
A97D

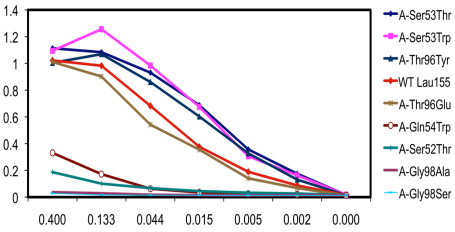




116

TCR engineering – The computational pipeline

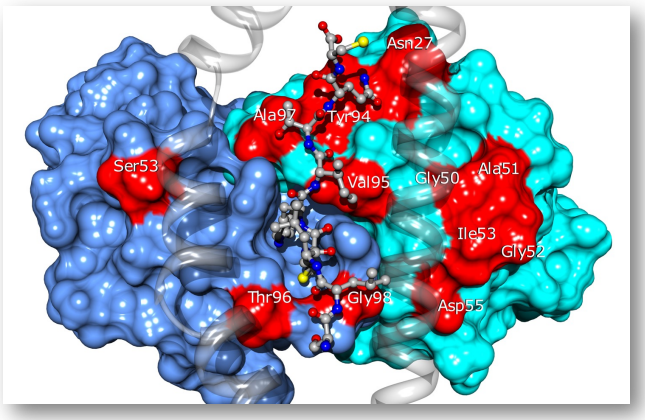




	CDR2							CDR3				
	48	49	50	51	52	53	54	55	93	94	95	96	97
	S	V	G	A	G	I	T	D	S	Y	V	G	A
VAL49ILE	S	I	G	A	G	I	T	D	S	Y	V	G	A
GLY50VAL	S	V	V	A	G	I	T	D	S	Y	V	G	A
GLY50ALA	S	V	A	A	G	I	T	D	S	Y	V	G	A
ALA51GLU	S	V	G	E	G	I	T	D	S	Y	V	G	A
50ALA51GLU	S	V	A	E	G	I	T	D	S	Y	V	G	A
ALA51ASP	S	V	G	D	G	I	T	D	S	Y	V	G	A
ALA51VAL	S	V	G	V	G	I	T	D	S	Y	V	G	A
50ALA51ASP	S	V	A	D	G	I	T	D	S	Y	V	G	A
GLY52GLN	S	V	G	A	Q	I	T	D	S		V	G	A
ILE53PHE	S	V	G	A	G	F	T	D	S	Y	V	G	A
ILE53TRP	S	V	G	A	G	W	T	D	S	Y	V	G	A
ILE53GLU	S	V	G	A	G	E	T	D	S	Y	V	G	A
ASP55GLU	S	V	G	A	G	I	T	E	S	Y	V	G	A
TYR94ASN	S	V	G	A	G	I	T	D	S	N	V	G	A
VAL95LEU	S	V	G	A	G	I	T	D	S	Y	L	G	A
GLY96SER	S	V	G	A	G	I	T	D	S	Y	V	S	A
ALA97VAL	S	V	G	A	G	I	T	D	S	Y	V	G	V
ALA97LEU	S	V	G	A	G	I	T	D	S	Y	V	G	L

TCR engineering – Application to BC1 TCR targeting NY-ESO-1

- 24 single/double mutants tested (M. Irving)
- 13 (54 %) were more active than the wt TCR
- up to 56 fold increase for single mutations
- 150 fold increase for TCR Vβ G50A/A51E/A97L + Vα S53W

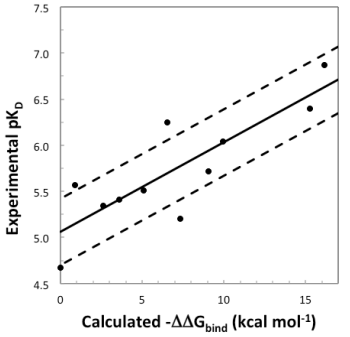


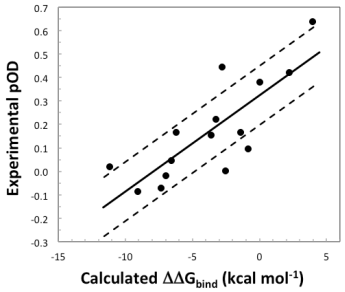
TCR engineering – Application to BC1 TCR targeting NY-ESO-1

- 24 single/double mutants tested (M. Irving)
- 13 (54 %) were more active than the wt TCR
- up to 56 fold increase for single mutations
- 150 fold increase for TCR Vβ G50A/A51E/A97L + Vα S53W
- good correlation between calculated binding free energies and experimental results

$$pOD = 0.32\ (\pm 0.04) + 0.041\ (\pm 0.008)\ \Delta\Delta G_{bind} \quad R=0.81$$

$$pK_D = 5.05\ (\pm 0.18) - 0.098\ (\pm 0.021)\ \Delta\Delta G_{bind} \quad R=0.82$$





Unil UNIL | Université de Lausanne 119

119

TCR engineering – Application to BC1 TCR targeting NY-ESO-1

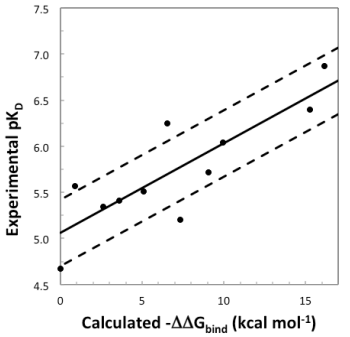
- 24 single/double mutants tested (M. Irving)
- 13 (54 %) were more active than the wt TCR
- up to 56 fold increase for single mutations
- 150 fold increase for TCR Vβ G50A/A51E/A97L + Vα S53W
- good correlation between calculated binding free energies and experimental results

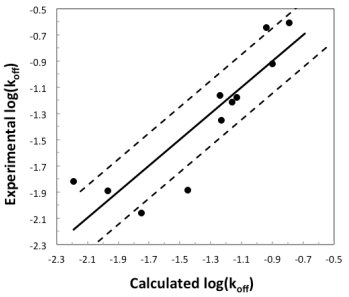
$$pOD = 0.32\ (\pm 0.04) + 0.041\ (\pm 0.008)\ \Delta\Delta G_{bind} \quad R=0.81$$

$$pK_D = 5.05\ (\pm 0.18) - 0.098\ (\pm 0.021)\ \Delta\Delta G_{bind} \quad R=0.82$$

- good correlation between calculated energies and experimental k_{off} (R=0.88)

$$\log(k_{off}) = -0.94\ (\pm 0.13) + 0.13\ (\pm 0.02)\ \Delta\Delta G_{non-polar} + 0.044\ (\pm 0.016)\ \Delta\Delta G_{polar}$$





Unil UNIL | Université de Lausanne 120

120

TCR engineering – Application to BC1 TCR targeting NY-ESO-1

- unfitted approach: can be applied to other systems

→ Applied to Melan-A antigen with 73% success rate

- step-by-step modeling approach: incremental improvements in TCR affinity, while minimizing the loss of specificity

-mutations toward both non polar and polar mutants

- up to 56 fold affinity increase for single mutants. Some single mutations are compatible. Combinations of them led to 150 fold increase in affinity compared to WT

- correlation between experimental affinity and calculated binding free energy

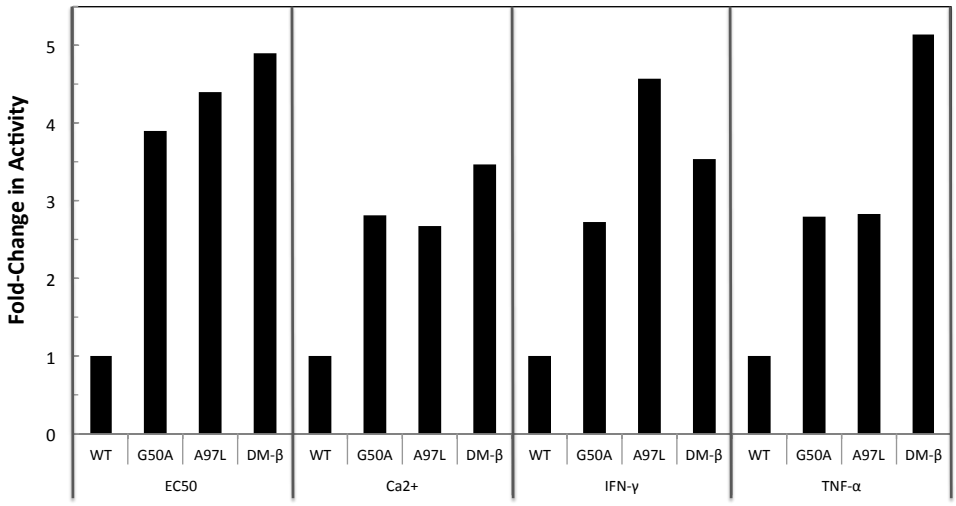
- correlation between k_{off} and calculated energy terms

- no cross reactivity

121

TCR engineering – Application to BC1 TCR targeting NY-ESO-1

- both T-cell proliferation after antigenic challenge and tumor cell killing were significantly improved



→ Mouse model / Clinical trial at CHUV

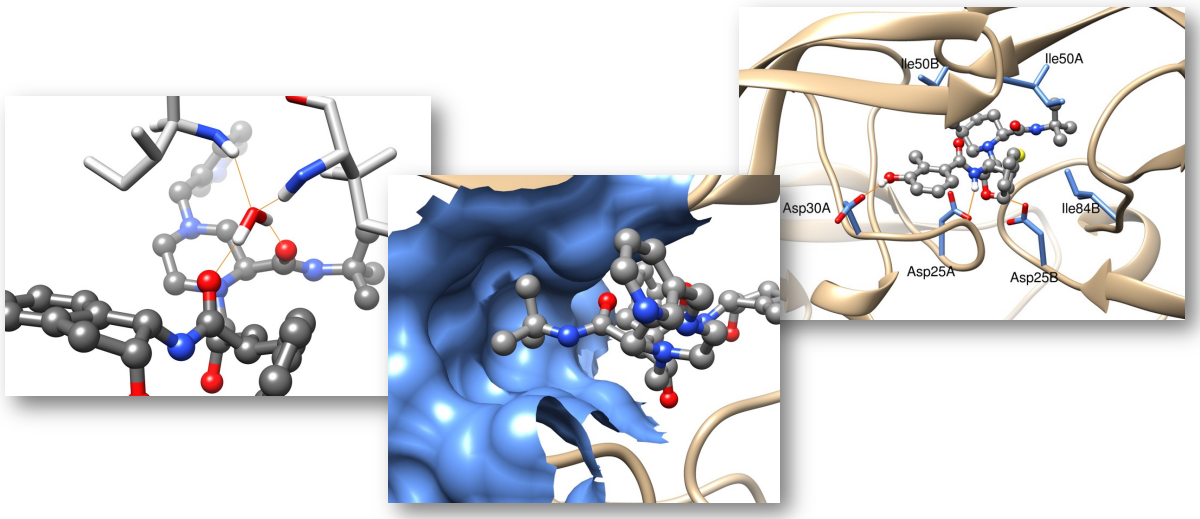
122

The End!

If you have questions: vincent.zoete@unil.ch

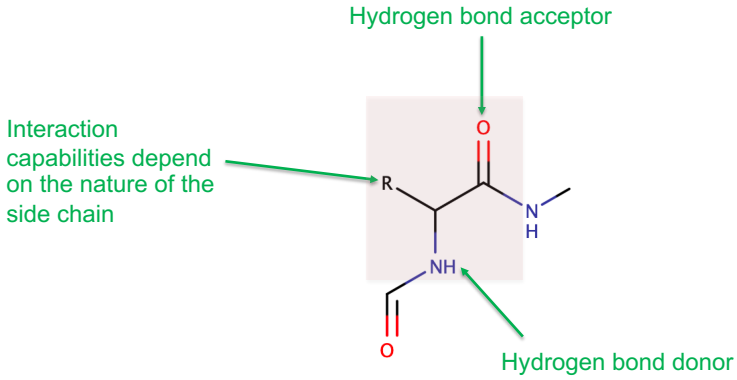
123

Molecular Recognition



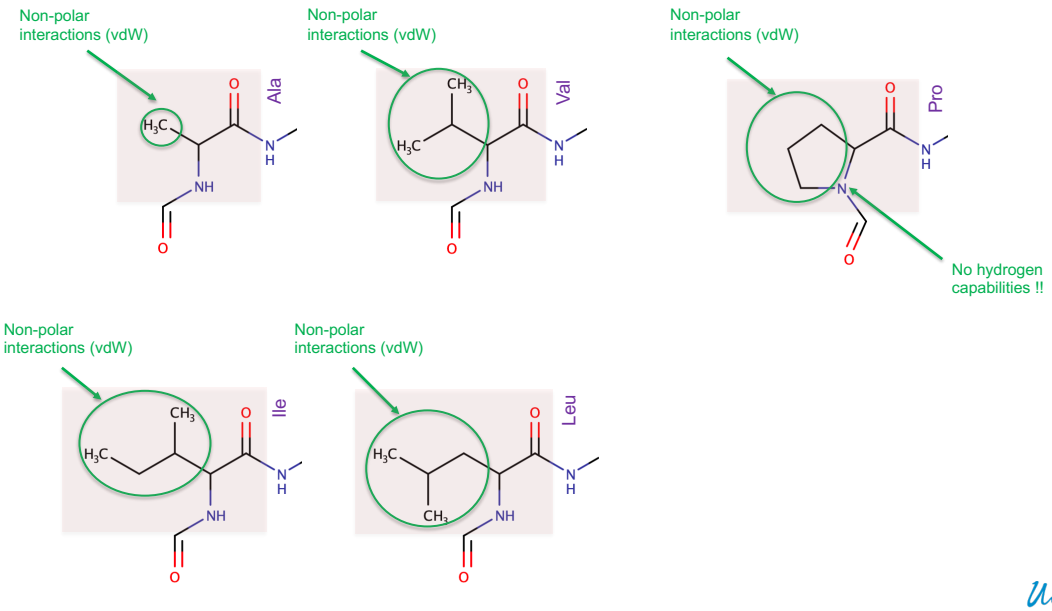
124

Molecular recognition – Possible interactions per amino acids



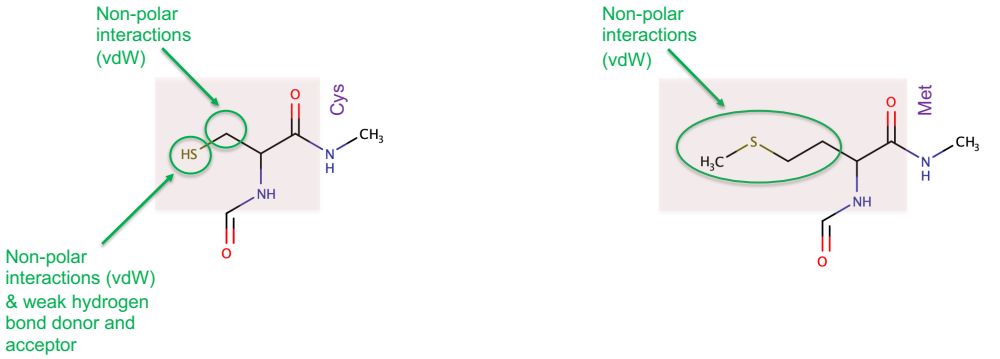
125

Molecular recognition – Possible interactions per amino acids



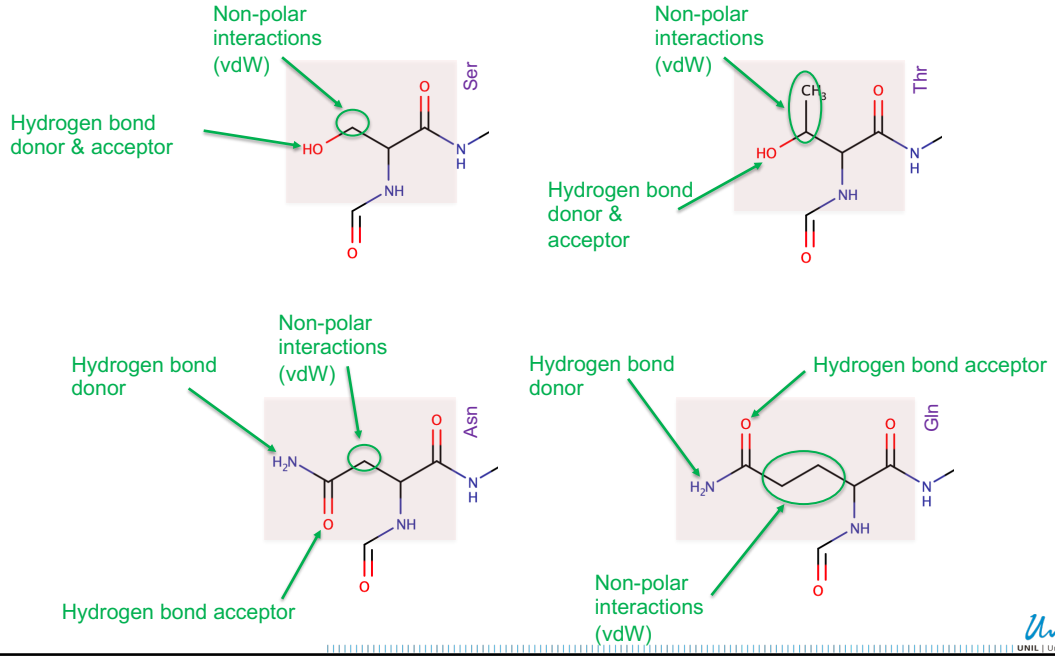
126

Molecular recognition – Possible interactions per amino acids



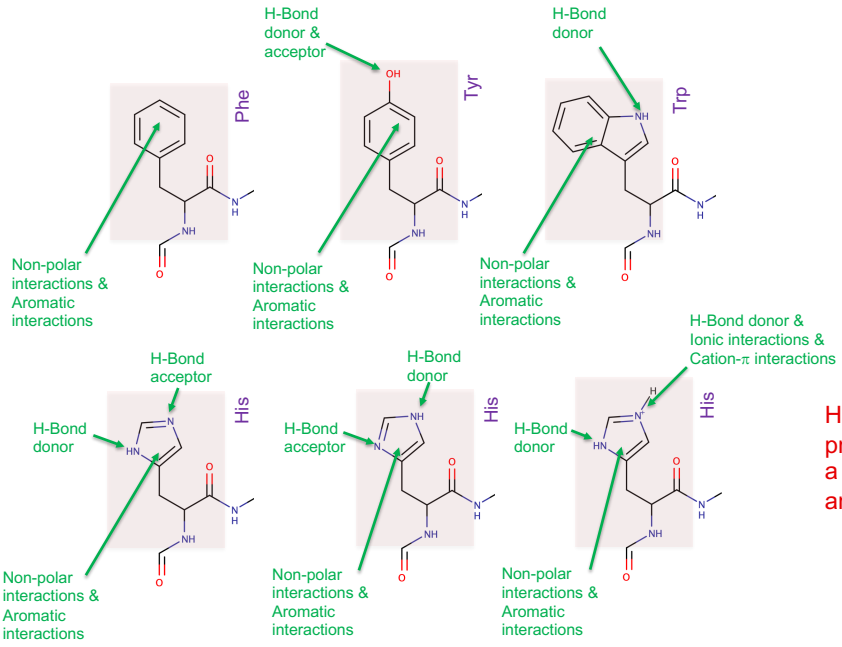
127

Molecular recognition – Possible interactions per amino acids



128

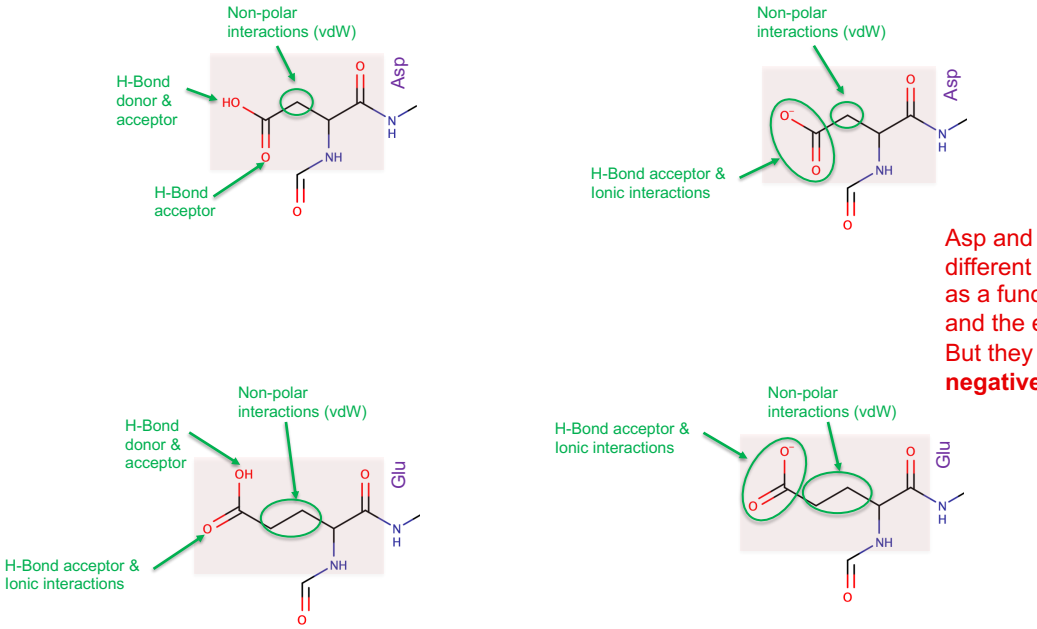
Molecular recognition – Possible interactions per amino acids



His exists in 3 different protonation states as a function of the pH and the environment

129

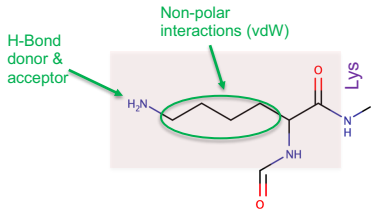
Molecular recognition – Possible interactions per amino acids

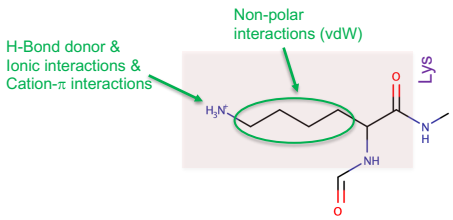


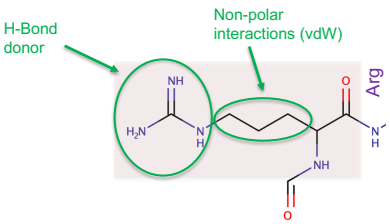
Asp and Glu exist in 2 different protonation states as a function of the pH and the environment
But they are **generally negative**

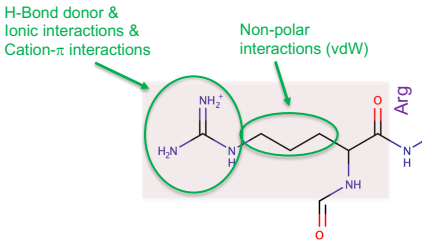
130

Molecular recognition – Possible interactions per amino acids









Arg and Lys exist in 2 different protonation states as a function of the pH and the environment
But they are **generally positive**

131