



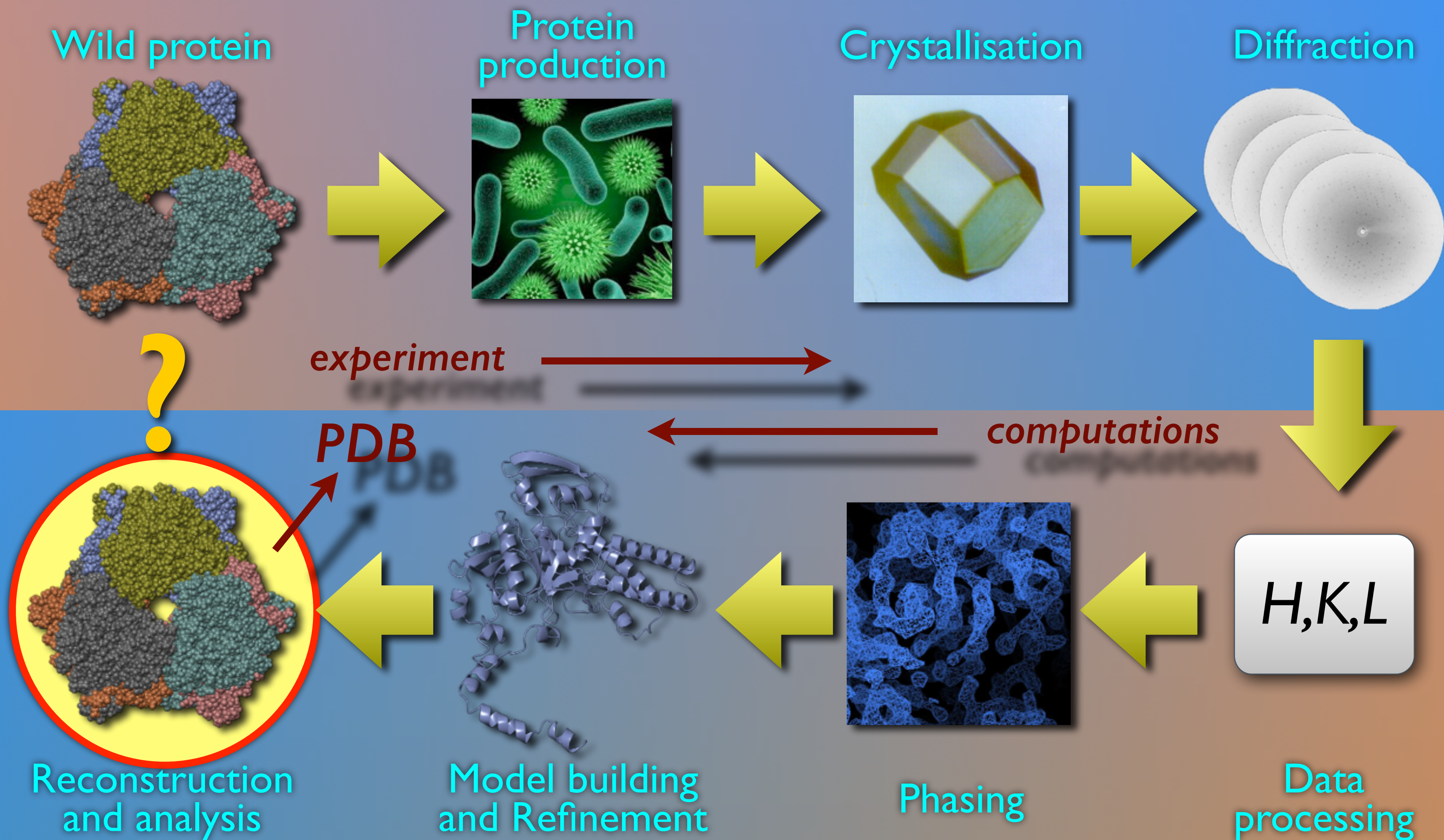
Macromolecules in Crystals and Solutions

Eugene Krissinel

CCP4, STFC Research Complex at Harwell
Didcot, United Kingdom

eugene.krissinel@stfc.ac.uk

MX Schematic Loop



What do we solve structures for?

★ In general, for learning about interaction between biomolecules:

- *enzyme functioning*
- *transport mechanisms*
- *chemical signalling*
-

★ Why structure? There are better methods to study interactions.

- *hope to understand important interaction in fine details*
- *hope to learn key features and optimise experimentation*
- *hope to learn how to predict interactions*
- *hope to learn how to control interactions (drugs, medicine)*

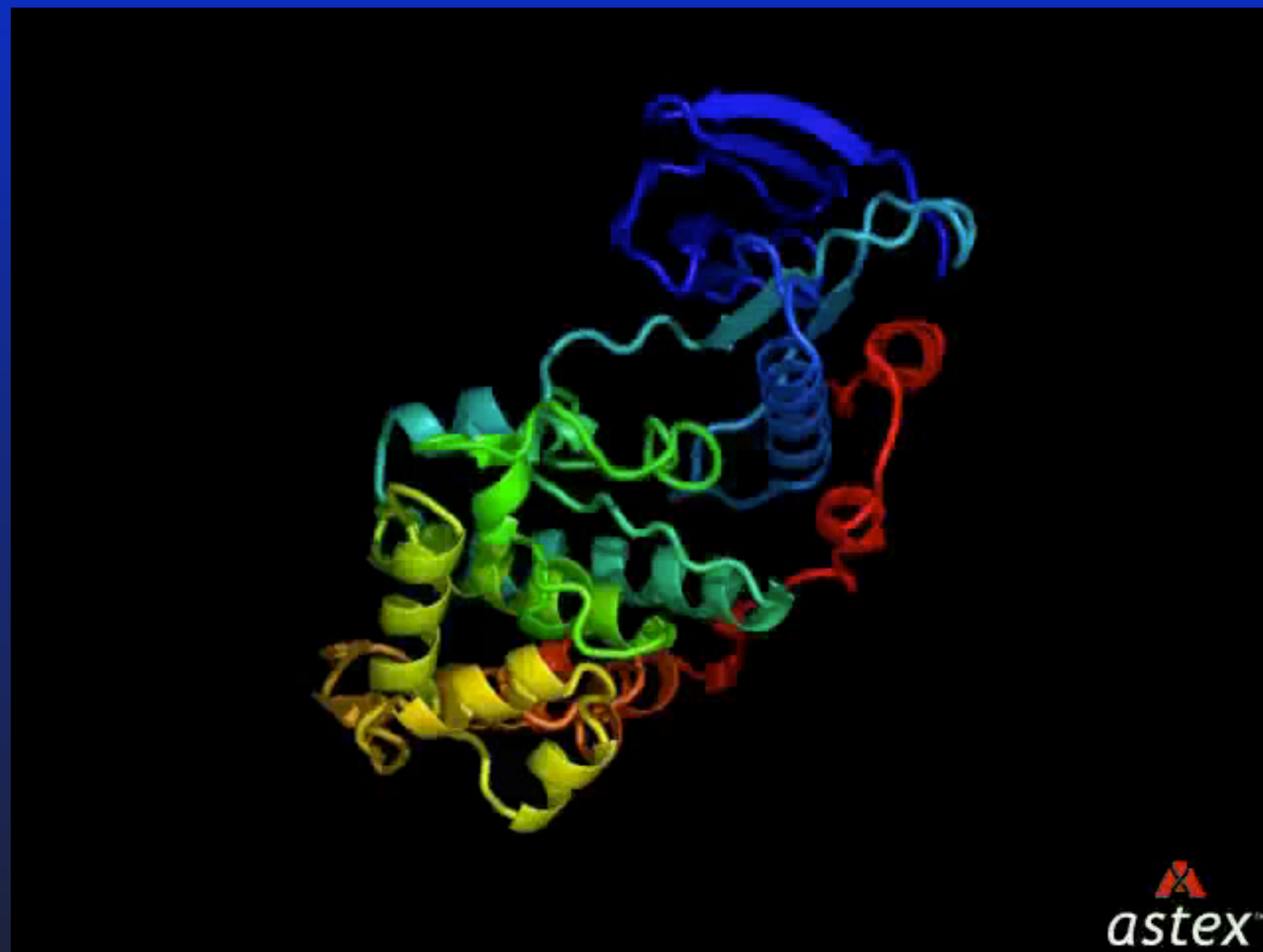
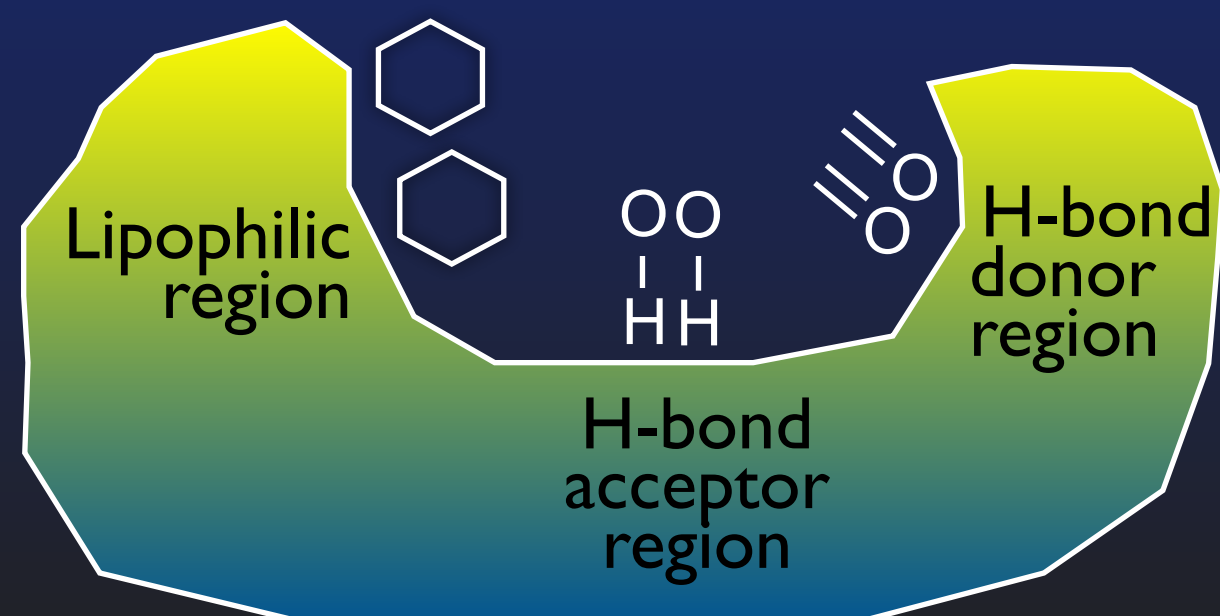


Intelligent Drug Design

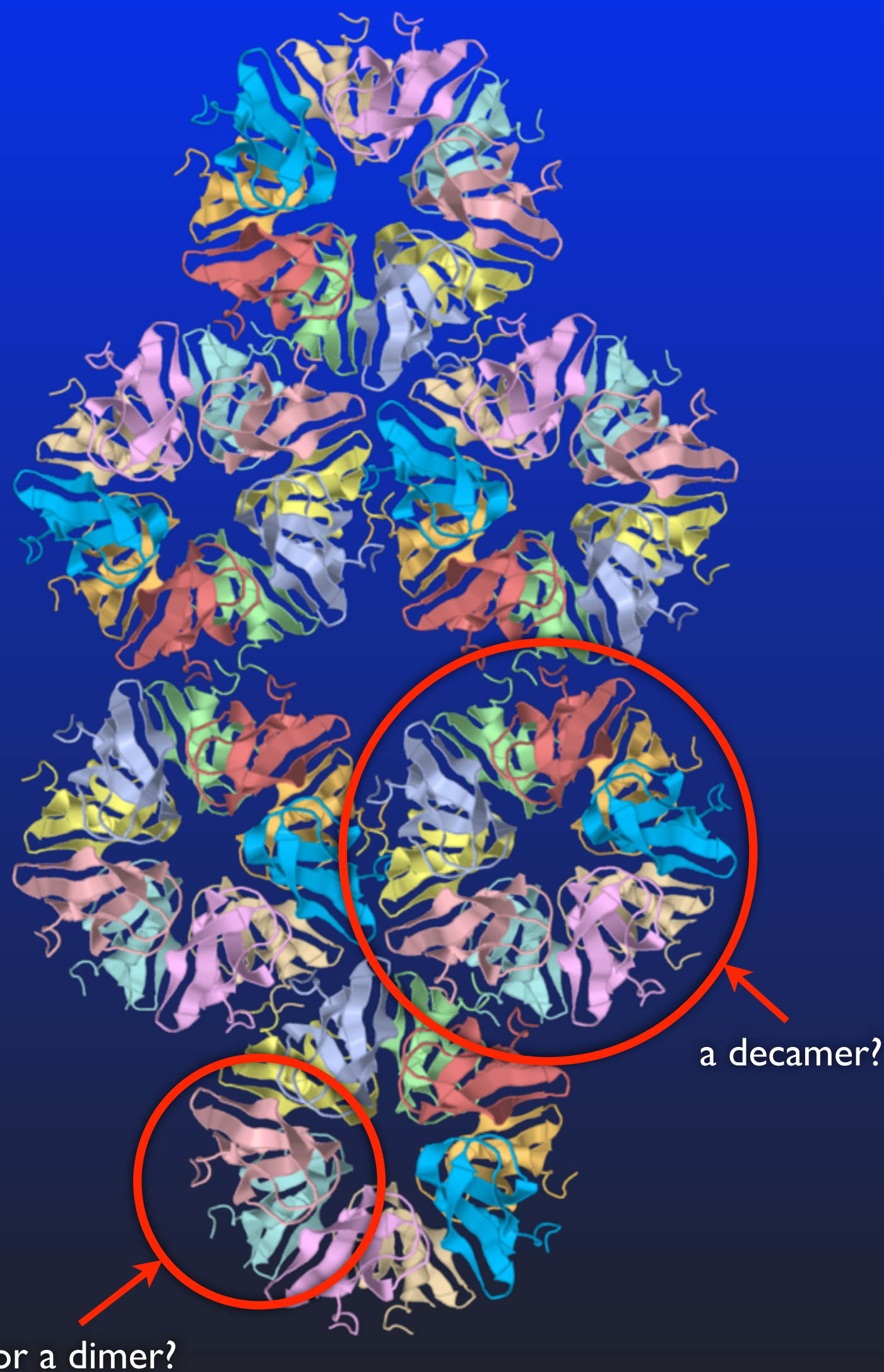
★ A generic name for many methods aimed to discover new drugs by means better than at random

- *Bioinformatics*
- *Directed Combinatorial Chemistry*
- *Computer-Assisted Drug Design*
- *Structure-Based Drug Design*
- *Fragment-Based Drug Design*

★ Basic idea: find a molecule that blocks the “right” protein’s active site, or suggest how such a molecule should look like



From web-site of ASTEX Pharmaceutical



★ Macromolecular crystals present us with **models** of biological structures and interactions between them

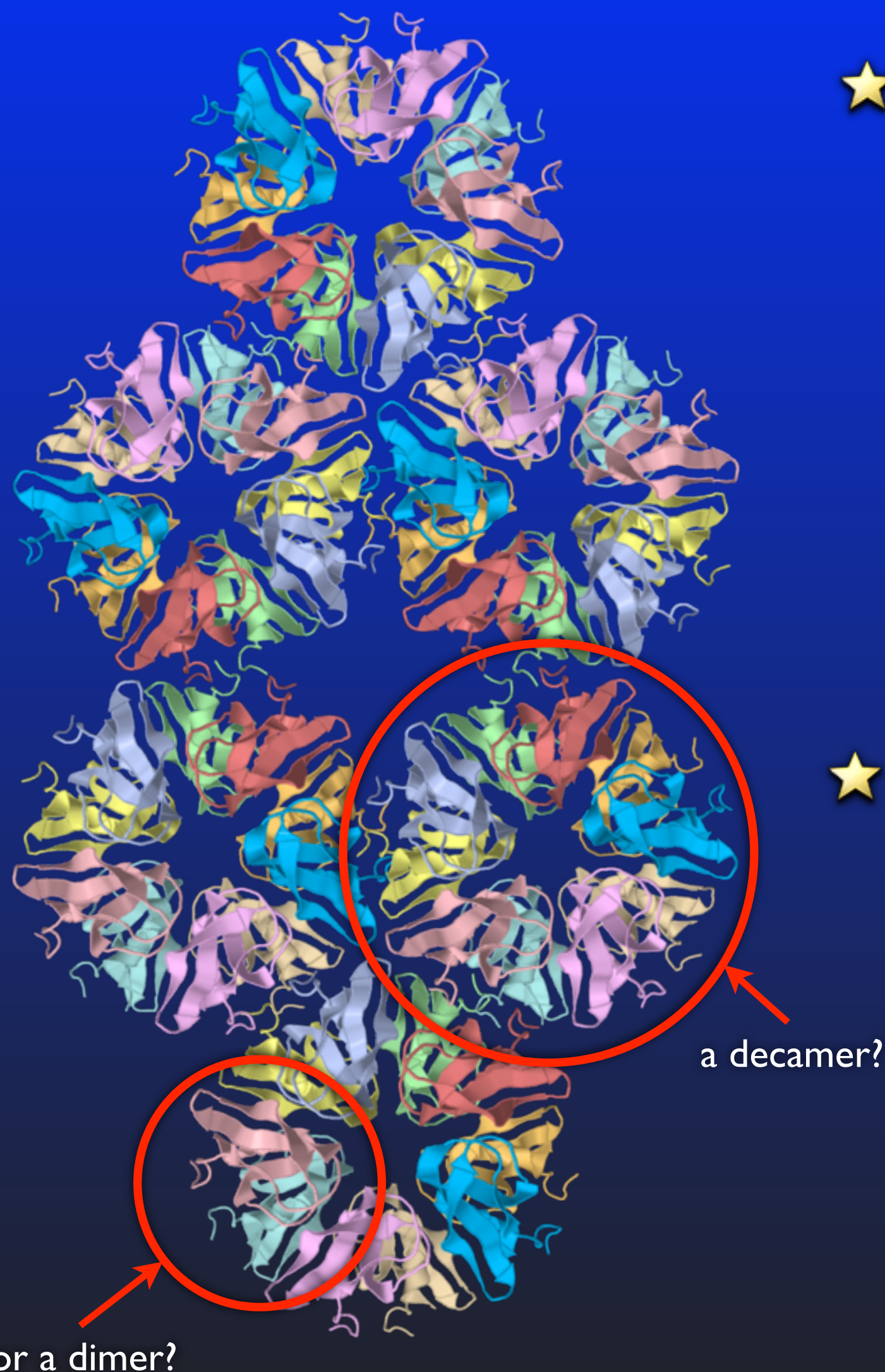
→ “if you want to know how A interacts with B - crystallise them together!” (crystallographer’s sweet dream, *but does this always work?*)

→ interactions make complexes

→ complexes make biology

→ biology tells which drug





- ★ Crystals present us with both real and artifactual interactions, which may be difficult to differentiate. Frequently used techniques:

Rules of thumb: e.g. manifestation in different crystal forms

Experimental: complementing studies (MS, EM, NMR, scattering)

Bioinformatical: homology and interface similarity analysis

Computational: energy estimates and modelling

- ★ PISA infers significant interactions and macromolecular assemblies from crystal data by evaluating their free Gibbs energy:

$$\Delta G_0 = - \Delta G_{\text{int}} - T \Delta S > 0$$

<http://ccp4serv7.rc-harwell.ac.uk/srv/pisa/pisa.php>



The PDB does indeed contain a wealth of experimental data on macromolecular complexes

More than 80% of macromolecular structures are solved by means of X-ray diffraction on crystals.

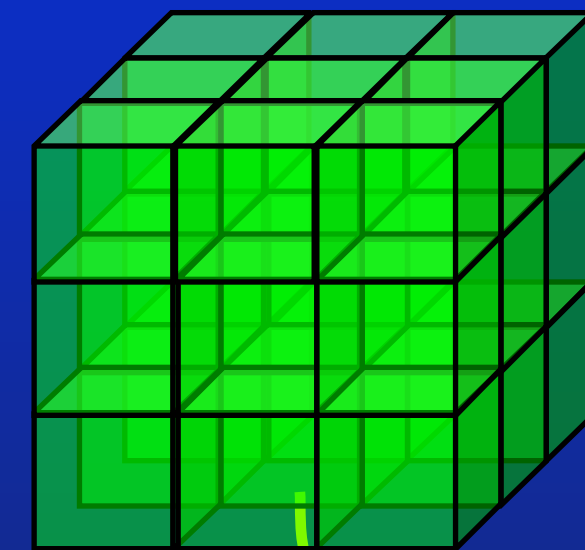
Any crystal represents macromolecular interactions and associations through inter-molecular interfaces

An X-ray diffraction experiment produces atomic coordinates of the Asymmetric Unit (ASU), which is stored as a PDB file.

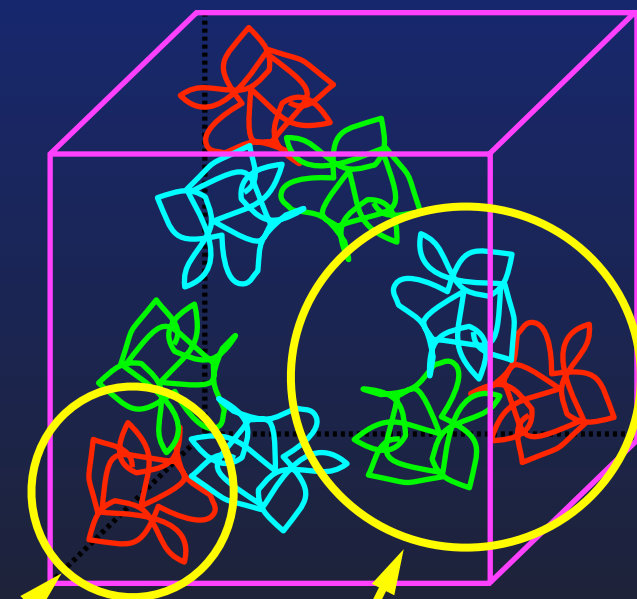
In general, neither ASU nor Unit Cell has any direct relation to PQS. The PQS may be made of

- a single ASU
- several ASU
- a part of ASU
- several ASU parts

Crystal = translated Unit Cells



Unit Cell = all space symmetry group mates of ASU



PDB file
(ASU)

Biological
Unit



Detection of Biological Units in Crystals: PISA summary

1. Enumerate all possible multimeric assemblies in crystal packing, subject to crystal properties: space symmetry group, geometry and composition of the Asymmetric Unit

- Achieved with graph-theoretical techniques, by representing crystal as an infinite periodic graph of connected macromolecules

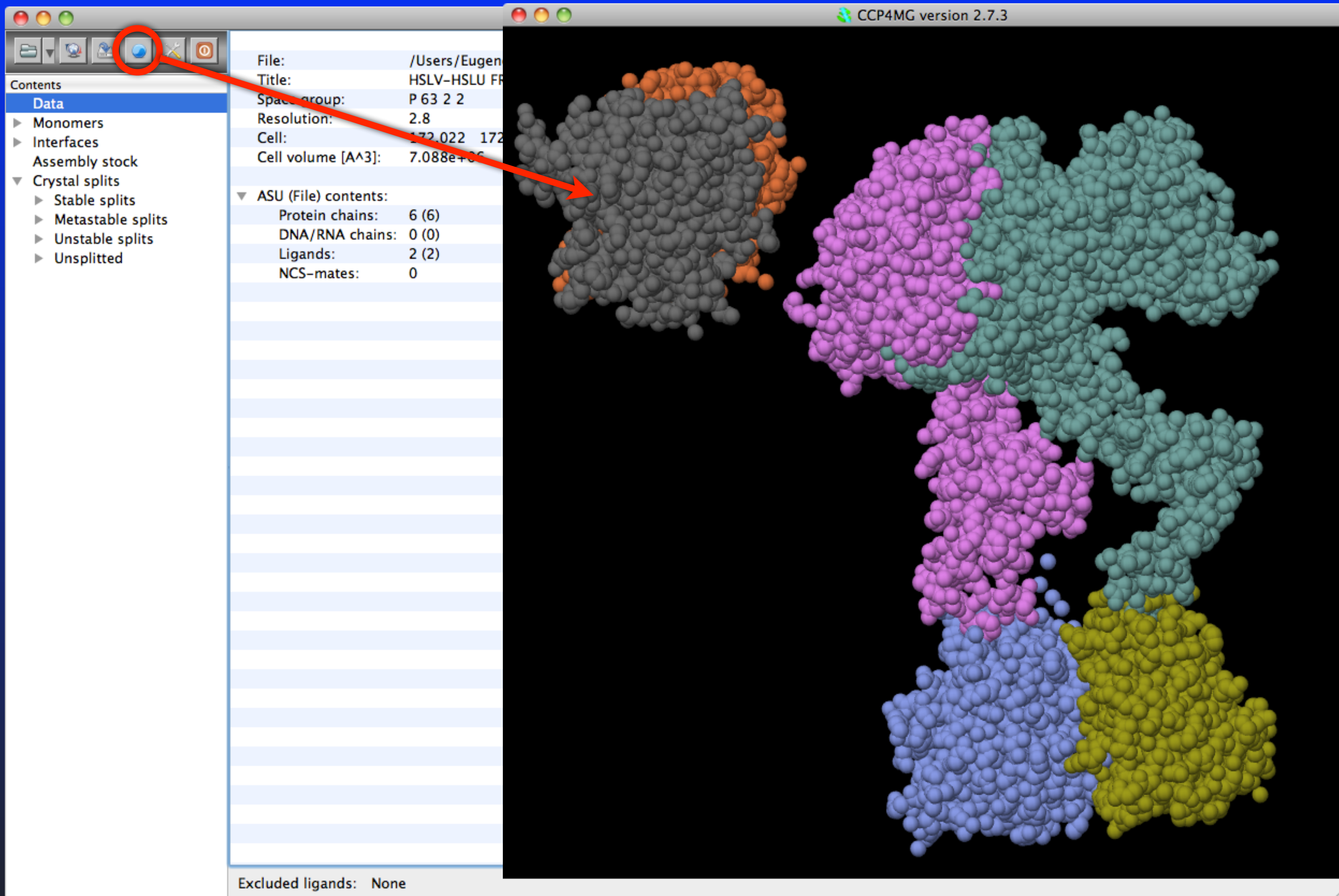
2. Evaluate assemblies for chemical stability:

$$\Delta G_0 = - \Delta G_{\text{int}} - T \Delta S > 0$$

3. Leave only sets of stable assemblies in the list, and range by chances to be a biological unit:

- Larger assemblies take preference
- Single-assembly sets take preference
- Otherwise, assemblies with higher ΔG_0 take preference





The screenshot displays the CCP4MG version 2.7.3 interface. At the top, a table lists quaternary structures. A red circle highlights the 'Data' button in the top-left toolbar. Two red arrows point from the table to 3D molecular models: one from the row with Split No. 3 to a model on the right, and another from the row with Split No. 2 to a model on the left.

	Split No.	Size	Type	ASA	BSA	dG_diss
1	1	12	1	77218.4	25619.7	31.6
2	12	1	77204.3	25139.2	22.4	
3	6	2	104504.1	32216.1	69.2	
4	2	12	1	77218.4	25619.7	31.6
5	12	1	77204.3	25139.2	22.4	

jsPISA 2.0.1 [1E94.pdb]

  ccp4serv7.rc-harwell.ac.uk/srv/pisa/sessions/US-568-SD/index.html

 Google


jsPISA 2.0.1 [1E94.pdb]



Input

Monomers

Interfaces

Stock

Crystal Splits

Log file

Interfaces

▼ List of interfaces

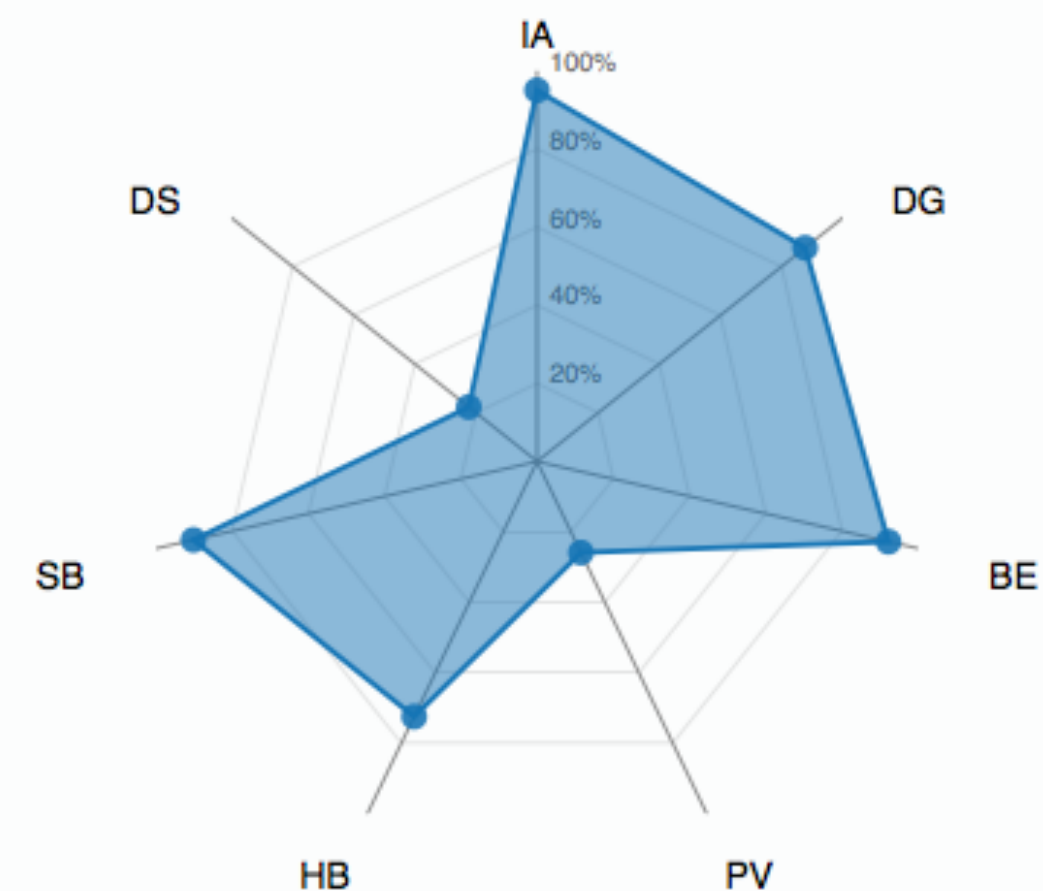
- ▶ 1. F || E
- ▶ 2. E || F
- ▶ 3. E || E
- ▶ 4. A || B
- ▶ 5. B || A
- ▶ 6. D || C
- ▶ 7. D || C
- ▶ 8. B || A
- ▶ 9. D || C
- ▶ 10. C || C
- ▶ 11. A || A
- ▶ 12. B || B
- ▶ 13. D || D
- ▶ 14. [ANP]E:500 || E
- ▶ 15. [ANP]F:501 || F
- ▶ 16. F || B
- ▶ 17. E || A
- ▶ 18. C || F
- ▶ 19. E || A
- ▶ 20. [ANP]E:500 || F
- ▶ 21. E || [ANP]F:501
- ▶ 22. F || E
- ▶ 23. E || E
- ▶ 24. F || F

Interface F || E

Summary

	Monomer 1		Monomer 2	
Monomer ID	F		E	
Class	Protein		Protein	
Symmetry operation	X,Y,Z		X,Y,Z	
Symmetry ID	1_555		1_555	
Interface atoms	232	7.3%	221	6.9%
Surface atoms	1925	60.5%	1921	60.4%
Total atoms	3184	100.0%	3183	100.0%
Interface residues	66	16.1%	65	15.9%
Surface residues	380	92.9%	385	94.4%
Total residues	409	100.0%	408	100.0%
BSA, Å ²	2146.6	9.6%	2345.5	10.6%
ASA, Å ²	22253.4	100.0%	22072.6	100.0%
Solvation energy, kcal/mol	-331.9		-334.3	
SE gain, kcal/mol	-7.5		-5.0	

Interaction radar



Interface parameters

IA : Interface area, Å² 2246
DG : Solvation Energy, kcal/mol -12.46
BE : Total Binding Energy, kcal/mol -26.65
PV : Hydrophobic P-value 0.3542
HB : Number of Hydrogen Bonds 11
SB : Number of Salt Bridges 25
DS : Number of Disulphide Bonds 0

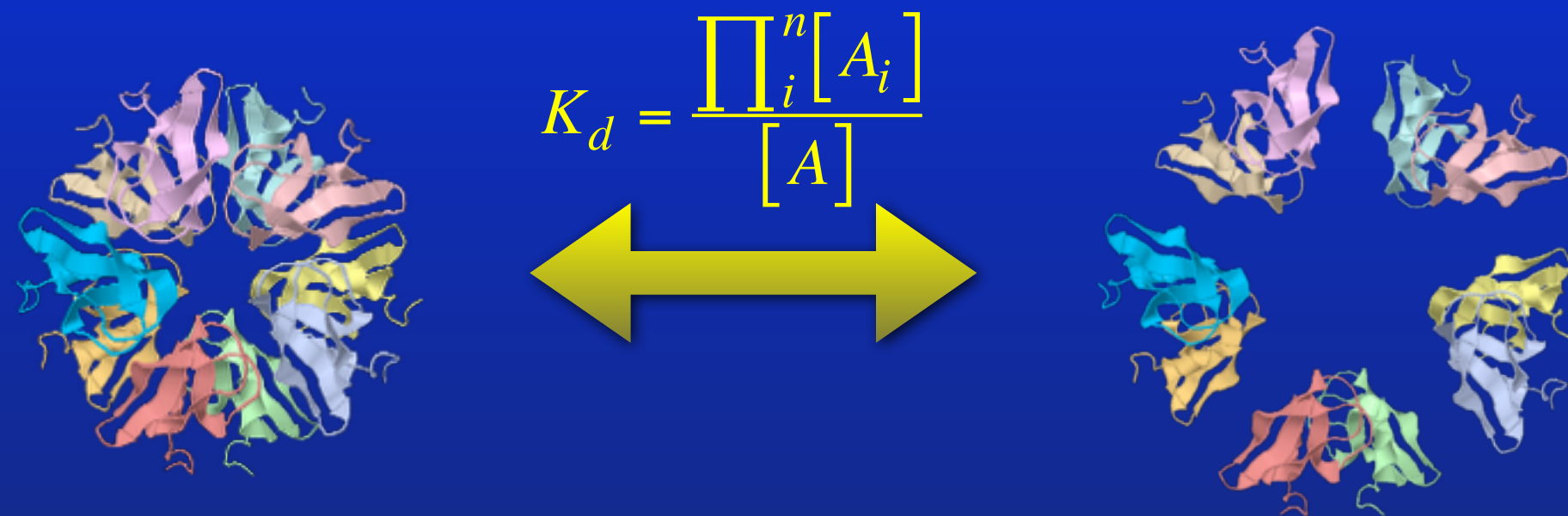
Download PDB

Download PDBx

View in JSMol

What is “A Stable Complex?”

- ★ Chemical systems always move towards equilibrium:



- ★ PISA reports $\Delta G_0 = -RT \log K_d$, how to interpret?

✦ *In general, if equilibrium is shifted to the left ($K_d < 1$), the complex is stable.*

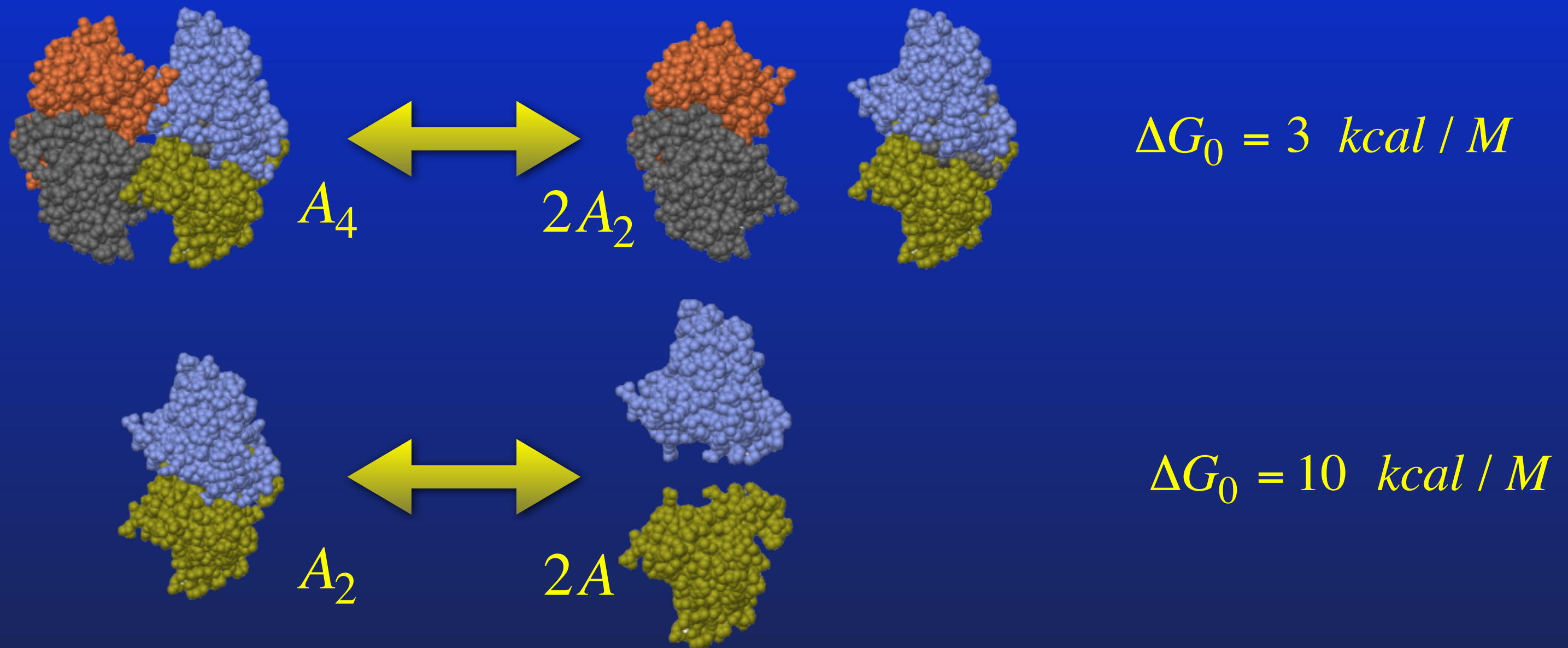
- But does this always mean that stable complex has higher concentration than the dissociates? - no it does not
- And this depends on protein concentration anyway? - yes it does
- And it also depends on the dissociation pattern (dissociation into monomers, dimers, trimers etc.)? How to know the pattern?

- *the pattern is, in essence, the minimum free energy route*
 - *is the minimum free energy route always unique?*
 - *does it not depend on concentration (temperature, pH, etc.), too?*



Is ΔG_0 Sufficient An Indicator?

★ Consider PDB entry 3LT5:



The tetramer is weaker than the dimer, so one may think that the structure is dimeric

But the tetramer is equilibrated with the dimer, so that their concentrations can be comparable

What is the correct answer?



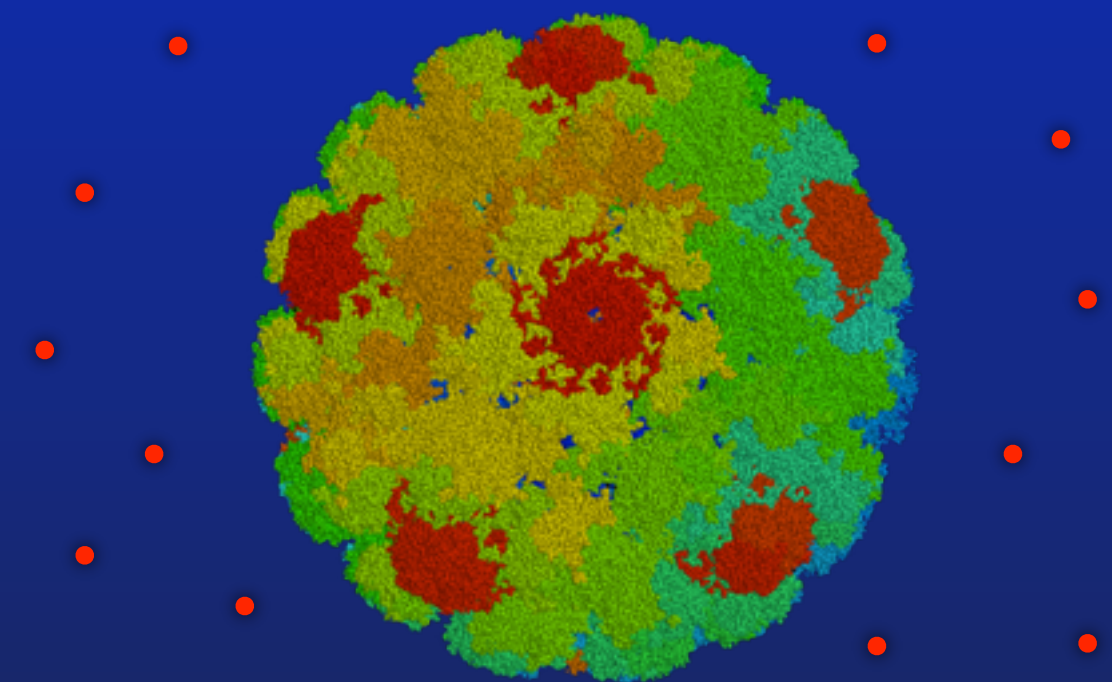
The Stock

- ★ All possible complexes co-exist in dynamic equilibrium and form a “stock”
- *PISA's Stock is limited to complexes formed by crystal interfaces*
- ★ Their stock concentration do vary
- ★ Concentrations depend on free energy of dissociation *and stock composition*
- ★ Concentration-based analysis is not very indicative:

for the equilibrium between large complex and its monomeric units on the right,

$$[A_{360}] \ll [A]$$

*from which one could conclude that the complex is unstable;
but obviously, the protein is highly aggregated*



- ★ Aggregated states are better indicated by the aggregation index:

$$A_i = \frac{m_i}{\sum_j m_j}$$

m_i mass of i th species in the Stock

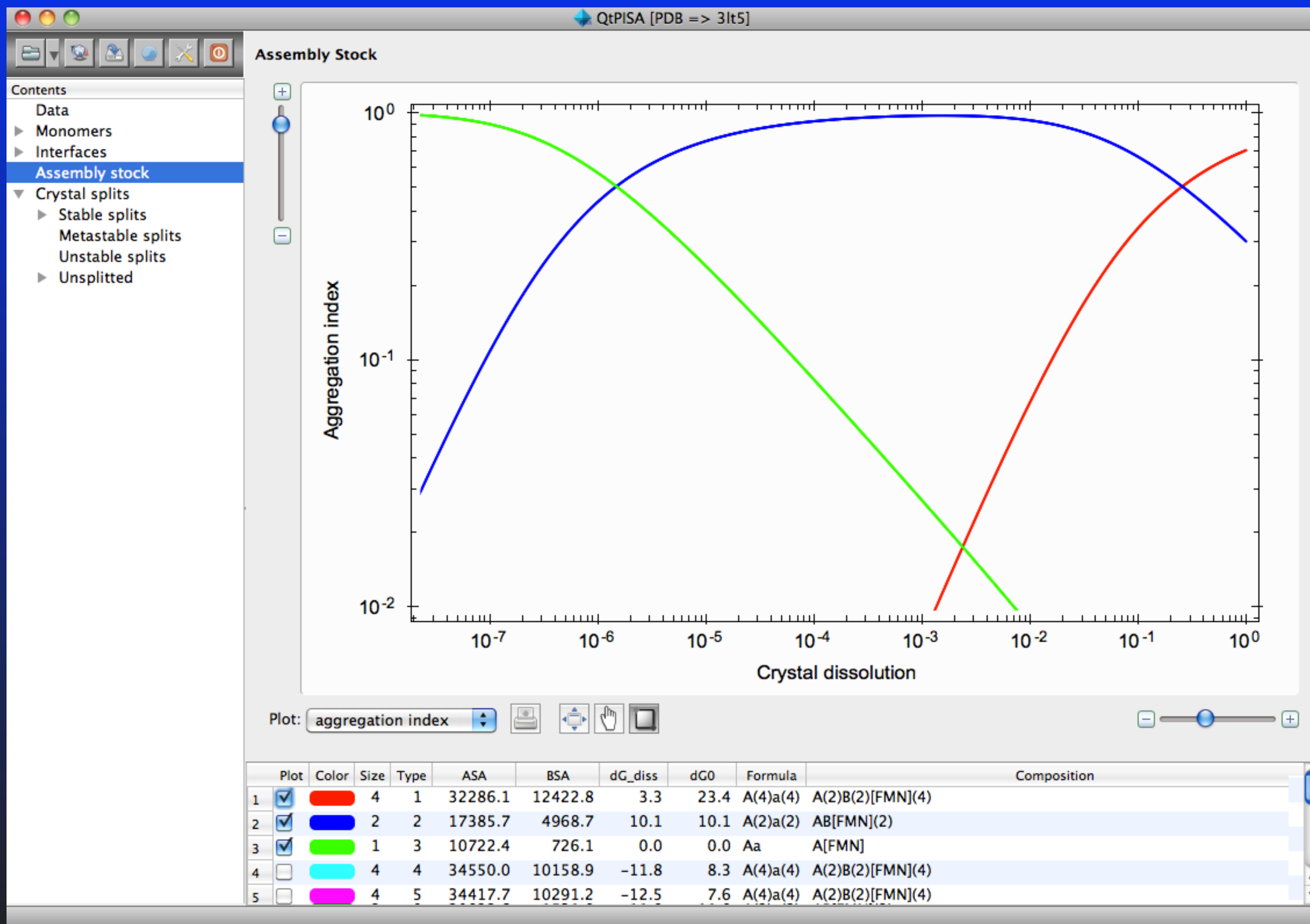
$$0 < A_i < 1$$

fully dissolved

fully aggregated



Assembly Stock for 3LT5 ($A_4 \rightleftharpoons A_2 \rightleftharpoons A$)



Classification of Protein Assemblies

Assembly classification on the benchmark set of 218 protein structures published in

Ponstingl, H., Kabir, T. and Thornton, J. (2003) Automatic inference of protein quaternary structures from crystals. J. Appl. Cryst. 36, 1116-1122.

	1mer	2mer	3mer	4mer	6mer	Other	Sum	Correct
1mer	49	3	0	1	1	1	55	89%
2mer	3	71+11	0	2+1	0	0	76+12	93%
3mer	1	0	22	0	1	0	24	92%
4mer	2	2+1	0	26+6	0	1	31+7	84%
6mer	0	0	0	1	10+2	0	10+3	92%
196+22 $\Leftrightarrow$ 196 homomers and 22 heteromers							Total: 196+22	90%

Classification error in ΔG_0 : ± 5 kcal/mol



Classification of Protein-DNA Complexes

Assembly classification on the benchmark set of 212 protein-DNA complexes published in

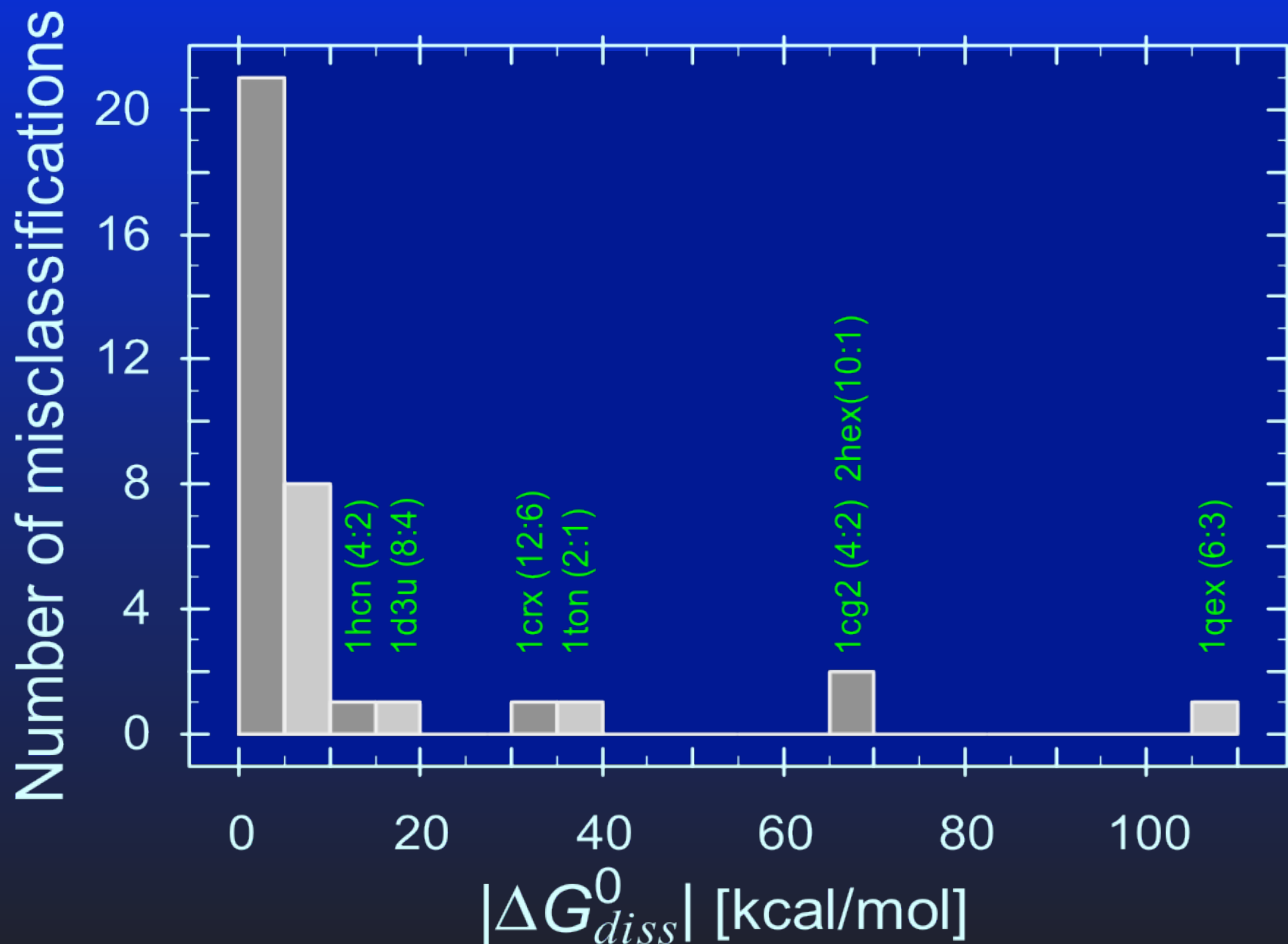
Luscombe, N.M., Austin, S.E., Berman, H.M. and Thornton, J. (2000) An overview of the structures of protein-DNA complexes. Genome Biol. 1, 1-37.

	2mer	3mer	4mer	5mer	6mer	10mer	Other	Sum	Correct
2mer	1	0	0	0	0	0	0	1	100%
3mer	6	96	0	0	1	0	2	105	91%
4mer	0	2	83	0	0	0	0	85	98%
5mer	0	0	2	3	0	0	0	5	60%
6mer	1	0	0	0	13	0	1	15	87%
10mer	0	0	0	0	0	1	0	1	100%
Total:								212	93%

Classification error in ΔG_0 : ± 5 kcal/mol

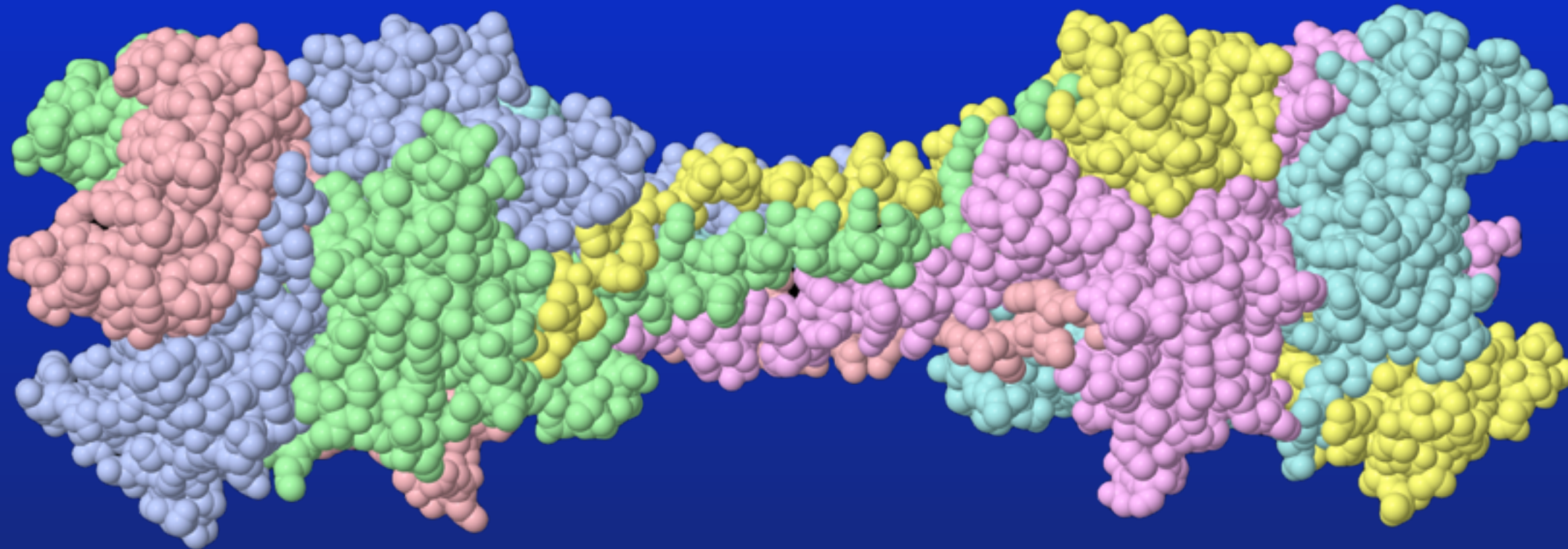


Free Energy Distribution of Misclassifications

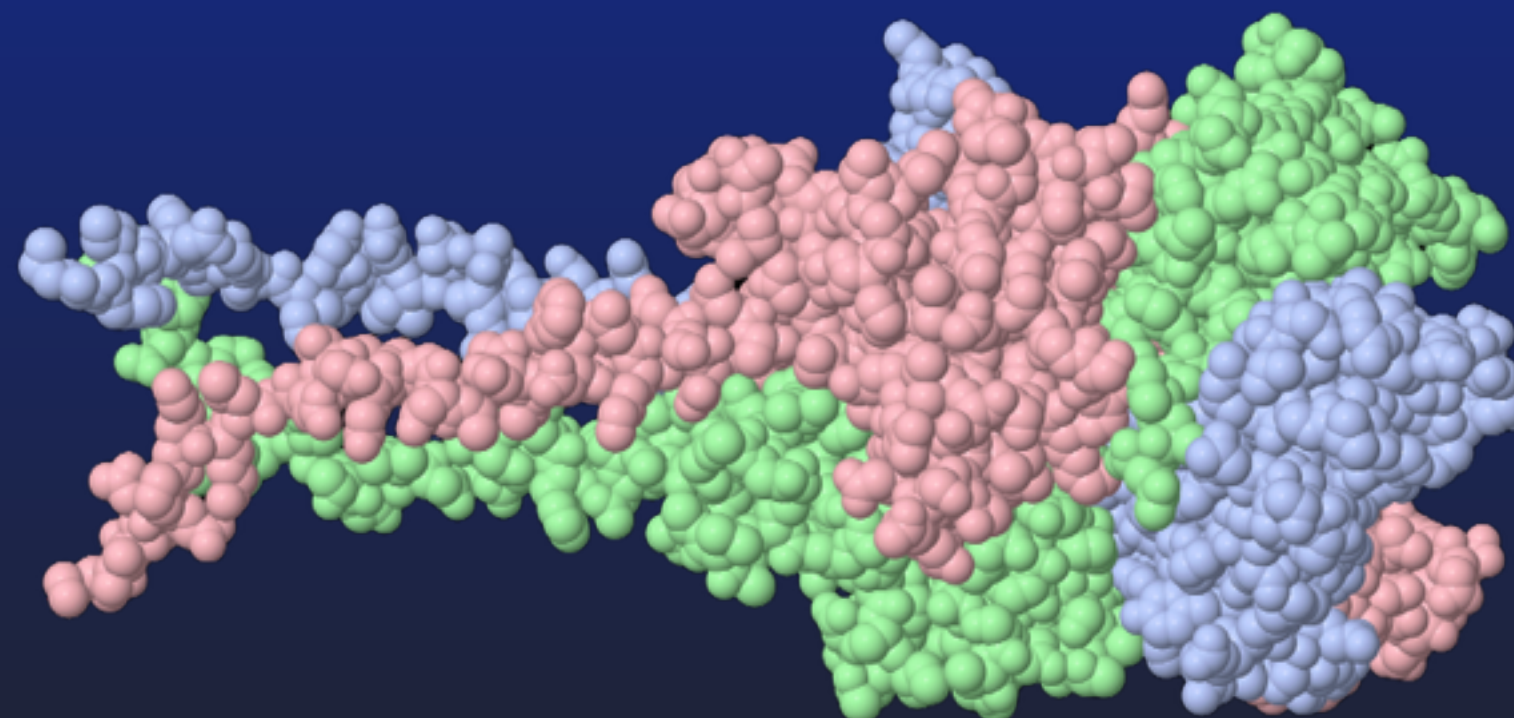


Example of misclassification: 1QEX

BACTERIOPHAGE T4 GENE PRODUCT 9 (GP9), THE TRIGGER OF TAIL CONTRACTION AND THE LONG TAIL FIBERS CONNECTOR



Predicted: homohexamer
Dissociates into 2 trimers
 $\Delta G_0 \approx 106$ kcal/mol

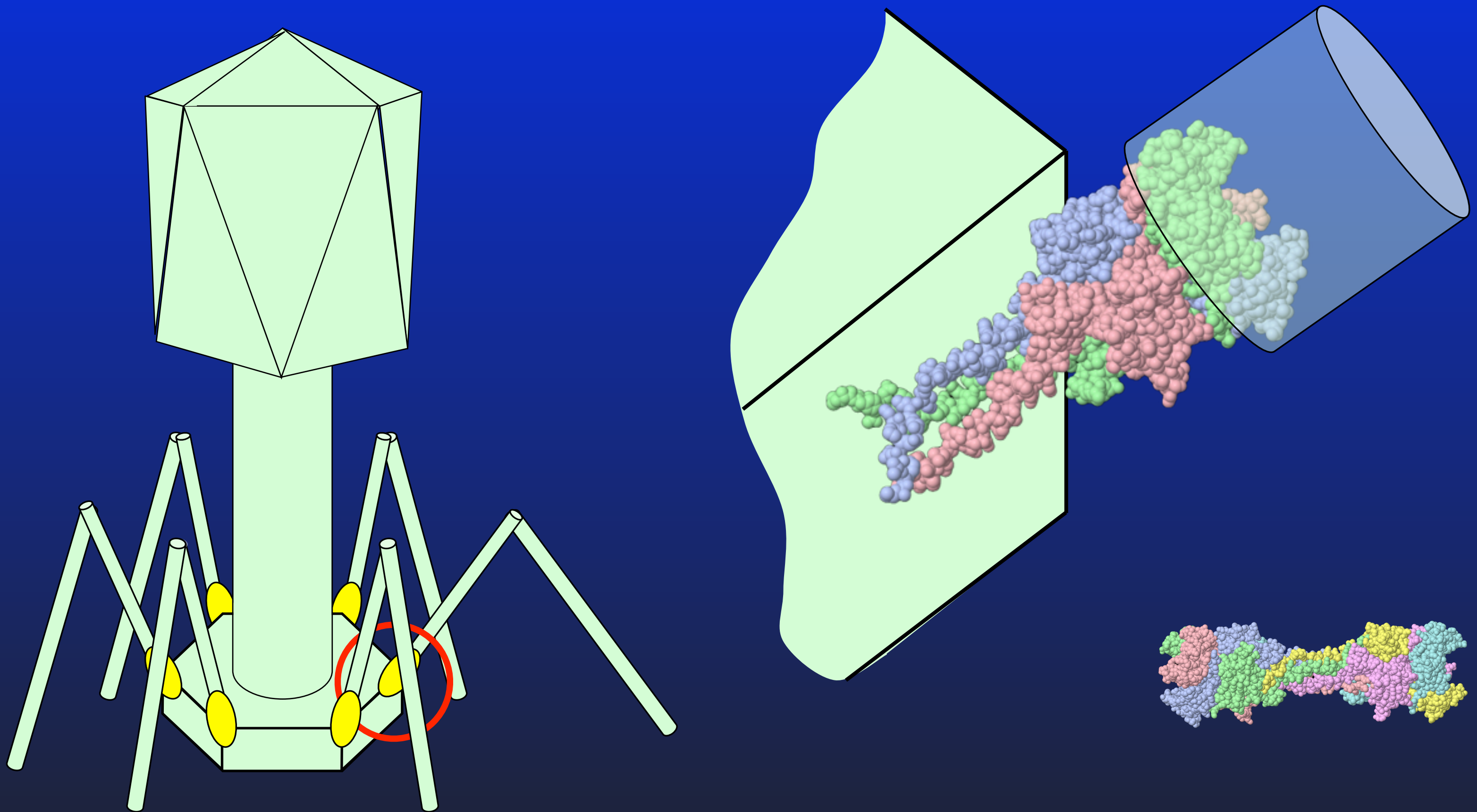


Biological unit: homotrimer
Dissociates into 3 monomers
 $\Delta G_0 \approx 90$ kcal/mol



Example of misclassification: 1QEX

BACTERIOPHAGE T4 GENE PRODUCT 9 (GP9), THE TRIGGER OF TAIL CONTRACTION AND THE LONG TAIL FIBERS CONNECTOR

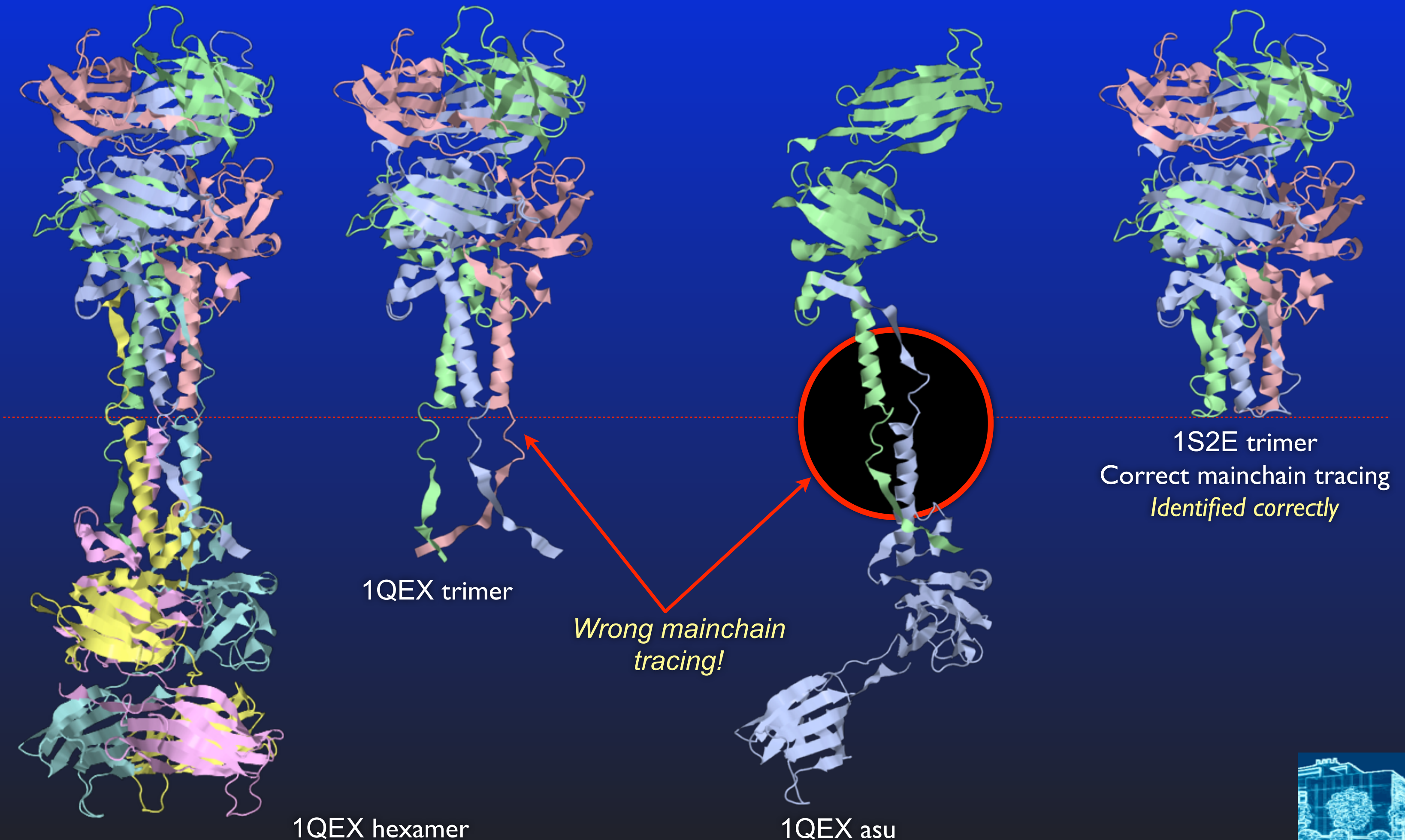


Rossmann M.G., Mesyanzhinov V.V., Arisaka F and Leiman P.G. (2004) *The bacteriophage T4 DNA injection machine*.
Curr. Opin Struct. Biol. 14:171-180.



Example of misclassification: 1QEX

BACTERIOPHAGE T4 GENE PRODUCT 9 (GP9), THE TRIGGER OF TAIL CONTRACTION AND THE LONG TAIL FIBERS CONNECTOR



Example of misclassification: 1D3U

TATA-BINDING PROTEIN / TRANSCRIPTION FACTOR

Predicted: octamer

Dissociates into 2 tetramers

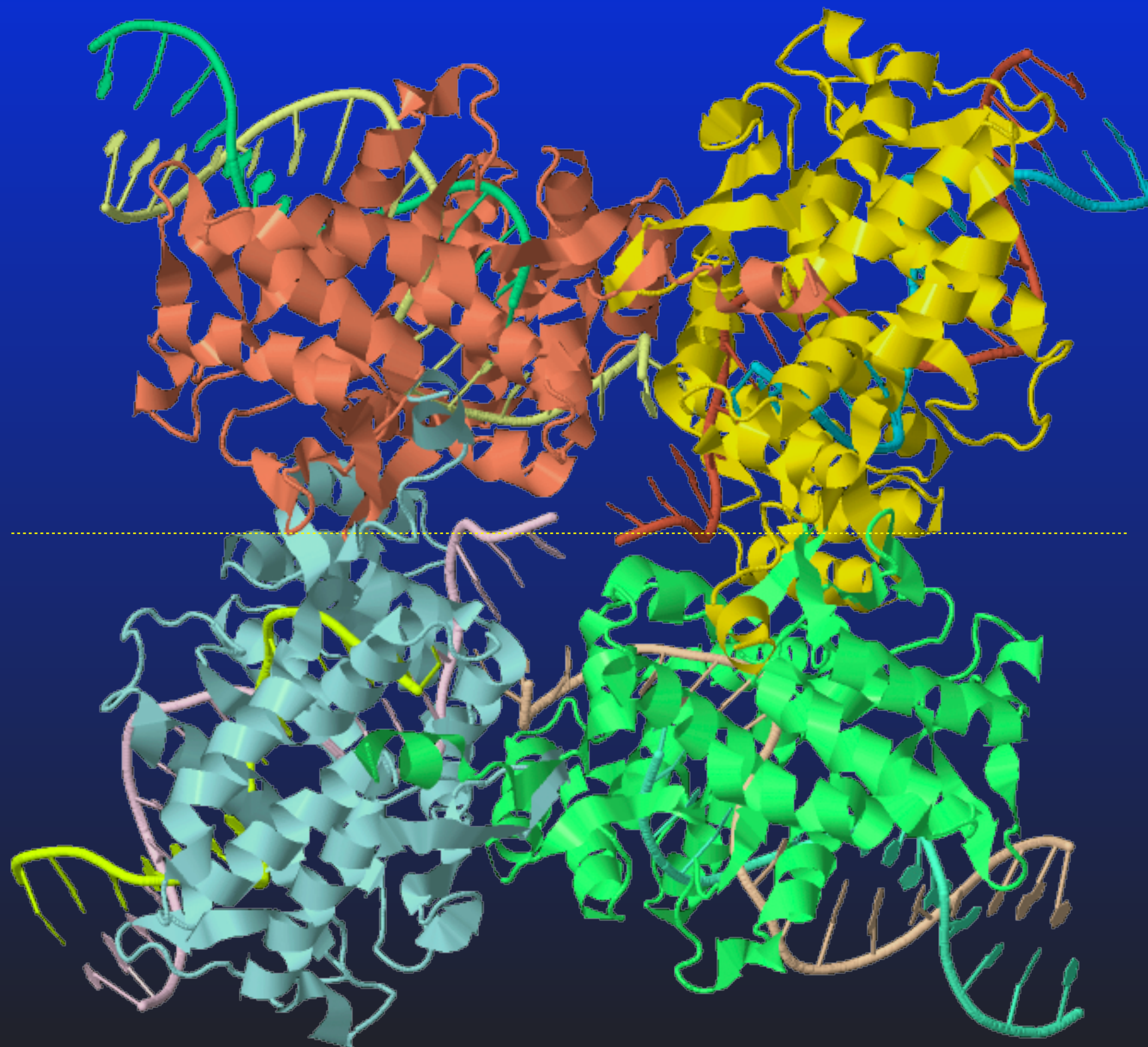
$$\Delta G_0 \approx 20 \text{ kcal/mol}$$

Functional unit:
tetramer



Example of misclassification: 1CRX

CRE RECOMBINASE / DNA COMPLEX REACTION INTERMEDIATE



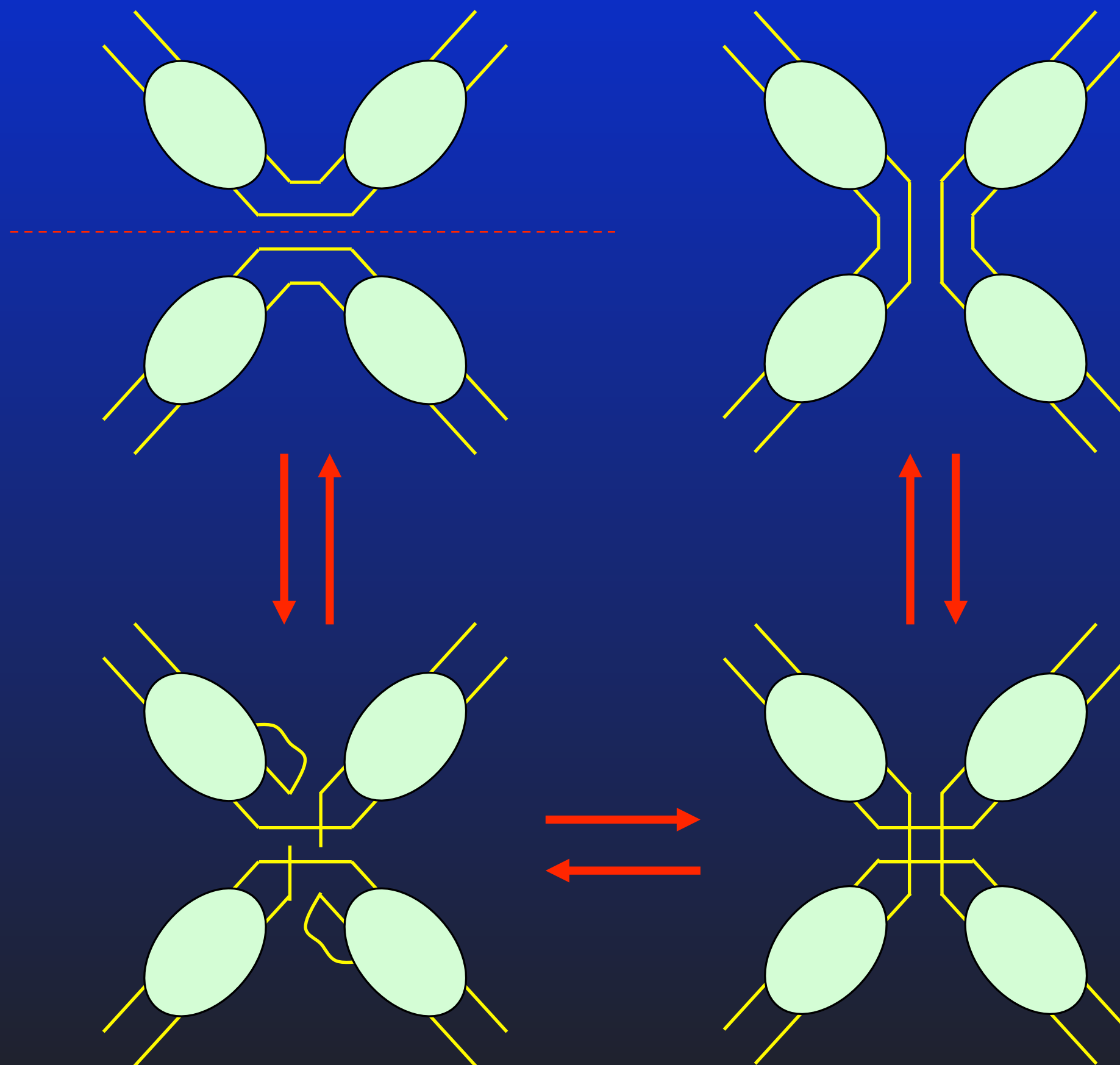
Predicted: dodecamer
Dissociates into 2 hexamers
 $\Delta G_0 \simeq 28$ kcal/mol

Functional unit: trimer



Example of misclassification: 1CRX

CRE RECOMBINASE / DNA COMPLEX REACTION INTERMEDIATE



Guo F., Gopaul D.N. and van
Duyne G.D. (1997)

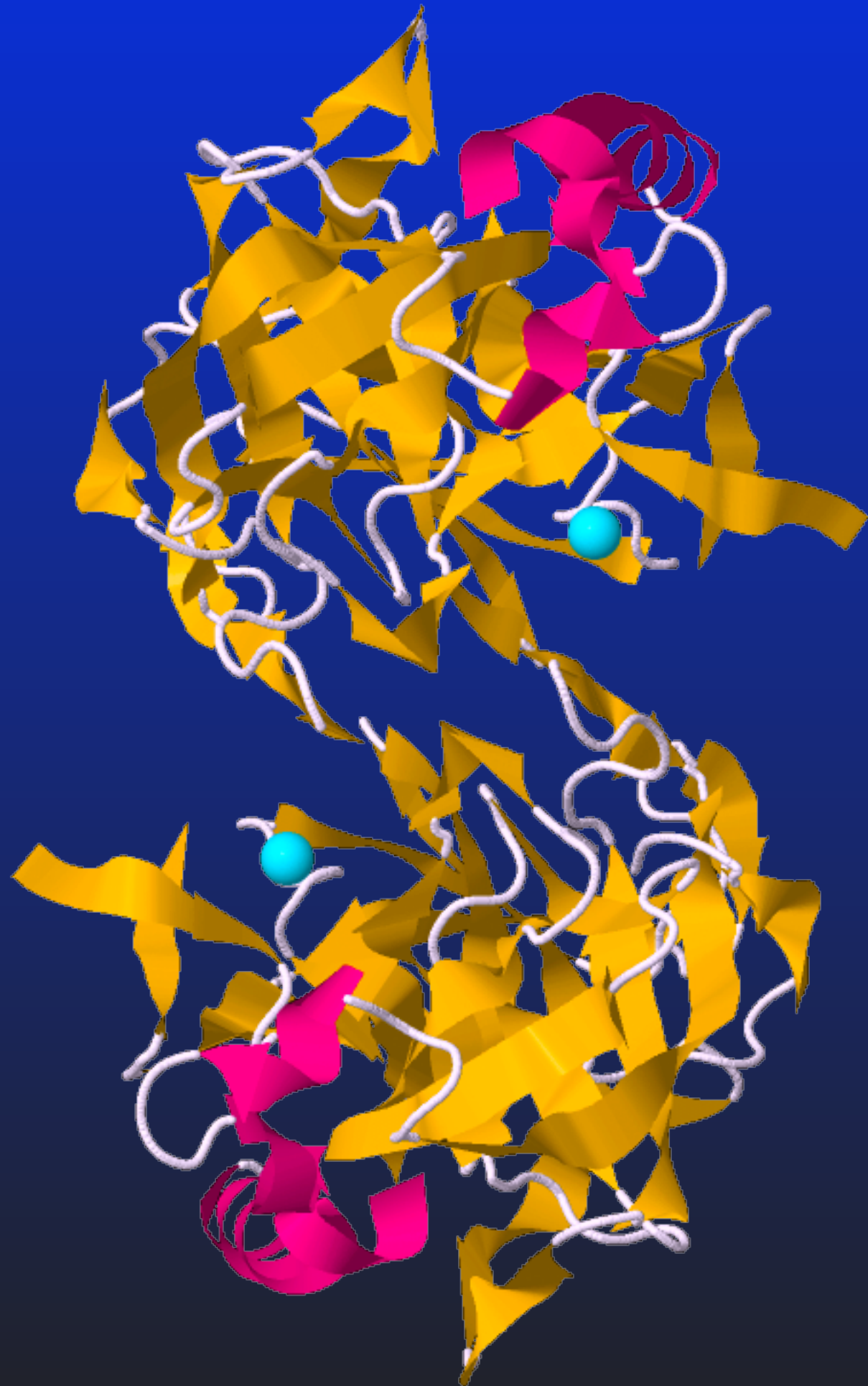
*Structure of Cre recombinase
complexed with DNA in a site-
specific recombination
synapse.*

Nature 389:40-46.



Example of misclassification: 1TON

TONIN



Predicted: dimer

Dissociates at

$$\Delta G_0 \simeq 37 \text{ kcal/mol}$$

Biological unit: monomer

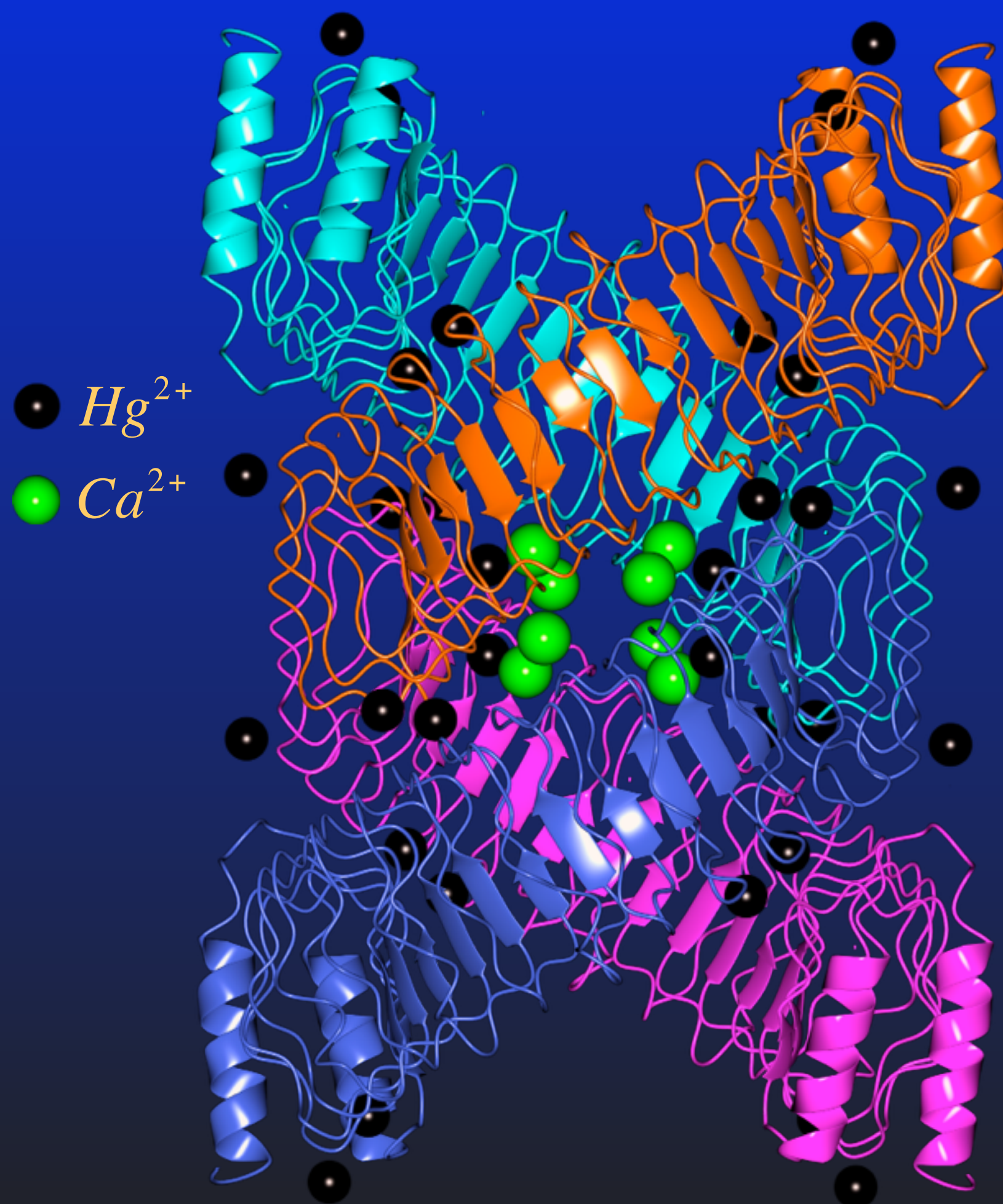
Apparent dimerization is an artefact due to the presence of Zn^{+2} ions added to the buffer to aid crystallization. Removal of Zn from the file results in $\Delta G_0 \simeq 3 \text{ kcal/mol}$

Fujinaga M., James M.N.G. (1997) *Rat submaxillary gland serine protease, tonin structure solution and refinement at 1.8 Å resolution.* J.Mol.Biol. 195:373-396.



Example of ion effect: 1G9U vs 1JL5

Y. PESTIS CYTOXIN YopM



Predicted: homotetramer in form of a superhelix featuring a hollow cylinder with an inner diameter of ~ 35 Å.

	1G9U	1JL5
Space Group	$P4_222$	$I4_122$
ΔG_0 , kcal/mol	37	3
Number of ions	40	16

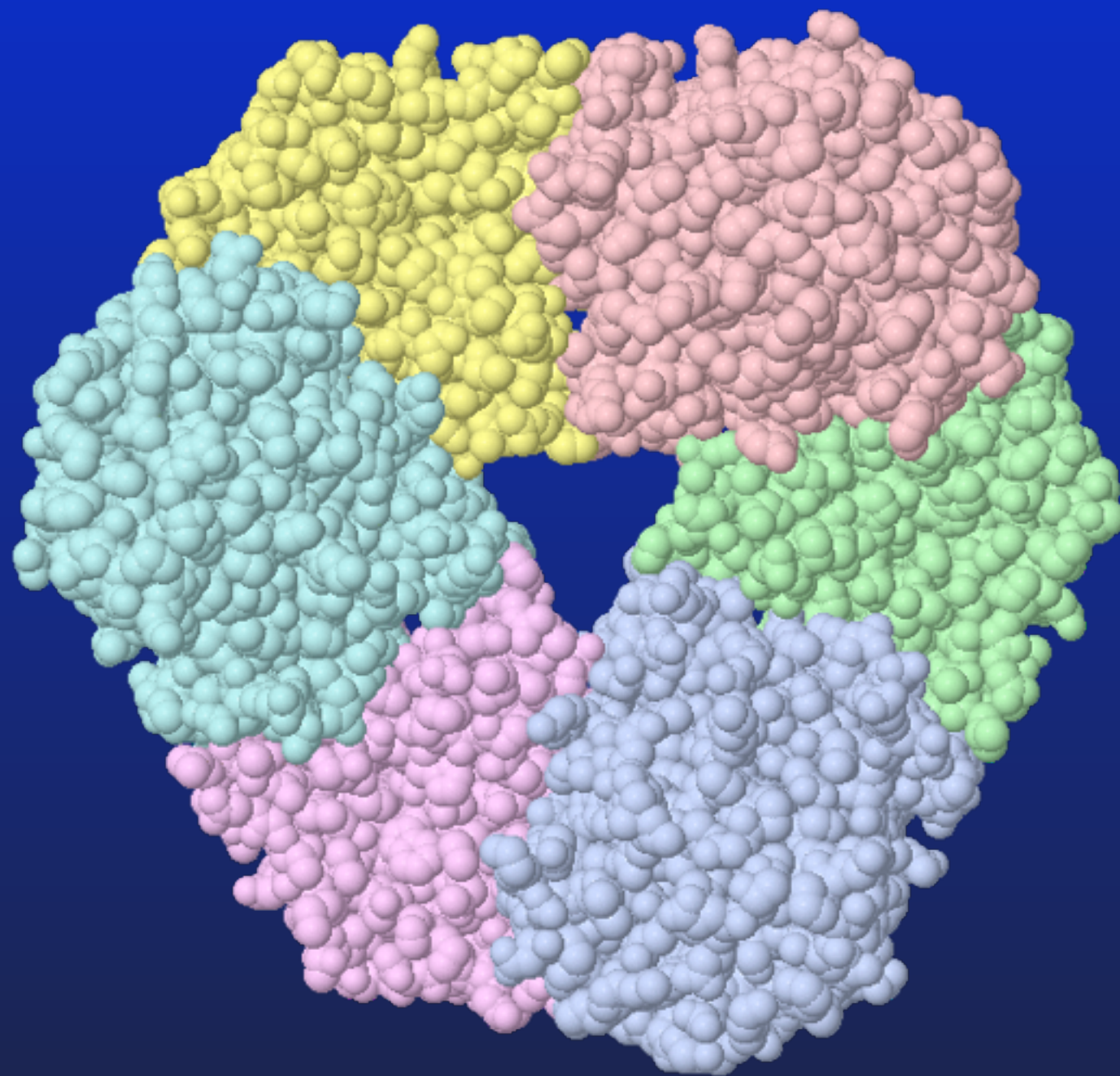
Biological unit: monomer

Evdokimov, A. G., Anderson, D. E., Routzahn, K. M. & Waugh, D. S. (2001). J. Mol. Biol. 312, 807–821

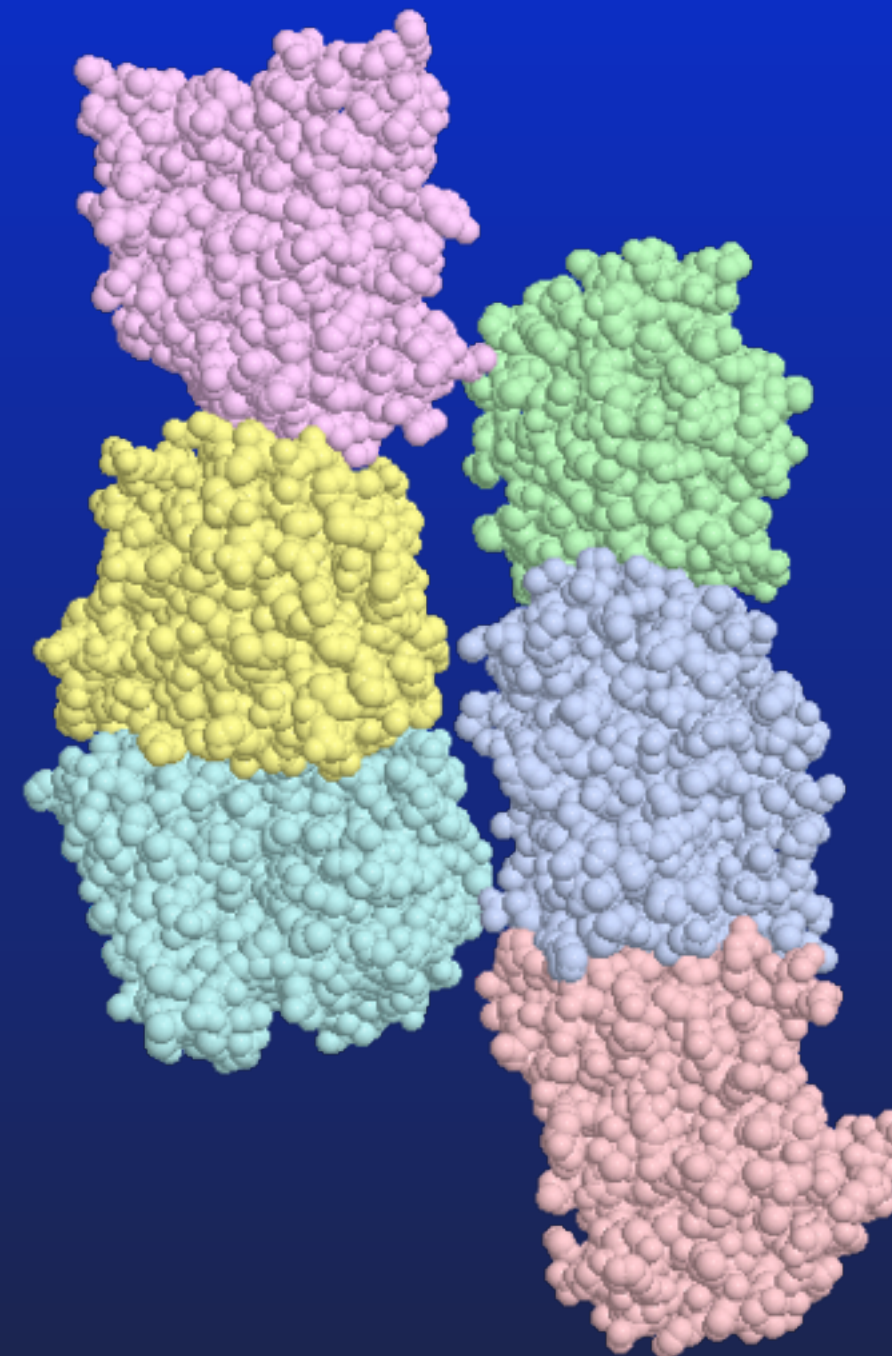
Removal of ions makes the structure monomeric in PISA estimates



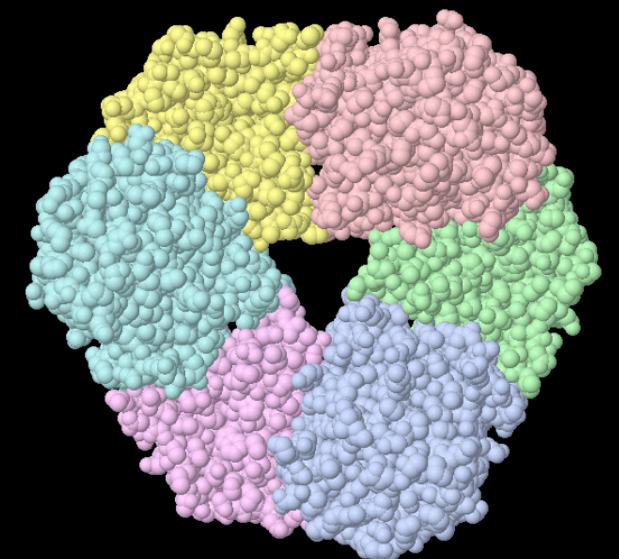
Example of misclassification: 1YWK



Predicted: homohexameric
Dissociates into 3 dimers at
 $\Delta G_0 \simeq 4.4$ kcal/mol



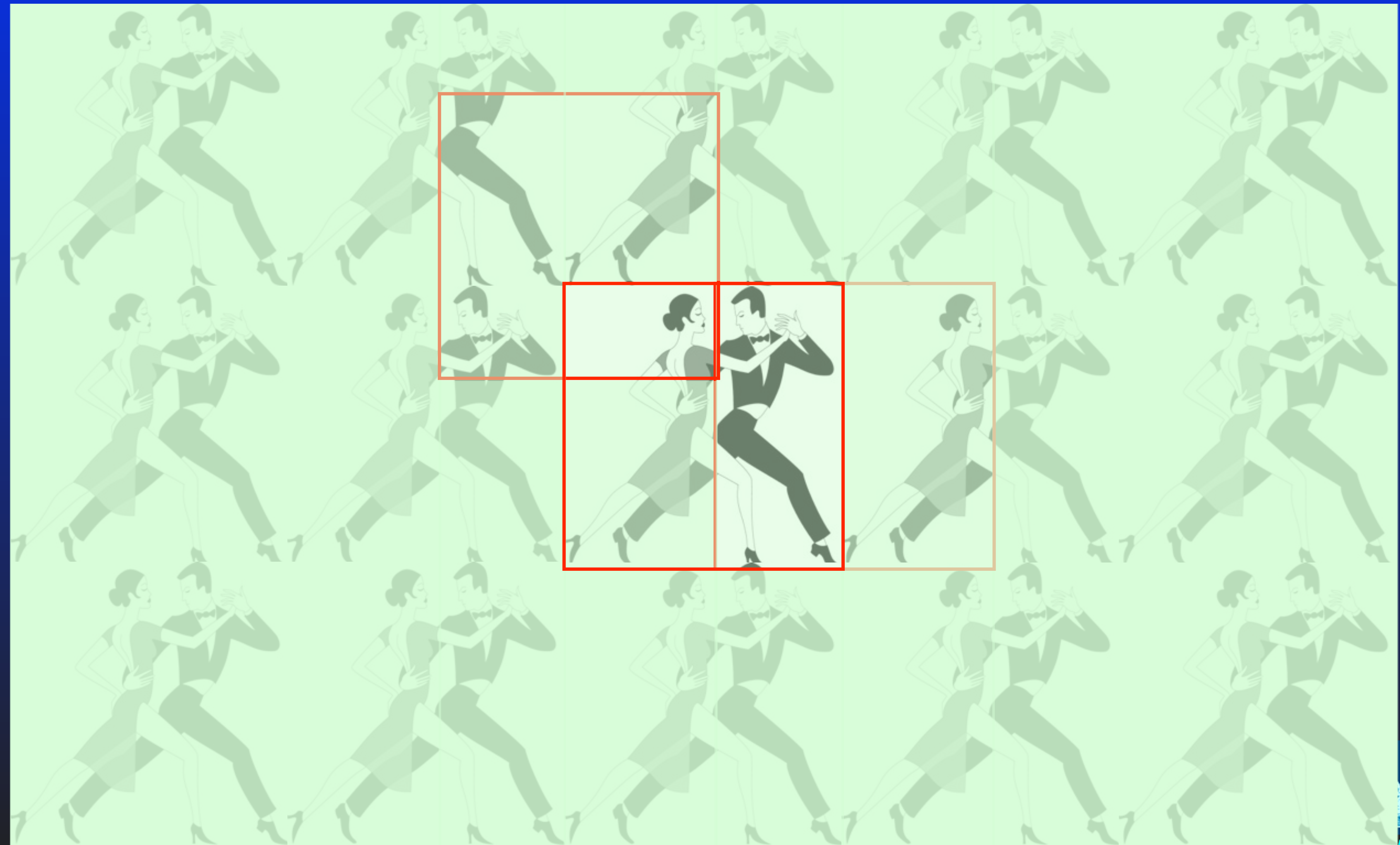
Believed to be:
monomeric,
6 units in ASU



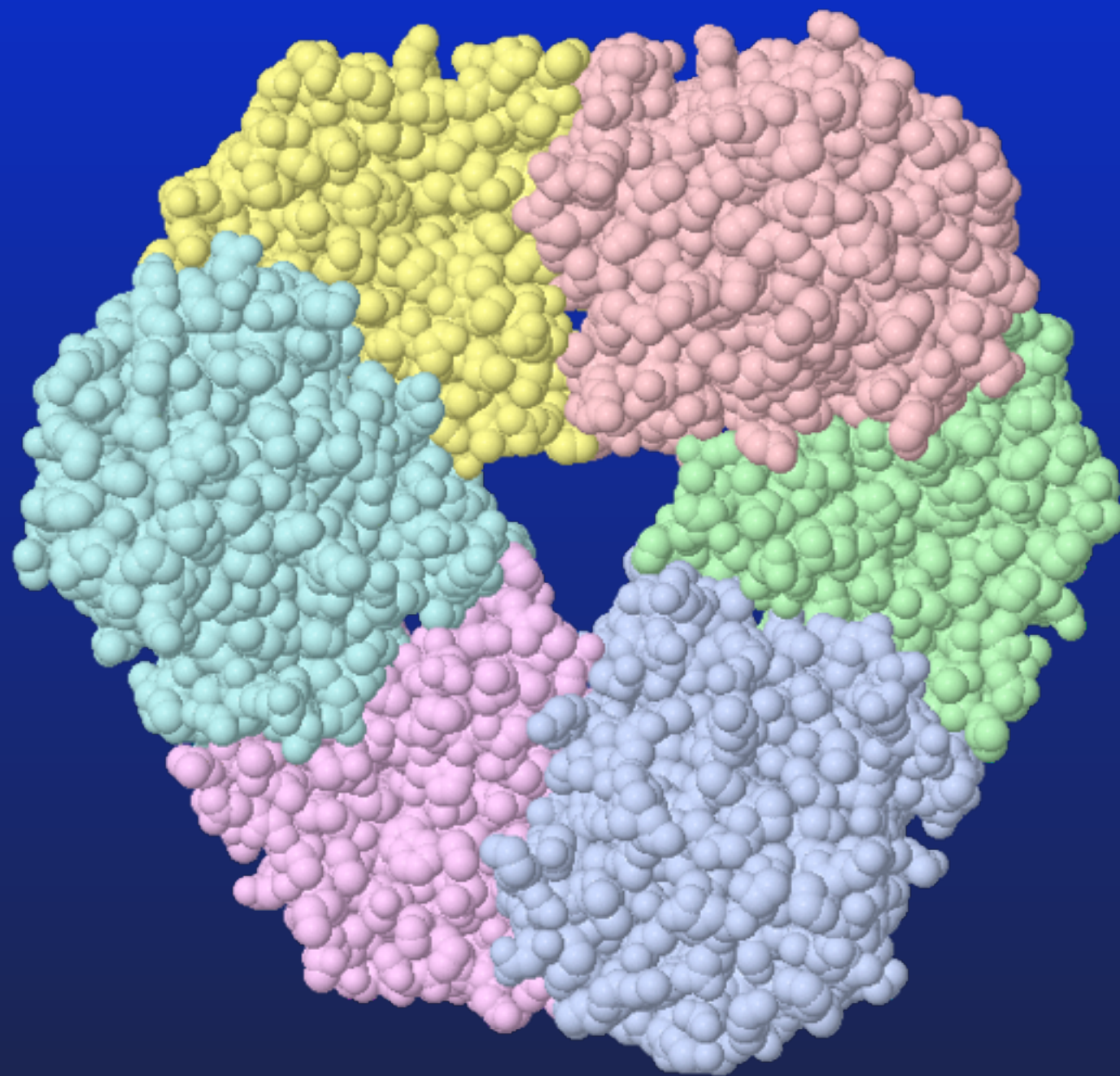
Structural homologue
1XRU:
RMSD = 0.9 Å
Seq.Id = 50%
Homohexameric with
 $\Delta G_0 \simeq 20$ kcal/mol



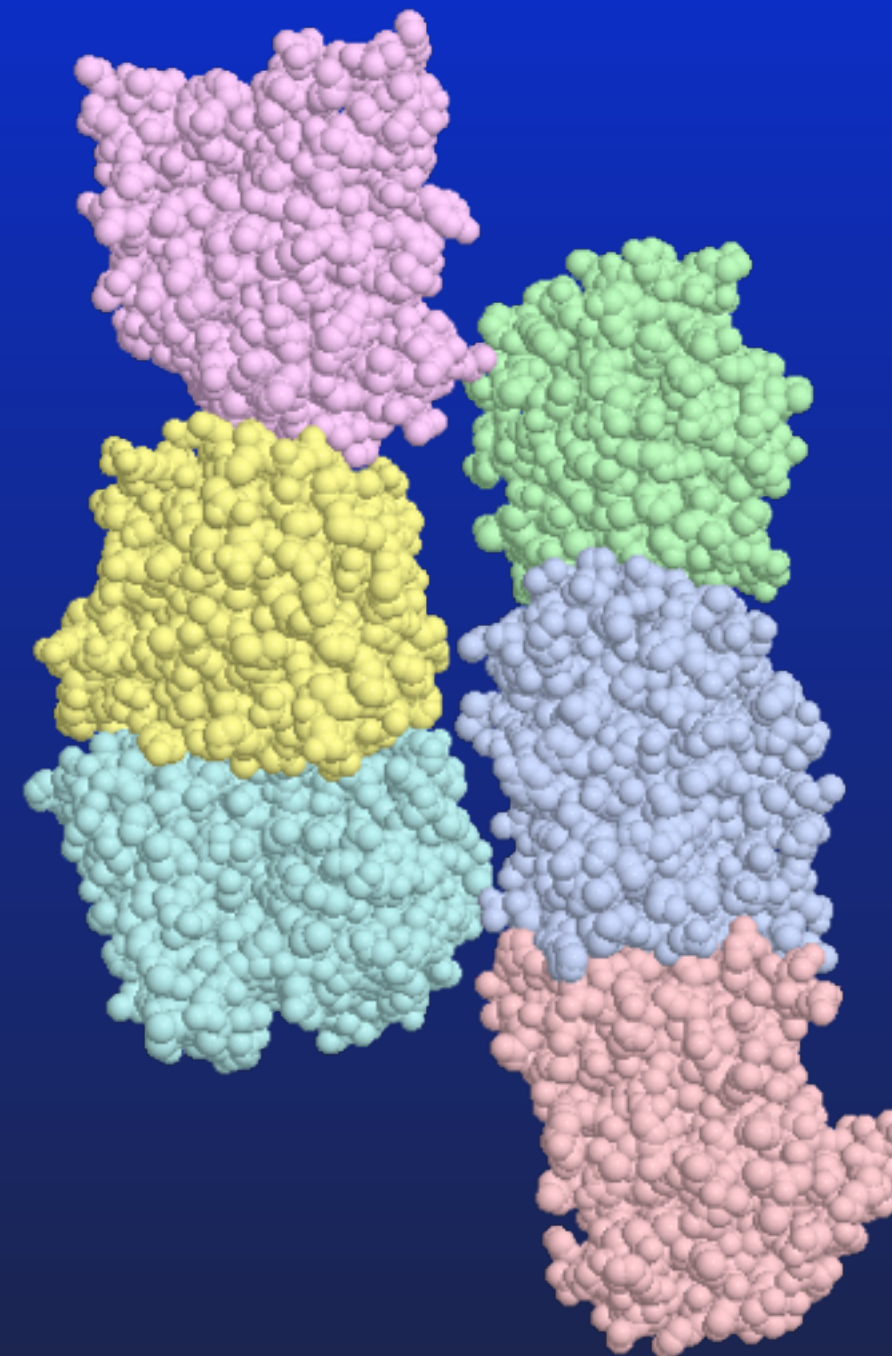
Choice of ASU



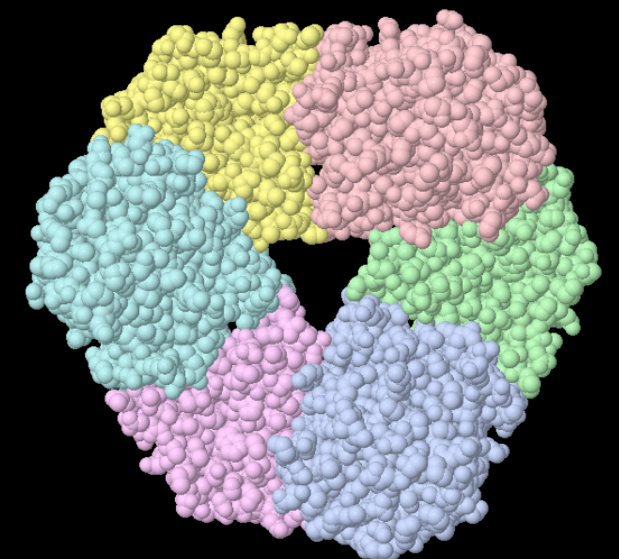
Example of misclassification: 1YWK



Predicted: homohexameric
Dissociates into 3 dimers at
 $\Delta G_0 \simeq 4.4$ kcal/mol



Believed to be:
monomeric,
6 units in ASU



Structural homologue
1XRU:
RMSD = 0.9 Å
Seq.Id = 50%
Homohexameric with
 $\Delta G_0 \simeq 20$ kcal/mol



Does it really work?

- ★ PISA appears to work quite well, which seems to be a “problem”
 - ➔ 90% success rate achieved on the benchmark set
 - ➔ in 2007, wwPDB adopted PISA as a mandatory processing tool for all depositions
 - ➔ since that, feedback from wwPDB curators suggests that up to 95% of classifications made by PISA agree with experimental data on oligomeric state, where available, and with intuitive and common-sense considerations where experimental evidence is not given

- ★ Why it should work well? Two reasons:

Energy models and calculations are quite accurate

PISA relies on geometry of interactions given by crystal packing. PISA does not dock monomeric units; rather, it uses crystal contacts as “nature’s dockings” assuming that they are correct.

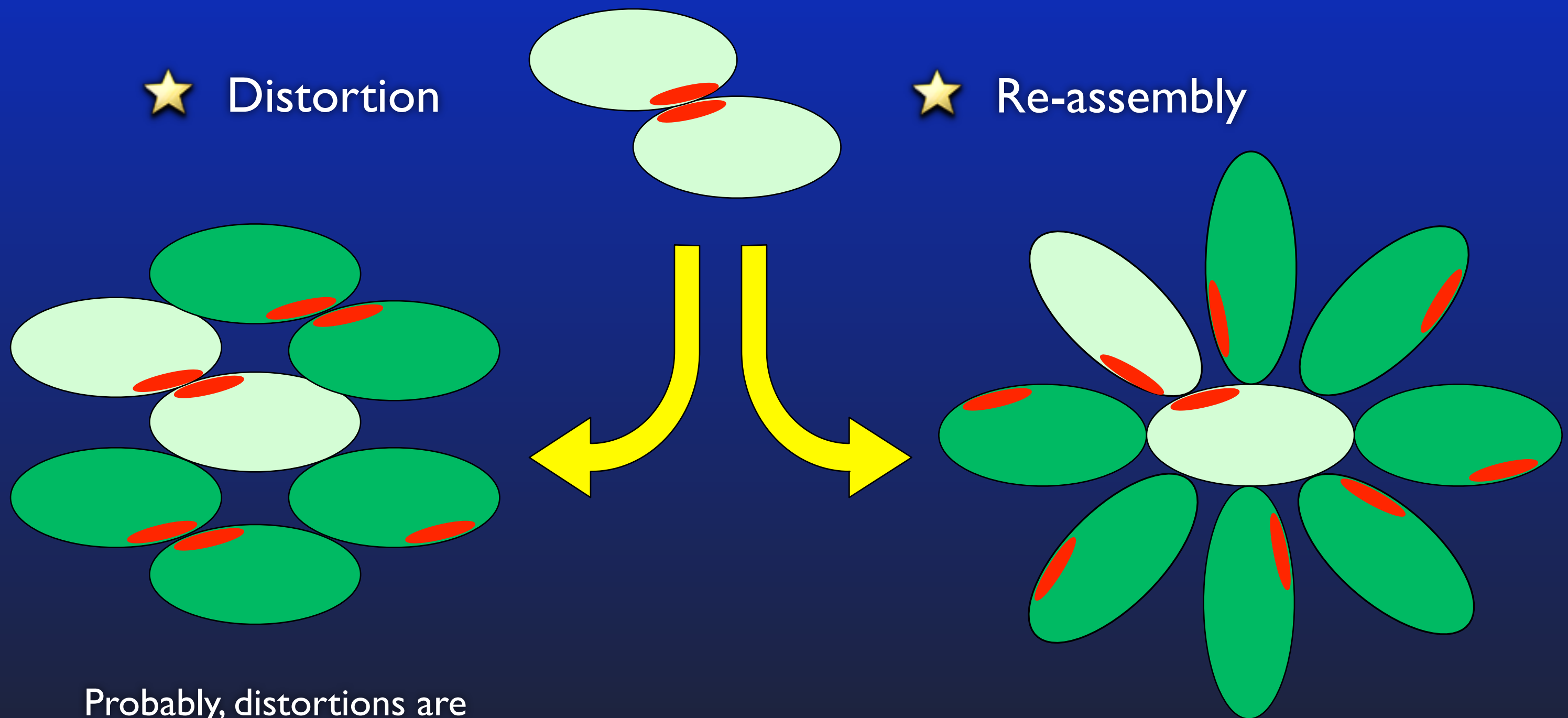
Obviously wrong

Probably correct



Distortions and Re-assembly

- ★ Crystal optimizes energy globally, therefore it may sacrifice biologically relevant interaction in favour of unspecific crystal contacts



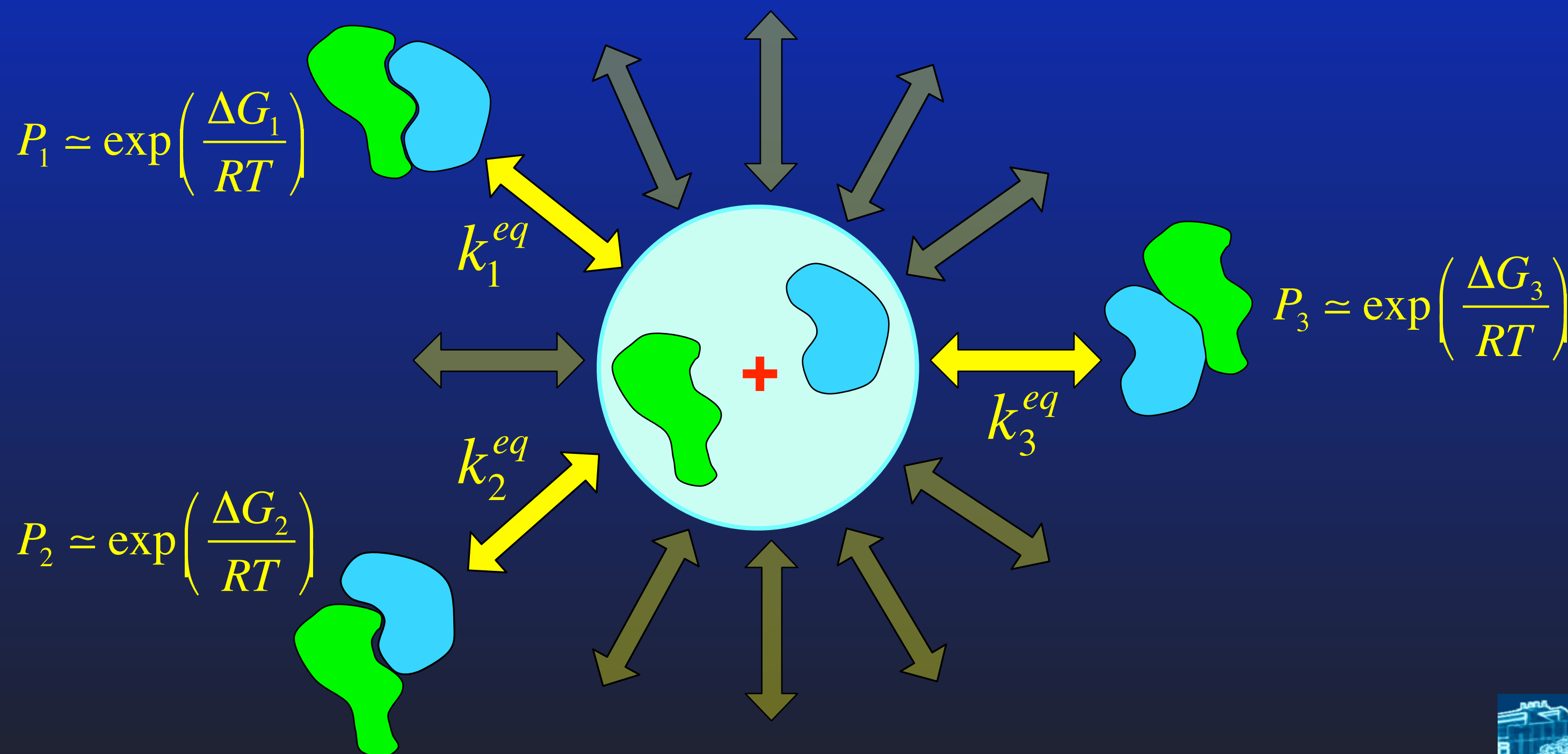
Probably, distortions are always there

There is a chance for re-assembly if interaction is weak



Alternative assemblies

- ★ All complexes (assemblies) have right to exist in solvent, however with different occurrence probabilities. These probabilities may differ of those in crystal environment, e.g., in case of substantially assisted crystallisation.



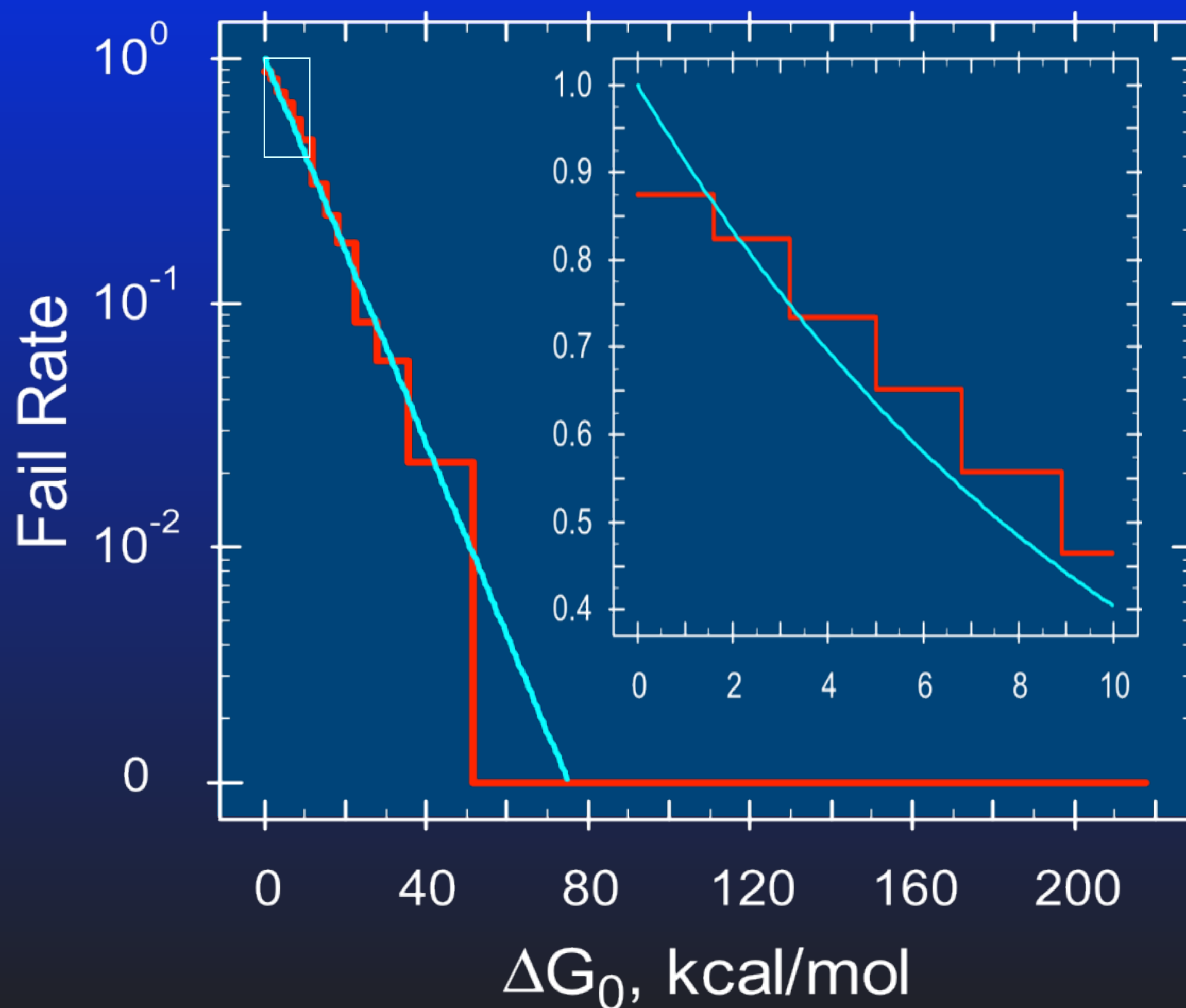
Real and superficial crystal contacts

- ★ If a crystal contact remains thermodynamically preferential in solution, the chances are that it represents a biochemically relevant interaction
- ★ Experimental data on structure of complexes in solution is **very** sparse
- ★ One can hope to get some clues using computational docking, assuming that docking approximates in-solvent situation
- ★ Being applied to 4065 non-redundant dimers from the PDB, docking **fails** to arrive at crystal interface in **38%** of instances

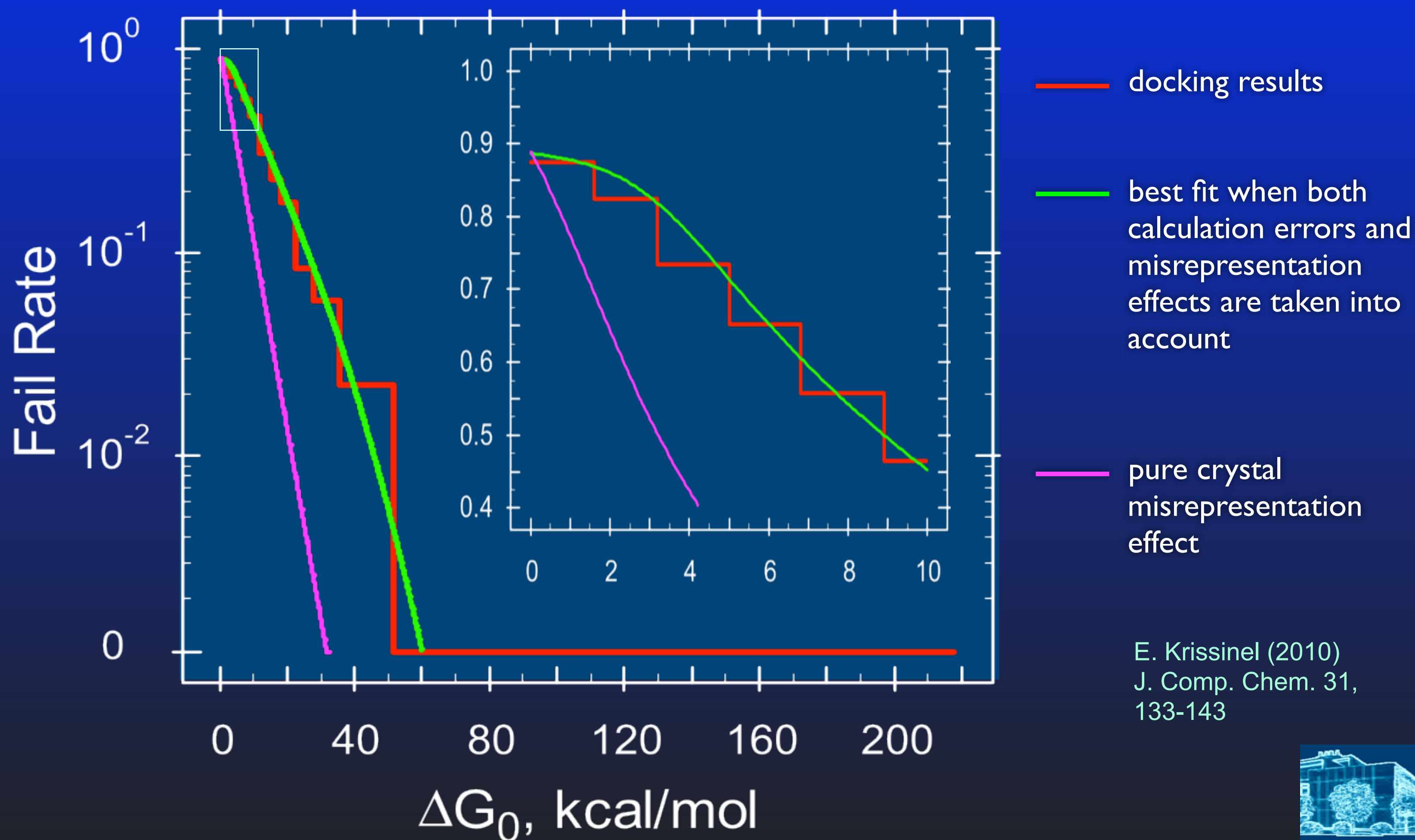
E. Krissinel (2010) J. Comp. Chem. 31, 133-143



Fail rate of docking



Calculation errors and crystal misrepresentation effects



E. Krissinel (2010)
J. Comp. Chem. 31,
133-143



So what is the practicality of all this?

- ★ PISA *is not* a substitution for experiments on the identification of protein's oligomeric state
 - both the software and (much less likely) experiment may give wrong results
 - in difference of experimental results, calculations do not make a scientific evidence!
- ★ PISA may be used for choosing complex models for molecular replacement
 - already done in BALBES automatic molecular replacement pipeline
- ★ PISA may be used for the interpretation of experimental results when evidence is not sufficient for a definite answer
 - which dimer?
 - inconclusive evidence (e.g. oligomeric state highly dependent on concentration/temperature/ion presence etc.)
- ★ PISA may be used for sanity checks, comparative analysis and flag raising
 - is proposed complex structure compatible with crystal packing?
 - is proposed complex different from close homologs?
 - is there a strong disagreement with biological/biochemical expectations?



Acknowledgements

Kim Henrick <i>European Bioinformatics Institute</i>	General introduction and PQS expertise
Mark Shenderovich <i>Structural Bioinformatics Inc.</i>	Helpful discussion
Hannes Ponstingl <i>Sanger Centre</i>	Sharing expertise and benchmark data
Sergei Strelkov <i>University of Leuven</i>	“Mystery” of bacteriophage T4
MSD & PDB teams <i>EBI & Rutgers</i>	Everyday use of PISA, examples, verification and feedback
CCP4 <i>Daresbury-York-Oxford</i>	Encouragement and publicity
~10,000 PISA users <i>Worldwide</i>	Using PISA and feedback
Biotechnology and Biological Sciences Research Council <i>(BBSRC) UK</i>	Research grant No. 721/B19544

