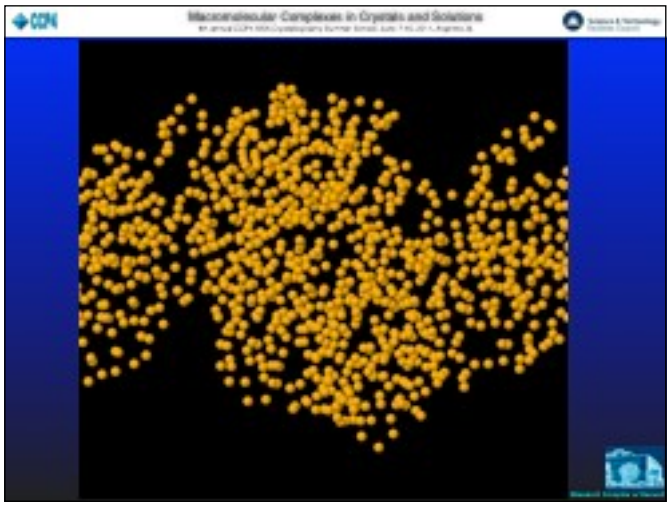
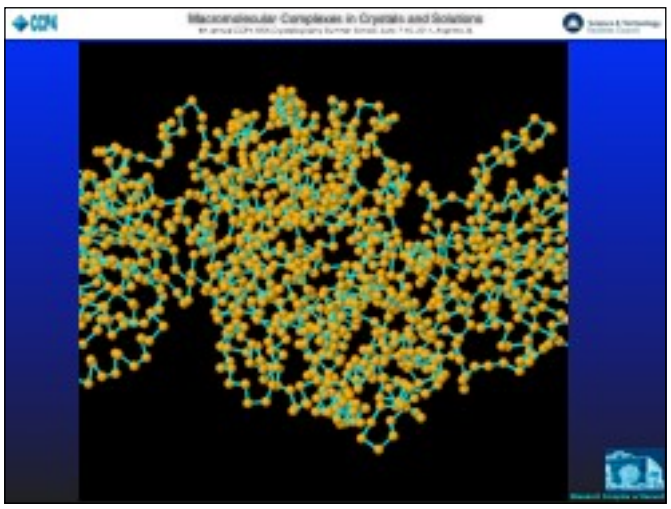
Macromolecular Complexes in Crystals and Solutions

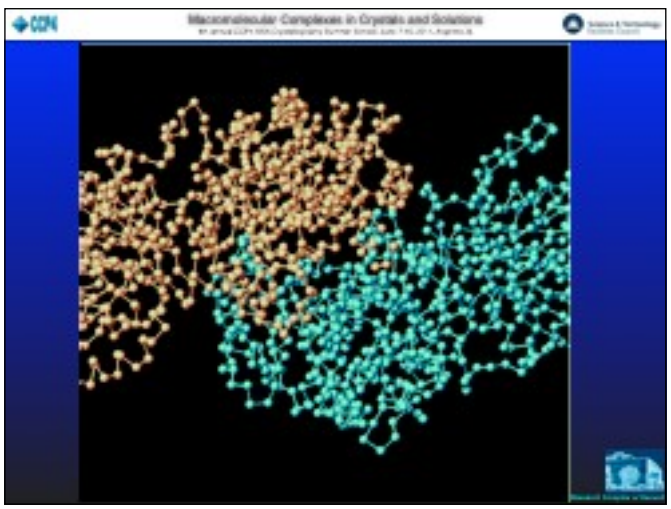
Eugene Krissinel
CCP4, STFC Research Complex at Harwell
Didcot, United Kingdom
eugene.krissinel@stfc.ac.uk



Macromolecular Complexes in Crystals and Solutions



Macromolecular Complexes in Crystals and Solutions



- Why would we want to know structure of a macromolecule?
  - for many reasons, but probably firstly for finding out how it interacts with other molecules
- Macromolecular crystals present us with models of biological structures and interactions between them
  - "if you want to know how A interacts with B – crystallize them together" (crystallographer's sweet dream)

a dimer?

or a dimer?

3bxc

"Monomer Association—According to gel filtration data, mKate exists in solution in the monomeric state at concentration as high as 10 mg/ml (13). However, in the crystalline state, which corresponds to a much higher protein concentration, mKate adopts at all pH values tetrameric arrangement with 222 symmetry, typically seen in GFP-like proteins."

Pletnev et al. (2009) J. Biol. Chem. 283, 28980-7

a dimer?

or a dimer?

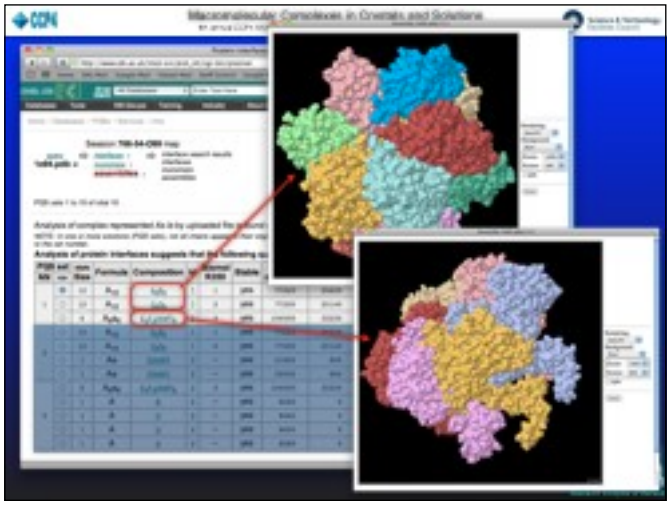
- Crystals present us with both real and artifactual interactions, which may be difficult to differentiate. Frequently used techniques:

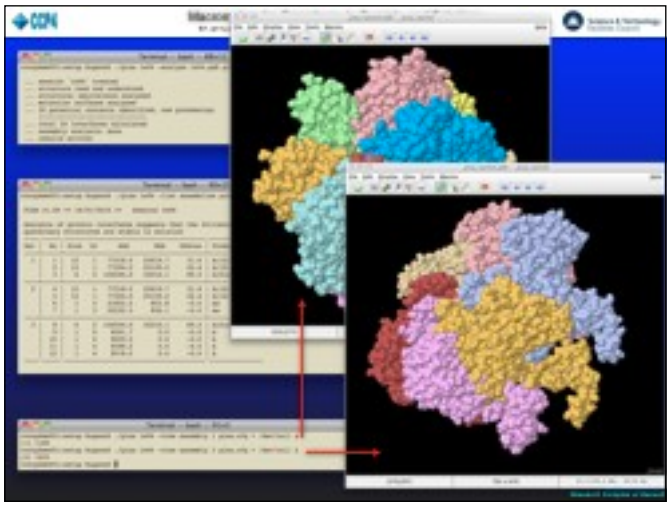
Theoretical:	sharp eye and scientific authority
Rules of thumb:	e.g. manifestation in different crystal forms
Experimental:	complementary studies (MS, EM, NMR, scattering)
Bioinformatical:	homology and interface similarity analysis
Computational:	energy estimates and modelling

- PISA software infers significant interactions and macromolecular assemblies from crystals by evaluating their free Gibbs energy:

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

http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html





Detection of Biological Units in Crystals: PISA summary

1. Enumerate all possible 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:

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

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

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	28+6	0	1	31+7	84%
6mer	0	0	0	1	10+2	0	10+3	92%
	196+22 <= 196 homomers and 22 heteromers					Total:	196+22	90%

Classification error in ΔG_0: ±5 kcal/mol

Classification of Protein-DNA Complexes

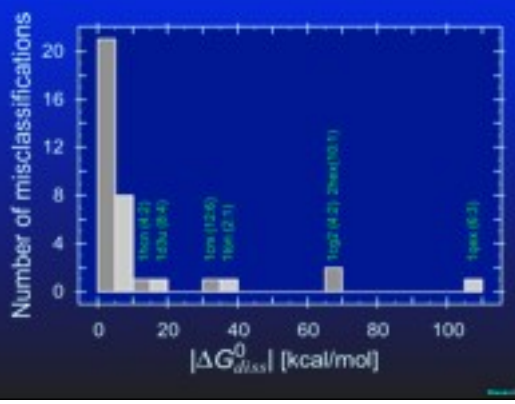
Assembly classification on the benchmark set of 212 protein-DNA complexes published in

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	98	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%

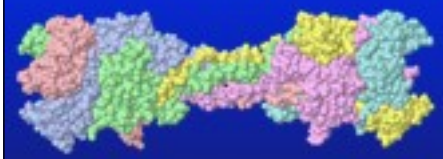
Classification error in ΔG_0: ±5 kcal/mol

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

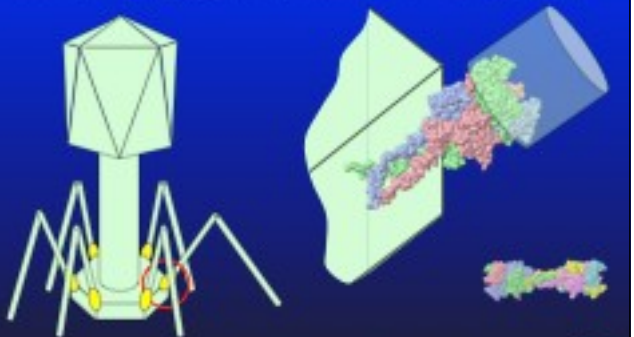


Predicted: homohexamer
Dissociates into 2 trimers
ΔG_0 ≈ 106 kcal/mol

Biological unit: homotrimer
Dissociates into 3 monomers
ΔG_0 ≈ 90 kcal/mol

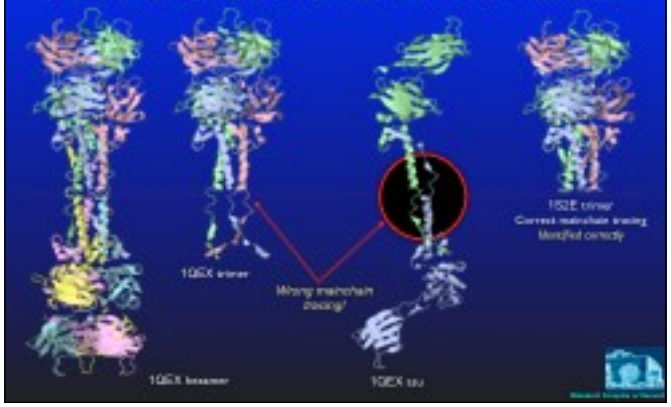
Example of misclassification: 1QEX

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

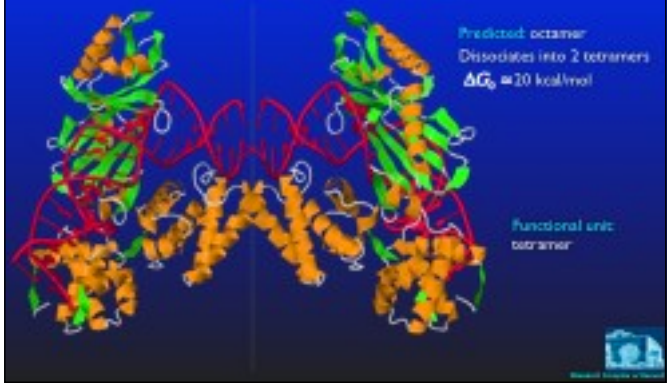


Rossmann M.G., Mesyanzhinov V.V., Arisaka F. and Leiman P.G. (2004) The bacteriophage T4 DNA injection machine. Curr. Opinion Struct. Biol. 14:171-180.

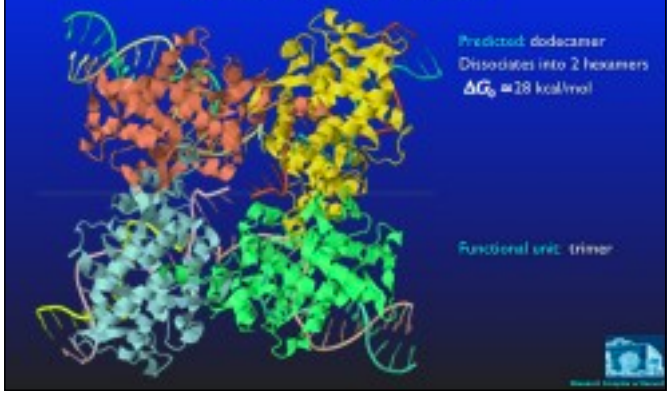
Example of misclassification: 1QEX



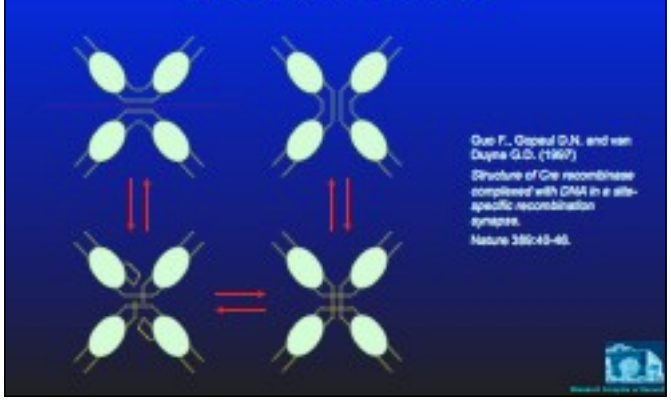
Example of misclassification: 1D3U



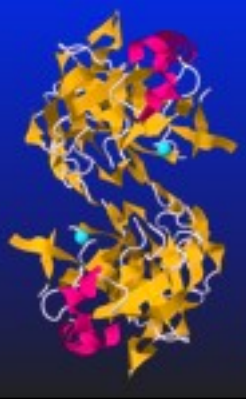
Example of misclassification: 1CRX



Example of misclassification: 1CRX



Example of misclassification: 1TON



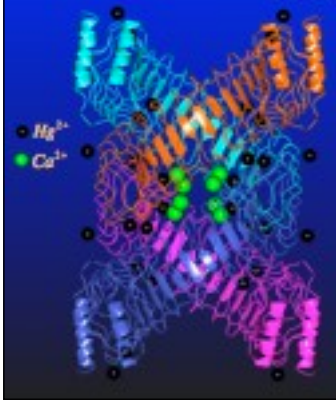
Predicted: dimer
Dissociates at
ΔG_0 ≈37 kcal/mol

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 ΔG_0 ≈ 3 kcal/mol

Fujinaga M., James M.N.G. (1987).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



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

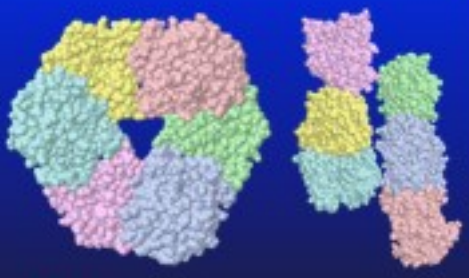
	1G9U	1JL5
Space Group	$P4_222$	$I4_122$
ΔG_0, kcal/mol	37	3
Number of ions	40	16

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 estimator

Example of misclassification: 1YWK

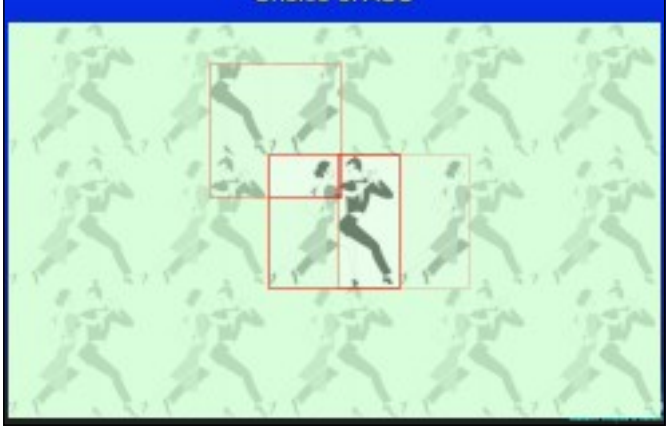


Structural homologue
1XBU:
RMSD = 0.9 Å
Seq.Id = 50%
Homohexameric with
ΔG_0 ≈ 4.4 kcal/mol

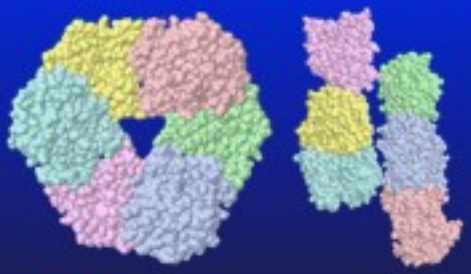
Predicted: homohexameric
Dissociates into 3 dimers at
ΔG_0 ≈4.4 kcal/mol

Believed to be:
monomeric,
6 units in ASU

Choice of ASU



Example of misclassification: 1YWK



Structural homologues (1XKU):
RMSD = 0.9 Å
Seq.Id = 30%
Homohexameric with
ΔG_D = 4.4 kcal/mol

Predicted: homohexameric
Dissociates into 3 dimers at
ΔG_D = 4.4 kcal/mol

Believed to be:
monomeric,
6 units in ASU

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 — Obviously wrong

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. — 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

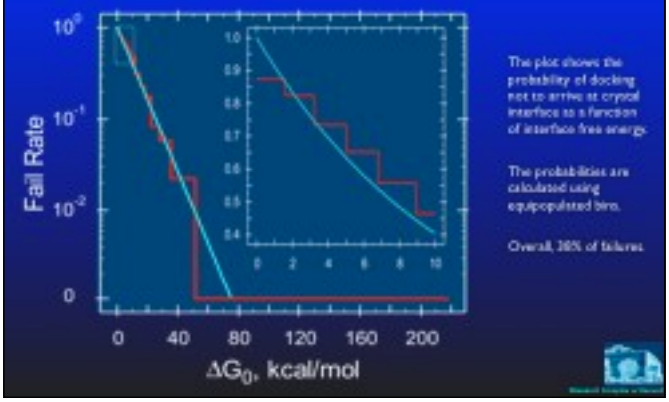
- All complexes (assemblies) have right to exist in solvent, however with different occurrence probabilities. These probabilities may differ of those in crystal environment if, e.g., crystallisation was substantially assisted.

Real and superficial interfaces

- If a crystal interface remains thermodynamically preferential in solvent, 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

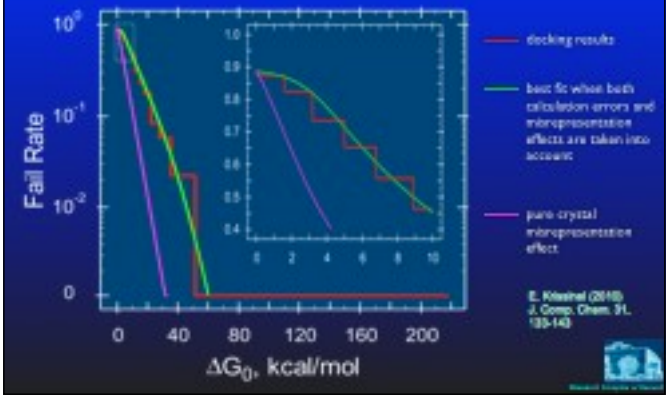


The plot shows the probability of docking not to arrive at crystal interface as a function of interface free energy.

The probabilities are calculated using equipopulated bins.

Overall, 38% of failures

Calculation errors and crystal misrepresentation effects



docking results

best fit when both calculation errors and misrepresentation effects are taken into account

pure crystal misrepresentation effect

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 identification of protein's oligomeric state
  - both 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 automatically in BALBES molecular replacement pipeline
- PISA may be used for interpretation of experimental results when evidence is not sufficient
  - 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 *European Bioinformatics Institute*	General introduction and PQS expertise
Mark Shenderovich *Structural Bioinformatics Inc.*	Helpful discussion
Hannes Ponstingl *Sanger Centre*	Sharing expertise and benchmark data
Sergei Strelkov *University of Leuven*	"Mystery" of bacteriophage T4
MSD & PDB teams *Eibi & Rutgers*	Everyday use of PISA, examples, verification and feedback
CCP4 *Daresbury-York-Oxford*	Encouragement and publicity
~10,000 PISA users *Worldwide*	Using PISA and feedback
Biotechnology and Biological Sciences Research Council (BBSRC) UK	Research grant No. 721/B19544