IIT Guwahati

Lecture 24

Course BT 631

Protein Structure, Function and Crystallography

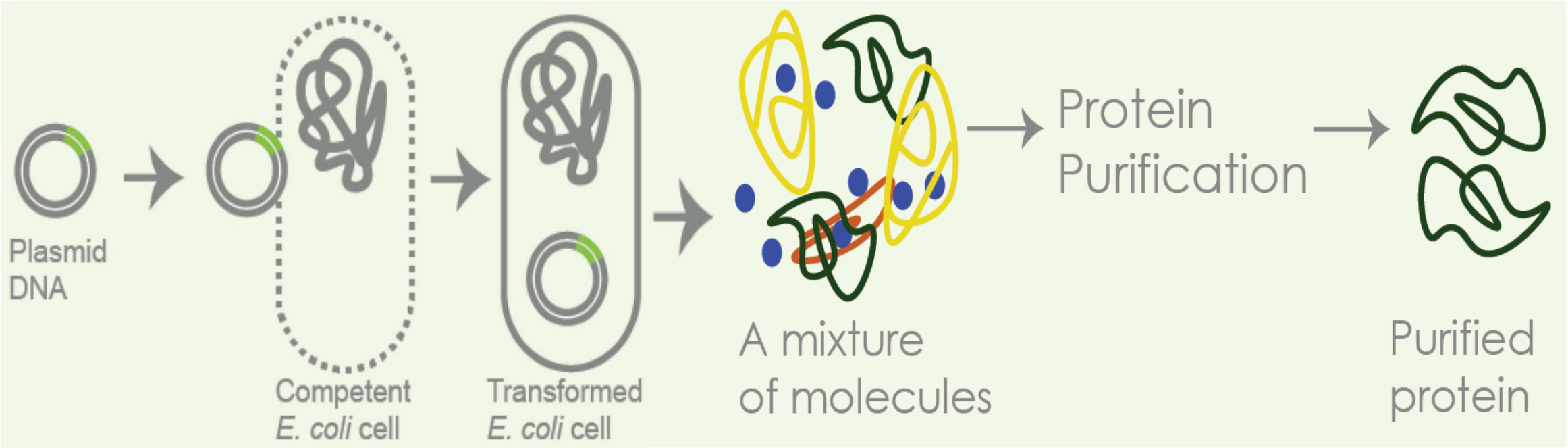
Prof. Arun Goyal

Dept. of Biosciences and Bioengineering

Protein Production for X-ray Crystallography

- The foremost requirement of any structure determination is availability of the protein under study in good amount and high concentration.
- Protein production requires a great deal of experimentation and simultaneous screening of various expression conditions.
- Initial phase of a protein production process requires gathering information about the protein or its related proteins that have already been studied.
- This initial phase includes designing of suitable primers and cloning.
- Efficient protein expression, purification is aided by recombinant DNA technologies by allowing addition of affinity tags (e.g. His_6-Tag).

Protein Production for X-ray Crystallography



- **Cloning of the gene construct into a suitable expression vector (pET-28a) followed by transformation of the vector into a protein expression host (*E. coli.* BL-21 DE3).**
- **The production phase involves over-expression of target protein in the host system.**
- **The production phase is concluded by isolation, purification and polishing of the purified protein followed by its concentration.**

Protein Engineering for X-ray Crystallography

Protein Engineering for Crystallisation

- The protein itself is the most crucial determinant for the success of any crystallographic study.
- The properties of a protein dictate, if it would be possible to express it in a given host system, purify and finally crystallise.
- To achieve this, several modifications may be required to produce a protein that expresses and crystallizes successfully.
- One of the most crucial property of a protein is its solubility in the host system.

Protein Engineering for X-ray Crystallography

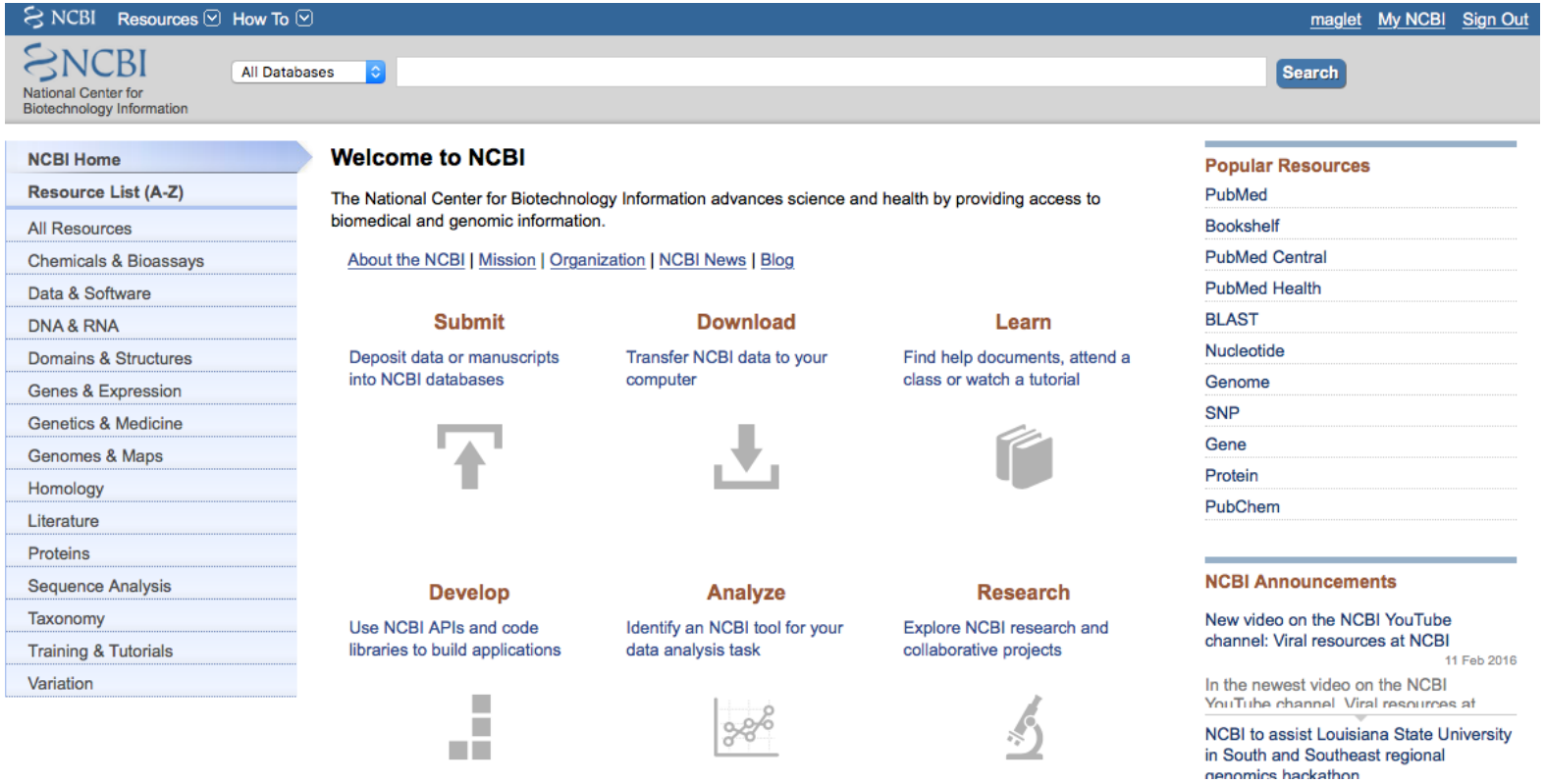
Protein engineering can be operated at the DNA level or the level of the protein.

- **At both these levels, the targeted protein design or combinatorial designs may be employed.**
- **Targeted protein design is based on the analysis of properties of the protein construct. For example, analysing a potential disordered termini in a protein sequence and removing it, is an example of targeted design.**
- **Combinatorial design is based on random mutagenesis and extensive screening.**
- **Targeted protein design can be combined with combinatorial design when trying to remove termini of various lengths.**
- **Both of these design strategies may be employed at either DNA level, by modifying nucleotide sequence OR at protein level, by chemical or biochemical modification of the expressed protein.**

Bioinformatic tools for X-ray Crystallography

Use of bioinformatics tools for Targeted design strategies at the DNA level

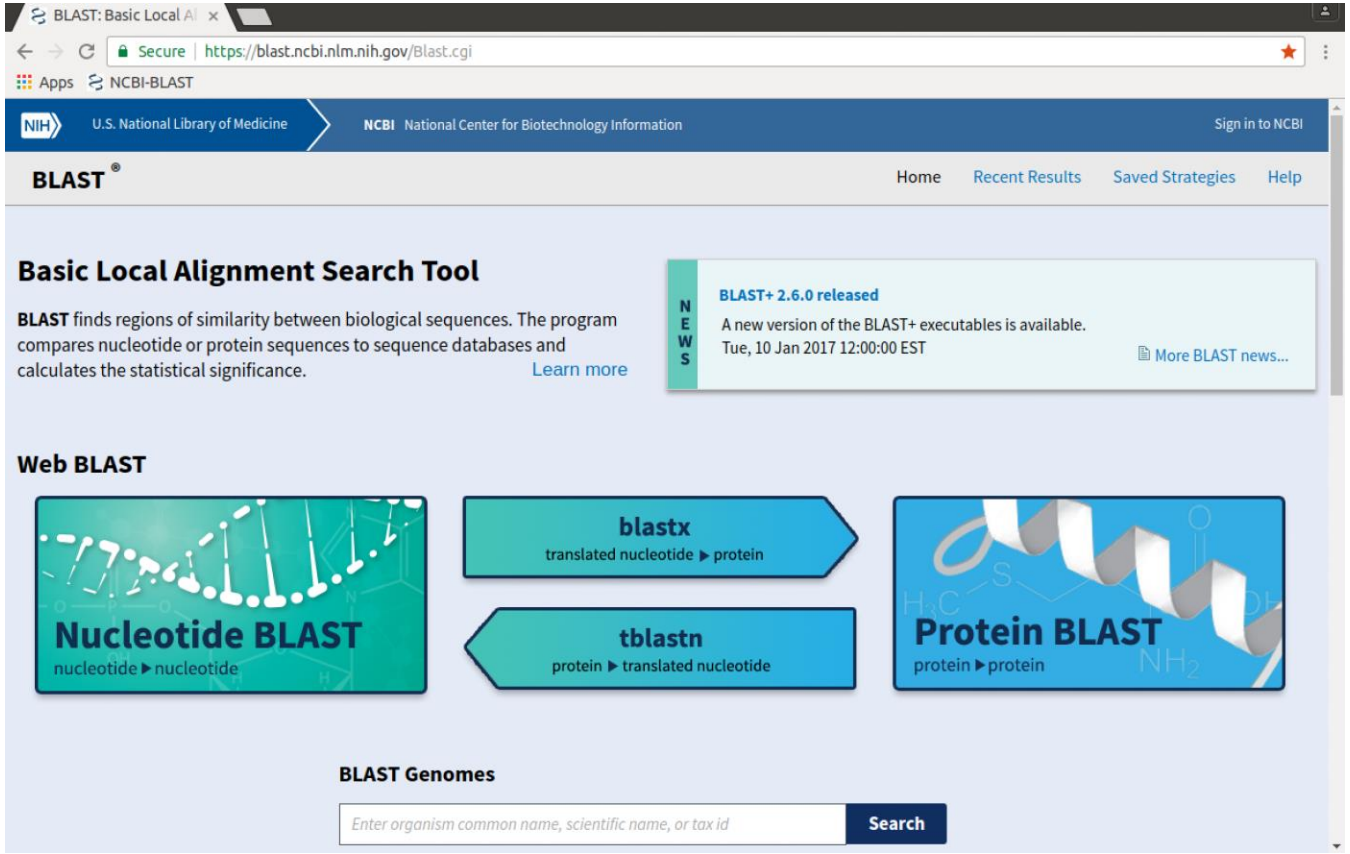
- The primary requirement of any protein structure study is the information about the primary sequence of the protein.
- A variety of bioinformatics tools have been developed for the analyses of protein sequences and other purposes. These can be accessed through Bioinformatics Links Directory Meta Server (NCBI, National Center for Biotechnology Information).



Bioinformatic tools for X-ray Crystallography

- Sequence analysis involves finding similar sequences. The similarity at the sequence level indicates structural relatedness to some extent. Thus it can be done by using the BLAST tool.

- In case a 3D model of a protein with similar sequence is available in the PDB, ***molecular replacement may be used while solving the structure from diffraction data***. The results from the BLAST may also be used to obtain a multiple sequence alignment (MSA) using the CLUSTALW2 program.



- Multiple Sequence Alignment can be used for identification of matching motifs or unidentified domains that can be separately expressed and studied. If the sequence analysis indicates a similar sequence, it is even possible to assign secondary structure elements to the query sequence using program Dictionary of Secondary Structures of Proteins (DSSP).

Bioinformatic tools for X-ray Crystallography

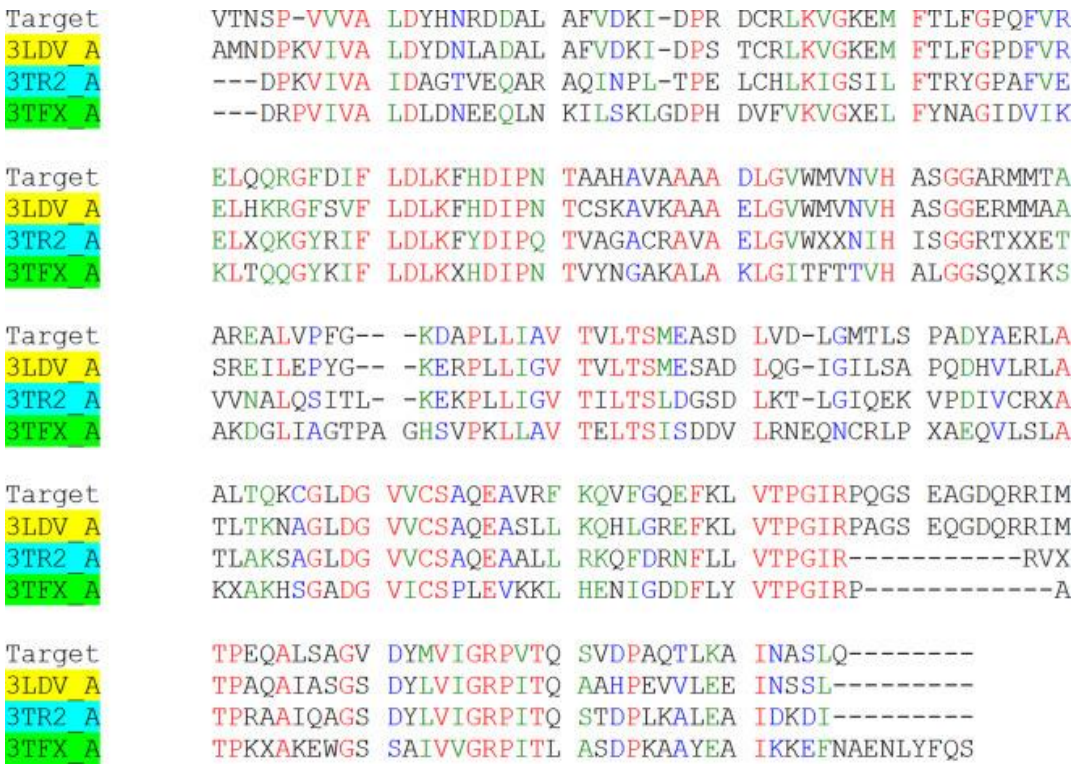
- **Multiple sequence alignment (MSA) may suggest some protein engineering approaches like**

  i) truncation of a flexible terminal or

  ii) expression of a domain separately or

  iii) modification of linker regions.

- **While dealing with a novel protein, when similarities are not detected, we may get the information from protein sequence like its**
  - **Globularity (using GLOBE),**
  - **Solvent accessibility and secondary structure prediction (SABLE -relative Solvent AccessiBiLitiEs of amino acid residues of protein, PSIPRED),**
  - **Dilsulphide bonds (DISULFIND),**
  - **Disordered regions (GlopPlot, PONDR),**
  - **Transmembrane regions and signal peptides (Phobius, SignalP).**



Multiple Sequence Alignment

Bioinformatic tools for X-ray Crystallography

Crystallisation propensity estimation

- It is always desirable to have a qualitative estimate of crystallisation propensity of a given protein, or how much, resource would be required. However, a crystallisation experiment can give the ultimate word.

- Balanced data sets including positive and negative results are important for predicting crystallisation propensity. High-throughput facilities sponsored by NIH-Protein Structure Initiative record failed trials for statistical analysis. However the data are biased towards prokaryotic and highly soluble proteins, but the other projects with high impact results, involved eukaryotic and membrane proteins.

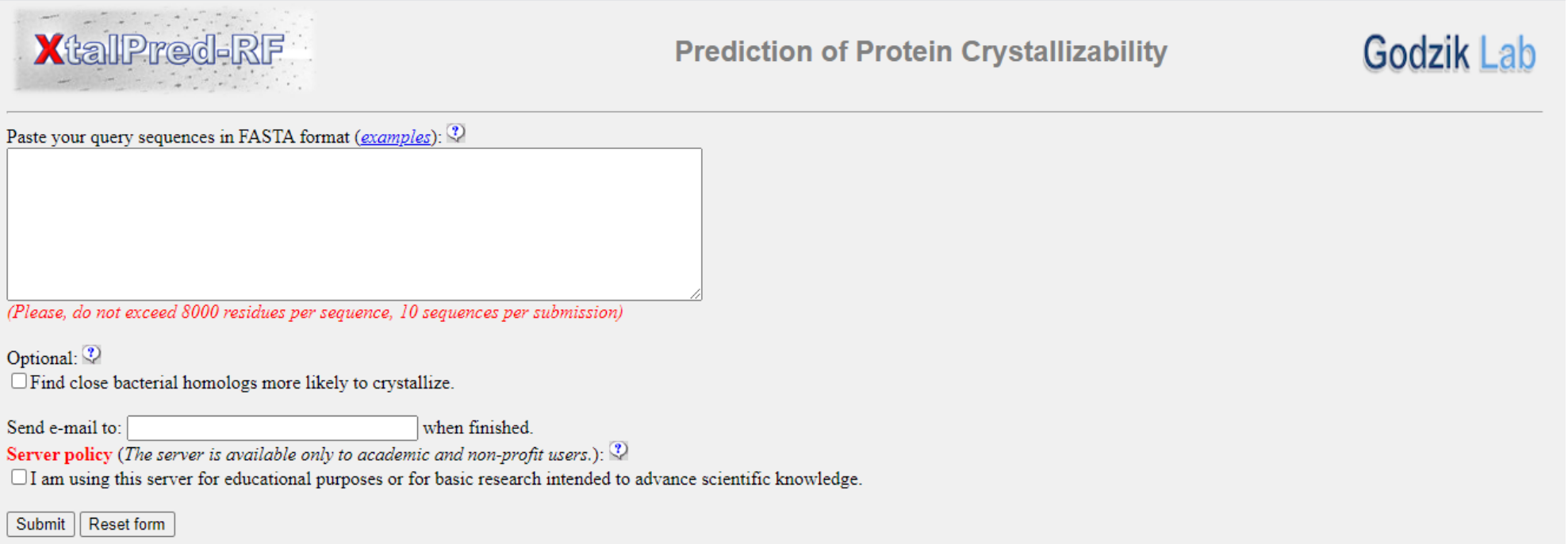
Bioinformatic tools for X-ray Crystallography

Crystallisation feasibility score (a web server for prediction of protein crystallizability)

The XtalPred server uses following information to calculate a 'Crystallisation feasibility score' (1 to 5)

i) It is based on Sequence derived information like sequence length, molecular weight, hydropathy index, instability index, isoelectric point and Cys, Met, Trp and Phe content.

ii) XtalPred can also give information about disordered regions, transmembrane regions etc.

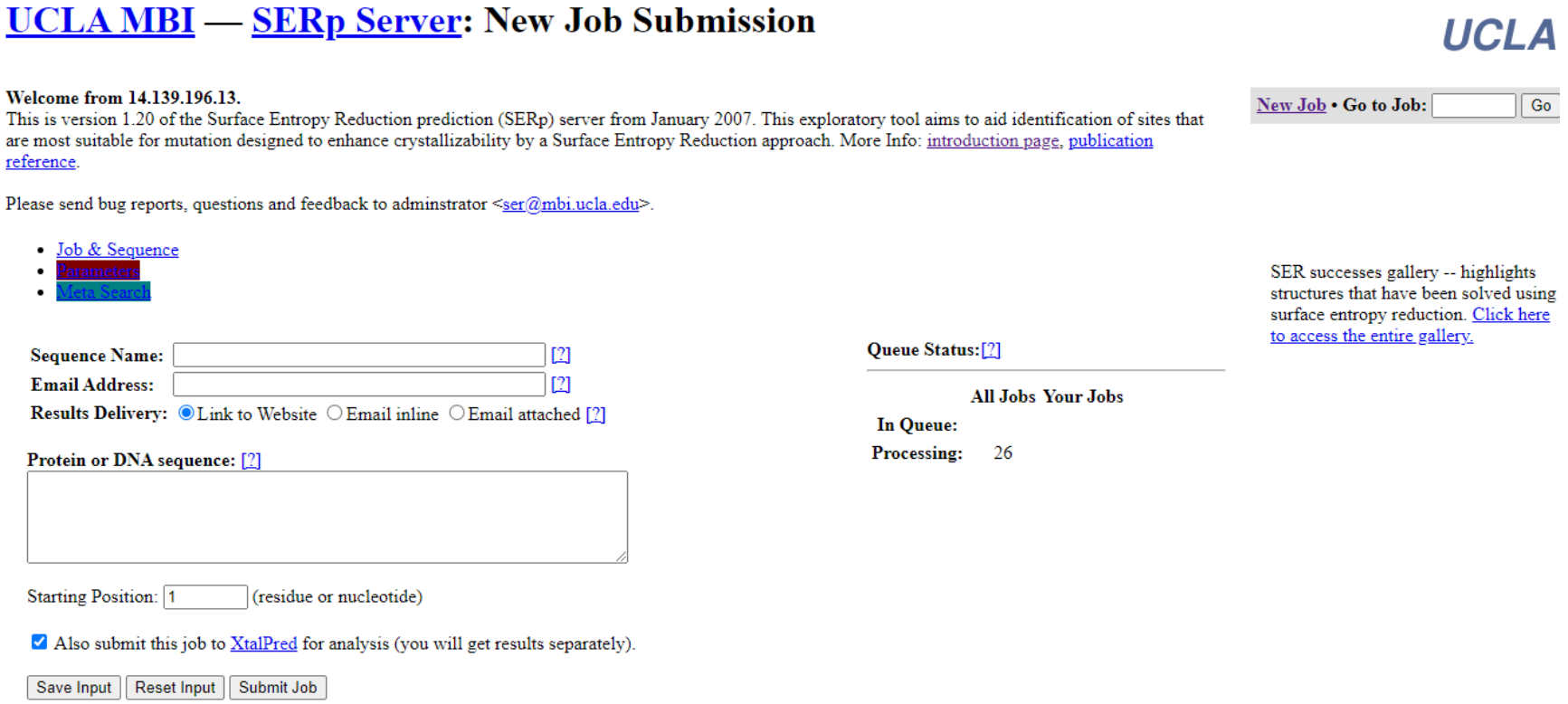


- XtalPred also searches for homologues that may have a better chance of crystallisation.

Bioinformatic tools for X-ray Crystallography

Surface Entropy Reduction (SER) method

SER is a site-specific protein engineering method, which identifies the long side-chain surface residues having high entropy like Lys, Glu, Gln. These residues may be mutated to short side-chain residues, such as Ala, Thr or Val. e.g. Organic hydroperoxide resistance protein (OhrB) structure has been solved by using SER method.



UCLA MBI — SERp Server: New Job Submission

UCLA

Welcome from 14.139.196.13.
This is version 1.20 of the Surface Entropy Reduction prediction (SERp) server from January 2007. This exploratory tool aims to aid identification of sites that are most suitable for mutation designed to enhance crystallizability by a Surface Entropy Reduction approach. More Info: introduction page, publication reference.

Please send bug reports, questions and feedback to adminstrator <ser@mbi.ucla.edu>.

New Job • Go to Job: Go

- Job & Sequence
- Parameters
- Meta Search

SER successes gallery -- highlights structures that have been solved using surface entropy reduction. Click here to access the entire gallery.

Sequence Name: [?]
Email Address: [?]
Results Delivery: Link to Website / Email inline / Email attached [?]

Queue Status:[?]

All Jobs Your Jobs
In Queue:
Processing: 26

Protein or DNA sequence: [?]

Starting Position: 1 (residue or nucleotide)

Also submit this job to XtalPred for analysis (you will get results separately).

Save Input | Reset Input | Submit Job

These conditional prediction methods pose a problem that they are sufficiently accurate at the extreme cases (score1, score 5). However, the word on the intermediate scoring proteins can be established only by setting crystallisation trials.

The targeted design at DNA level is commonly achieved by directional primer based cloning or site-directed mutagenesis.