Sabancı Üniversitesi

APPLICATIONS of SMALL ANGLE X-RAY SCATTERING and BIOINFORMATICS TOOLS

Zehra Sayers

Faculty of Engineering and Natural Sciences,
Sabanci University, Istanbul Turkey

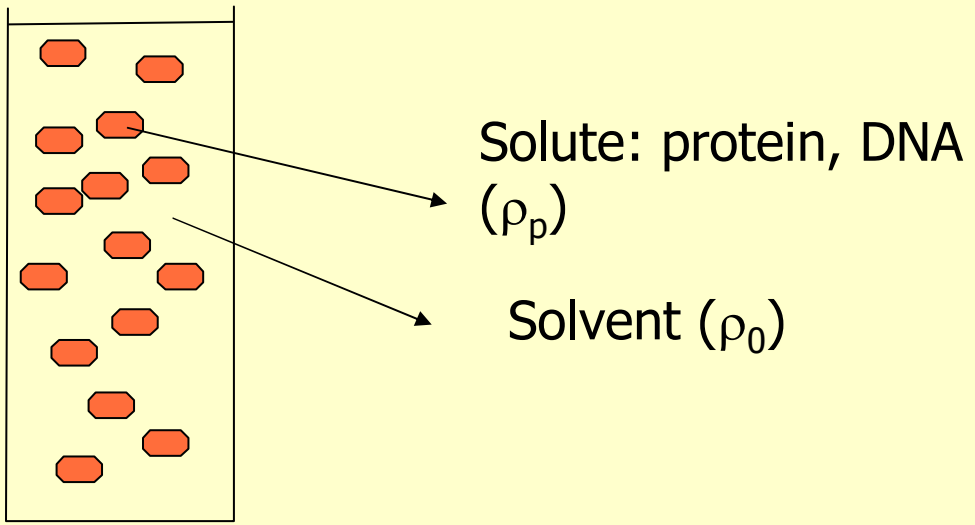
SMALL ANGLE SOLUTION X-RAY SCATTERING

- A method for investigating structure of macromolecules in solution.
- Offers the advantage of having the material in "native" conditions.
- Allows introduction of perturbations, e.g. rapid mixing, temperature jump, activation by laser light, pressure change.
- ***Time resolved measurements*** in sub-millisecond range are made possible by synchrotron radiation. These provide insights into mechanisms of reactions, interactions, folding and unfolding of biological macromolecules.
- Can be applied to molecules with sizes from a few kD to several MD.

SMALL ANGLE SOLUTION X-RAY SCATTERING

- Small angle X-ray scattering results from inhomogeneities in the electron density in a solution due to macromolecules dispersed in the uniform electron density of the solvent (ρ_0).

A solution of macromolecules



- Scattering pattern is determined by the excess electron density of the solute, $\rho(\mathbf{r})$

$$\rho(\mathbf{r}) = (\rho_p - \rho_0)\rho_c(\mathbf{r}) + \rho_s(\mathbf{r})$$

$$= \rho_{av}\,\rho_c\,(\mathbf{r}) + \rho_s\,(\mathbf{r}) \qquad (1)$$

Where

ρ_p = the average electron density of the particle.

ρ_{av} = the average electron density of the particle above the level of the solvent (contrast).

$\rho_c\,(\mathbf{r})$ = dimensionless function describing the volume of the solute (with the value 1 inside the particle and 0 elsewhere).

$\rho_s\,(\mathbf{r})$ = fluctuations of the electron density above and below the mean value (independent of the contrast).

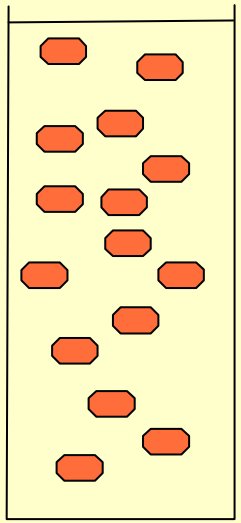
- For a solution of chromatin

  ρ_0= the average electron density of the buffer; 330 e/nm^3.

  ρ_p= the weighted average electron density of the DNA and histones; 484 e/nm^3.

  $\rho_c(\mathbf{r})$ depends on the shape of the fiber

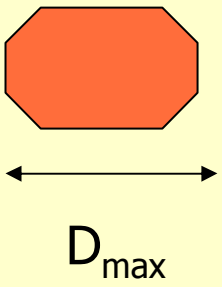
  $\rho_c(\mathbf{r})$ represents deviations from the average electron density due to regions of linker DNA and the regions with the nucleosome core particles. Excess scattering mass of the linker DNA is about 7X10^3 electrons against 4X10^4 electrons of the nucleosome. The scattering pattern at low angles is dominated by the contribution from nucleosomes.



- In an ideal solution all particles are identical and randomly positioned and oriented in the solvent.

- Scattering pattern contains information about the spherically averaged structure of the solute described by a *distance probability function p(**r**)*

- p(**r**) is the spherically averaged autocorrelation function of $\rho(\mathbf{r})$ and
$r^2p(\mathbf{r})$ is the probability of finding a point inside the particle at a distance between r and r+dr from any other point inside the particle



- For a globular particle p(**r**) has two main regions

  a. A region of sharp fluctuations due to neighbouring atom pairs (0.1 nm$\leq$r$\leq$ 0.5 nm) and of damped oscillations due to structural domains
  (i.e α-helices in proteins)

  b. A smooth region corresponding to intramolecular vectors.

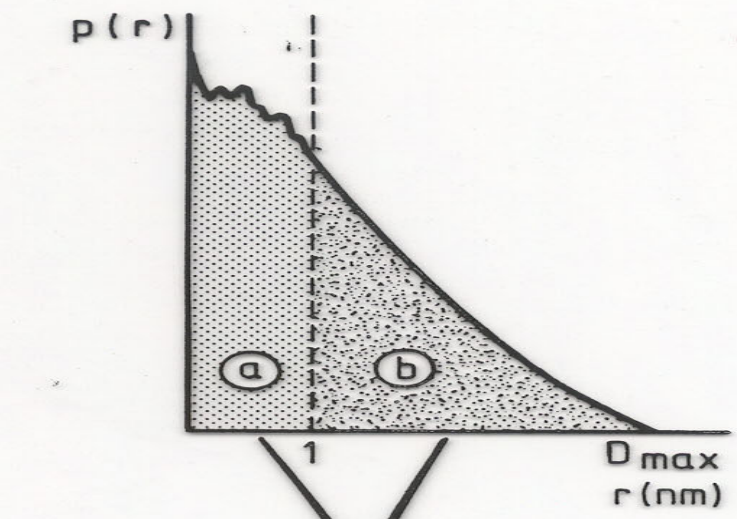
- Beyond D_{max} p(r) vanishes

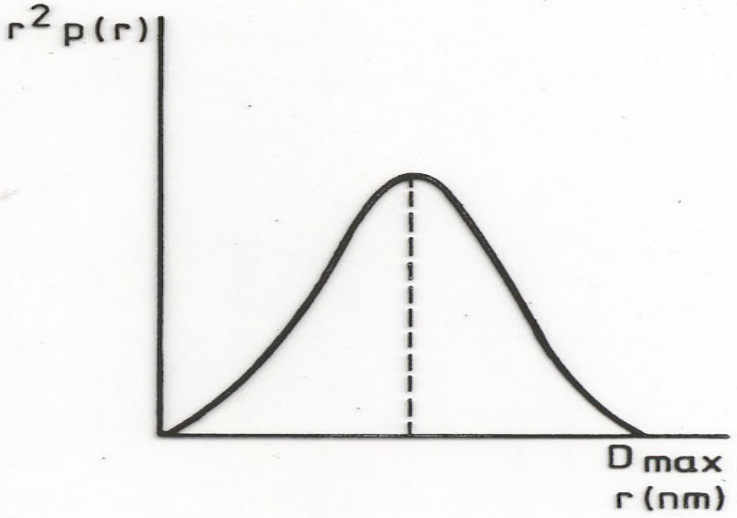
- The scattering curve also contains two regions:

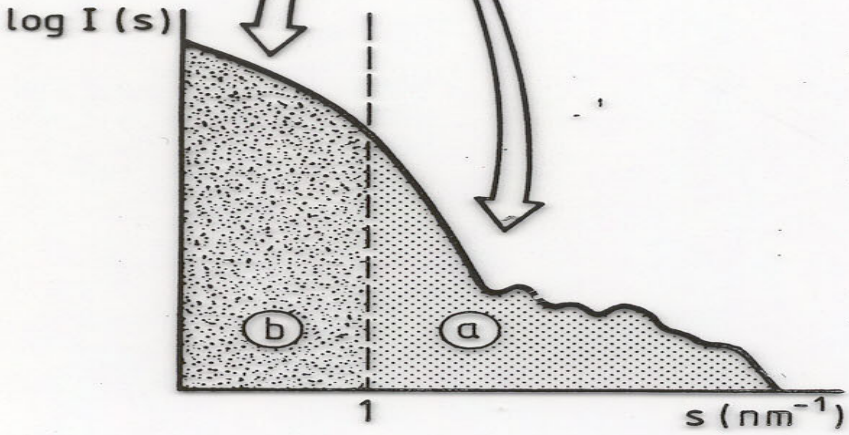
  a. Small angle region; information on the long range organization (shape) of the particle

  b. Large angle region; internal structure of the particle (deviations from ρ_p)

Scattering pattern contains information on the distance distribution function p(**r**)







Scattering intensity and the distance distribution function are related by a Henkel transformation.

$$I(s) = 4\pi \int_0^\infty r^2 p(r) \frac{\sin(2\pi sr)}{2\pi sr} dr$$

THE PRINCIPLE OF A SMALL ANGLE X-RAY SOLUTION SCATTERING EXPERIMENT

- The optical system selects X-rays with a wavelength of 0.15 nm and a narrow band-width
- The beam is focused on a position sensitive detector with an adequate cross section at the sample position
- The incident beam intensity I_0 is monitored using an ion chamber

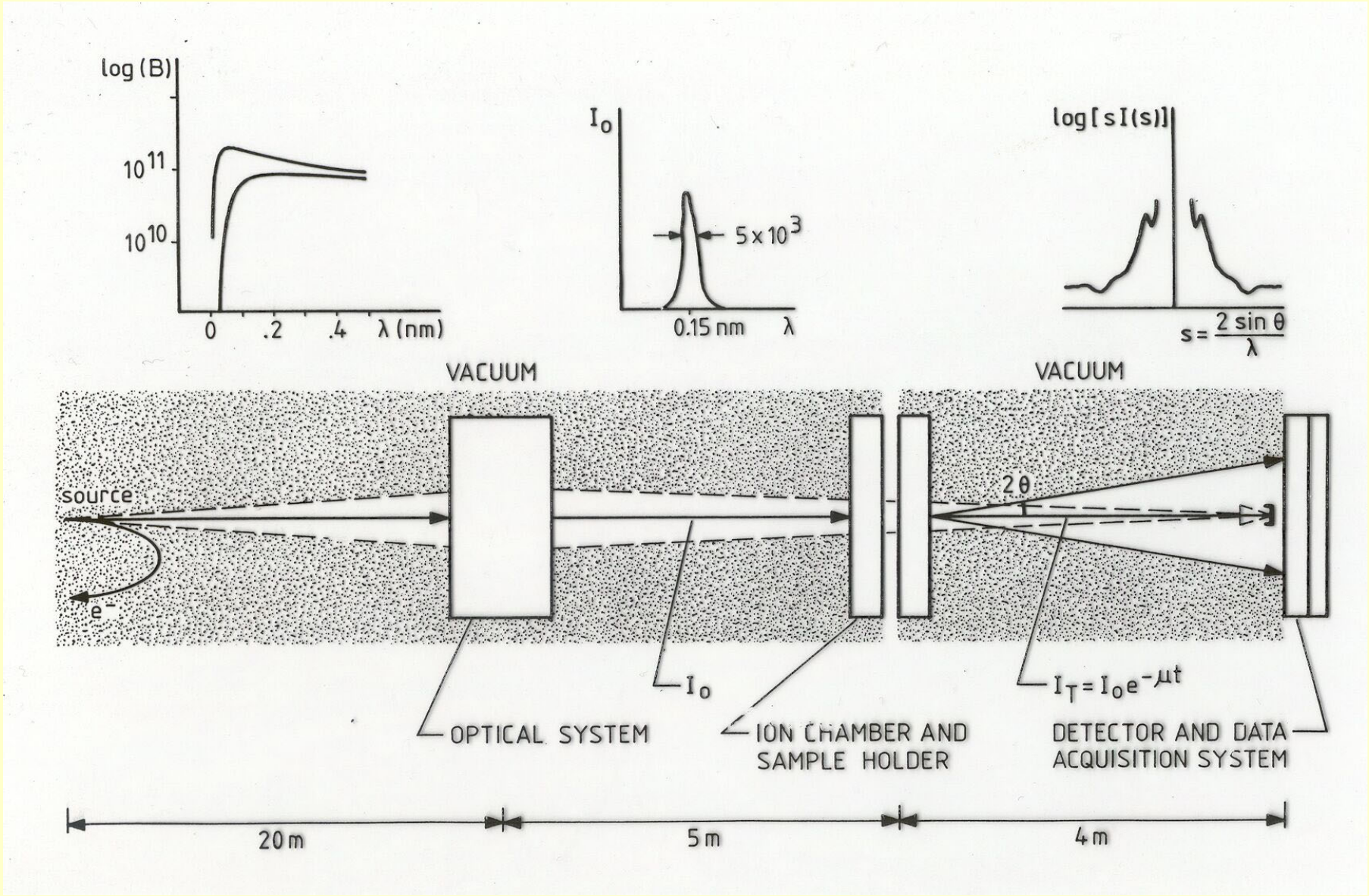
- I_T is the intensity of the beam transmitted through the sample and $I_T = I_0 \exp(-\mu t)$, where the factor $(-\mu t)$ represents the absorbance of a solution of thickness t

- I(s) is the scattered intensity which depends on the scattering vector s defined as

  $$s = 2\sin\theta/\lambda$$

  where 2θ is the scattering angle and λ is the wavelength

Set-up for small angle X-ray scattering experiments on the synchrotron



STRATEGY FOR INVESTIGATIONS

Preparation of material:

- Clone genes of proteins of interest and overexpress proteins.
- Purify proteins and biochemically characterize them.
- Try to crystallize.
- Prepare concentrated solutions of the material.

Measurements and Experiments:

- Data collection for crystal structure determination.
- Solution X-ray scattering experiments for determination of shape and for monitoring structural changes during activity (function).

Data Analysis, Evaluation and Modelling:

- Use of new methods for small angle scattering data analysis to obtain structural details.
- Modelling to support information on mechanisms relating to function of the molecule.

Complementation of Experimental Work using Bioinformatics Tools

- Use of computational methods based on sequence alignment, secondary structure prediction, homology modelling etc. Development of models of the 3D structure and models for mechanisms which can be tested through experiments

APPLICATIONS

- Determination of radius gyration, radius gyration of the cross section, molecular weight.
- Shape determination; at low angle (2-3 nm) the scattering curve is dominated by the shape of the particle.
- Modern methods allow domain structure analysis, possibility of modelling loop domains, analysis of non-equilibrium systems (Svergun and Koch 2002, Current Opinion in Structural Biology, 12:654-660).

- Two systems under investigation:

Heterotrimeric G-proteins from *A. Thaliana,*

Mert Sahin, Suphan Bakkal and Ugur Sezerman

***T. Durum* metallothionein**

Kivanc Bilecen, Umit Ozturk and Ugur Sezerman

- These systems are particularly suitable for small angle X-ray scattering experiments because,

  Changes in the heterotrimer structure during interaction of the subunits and when the trimer is activated can only be investigated in solution.

  Metallothioneins are hard to crystallize and the formation of the two metal binding domains can be studied by solution scattering.

A. thaliana G-PROTEIN

- Heterotrimeric G-proteins: a major component of signal transduction pathways in several organisms from yeast to mammals.
- The heterotrimer consists of α-, β– and γ- subunits forming a tight complex at the interior of the cell membrane.
- The α-subunit has two domains;

  -a helical domain

  -the GDP/GTP binding site, GTP hydrolysis activity and the covalently attached lipid for membrane association
- Upon activation by a signal, GDP is exchanged for GTP resulting in dissociation of the α-subunit from the $\beta\gamma$ complex. Both α- and $\beta\gamma$ complex then bind to their effectors and transmit the signal downstream in the cell.

G- PROTEIN SUBUNITS FOR X-RAY SCATTERING EXPERIMENTS

- The *GPA1* gene (Dr. Hong Ma, PennState University, USA) in pCIT757 vector.
- Amplification of *GPA1* using 3 sets of primers. Different restriction sites for cloning into expression vectors pGEX-4T2 pGFP-uv, pETM-11 and pETM-30 (EMBL).
- Subcloning of *GPA1* into into pCR® II-TOPO® vector system using *E. coli* XL1BLUE.
- Cloning *GPA1* insert after sequence verification 3'- MC site of pGFPuv expression vector using *Sac I* and *Spe I* restricition sites.

PRIMERS

pGFP-uv/5′ (*Sac*I):
5′- AAA CCC GAG CTC ATG GGC TTA CTC -3′

pGFP-uv/3′(*Spe*I):
5′- AAA CCC ACT AGT TCA TAA AAG GCC A-3′

pGEX-4T2/5′ (*Eco*RI):
5′- GCG TCG AAT TCC CAT GGG CTT ACT CTG-3′

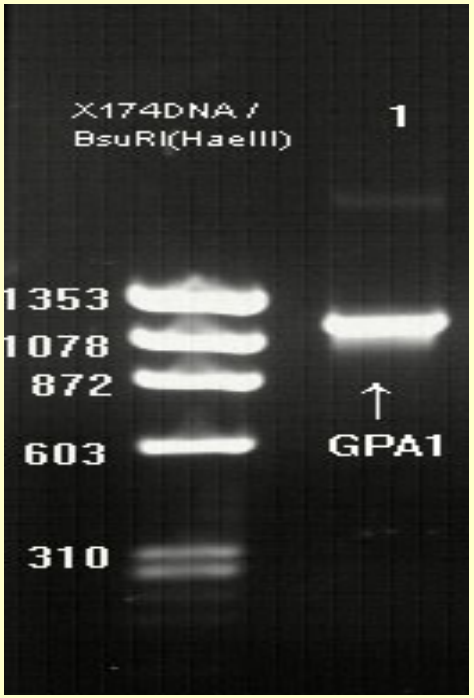
pGEX-4T2/3′(*Xho*I):
5′-AAA CCC CTC GAG TCA TAA AAG GCC A-3′

pETM-11&30/5′(*Eco*RI):
5′- GCG TCG AAT TCG ATG GGC TTA CTC TG-3′

pETM-11&30/3′(*Xho*I):
5′-AAA CCC CTC GAG TCA TAA AAG GCC A-3′

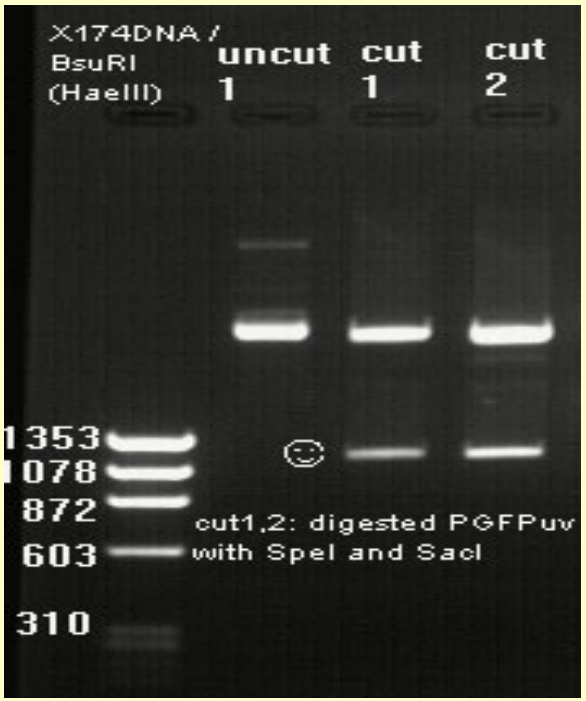
PCR

Amplificaton of *GPA1* with primers containing *SacI* and *SpeI* at 5'- and 3'-ends respectively. (*GPA1* is 1149 bp)

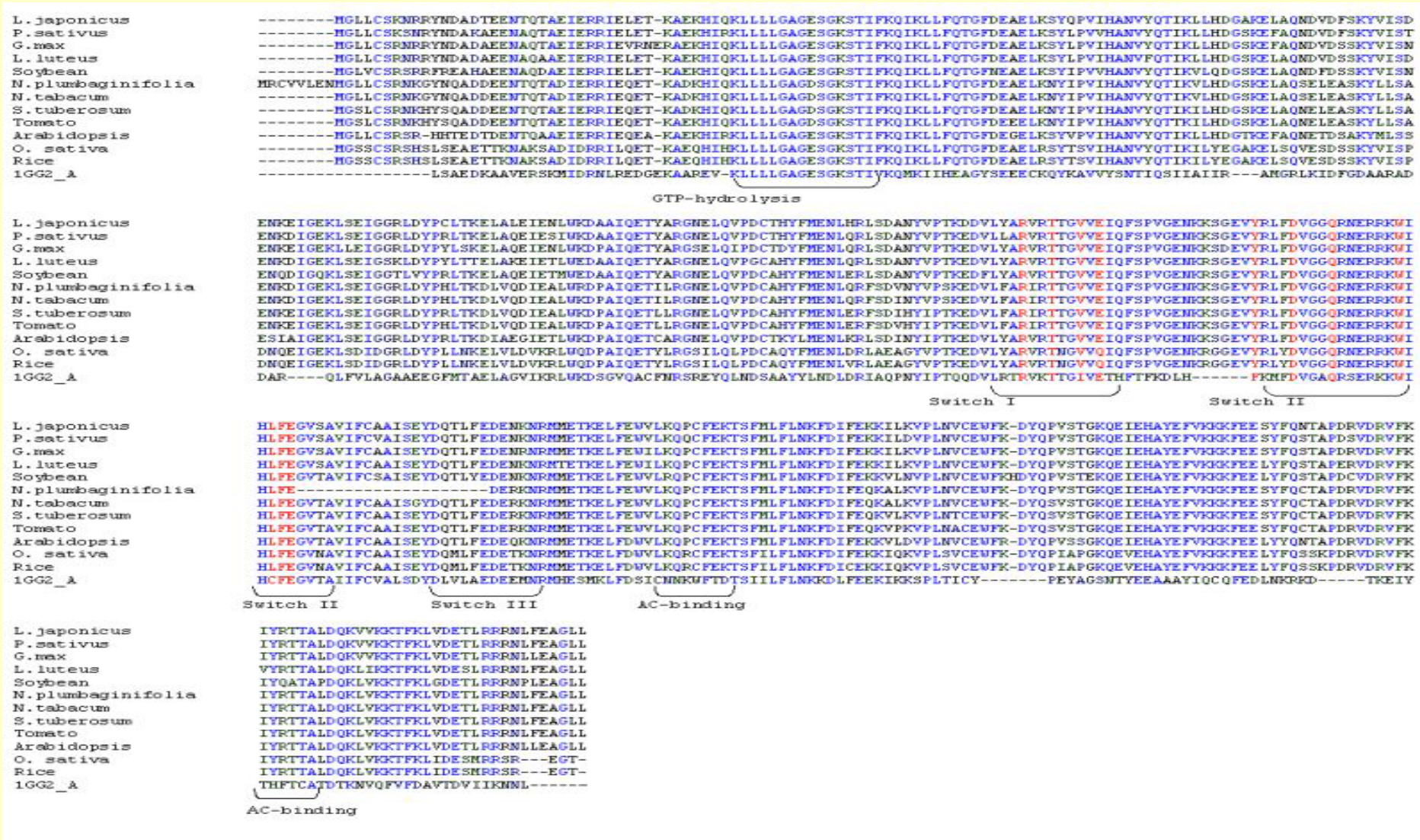


CLONING

***GPA1* + pGFPuv construct in E. coli XL1BLUE (Screening by digestion with *SpeI* and *SacI*)**



Towards Modelling G-Protein Subunits



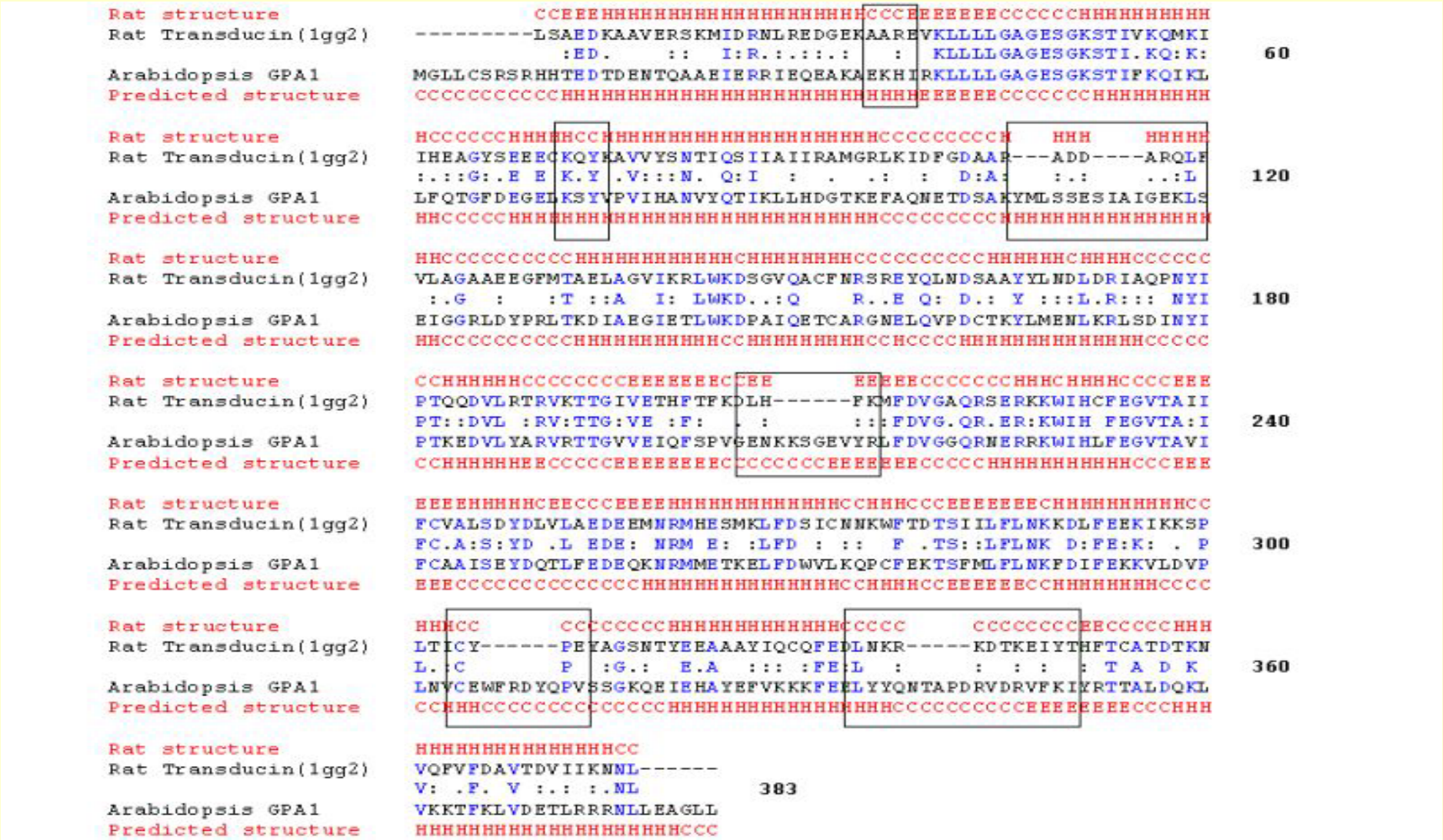
CLUSTALW alignment of all plant G-protein α-subunits and rat transducin α-subunit IGG2_A Residues in blue are identical, highly similar ones are in green and key residues are shown in red.

Comparison of plant α-subunit sequences with that of rat transducin

- The loop regions (G1 to G5) involved in Mg^{+2} coordination, GTP recognition and hydrolysis are highly conserved
- The switch regions involved in effector binding (and GTP hydrolysis) are highly conserved
- The switch II region involved in interaction with β-subunit is conserved
- The putative adenylate cyclase (AC) binding region although conserved among plant species shows little homology with mammalian sequences

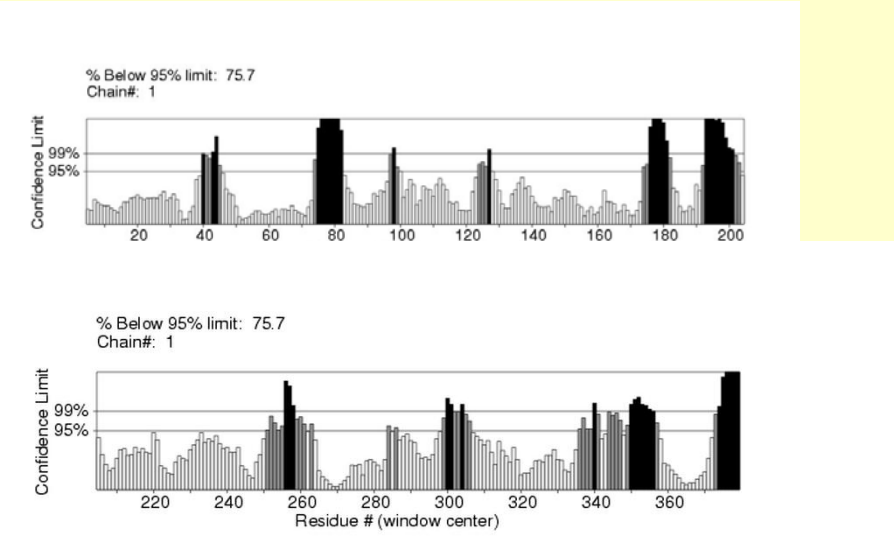
- Secondary structure prediction
  the META server
    PSIpred (Jones DT., 1999),
    PSSP (Raghava G., unpublished)
    SAM-T99 (Karplus K. et al.,1998).

- Generation of a consensus prediction



Sequence and secondary structure alignments of GPA1 and 1GG2_A. Secondary structure features (H; helix, C; coil, E; extended) are shown in red. Identical amino acids are shown in blue, and similar ones are shown in green. Boxes indicates erronuos predictions

MODELLING



Model:

CLUSTALW alignment of GPA1 with two rat transducin alpha subunits (PDB entries: 1GG2 and 1FQK),

Align2D alignment within the MODELLER program and high loop optimization conditions

Accuracy verified using ERRAT

Results:

76% of the structure is within confidence limits.

The two domain structure of GPA1, the GTP-binding pocket and the distiribution of helices and extended structures indicate that the loop regions do not strongly interfere with the structure of the functional sites.

The model can be improved by further optimizing the loop regions.

WHEAT METALLOTHIONEINS (MTs)

- Low molecular weight (6-7 kD) proteins.
- Rich cystein content and lack of aromatic amino acids
- Involved in:
  - >>> heavy metal (Cd, Hg, etc.) detoxification
  - >>> Zn and Cu regulation
  - >>> ROS scavenging
  - >>> metabolism of metallo-drugs & alkylating agents

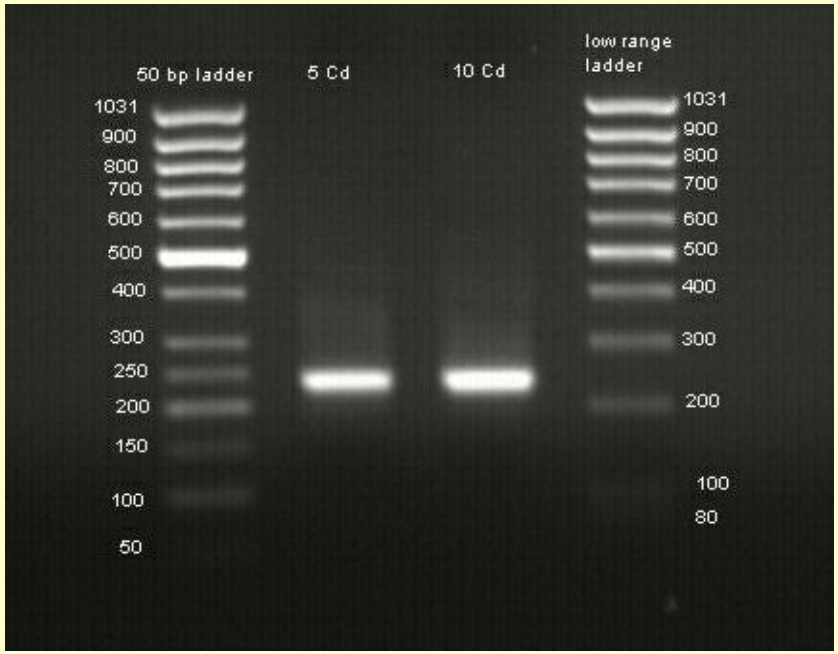
- Inducible by a variety of transcription factors and signals e.g. glucocorticoids, cytokines, ROS, metal ions.

- Possibly involved in;
  - Copper related diseases
  - Menkes and Wilson disease
  - Alzheimer disease

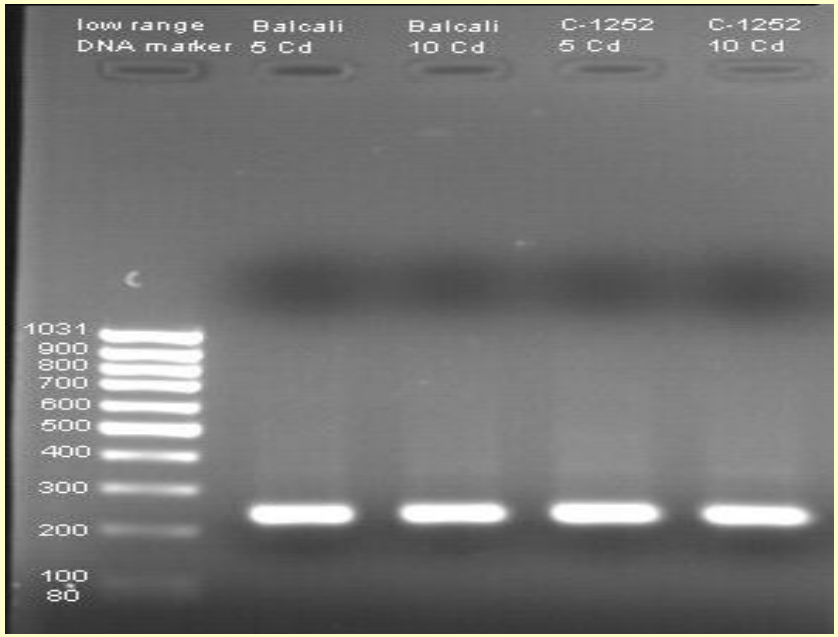
CLONING & CHARACTERIZATION of *d-mt*

- Total RNA and mRNA were isolated from *T. durum* cultivars (Balcali & C-1252) grown in medium supplemented with 5 and 10µM Cd.

- RT-PCR was carried out using QIAGEN® One-Step RT-PCR Kit and primers designed according to the known sequence of *a-mt* (L11879).



(A)

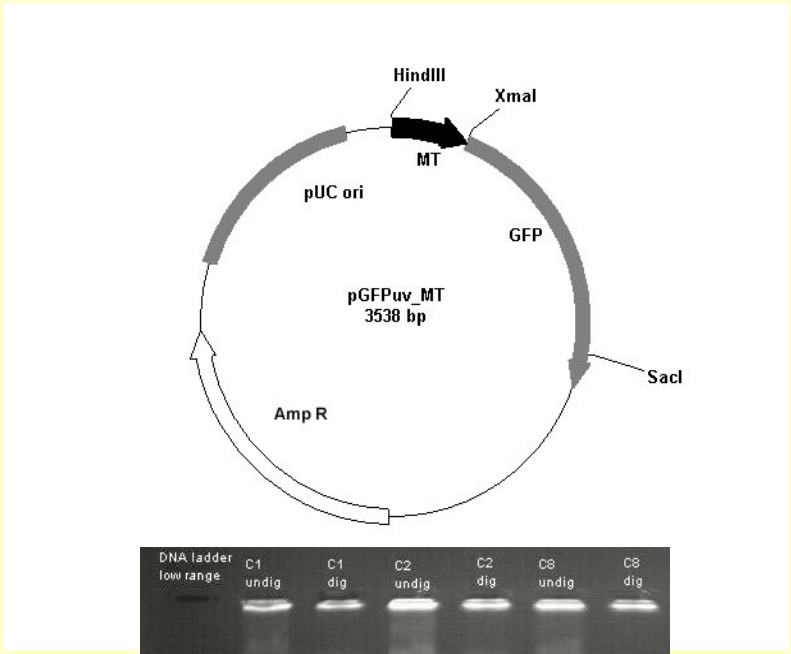


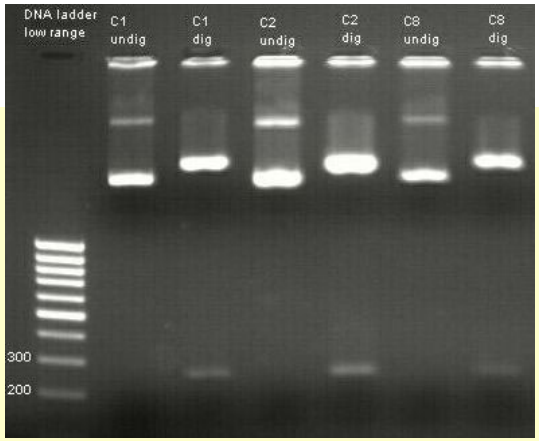
(B)

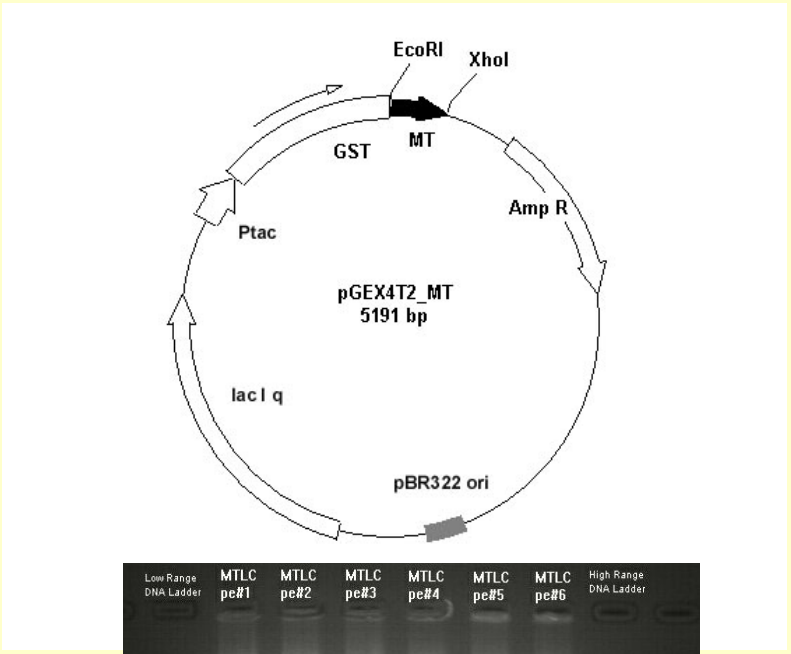
Agarose gel electrophoresis analysis of RT-PCR results

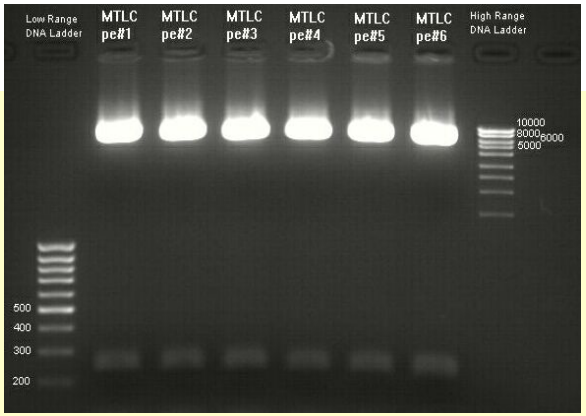
(A) ***T. aestivum,***
(B) ***T. durum*** **cultivars (Balcali and C-1252).**

The size of the PCR product was correlated with the size of the ***a-mt*** **which is 225 nucleotides.**





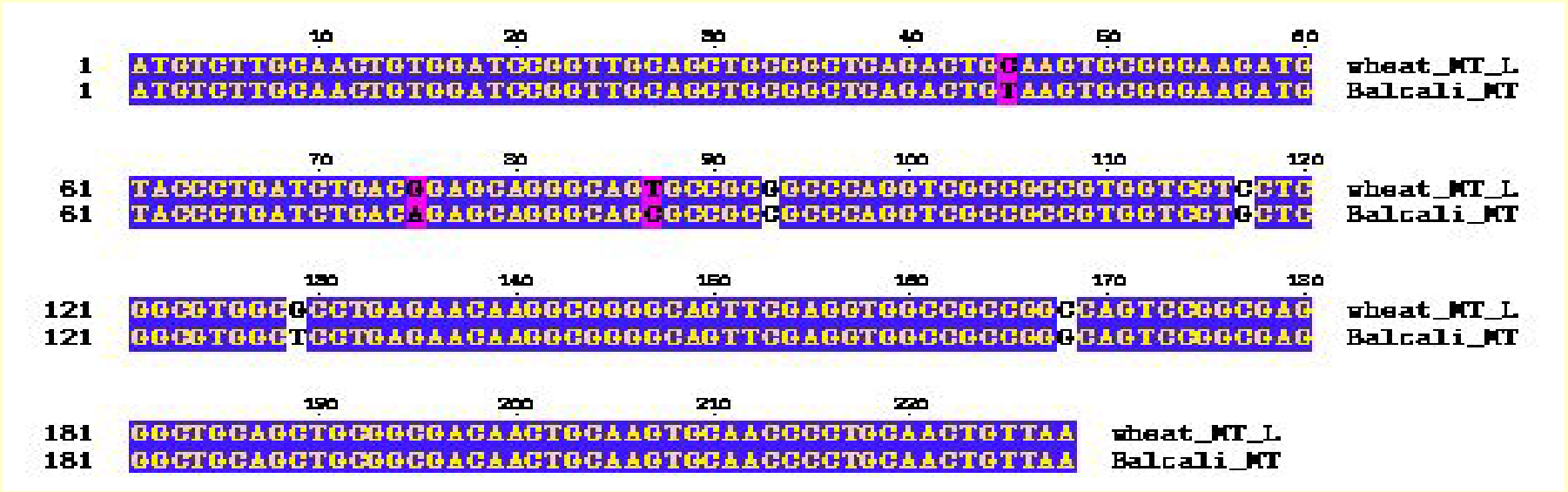




Screening for recombinant plasmids

Restriction enzyme digestion with *Hind* III and *Xma* I of pGFPuv-MT isolated from *E. coli* XL1-Blue.

Restriction enzyme digestion with *Eco* RI and *Xho* I of pGEX-4T-2 isolated from *E. coli* XL1-Blue.



Despite 7 possible mutations global alignment shows that *a-mt* and *d-mt* coded MT proteins are 100% identical.

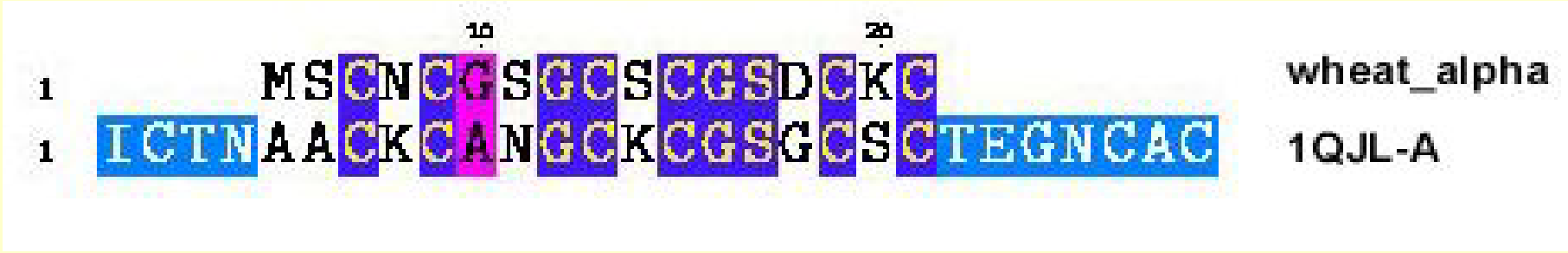
MODELING of w-MT PROTEIN USING "HOMOLOGY MODELING" & "*ab-initio*" APPROACHES

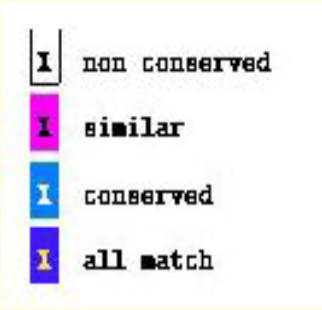
- High sequence similarity with the rat liver MT (4MT2)
  except in the hinge region connecting the two metal centers
- Plant MT hinge region contains up to 42 residues, whereas 2-3 in mammalians' MTs.
- For this reason w-MT structure was divided into 3 functional parts;
  alpha-domain
  beta-domain
  hinge region
- Each part was modeled separately

Homology modeling of the w-MT
α- and β- domains

- Lack of secondary structure features in metal centers
- Secondary structure prediction algorithms fail → presence of metals
- Modeling work was done using Deep View – The Swiss Pdb Viewer v3.7.

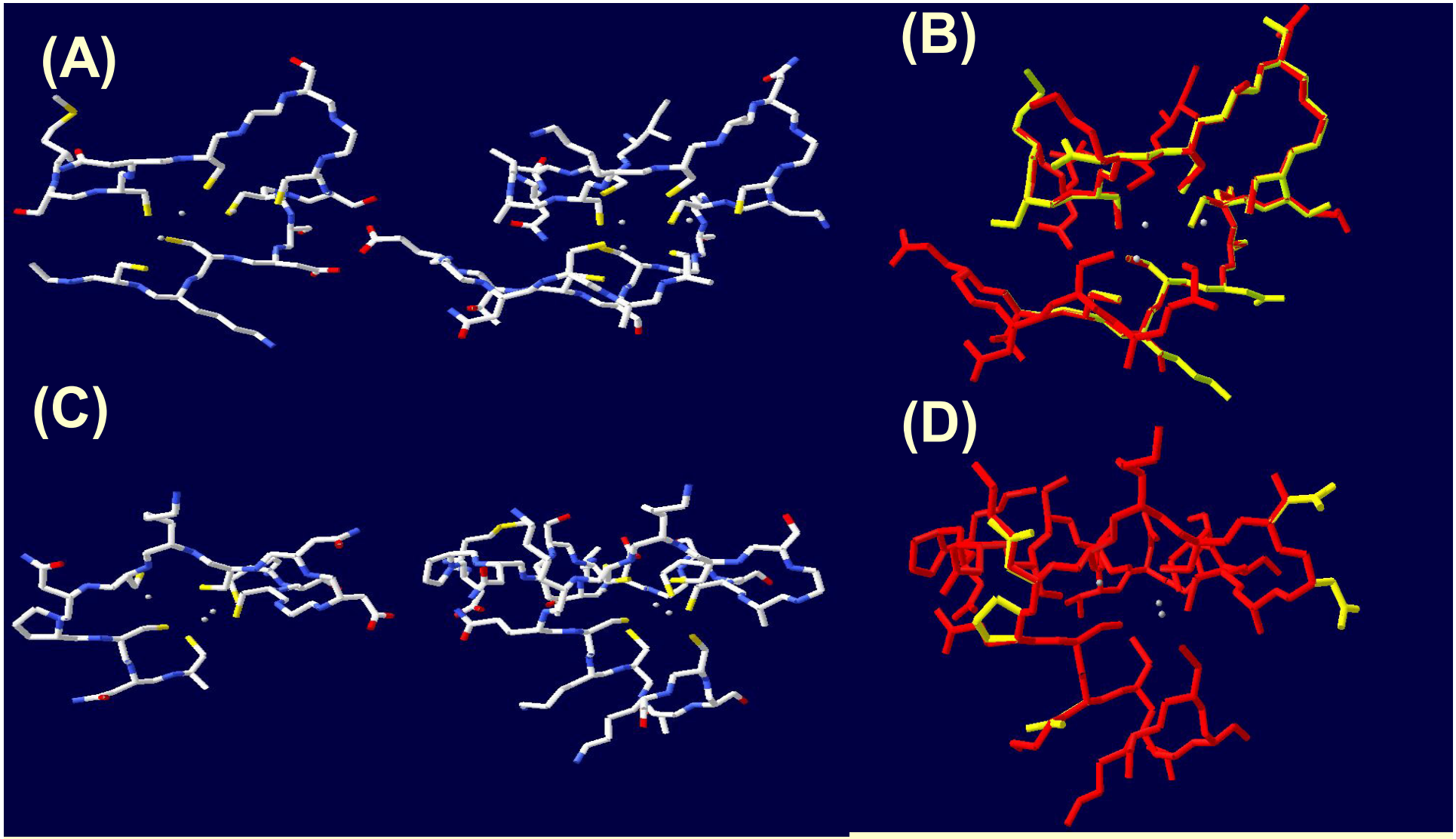
- Cystein residues in MT proteins form thiol bonds with metals,
- Cysteins are important and essential for the stabilization of the structure.
- In accordance with the alignment results metal-cystein distances were taken from 1QJL_A and 2MRT for homology modeling.





Best alignments are:

Sea Urchin MT beta domain (1QJL_A) for w-MT alpha domain.
Rat liver MT beta domain (2MRT) for w-MT beta domain.



(A) d-MT alpha domain and 1QJLA.
(B) Superimposed image of d-MT alpha and 1QJLA.
(C) d-MT beta domain and 2MRT.
(D) Superimposed image of d-MT beta and 2MRT.

Modeling of the w-MT hinge region

- There are no similar sequences within known MT structures.
- BLAST and FASTA searches in Protein Databank did not give an acceptable answer.
- The structure of the w-MT hinge region could not be modeled by using the same approach.

- Structure predictions were obtained through the
  "I-sites/Rosetta Prediction Server"

Predicted structures converge into two groups;

with larger loop regions that allow the metal centers to fold easily

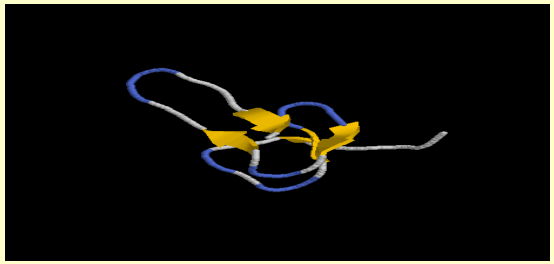
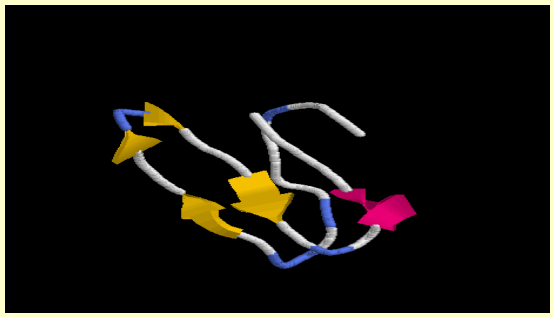
with beta sheets that result in rigid structures

STRUCTURE to FUNCTION ???

- Similarity searches through the fold recognition servers "3D-pssm" point to a possible protein-protein interaction function for the w-MT hinge region.
- The following are examples where conservation of hydrophobic residues (alanine, valine, proline) like those in w-MT hinge region are observed.

- w-MT hinge region sequence:

KMYPDLTEQGSAAAQVAAVVVLGVAPEN KAGQFEVAAGQSGE

Identity	e-value	Fold	Super family	PDB Id	Structure
24%	61.3	Protein binding	Trypsin inhibitor	1IW4	
18%	39	Small inhibitors, toxins	Leech antihemostatic proteins	1DEC	
16%	31	TNF receptor like	TNF receptor like	1D4V	
17%	48.9	Ribosome	Ribosomal protein 136	1KPJ	

CONCLUSIONS

- Small angle solution X-ray scattering is a powerful method for determination of molecular shape and folding features of biological macromolecules in native conditions.
- New experimental facilities and analysis methods allow elucidation of detailed structural features and provide the possibility to follow submillisecond conformational changes.
- Complementary use of bioinformatics tools, X-ray crystallography and scattering using synchrotron radiation facilitate a wide range of applications in studies of structure-function relationships.

- 3D structure of *A. thaliana* G-protein α-subunit has been modelled using computational (bioinformatics) tools.
- Modelling and sequence alignment results show that the two domain structure is valid also for plant G-protein α-subunits.
- Similarity in functional regions e.g. GTP recognizing and hydrolysis domains, indicate similar mechanisms for plants and mammalian systems.
- A new metallothionein has been identified in pasta wheat and the 3D structure is modelled using bioinformatics tools.
- Modelling of the hinge region indicate functional features that are different in plant MTs compared to mammalian MTs.
- Modelling results will be complementary X-ray crystallography and small angle scattering measurements.