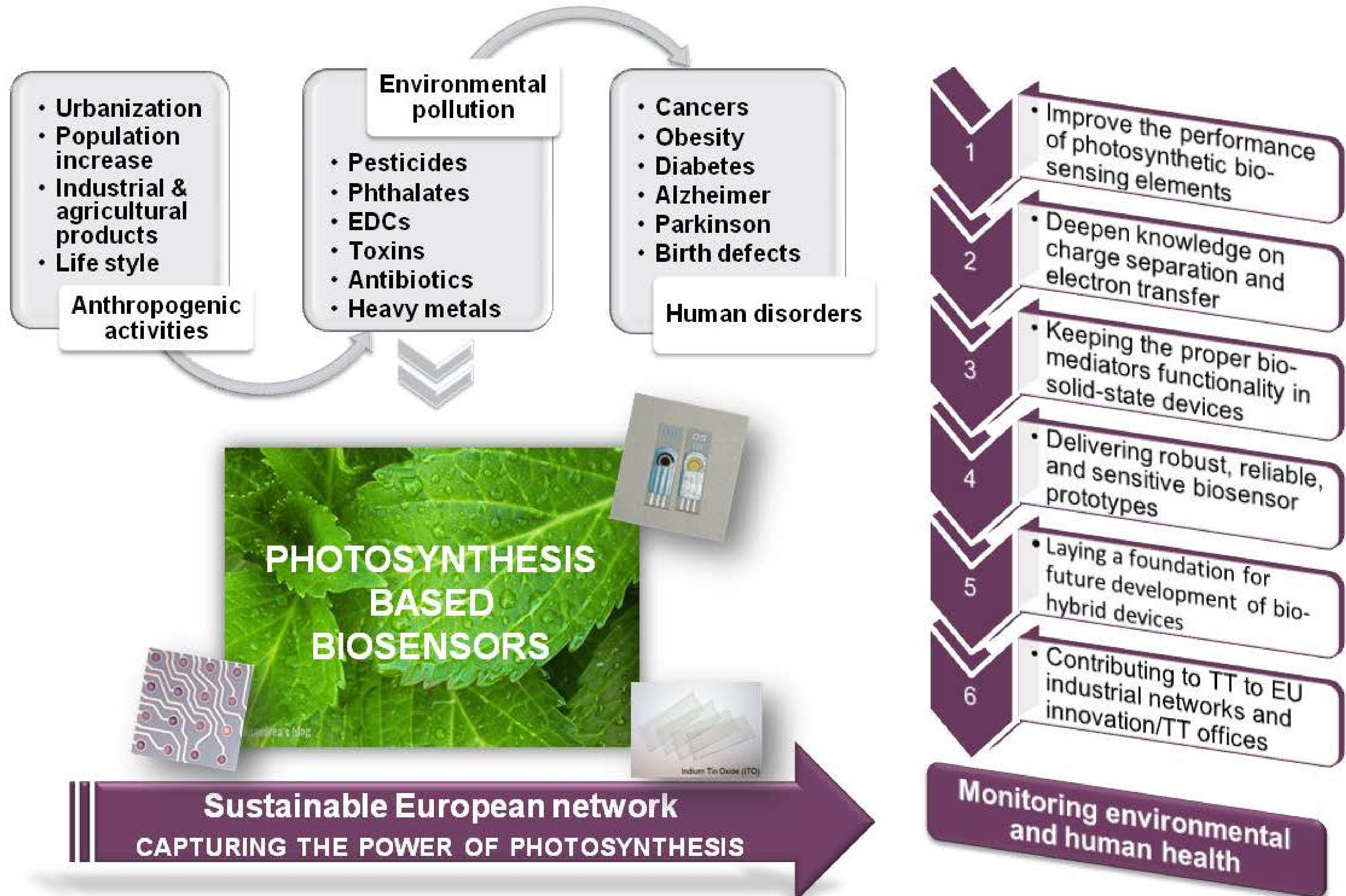
COST TD1102
PHOTOTECH: BIOSENSORS & BIOCHIPS

Molecular biotechnologies improving the bioreceptorial properties of the Photosystem II D1 protein

Giuseppina Rea, PhD
Institute of Crystallography
National Research Council of Italy
giuseppina.rea@ic.cnr.it

COST TD1102 PHOTOTECH: OVERVIEW



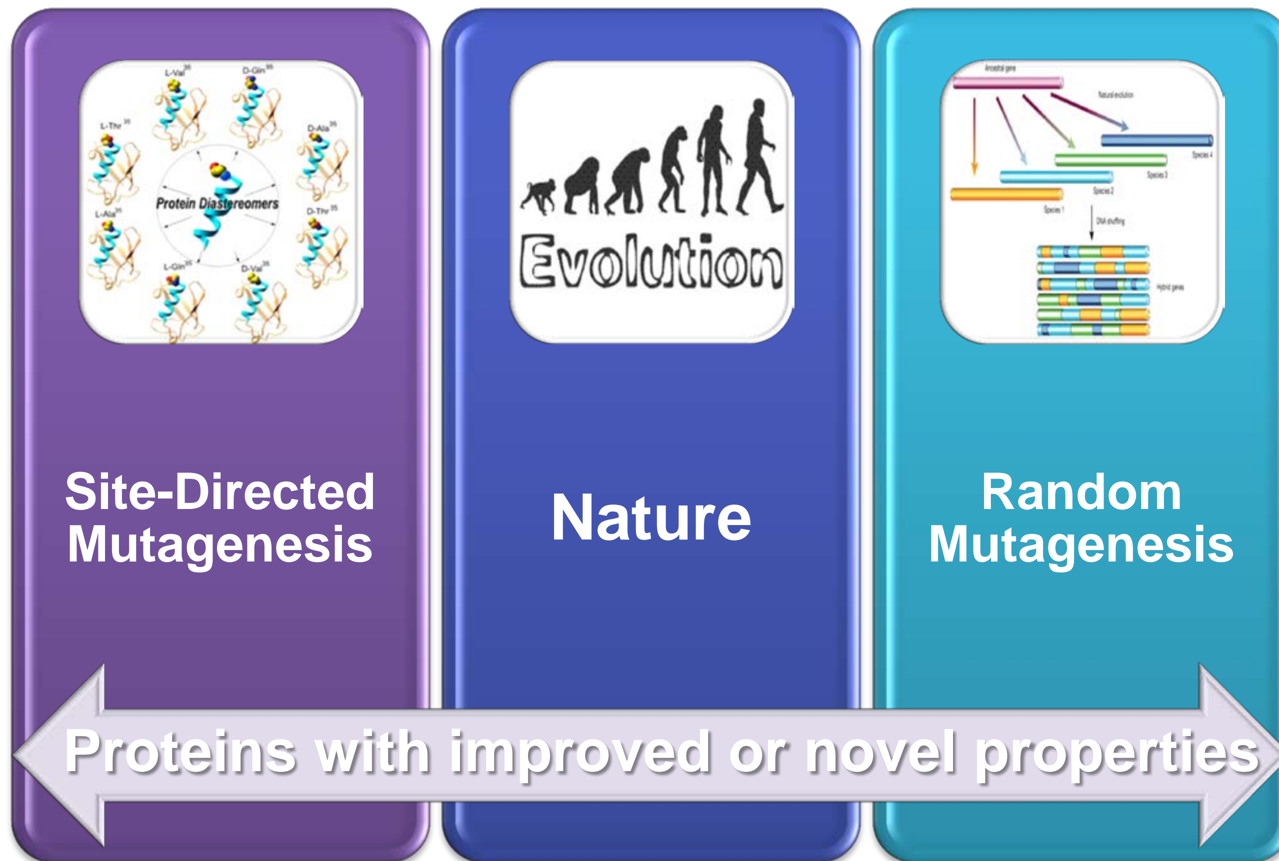
OUTLINE

- *PROTEIN ENGINEERING*
  - *Random Mutagenesis*
  - *Site-Directed mutagenesis*
  - *Mimicking peptides*

- *COMPUTATIONAL PROTEIN DESIGN*
  - *Binding energy calculations*
  - *Molecular docking*
  - *Molecular dynamics*
- *THE PSII D1 PROTEIN: A CASE STUDY*

PROTEIN ENGINEERING

Design of new proteins or enzymes with new or desiderable functions



THE GLAMOUR OF NATURE

The fascinating variety and complexity of life is a consequence of natural selection processes.

Natural selection process relies on mutations and interactions with environment

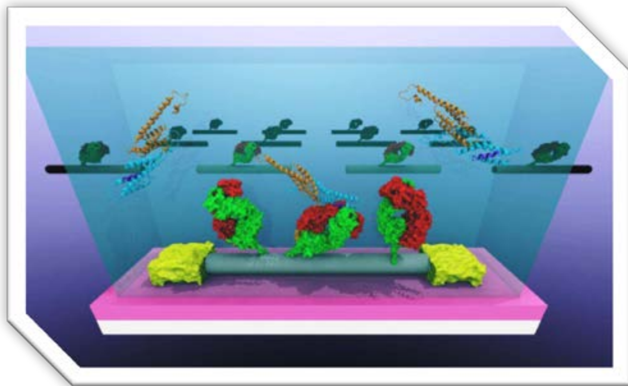
Proteins are the molecular machines mediating interaction of life with environment.

Protein engineering by mutations attempts to mimic nature’s recombination strategy

Differences between Natural & *In vitro* Evolution

Natural evolution is a gradual accumulation of changes based on environmental factors conferring successful traits to organisms possessing them.

In vitro evolution is a "guided" process towards a final goal that may or may not make biological sense



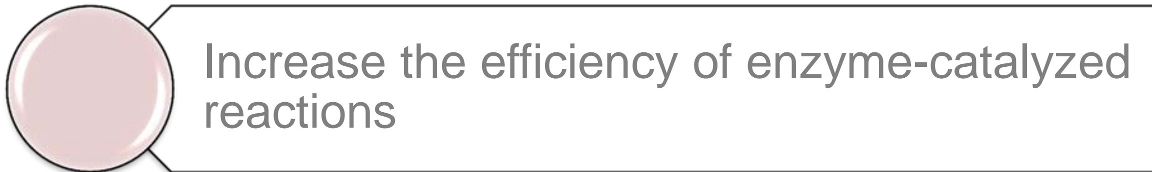
Combining Nanotube Technology and Genetically Engineered Antibodies to Detect Prostate Cancer Biomarkers. ACS nano 2012, 6(6):5143-5149.

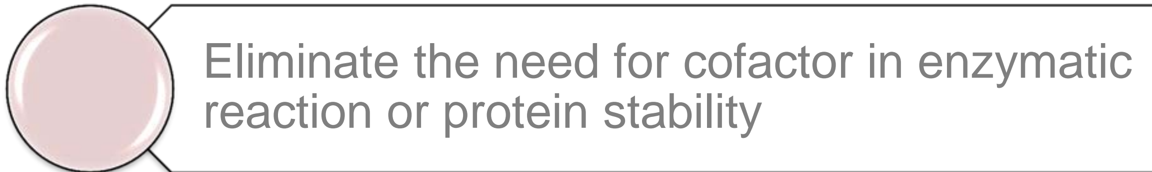
Mutagenesis: Why Mutate?

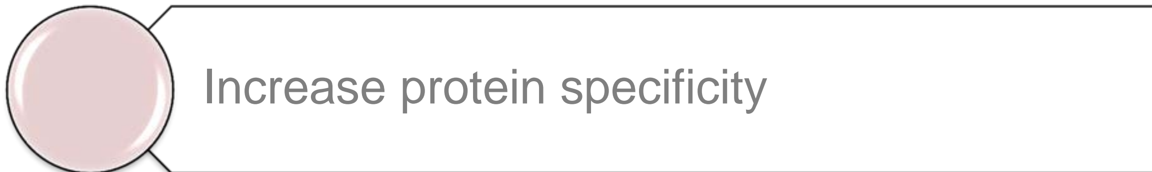
Native proteins are not well suited for biotechnological applications

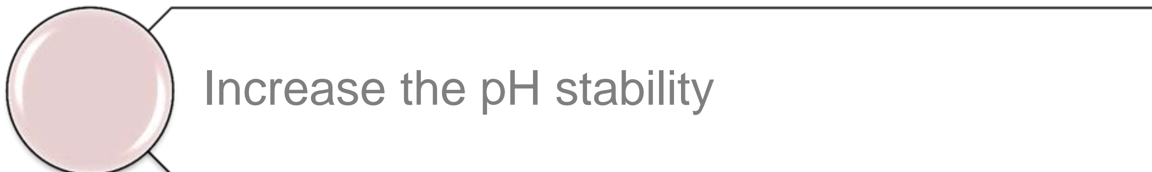
Although a variety of proteins and enzymes are now used in biotechnology and industry many of them have limited use because they are **denatured** on exposure to conditions which are encountered in industrial processes e.g. **high temperature, high pH, organic solvents and chemical solvents**.

Major Biotechnological Challenges









Biotechnological targets

Function

- **Binding**

  Interaction of a protein with its surroundings

  Improving/reducing binding affinity to specific substrates, or binding capabilities to additional substrates

  How many points are required to bind a molecule with high affinity?

- **Catalysis**

  a different form of binding – binding the transition state of a chemical reaction

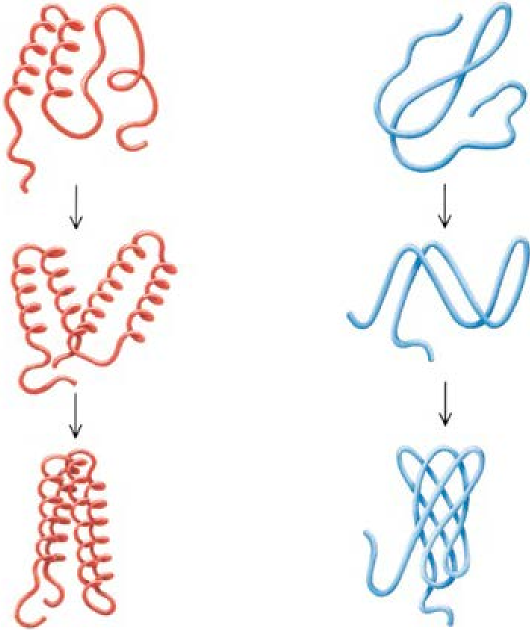
  Increased/decresed binding to the transition state ⇒ increased catalytic rates

  Requires: Knowledge of the Catalytic Mechanism → engineer Kcat and Km
  ***Michaelis constant or Km** is the **tightness of the substrate binding** to the enzyme.*
  (increases the specificity of the reaction and reduces side reactions).
  *The **Vmax** is the maximal **rate of conversion of the substrate** to the products.*
  (an increase in Vmax increase the amount of product produced).

Biotechnological targets

Folding/Structure

➢ Thermodynamic Stability



- Protein stability is the net balance of forces, which determine whether a protein will be in its native folded conformation or a denatured state.
- Protein stability normally refers to the physical (thermodynamic) stability, not the chemical stability.
- The net stability of a protein is defined as the difference in free energy between the native and denatured state
- Both G_N and G_U contribute to G

➢ Thermal and Environmental Stability

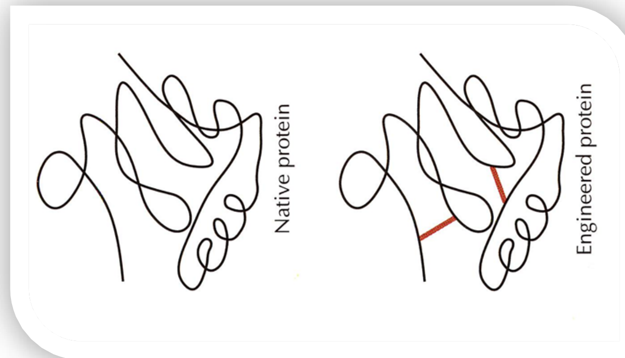
Temperature, pH, Solvent, Detergents, Salt …..

*An increase in **pH or thermal stability** may allow the protein to be used under conditions where it would normally be **denatured**.*

Protein Engineering Targets

Factors contributing to stability

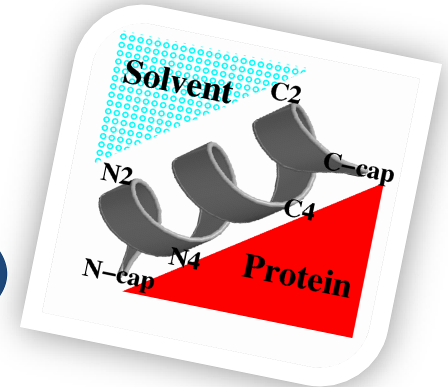
- **Hydrophobicity**
  hydrophobic core
- **Electrostatic Interactions**
  Salt Bridges
  Hydrogen Bonds
  Dipole Interactions
- **Disulfide Bridges**



- **Metal Binding** (Metal chelating site)

Protein Engineering Targets

Design of Thermal and Environmental stability



- Stabilization of α-Helix Macrodipoles
- Engineer Structural Motifs (like Helix N-Caps)
- Introduction of salt bridges
- Introduction of residues with higher intrinsic properties for their conformational state
  e.g. Ala replacement within a α-Helix
- Introduction of disulfide bridges
- Reduction of the unfolded state entropy with
  X → Pro mutations

Protein Engineering Targets

Cofactor Requirement

- The abolishment of the need for a **cofactor** may be beneficial where under certain industrial conditions a cofactor has to be constantly provided.

Specificity

- Increase **specificity** of the enzyme decreases undesirable side reactions.

Protein engineering: How does it work ?

- **Protein engineering** involves the use of genetic manipulations to **alter the coding sequence of a (cloned) gene** and thus **modify the properties of the protein** encoded by that gene.
- This **mutant gene** maybe expressed in a suitable system to produce unlimited quantities of the **modified protein**.

ASSUMPTION: Natural sequence can be modified to improve a specific protein function

IMPLICATION: Protein is NOT optimized for that function

Why Modify the Gene? Why not Modify the Protein?

- If the gene is modified by site directed mutagenesis then **each time** the host organism will produce the **modified protein**.
- However if the protein is modified each time the protein is **produced it has to be modified**.
- Further more **chemical modification** of protein is:

  Harsh
  Nonspecific
  Has to be repeatedly done

Protein engineering: Main guidelines

- Sequence changes should not disrupt the structure to avoid protein refold o misfolding
- New sequence should not be TOO different from the native sequence to avoid loss of function)

- **USUALLY:** Point mutations are good starting points

Protein engineering: Main approaches

How do we force something to change in the way we want ?

- ## RANDOM MUTAGENESIS

Creation of random libraries by applied molecular evolution. A target gene can be randomised exploiting one of the existing different methods), following by ligation of the library into a proper vector backbone and transformation into an appropriate host for selection and screening.

- ## SITE-DIRECTED MUTAGENESIS

Creation of single point mutations in a particular known area, obtaining only 2 species: wild-type and mutated DNA (site-specific). Ligation into a specific vector and transformation into an appropriate host for selection and screening.

Random mutagenesis: Error Prone PCR

Can create enzyme variants on scale of months/weeks/days by rounds of mutagenesis and screening

- Some heat stable DNA polymerases used during PCR can occasionally **insert the wrong nucleotide generating mutations** (Error Prone PCR).
- By **modifying PCR conditions** e.g.

  DNA template concentration

  Adding unequal concentration of each nucleotides

  Add Mn^{2+} instead of Mg→5-fold excess of dTTP and dCTP
- It is possible to introduce mutations into the PCR product.
- This product is then cloned and the modified protein expressed and tested for the desired properties

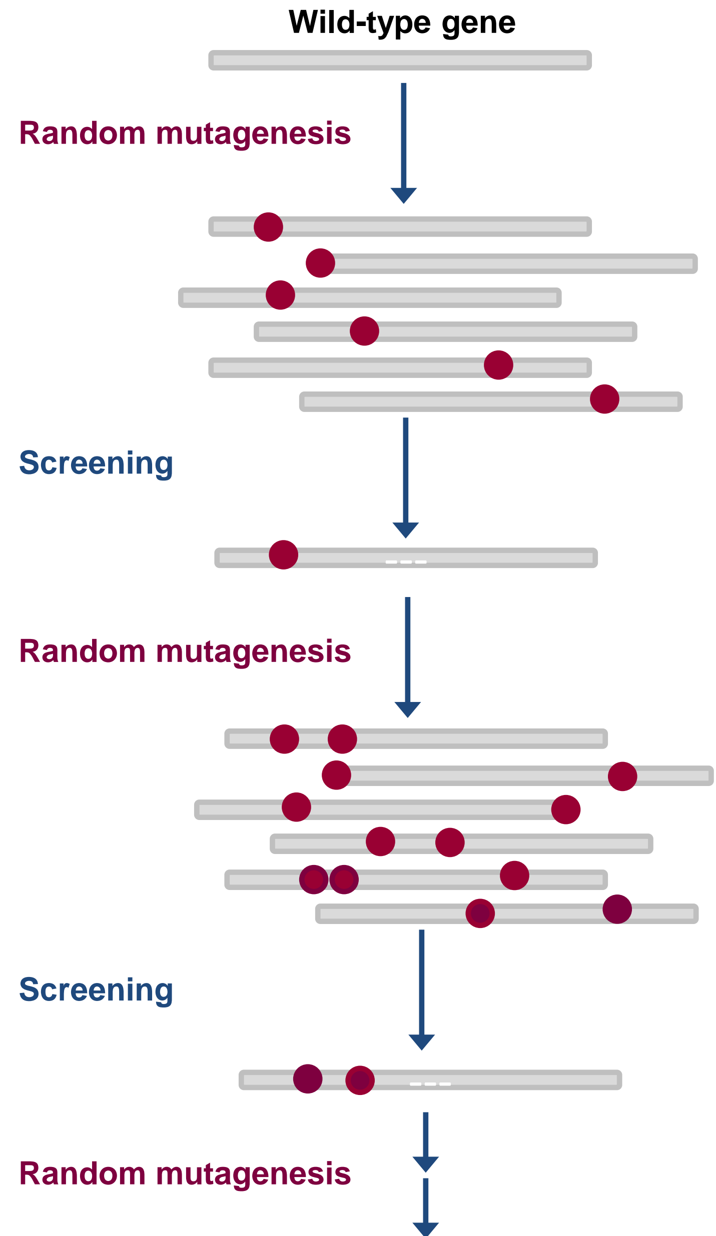
Random Mutagenesis

Error-prone PCR: method of choice if starting from single protein sequence

Mutation rate is 1/2 mutations per protein so all variants can be exhaustively evaluated - more mutations would create combinatorial challenges

Many created enzymes will be non/dysfunctional, evaluated through large screening libraries

Promising/improved variants subsequently subjected to additional rounds of mutagenesis



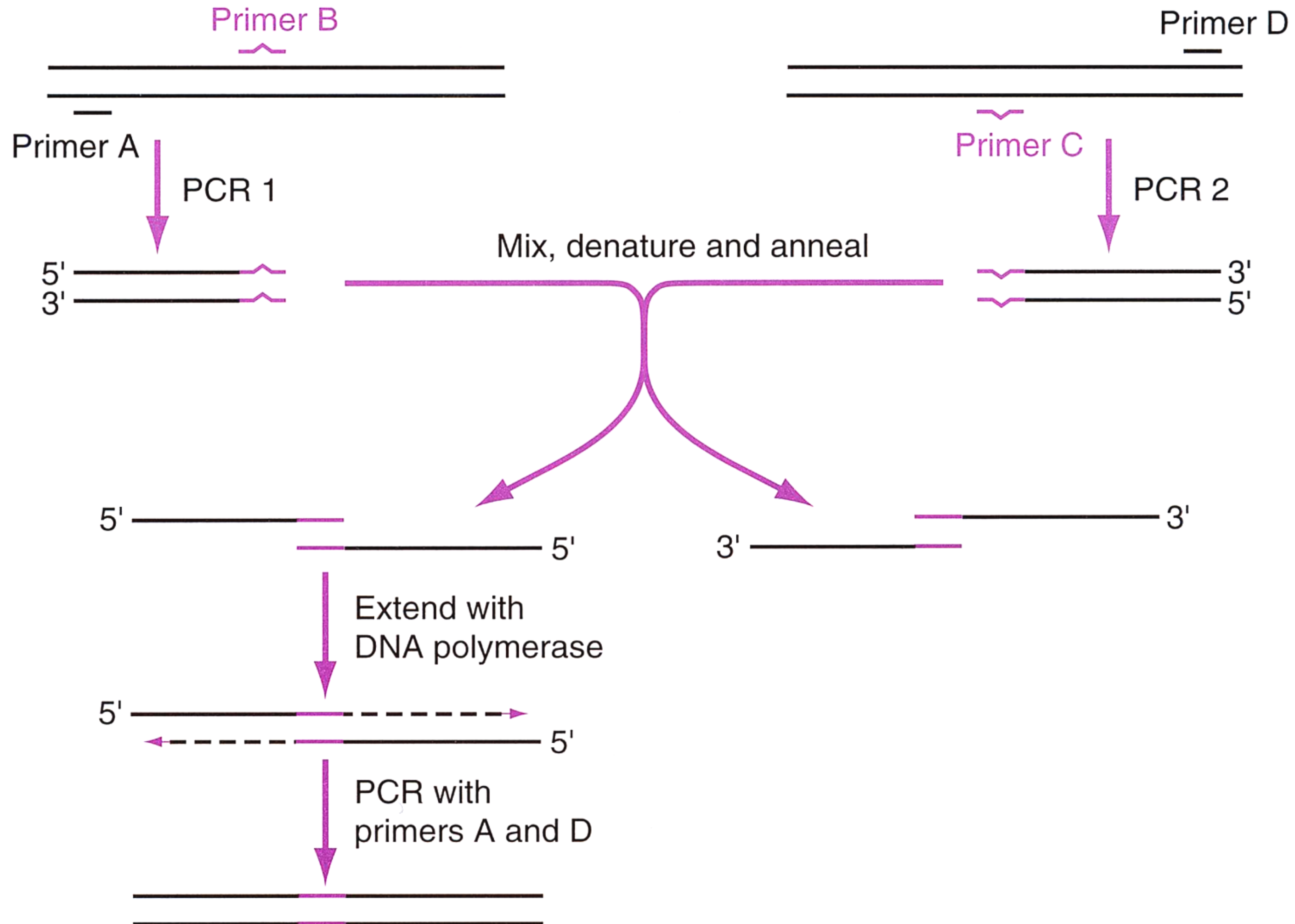
RATIONAL PROTEIN DESIGN

Site – directed mutagenesis

Requirements:

- Knowledge of sequence and preferable Structure (active site,….)
- Understanding of mechanism (knowledge about structure – function relationship)
- Identification of cofactors……..

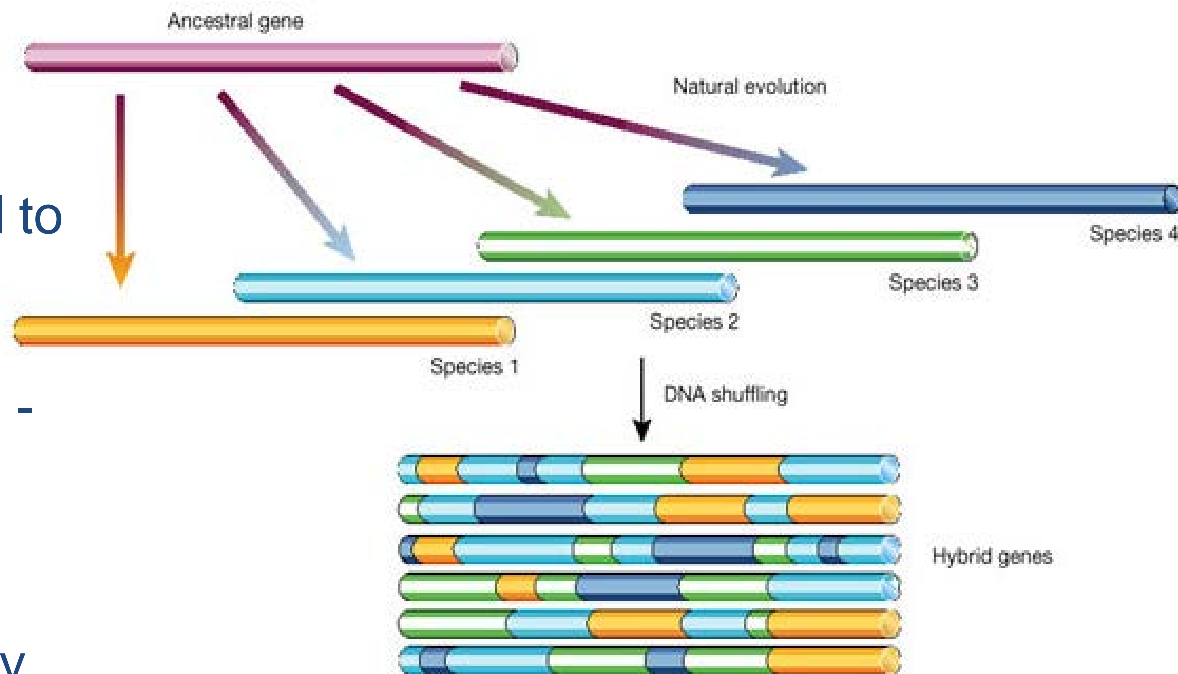
Site-directed mutagenesis PCR based–methods



GENE SHUFFLING

Homologous recombination used to create chimeric sequences containing multiple beneficial mutations

- **Family shuffling** of homologous genes
- **Synthetic shuffling** - oligonucleotides combined to create full lenght genes
- **Whole-genome shuffling** - accelerated phenotypic improvements
- **Drawback** - high homology required



Nature 409, 253-257 (January 2001)

Library of "chimeric genes" created that should fold in the same way as their precursors, but now there's variation present

Results of Mutagenesis

- Can successfully improve stability or activity of an enzyme - many specific solutions exist and mutations in iterative rounds are very additive.

- Drawback - genetic code is conservative, many similar codons code for same amino acid or another amino acid with same properties.

COMPUTATIONAL PROTEIN DESIGN

Binding energy calculations

Molecular dynamics

Molecular docking

Binding energy calculations

Molecular recognition is a central process in biology

- Any "interesting" phenomenon in biology requires recognition and binding between (macro)molecules
  - Protein-protein recognition and signaling
  - Enzyme-inhibition
  - Drugs action

A realistic model of molecular recognition processes has a big predictive and applicative impact

- Enzyme biotechnology
  - Improving enzyme-substrate recognition and catalysis
- Drug design
  - Design of novel ligands and prediction of binding energy

Binding energy calculations allow

- Structural and energetic determinants binding affinity
  - Which is the main driving force for substrate/substrate-analogs binding?
- Structural and energetic determinants of specificity
  - Which are the main interactions that determine specificity?

Binding free energy decomposition

- A physically realistic model allows to decompose and study the various energetic contributions to binding free energy

The "Honig" method (*Froloff et al. (1997) Protein Sci. 6, 1293)*

$$\Delta G_b = \Delta G_{el} + \Delta G_{np} + \Delta G_{strain} - T\Delta S_{sc} - T\Delta S_l - T\Delta S_{t,r}$$

$$\Delta G_{el} = \Delta G_{coul} + \Delta G_{sol}$$

$$\Delta G_{np} = \gamma_{aw} A_c \qquad \gamma_{aw} = 58.18 \text{ cal/Å}^2$$

$$-T\Delta S_{sc} ; T\Delta S_l = \Sigma RT \ln (N_{free}/N_{bound})$$

$$-T\Delta S_{t,r} = 7\text{-}10 \text{ kcal/mol}$$

ΔG_b, theoretical binding free energy
ΔG_{coul}, Coulomb contribution to binding
ΔG_{sol} reaction field (solvation) contribution to binding
ΔG_{np}, nonpolar (hydrophobic) contribution to binding
ΔG_{strain}, change in conformational free energy of both the receptor and the ligand upon binding
ΔS_{sc} loss of configurational entropy due to the side-chain torsional angles upon binding
$T\Delta S_{t,r}$, loss of translational and rotational degrees of freedom upon binding
γ_{aw} , microscopic surface tension associated with the transfer of alkane from liquid alkane to water

Molecular docking

Finding binding site and orientation of a molecular complex

- **Protein-protein complexes**
  - Usually rigid-body docking with a simplified representation of the macromolecules (backbone). Search for the relative orientation of the two molecules with lowest energy
  - Geometric (surface complementarity) scoring
  - Ranking of the complexes to find best solution

- **Protein-ligand complexes**
  - Flexible docking (at least for the ligand) with a detailed representation of the macromolecule (sidechains included)
  - Geometric (surface complementarity) scoring
  - Energetic scoring
  - Ranking

Molecular dynamics simulations

Use of a "force field" on each atom of the simulated system to simulate the time evolution of a macromolecular system

Covalent (bonds, angles, and dihedrals), van der Waals and electrostatic interactions energy used to evaluate forces acting on atoms

Classical mechanics equations used to calculate velocities and position of atoms from masses and forces

(Force = mass *times* acceleration)

Long simulation times not easily attainable on large macromolecular systems

- **One state-of-the-art CPU day to simulate 100 ps of a 500 amino acids protein in explicit water**

Computational studies of Photosystem II. Functional insights and biotechnological applications

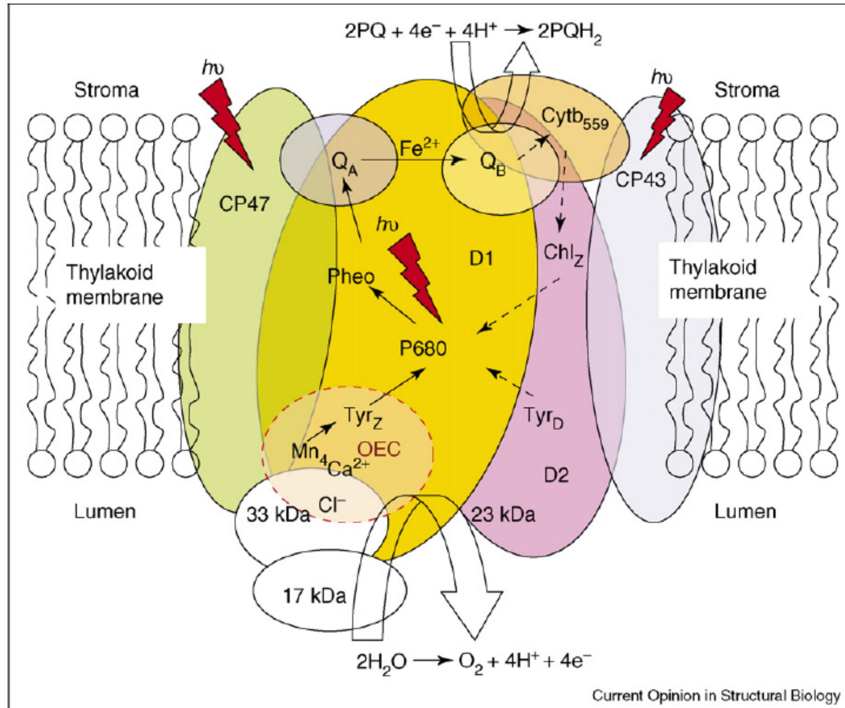
OUTLINE

- Design of PSII mutants for improved herbicide detection

- Design of PSII QB binding pocket mimics

- Docking simulations for virtual screening

Photosystem II (PSII)



Curr Opin Struct Biol. 2007 Apr;17(2):173-80. Epub 2007 Mar 28. Quantum mechanics/molecular mechanics structural models of the oxygen-evolving complex of photosystem II.
Sproviero EM, Gascón JA, McEvoy JP, Brudvig GW, Batista VS.

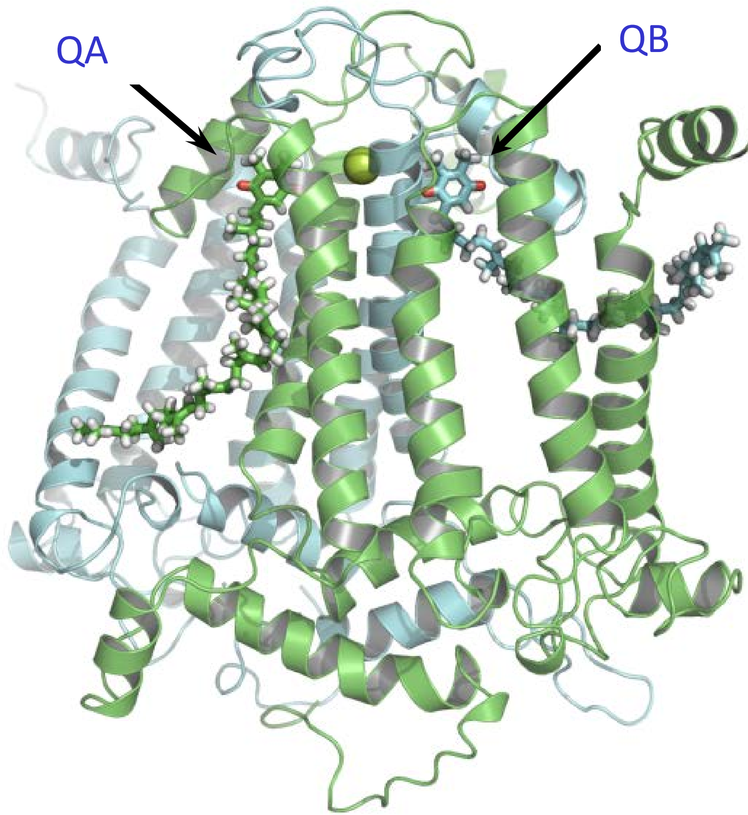
350 kDa protein-cofactors macromolecular complex located in the thylacoid membranes of the oxygenic photosynthetic organisms

Catalyzes the light-induced production of reducing equivalents in the form of plastoquinol molecules

The reaction byproduct is molecular oxygen and thus PSII is essential for higher organisms life on Earth.

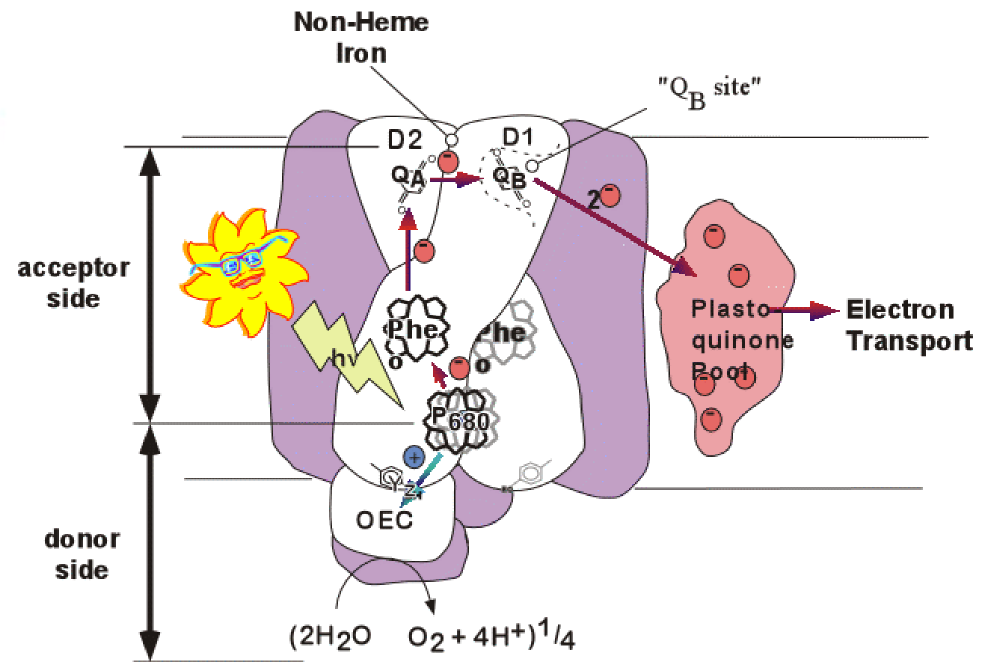
Plastoquinone reduction leads to its release from PSII with a still unclear molecular mechanism

PSII reaction center



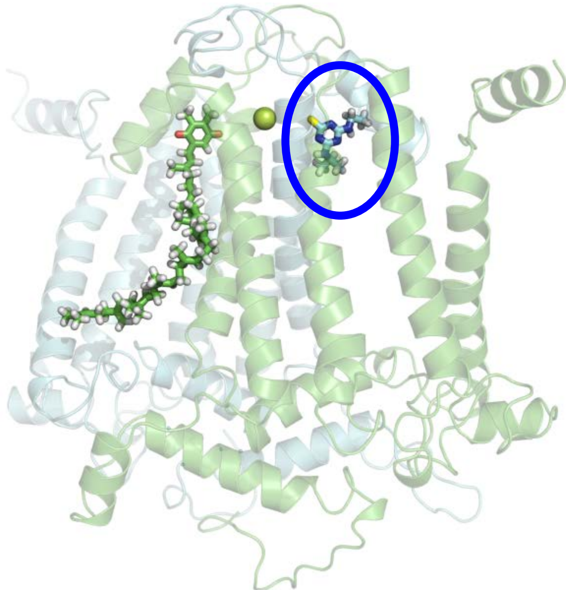
PSII core is made up by the D1 e D2 proteins (the "reaction center") which contain the plastoquinones QB and QA binding pockets and a non heme iron.

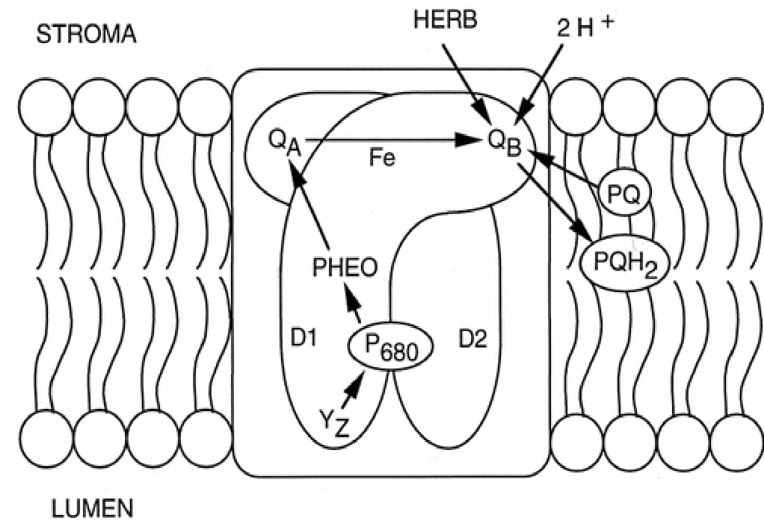
QA is stably bound to D2 while QB, bound to D1, once reduced via the iron atom is released, thus increasing the membrane plastoquinol pool.

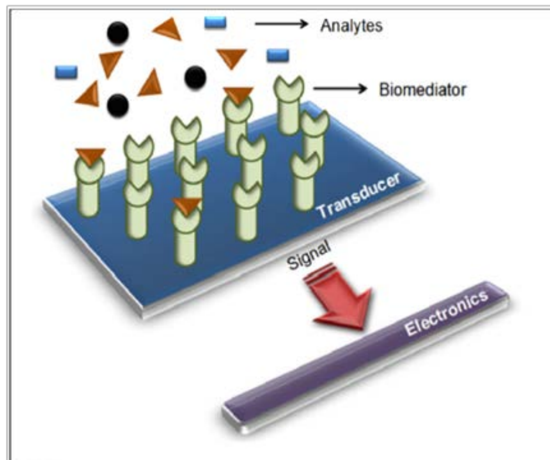


Design of PSII mutants for improved herbicide detection

PSII as a biomediator for the development of herbicides biosensors





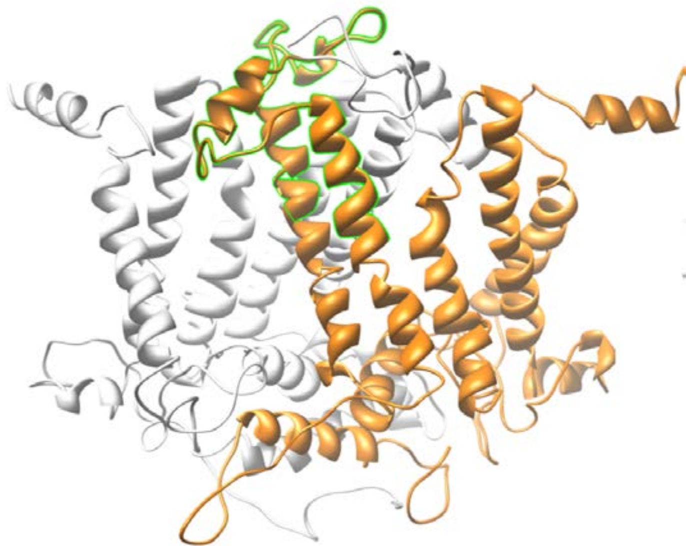


Binding of herbicides such as triazines to the QB site interrupts the electron flow causing oxidative damage and leading to cell death.

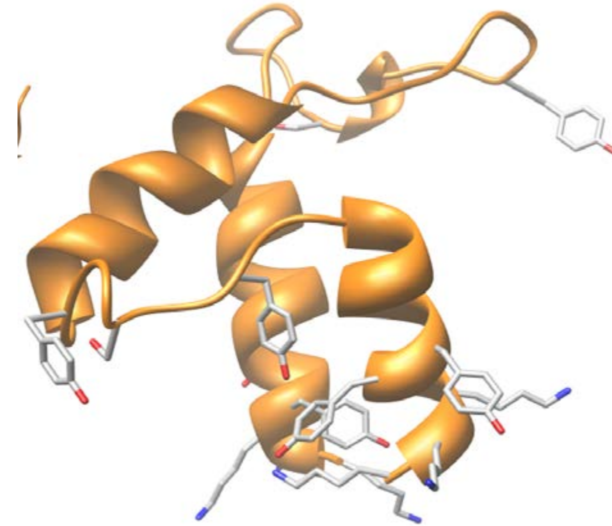
By immobilizing thylakoid membranes on electrodes it is possible to develop biosensors for the measurement of herbicides levels in water and ground.

Design of PSII QB binding pocket mimics

By computational modeling and automated protein synthesis, the D1 plastoquinone/atrazine binding niche in native and mutated forms was reconstituted and the structural and functional features analyzed in detail by circular dichroism, fluorescence spectroscopy and microcalorimetry.



D1/D2 protein



D1 biomimetic peptide

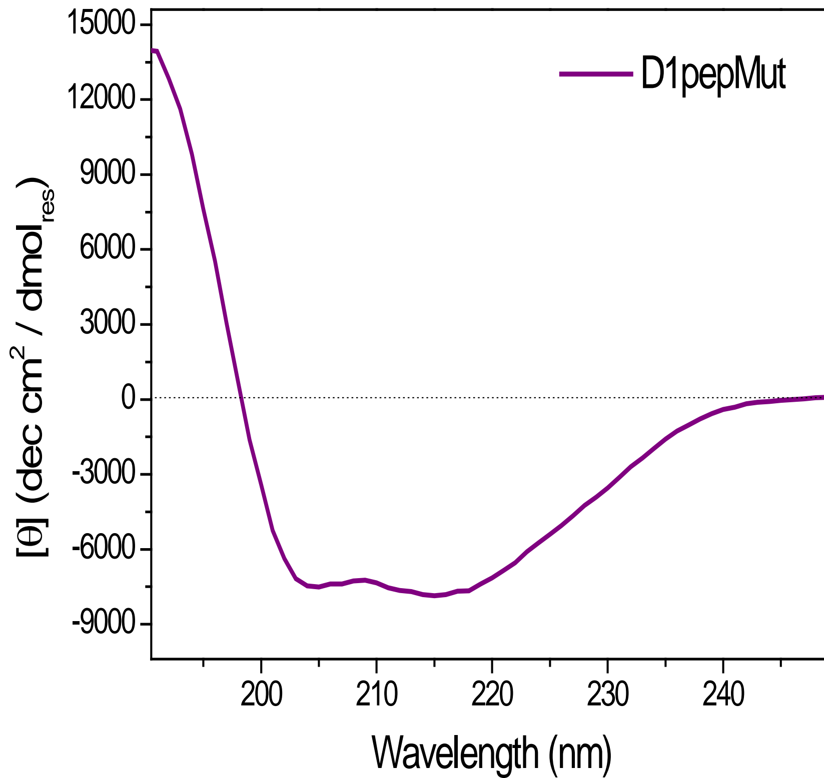
Grafting of the QB pocket: D1 residues 211-280

Wild-type and mutant D1 peptide sequences

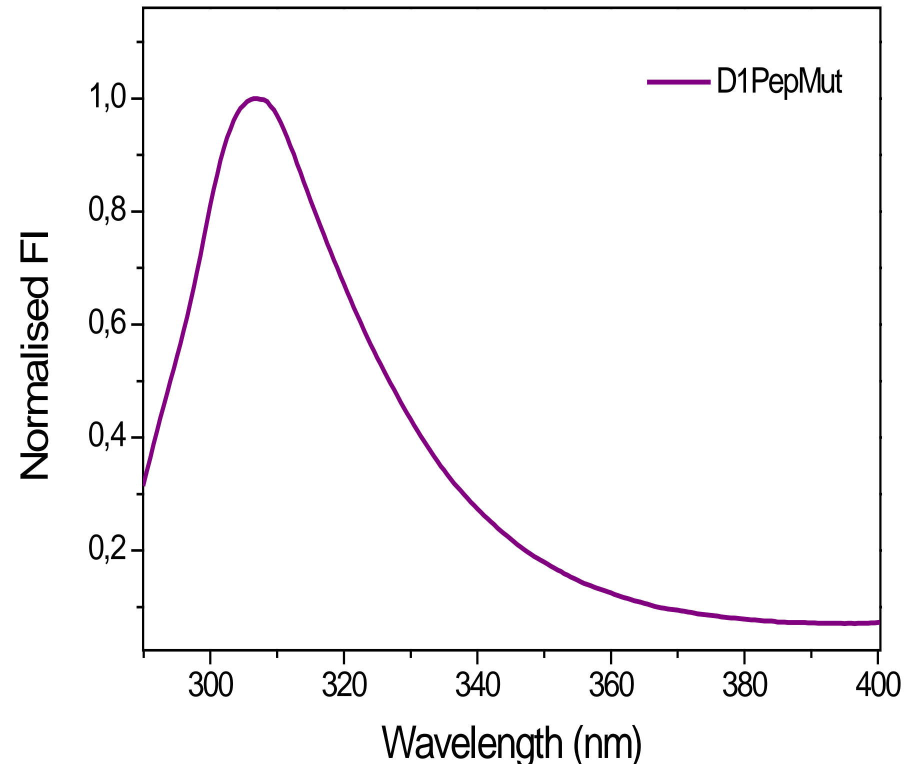
FSAMHGSLVT SSLIRETTEN ESANEGYRFG QEEETYNIVA AHGYFGRLIF QYASFNNSRS LHFFLAAWPV

KKK YSSMHGSLVT SSLIRETTEN ESSNEGYRYG QEEETYNIVS AHGYFGRLIY QYSSYNNSRS LHYYLA KKK

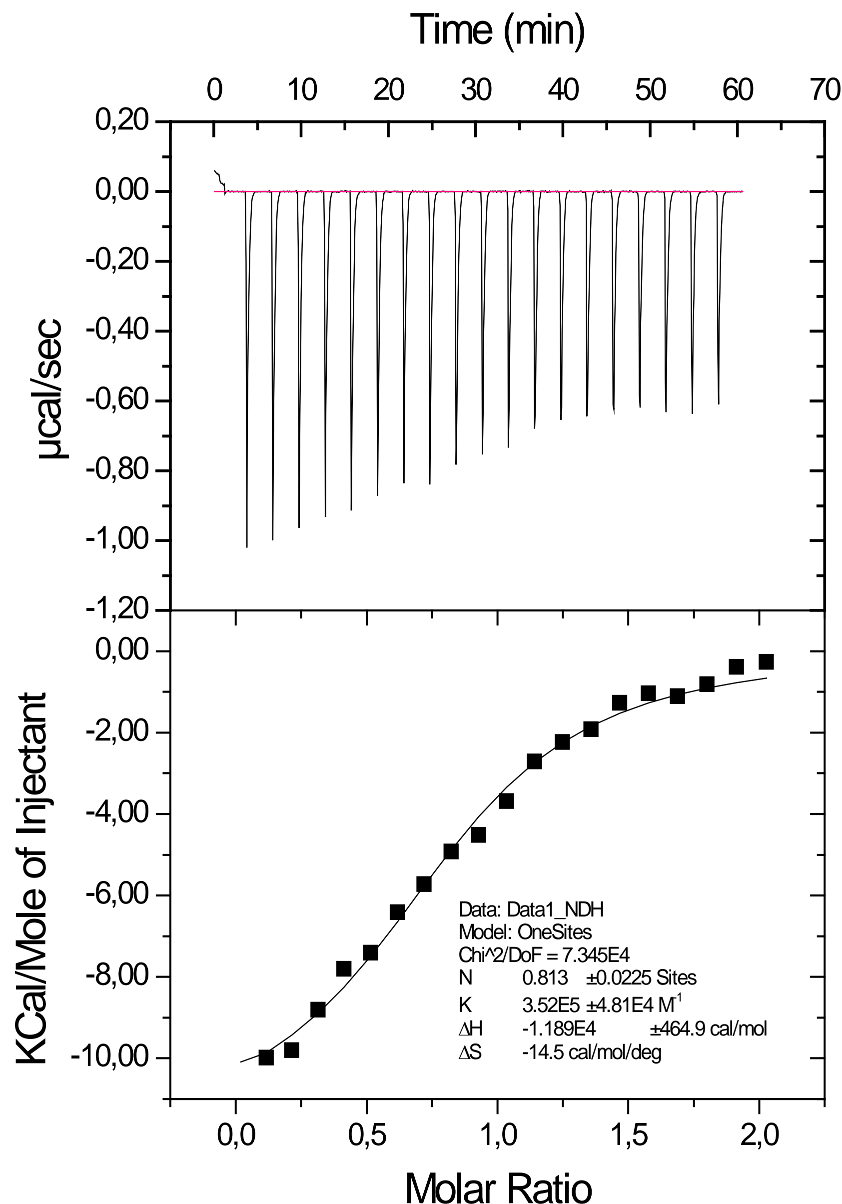
Structural characterisation by circular dichroism and fluorescence spectroscopy



Far-UV CD spectra of D1pepMut peptide indicated that the peptide has two markedly pronounced negative CD bands between 200 and 220 nm, typical of properly-folded α-helical proteins.



Fluorescence emission spectra are in agreement with CD data. The intrinsic emission profile is characteristic of a peptide lacking Trp residues but presenting 10 tyr residues.



MICROCALORIMETRY

Calorimetric data of D1pepMut with atrazine (raw and integration data)

Kd = 2.84 μM
(Atrazine binds to the peptide with an affinity constant of 3.52 x 10^5 M^{-1}).

n = 0.86 corresponding to a 1:1 binding mode.

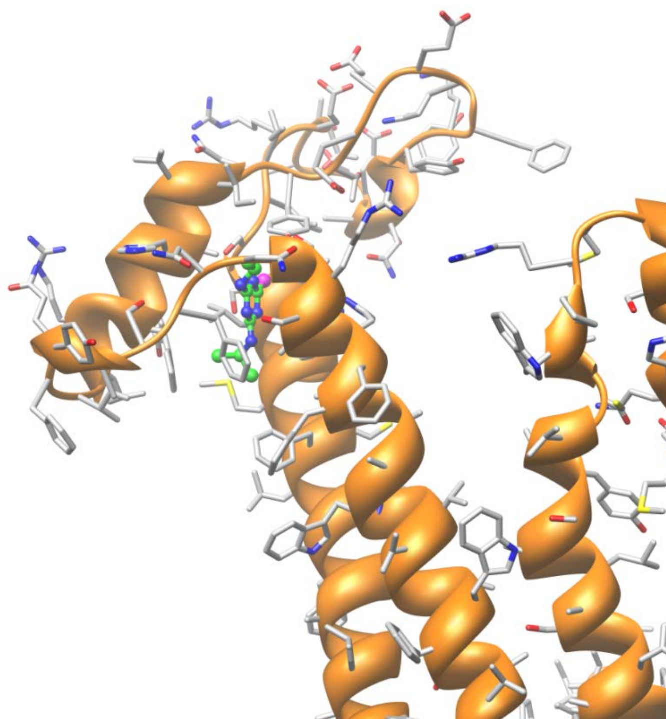
-TΔS = 4.3 kcal/mol
ΔH = -11.9 kcal/mol

negative free energy value of -7.6 kcal/mol

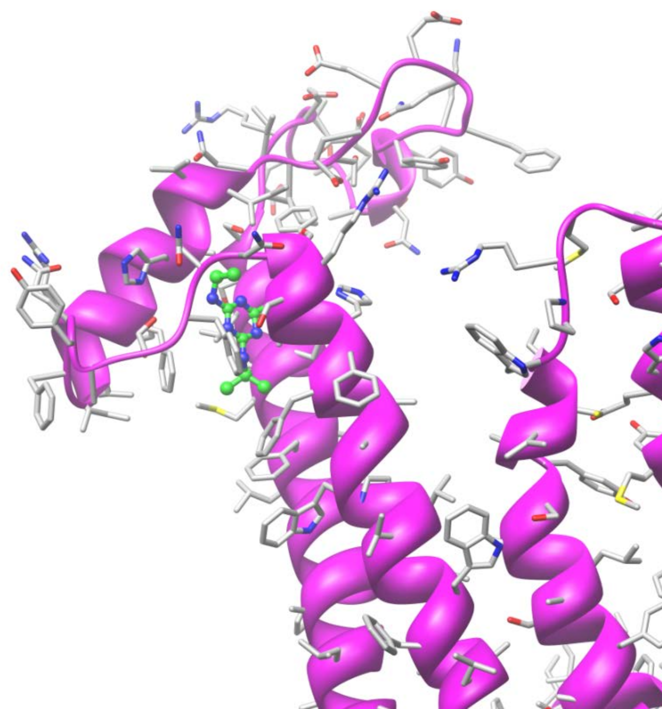
Despite a very unfavorable entropic term, the favorable enthalpic contribution suggests an enthalpy driven reaction with productive interactions between peptide and atrazine.

Docking simulations for virtual screening

ATRAZINE – MD SIMULATIONS *VS* DOCKING



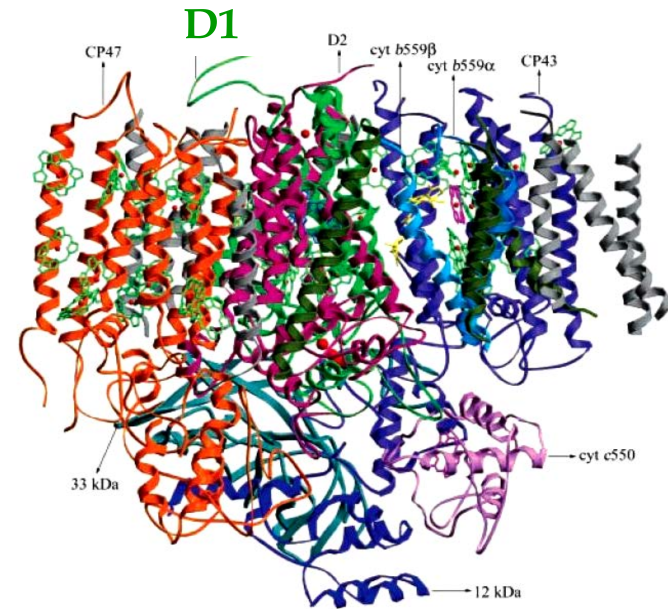
10 ns MD simulations

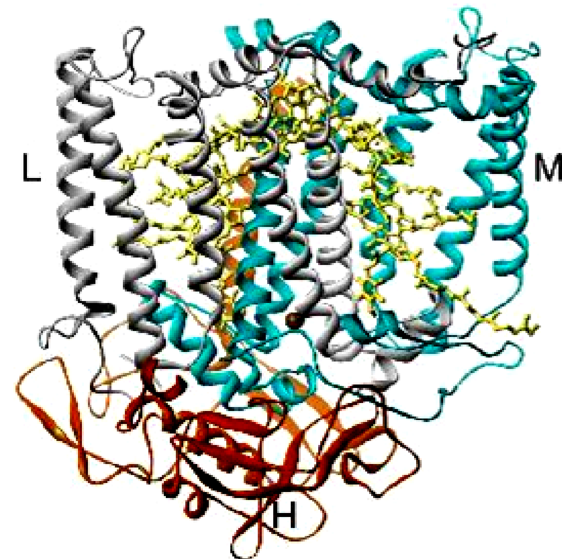


Docking simulations

- Very good agreement but docking is much faster
- Possibility to screen a significant number of compounds

Rational Chimera Design





Genetic Engineering of Reaction Centers: A Structure-Based Approach

Work Hypotheses:

Production of ***C. reinhardtii* mutant** strains carrying *puf*L gene (L protein) of *R. sphaeroides* in substitution of the *psb*A gene (D1 protein).

Production of ***R. sphaeroides* mutant** strains carrying *psb*A gene (D1 protein) of *C. reinhardtii* in substitution of *puf*L gene (L protein).

Production of ***R. sphaeroides* mutant** strains carrying a chimeric form of L protein encompassing the *C. reinhardtii* Q_B binding site.

Hybrid PSIIs

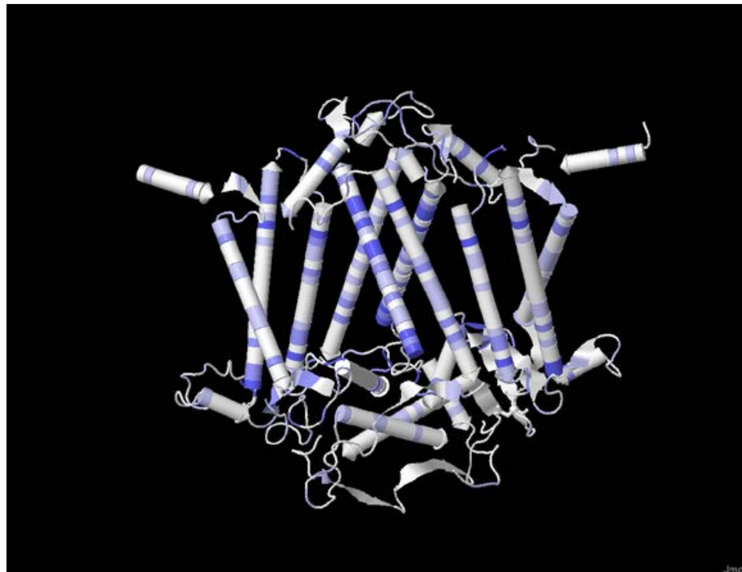
- Analysis of the homology between *Termosynechococcus elongatus* D1/D2 proteins and *Rhodobacter spheroides* L/M proteins

- Molecular modelling of hybrid PSII

- Analysis of the structural compatibility and success probability of hybrid forms production

In detail…….

```
T._elongatus_D1   1 MTTTLQRRESANLWERFCNWVTSTDNRLYVGWFGVIMIPTLLAATICFVIAFIAAPPVDIDGIRE 65
R._spheroides_L   1 ------------------------ALLSFERKYRVPGGTLVGGNLFDFWVGPFYVGFFGVATFFF 41

T._elongatus_D1  66 PVSGSLLYGNNIITGAVVPSSNAIGLHFYPIWEAAS------LDEWLYNGGPYQLIIFHFLLGAS 124
R._spheroides_L  42 AALGIILIAWSAVLQGTWNPQLISVYPPALEYGLGG--------APLAKGGLWQIITICATGAFV 98

T._elongatus_D1 125 CYMGRQWELSYRLGMRPWICVAYSAPLASAFAVFLIYPIGQGSFSDGMPLGISGTFNFMIVFQAE 189
R._spheroides_L  99 SWALREVEICRKLGIGYHIPFAFAFAILAYLTLVLFRPVMMGAWGYAFPYGIWTHLDWVSNTGYT 163

T._elongatus_D1 190 H-NILMHPFHQLGVAGVFGGALFCAMHGSLVTSSLIRETTETESANYGYKFGQEEETYNIVAAHG 253
R._spheroides_L 164 YGNFHYNPAHMIAISFFFTNALALALHGALVLSAANPEK-GKEMRTPDHEDT------------- 214

T._elongatus_D1 254 YFGRLIFQYASFNNSRSLHFFLAAWPVVGVWFTALGISTMAFNLNGFNFNHSVIDAKGN----VI 314
R._spheroides_L 215 FFRDLVGYSIGTLG---IHRLGLLLSLSAVFFSALCMIITGTIWFDQWVDWWQWWVKLP------ 270

T._elongatus_D1 315 NTWADIINRANLGMEVMHERNAHNFPLDLA                                    344
R._spheroides_L 271 --WWANIPGGING-----------------                                    281
```



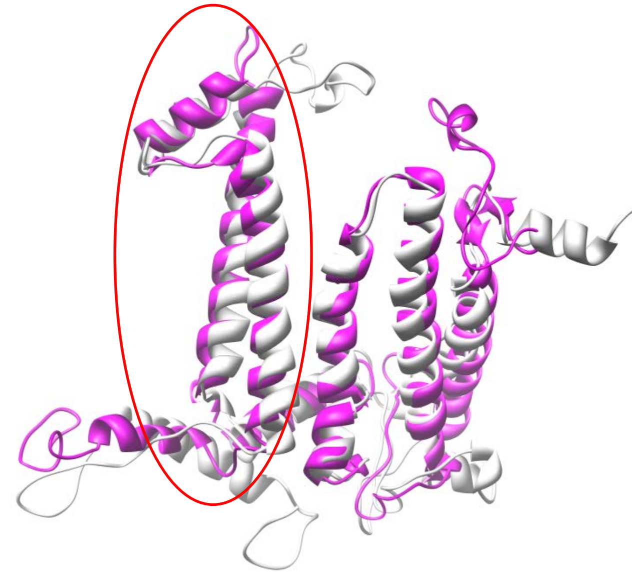
- **D1 displays longer N- and C-terminal regions**
- **Only 46 identical residues over 344**
- **Few conserved residues in QB pocket (regions 200-220, 250-270, 270-290) and interfacial regions.**

Structure-based approach

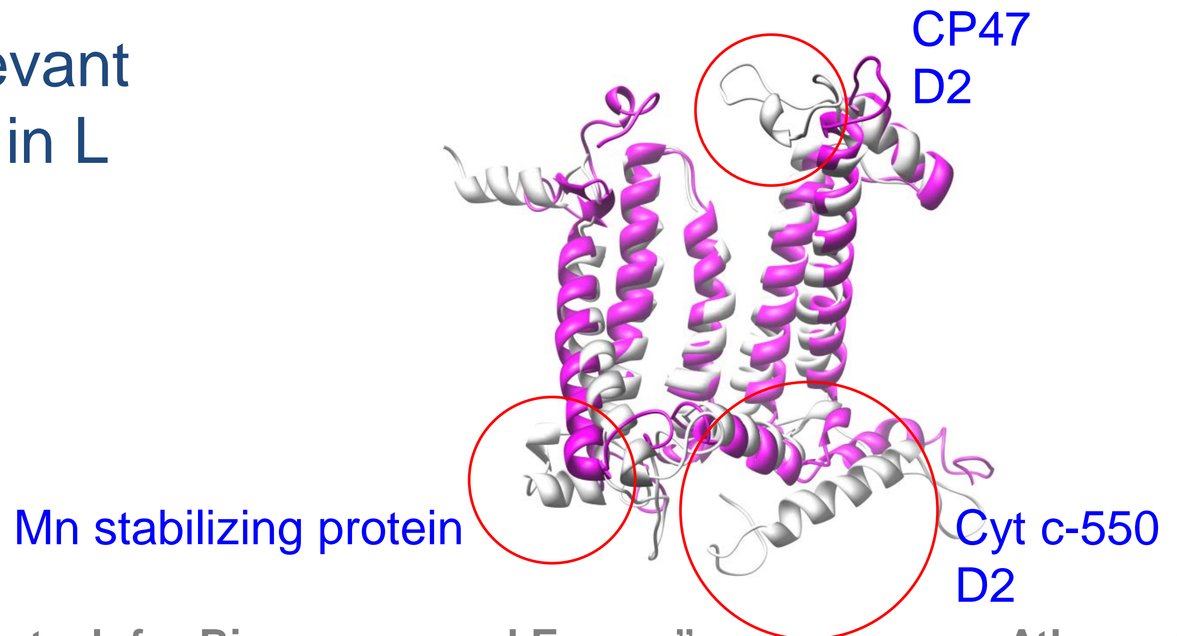
- Modeling hybrid L-D2 PSII core
- Analysis of stereochemical violations
- Analysis of interacting partners patterns

D1-L structural similarity

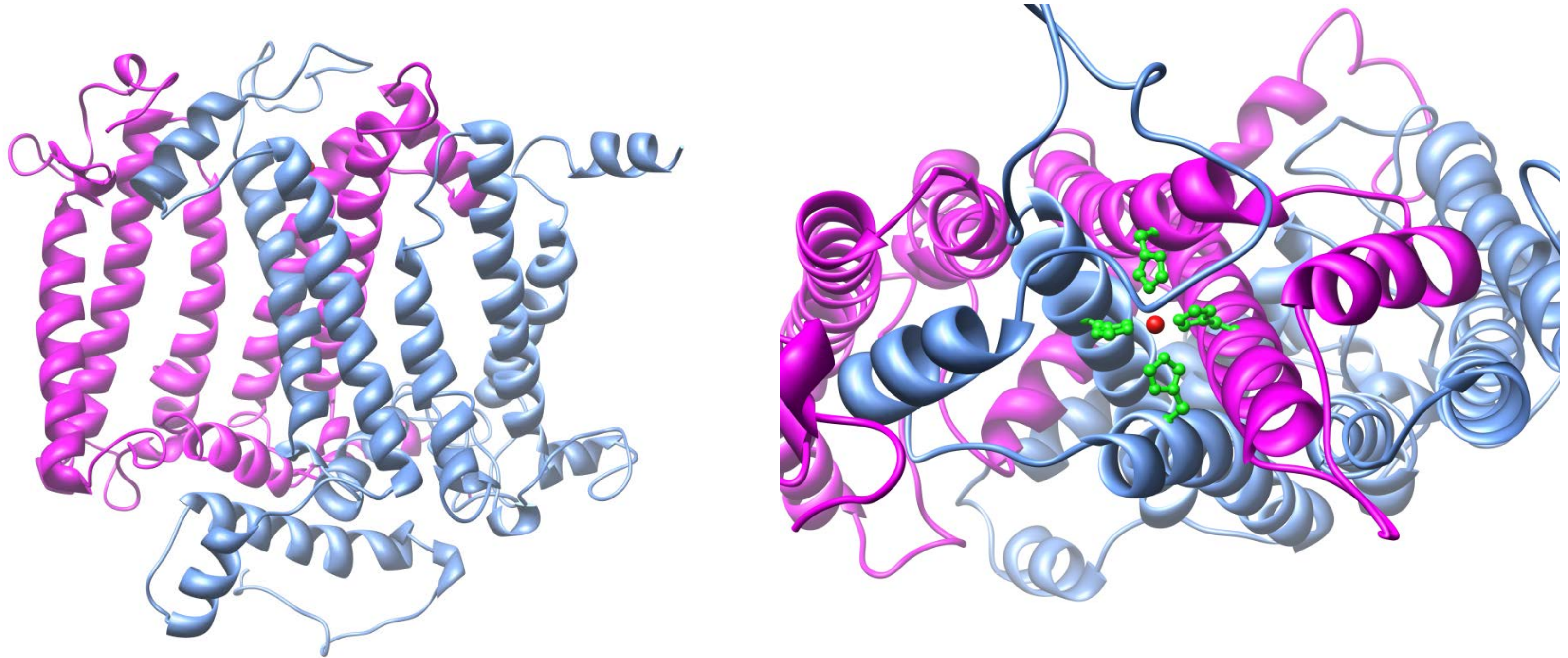
- Good fit of the QB binding niche
  - 2.03 Ǻ rmsd over 263 Cα atoms



- Functionally relevant regions missing in L



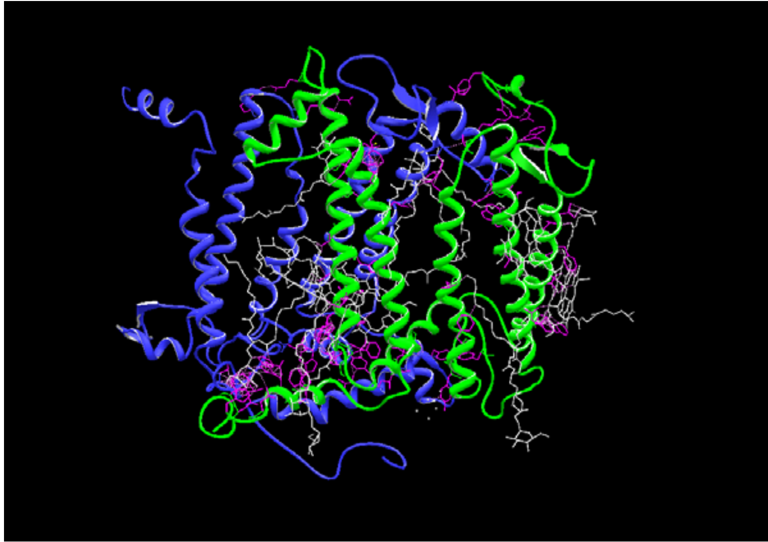
L-D2 molecular model

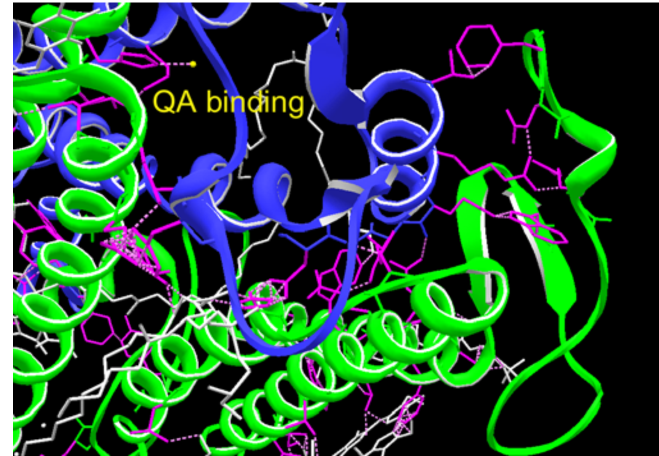


- Molecular modelling evidences general compatibility
- Feasibility of formation of a properly assembled iron site
- Feasibility of formation of properly connected Q_A and Q_B binding niches

HOWEVER……….

Several stereochemical violations





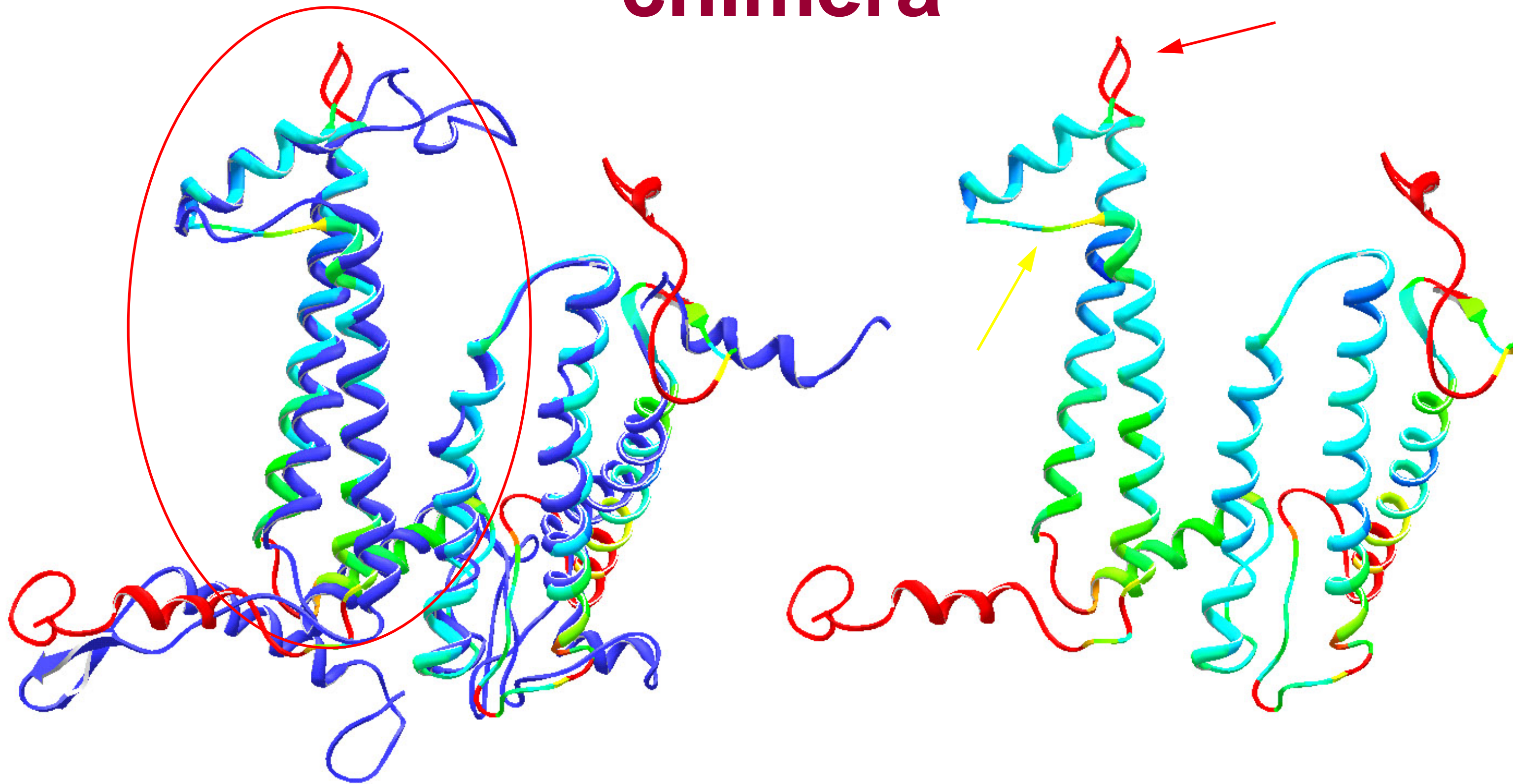
- Worse violations
  - the 250s helix, part of Q_A binding site
  - L-D2 interface nearby the Mn cluster
- Several violations with lipids and cofactors

CONCLUSIONS

- L-D2 hybrid
  - High probability of being unable to attain proper assembly and proper interaction with cofactors and macromolecular partners

  - Alternative strategies?
    - D1-L chimera by rational cut-and-paste
      - Faster but risky for possible packing defects
    - Site specific mutagenesis
      - More time-consuming but more likely to succeed

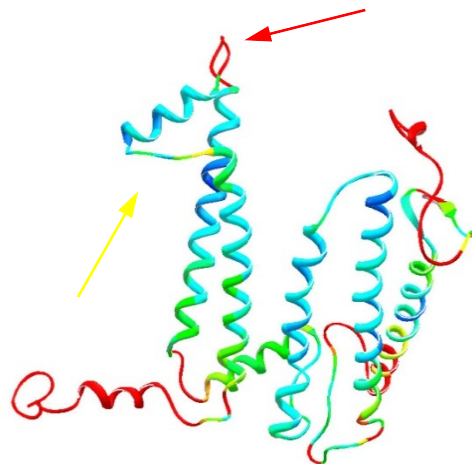
Feasibility of construction of a D1-L chimera



- Backbone fold highly compatible in the 190-291 region
  - Approx 25% identity
- Deviation in loop regions
  - warm colors indicate higher deviations

Feasibility of construction of a D1-L chimera

```
                                                                        !
D1 Chlamy 160   IYPIGQGSFS DGMPLGISGT FNFMI--VFQ AEHNILMHPF HQLGVAGVFG
L  Rhodoc 134   FRPVMMGAWG YAFPYGIWTH LDWVSNTGYT YGNFHYNPA- HMIAISFFFT
                . *.   *.  .   .  * **     .. .     .    .     .  *  ....   *

                  !   !!     !!                          ?     ?  ?!!??
D1 Chlamy 208   GALFCAMHGS LVTSSLIRET TETESANYGY KFGQEEETYN IVAAHGYFGR
L  Rhodoc 183   NALALALHGA LVLSAANPEK G--------- -----KEMRT PDHEDTFFRD
                 **   *.**.  ** *.    *                   *          .*

                !  !    !      !! ?!?   !
D1 Chlamy 258   LIFQYASFNN SRSL-HFFLA AWPVVGVWFT ALGISTMAFN LNG----FNF
L  Rhodoc 219   LVG--YSIG- -TLGIHRLGL LLSLSAVFFS ALCM-IITGT IWFDQWVDWW
                      *.         *  .      .. .* *.  **  .   ..     .
```



- Chimera made using blue and yellow regions of D1
- But....Several substitutions in the L-M contact surface!

Site-directed mutagenesis?

Thr	182	Gly	207
Leu	185	Leu	210
Ala	188	Ala	213
Leu	189	Met	214
Leu	193	Leu	218
Val	194	Val	219
Phe	215	Tyr	254
Phe	216	Phe	255
Leu	219	Leu	258
Val	220	Ile	259
Gly	221	Phe	260
Ile	224	Phe	265
Gly	228	Leu	271
Ile	229	deletion	
Leu	232	Phe	274
Leu	236	Trp	278

- Only three critical positions outside the 220-230 loop

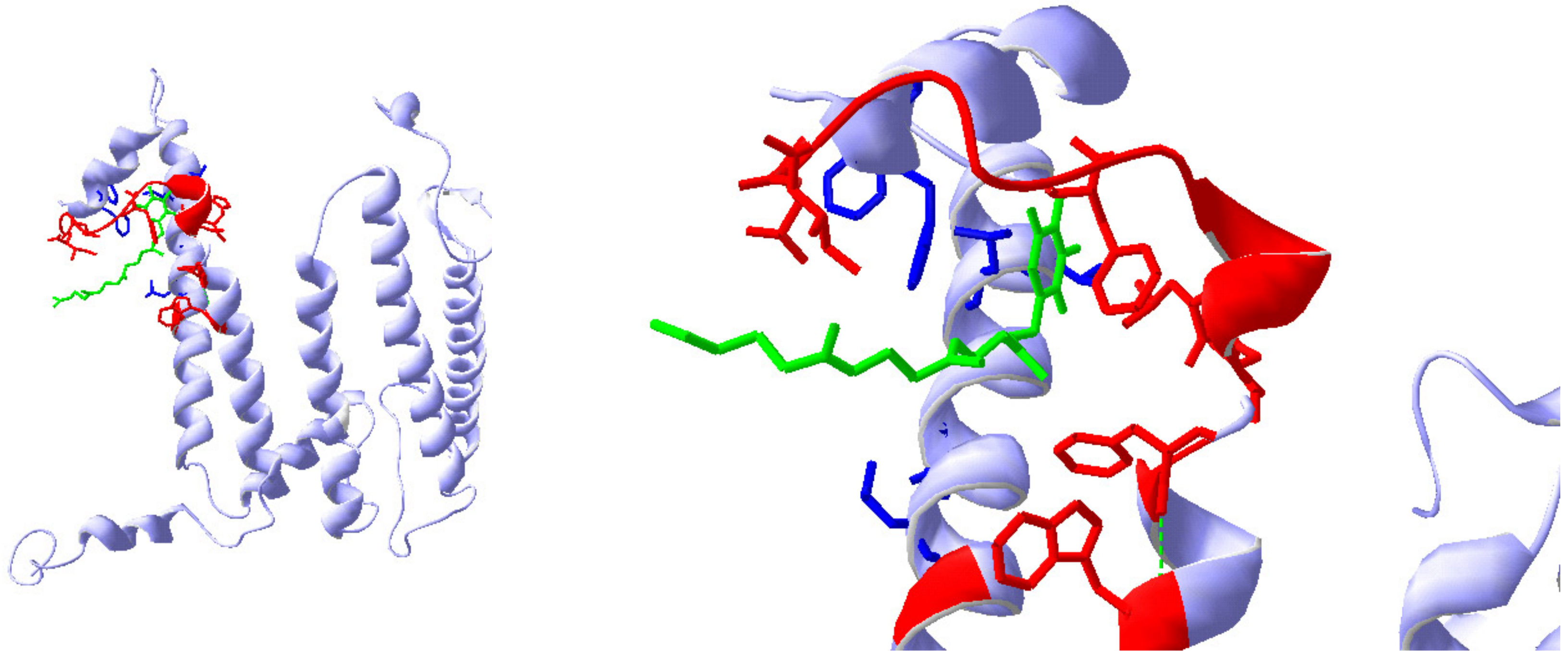
BUT……

- 220-230 loop critical for Q_B binding

SOLUTION…….

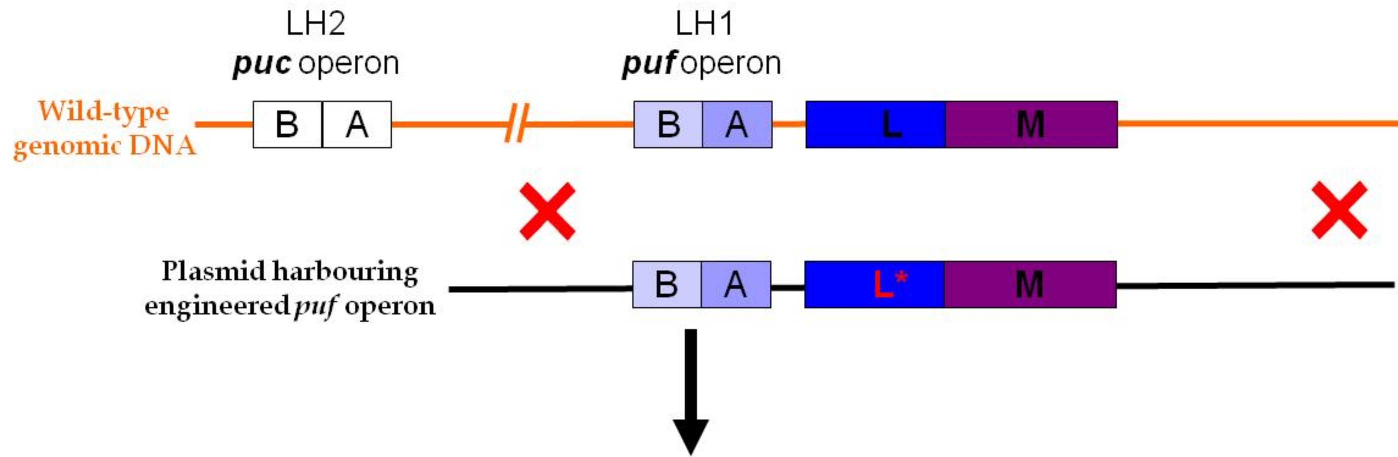
- Site-directed mutagenesis plus loop grafting

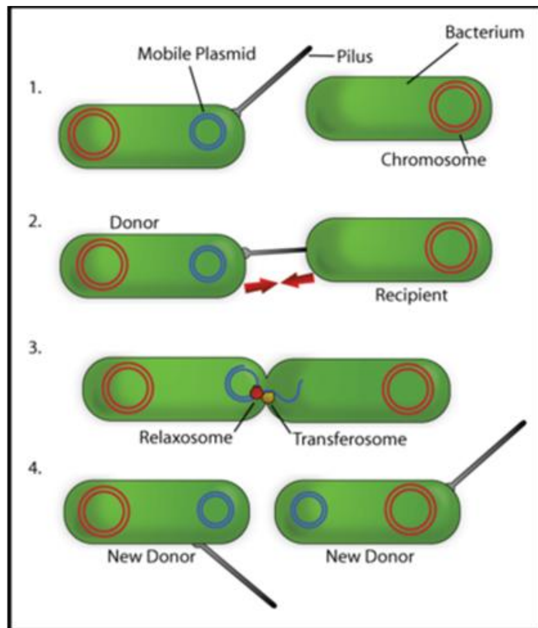
Result ?

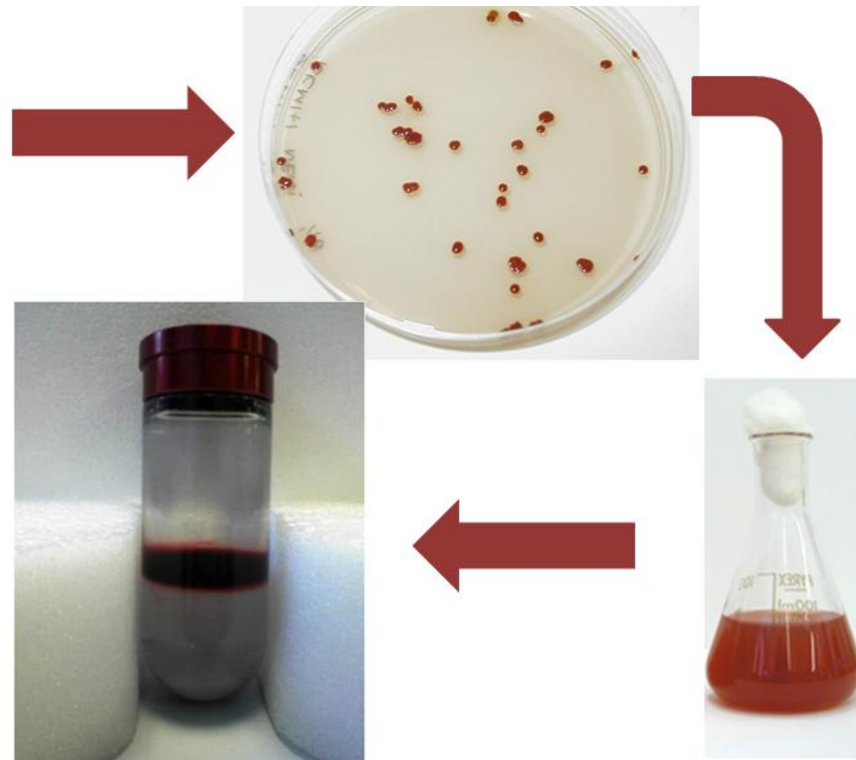


- Thr182Ala, Leu232Phe, Leu236Trp
- Substitution of the 220-230 loop with the 259-271 loop of D1

Creation of the Chimera

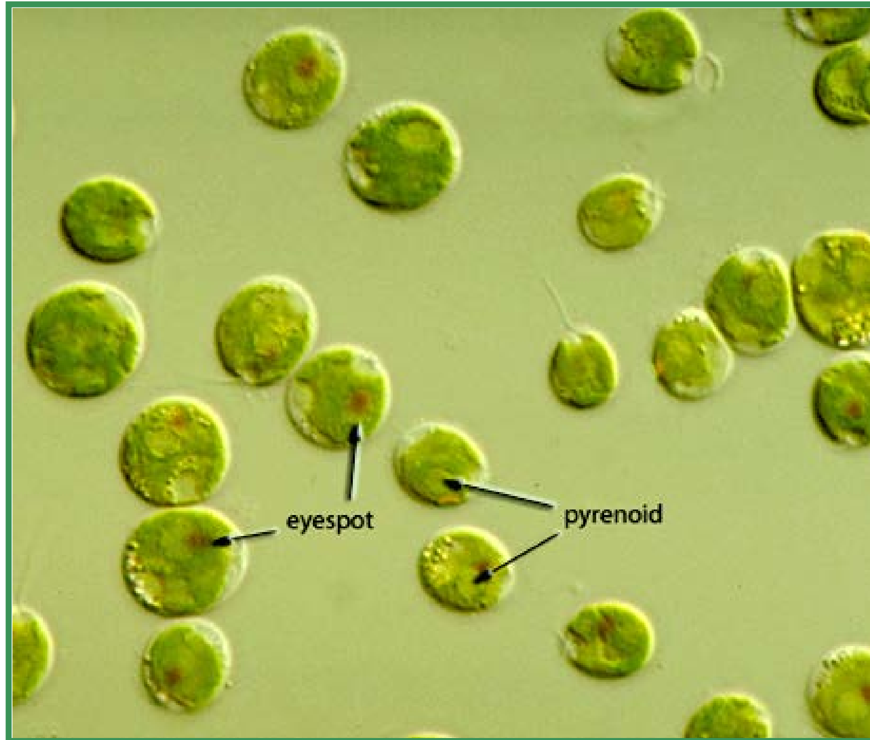






Directed evolution of *Chlamydomonas reinhardtii* PSII D1 protein for high stability biorecognition elements

Chlamydomonas reinhardtii is…



… a unicellular green alga having:

- short life cycle
- easy cultivation
- huge mutant collections
- low sensitivity to microgravity

WHY ALGAE IN SPACE?

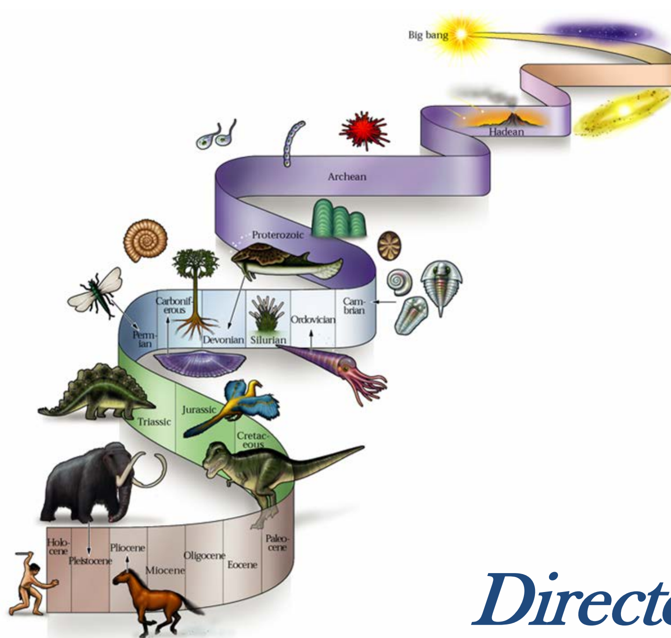
Oxygenic microalgae could provide:

- an oxygenic atmosphere
- edible biomass
- antioxidant and nutraceutic compounds

Microalgae are not adapted to the harsh conditions of solid-state device environment

What can we do using existing algae and modern techniques to obtain better adapted strains?

Natural evolution takes millions of years...

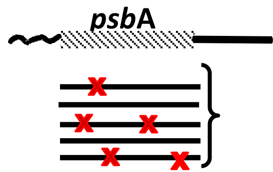


Directed evolution takes weeks...

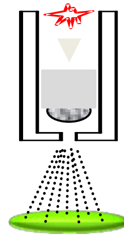
Improving radical-scavenging tolerance

1. Random mutagenesis

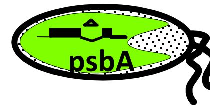
Error-Prone PCR on *psb*A gene



Transformation by particle gun



Primary pool of mutants

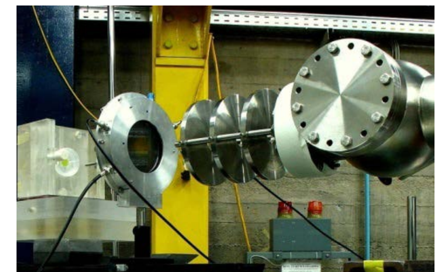


Selection for photo-autotrophic growth

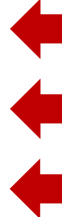
Cocktail of D1-random mutants

2. Selection by exposure to ionizing radiation

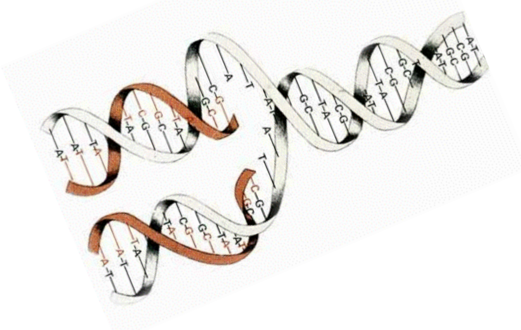
neutrons, 14 MeV (ENEA)
doses: 35&75 mGy



protons, 27 MeV (INFN)
doses: 0.5&5 Gy



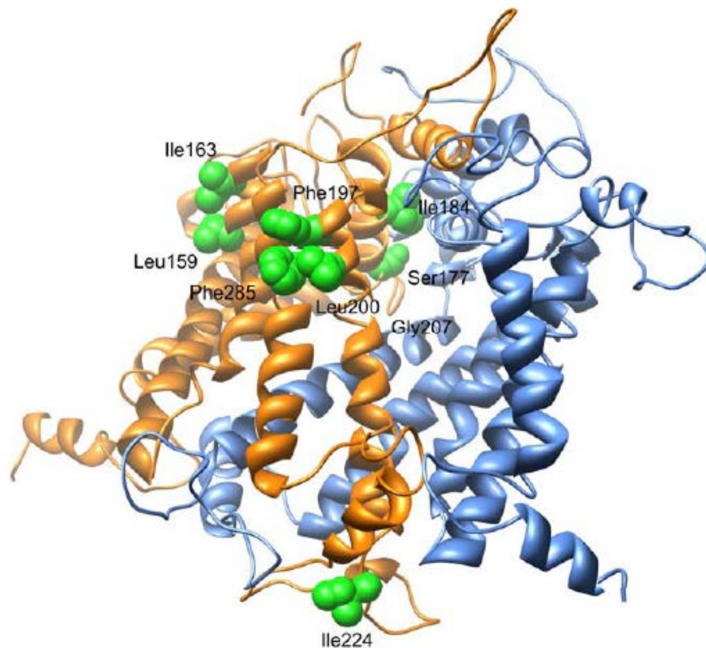
3. Identification of the D1 mutations in survived colonies

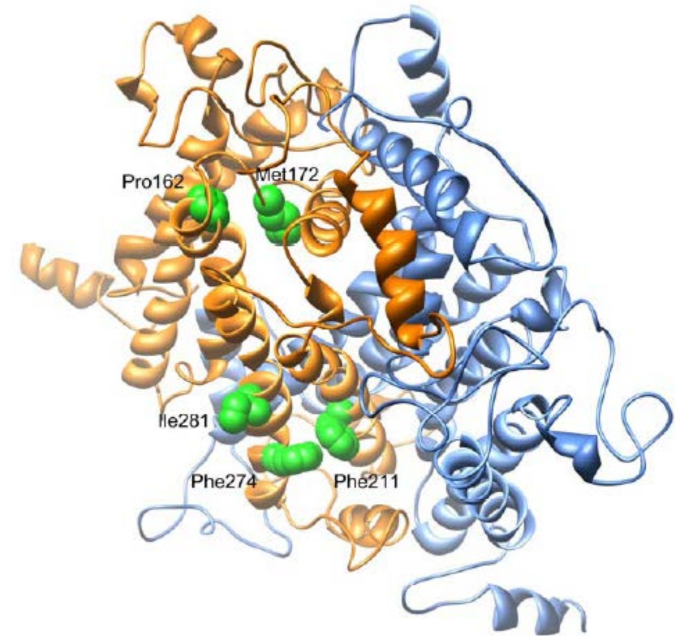


***psb*A gene sequencing analyses**

D1 random mutants tolerant to neutron/proton bombardments

Among the 2000 produced strains, 19 overcame the radiation induced stress.



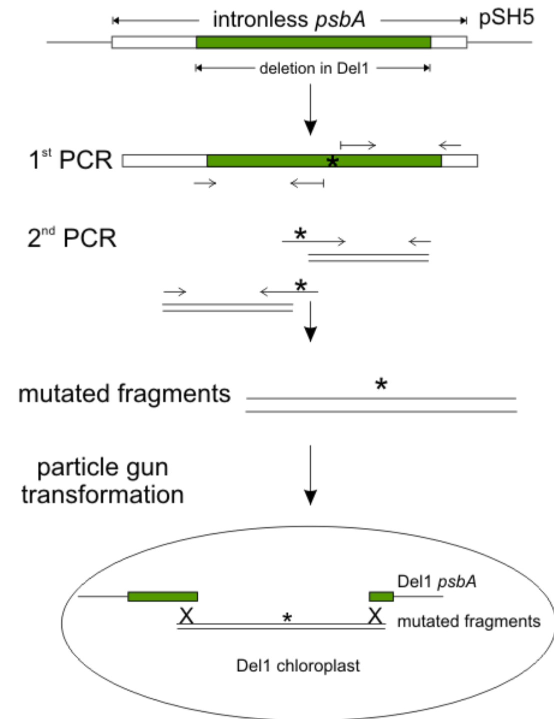


Under <u>neutron</u> mainly aliphatic and aromatic residues

Under <u>proton</u> only aliphatic and aromatic residues

D1 site-directed mutants of the random tolerant strains

Site-directed mutagenesis

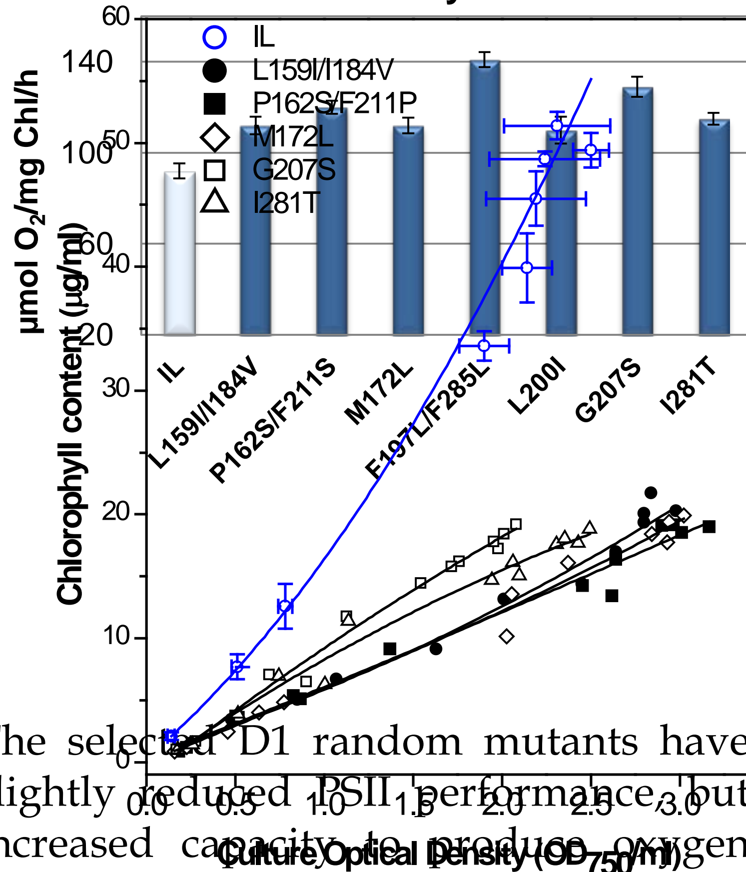


Most of the amino acid substitutions consisted of replacement of non-polar with polar residues that are less prone to oxidative damage.

Mutants	Amino acid substitutions wild type → mutated		Amino acid properties hydropathy index[a]/reactivity class/side chain polarity wild type → mutated				Localization of the mutation in the protein
P162S	proline	serine	-1.6 (III)	nonpolar	-0.8 (0)	polar	near to Tyr_{161}
I163T	isoleucine	threonine	4.5 (IV)	nonpolar	-0.7 (0)	polar	near to Tyr_{161}
M172L	methionine	leucine	1.9 (V)	nonpolar	3.8 (IV)	nonpolar	near to OEC
G207S	glycine	serine	-0.4 (I)	nonpolar	-0.8 (0)	polar	in the helix IV of D1
L200I	leucine	isoleucine	3.8 (IV)	nonpolar	4.5 (IV)	nonpolar	in the helix IV of D1
I281T	isoleucine	threonine	4.5 (IV)	nonpolar	-0.7 (0)	polar	in the helix V of D1

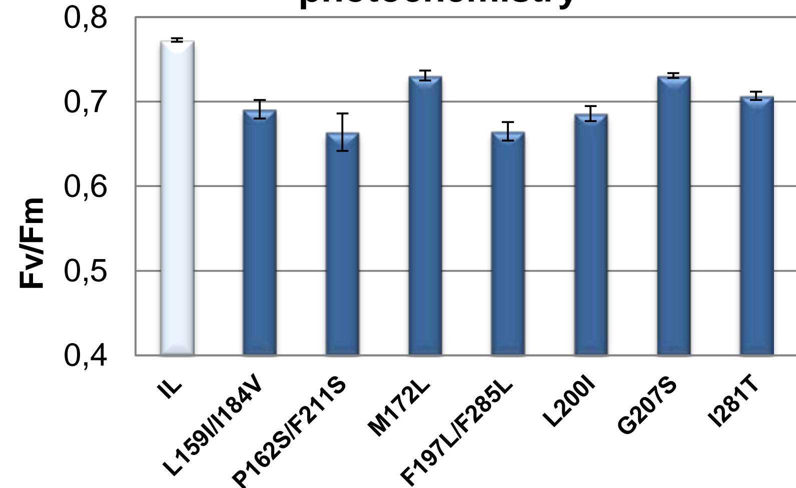
PHYSIOLOGICAL CHARACTERIZATION

Oxygen evolution rate under saturated light intensity

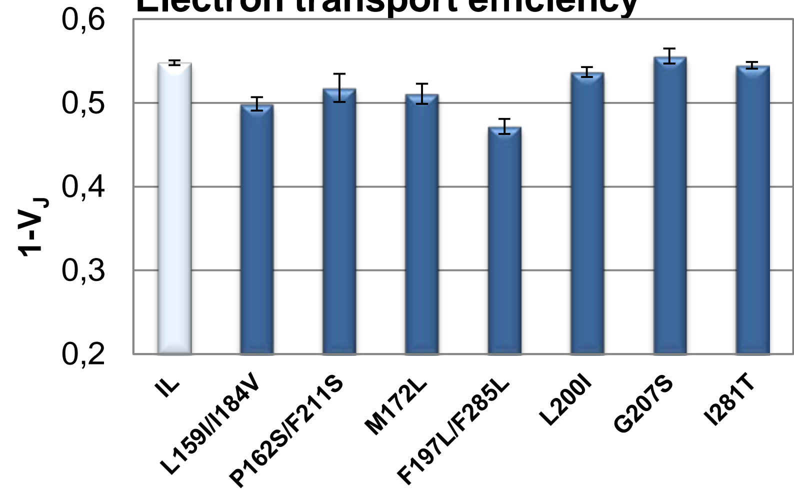


The selected D1 random mutants have slightly reduced PSII performance, but increased capacity to produce oxygen under saturated light intensity

Maximum quantum yield of PSII photochemistry



Electron transport efficiency

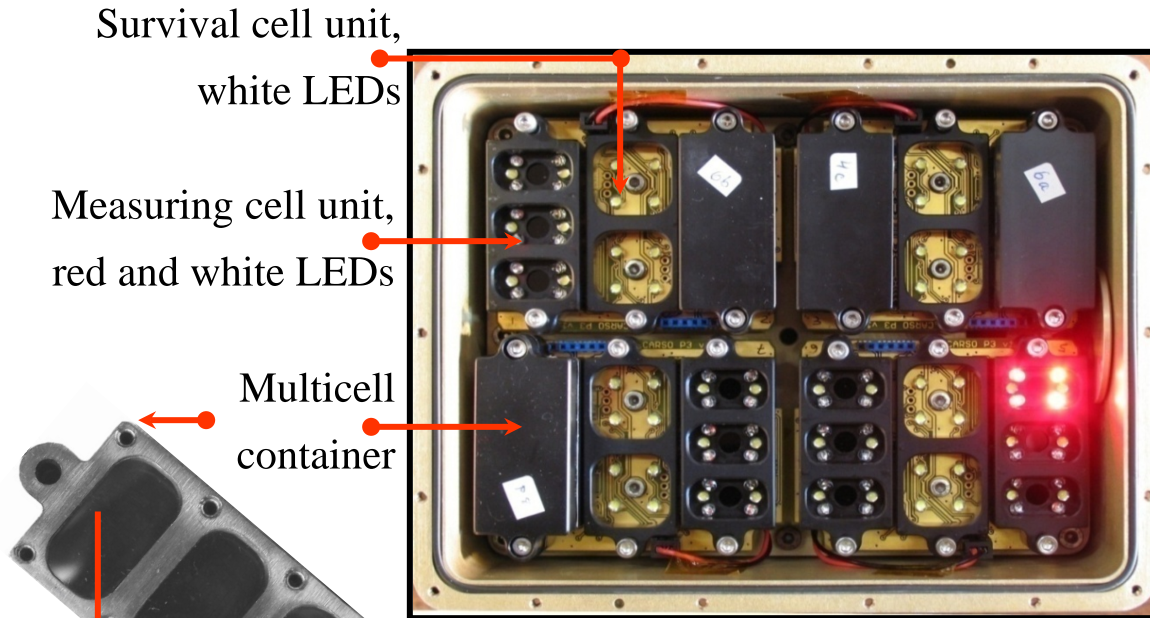


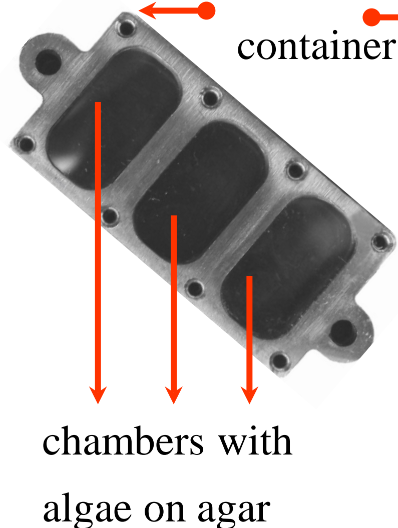
Long-term stability and tolerance to radical-generating conditions

Survival cell unit, white LEDs

Measuring cell unit, red and white LEDs

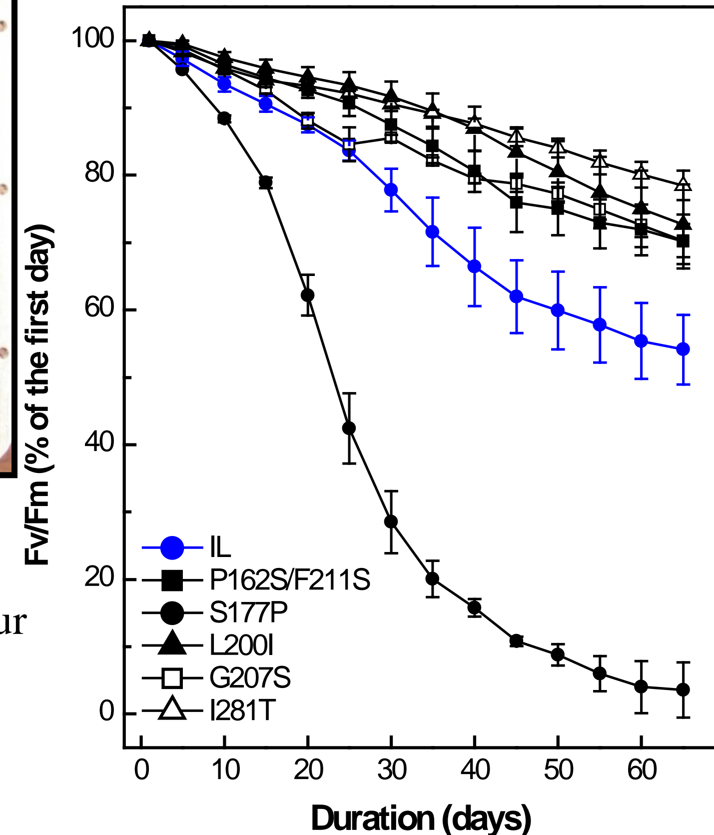
Multicell container





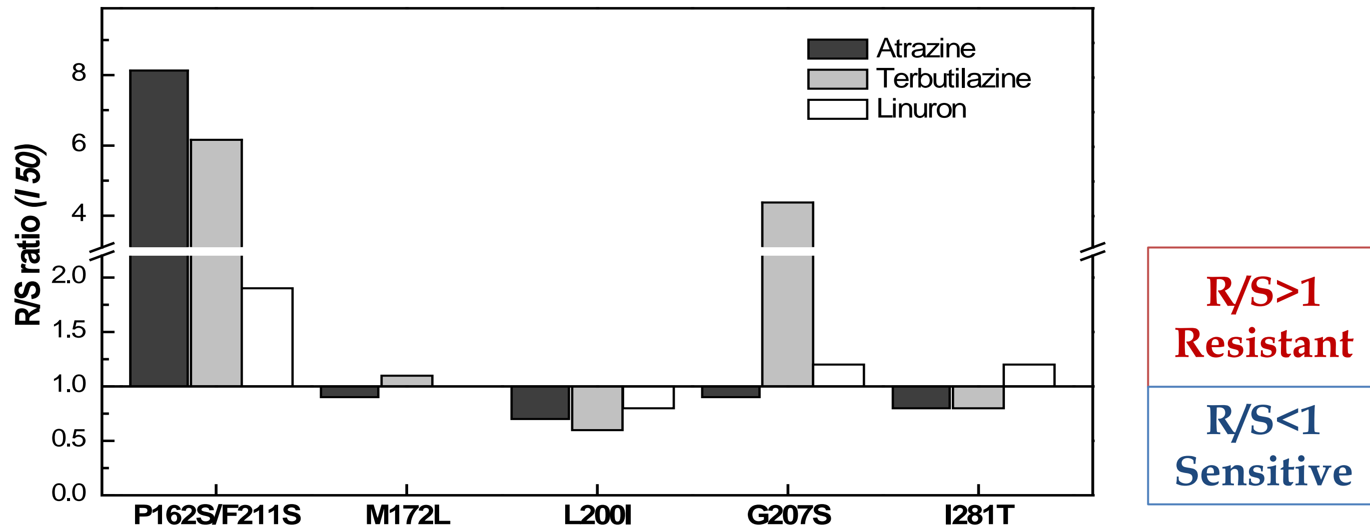
chambers with algae on agar

- 7 h light/17 h dark
- fluorescence measurement every hour
- thermo-sensors
- data storage for 1 month



Photosynthetic performance of immobilized cell cultures on TAP agar medium under 13 °C and 20 µmol/m^2/s light intensity for more than 2 months

STRAIN'S HERBICIDE RESISTANCE/SENSITIVITY



Mutants with modified pollutant response were selected

Institute of Crystallography,
National Council of Research,
Rome, *Italy*

Giuseppina Rea
Maya Lambreva
Viviana Scognamiglio
Amina Antonacci
Katia Buonasera
Ivan Husu
Giuseppe Rodio
Maria Teresa Giardi
Elefterious Touloupakis

Department of Biology,
University Roma Tre
Rome, *Italy*

Fabio Polticelli

National Institute for Biological Sciences, Bucharest, *Romania*

Simona Carmen Litescu

Martin-Luther-University,
Plant Physiology Institute,
Halle, *Germany*

Udo Johanningmeier
Ivo Bertalan

School of Biochemistry,
Bristol, United Kingdom

Mike Jones

Inst. Plant Physiology&Genetics
Bulgarian Academy of Sciences
Sofia, *Bulgaria*

Liliana Maslenkova
Violeta Peeva

Thank you!