Fellow Individual Report

Dipan Ghosh
BIGCHEM ESR-4

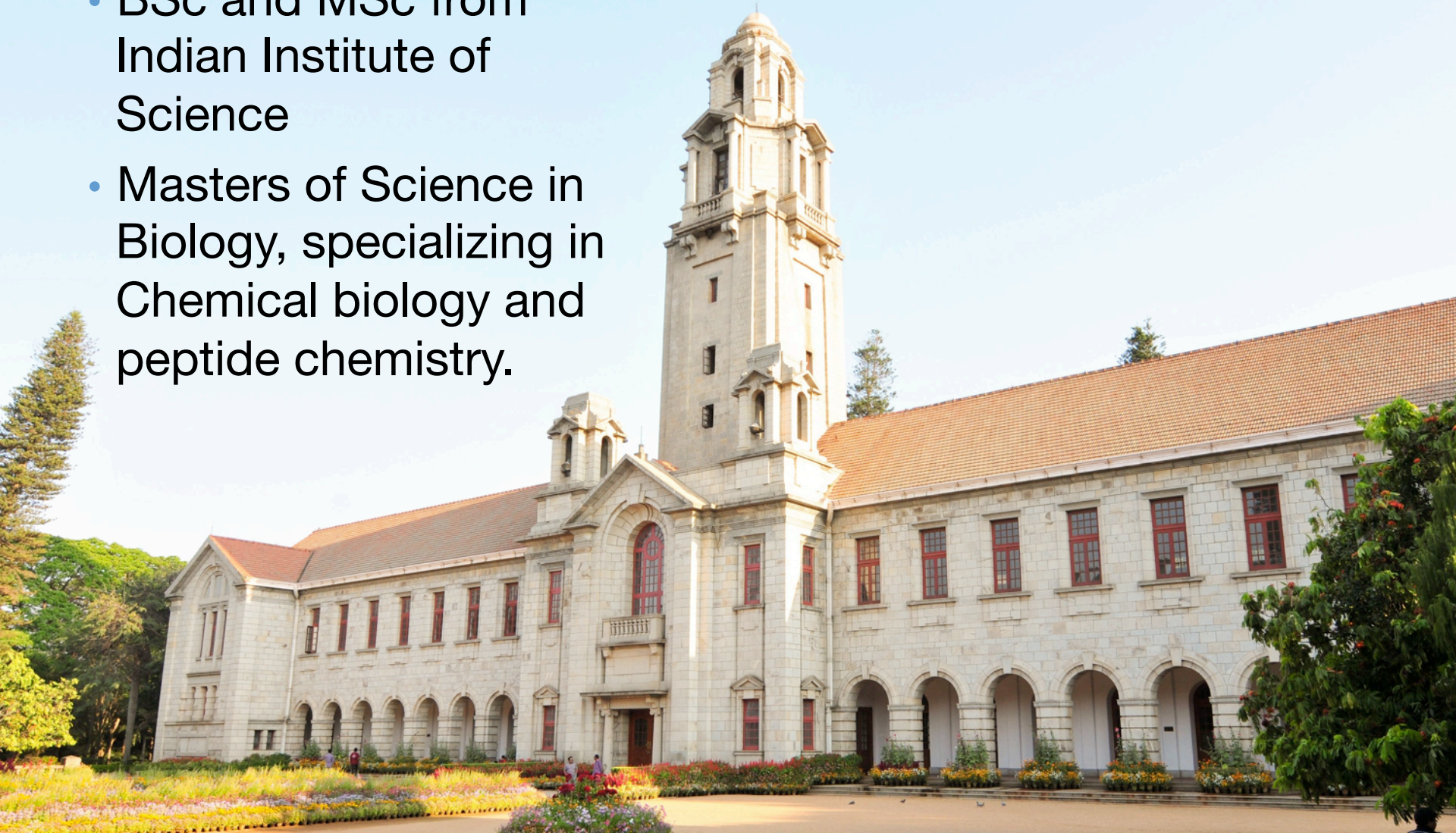
Dipan Ghosh

Indian

24 years old

Background

- BSc and MSc from Indian Institute of Science
- Masters of Science in Biology, specializing in Chemical biology and peptide chemistry.



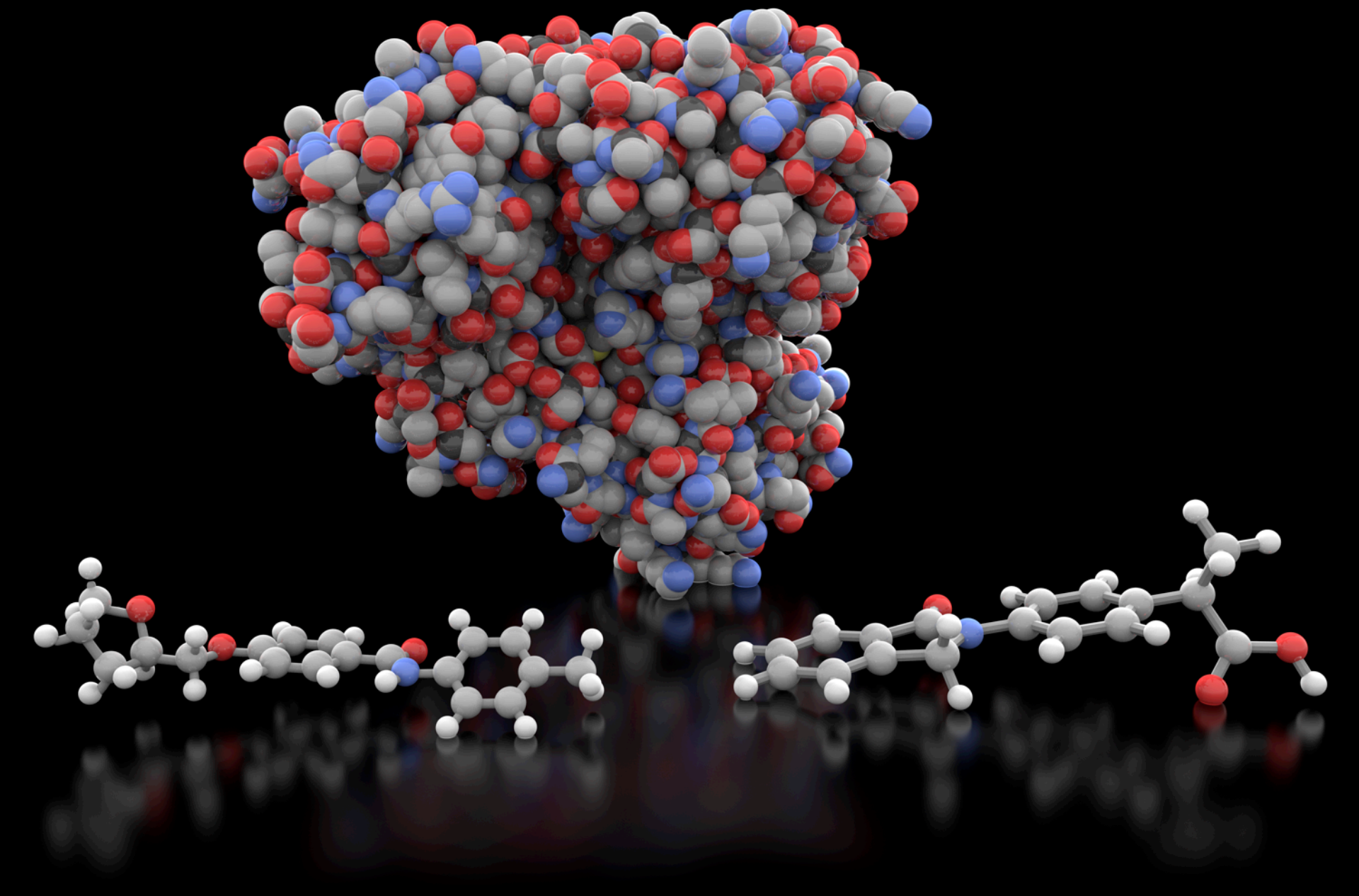
Training Experiences

- Participation in
  - two BIGCHEM Schools
  - BIGCHEM On-line courses
  - TUM kick-off seminar (3 days)
  - Monthly STB seminars and HMGU lectures
  - Triennial Congress of the World Association of Theoretical and Computational Chemists (WATOC) 2017
  - German Conference on Chemoinformatics (November 2017)
  - 4th Workshop of the Cross Program Activity Structural Biology of the Helmholtz Association (November 2017)

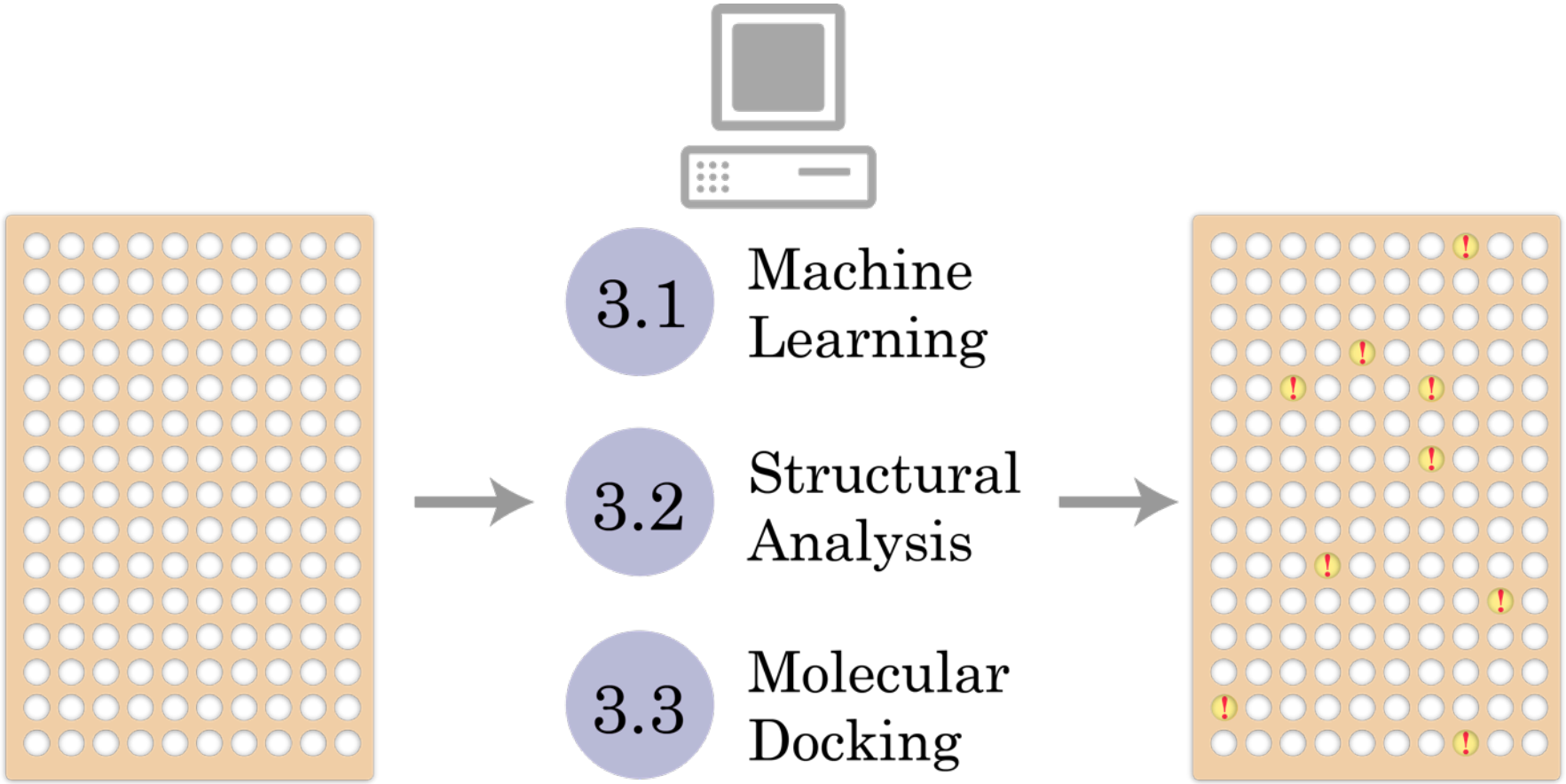
ESR4: Main Objectives

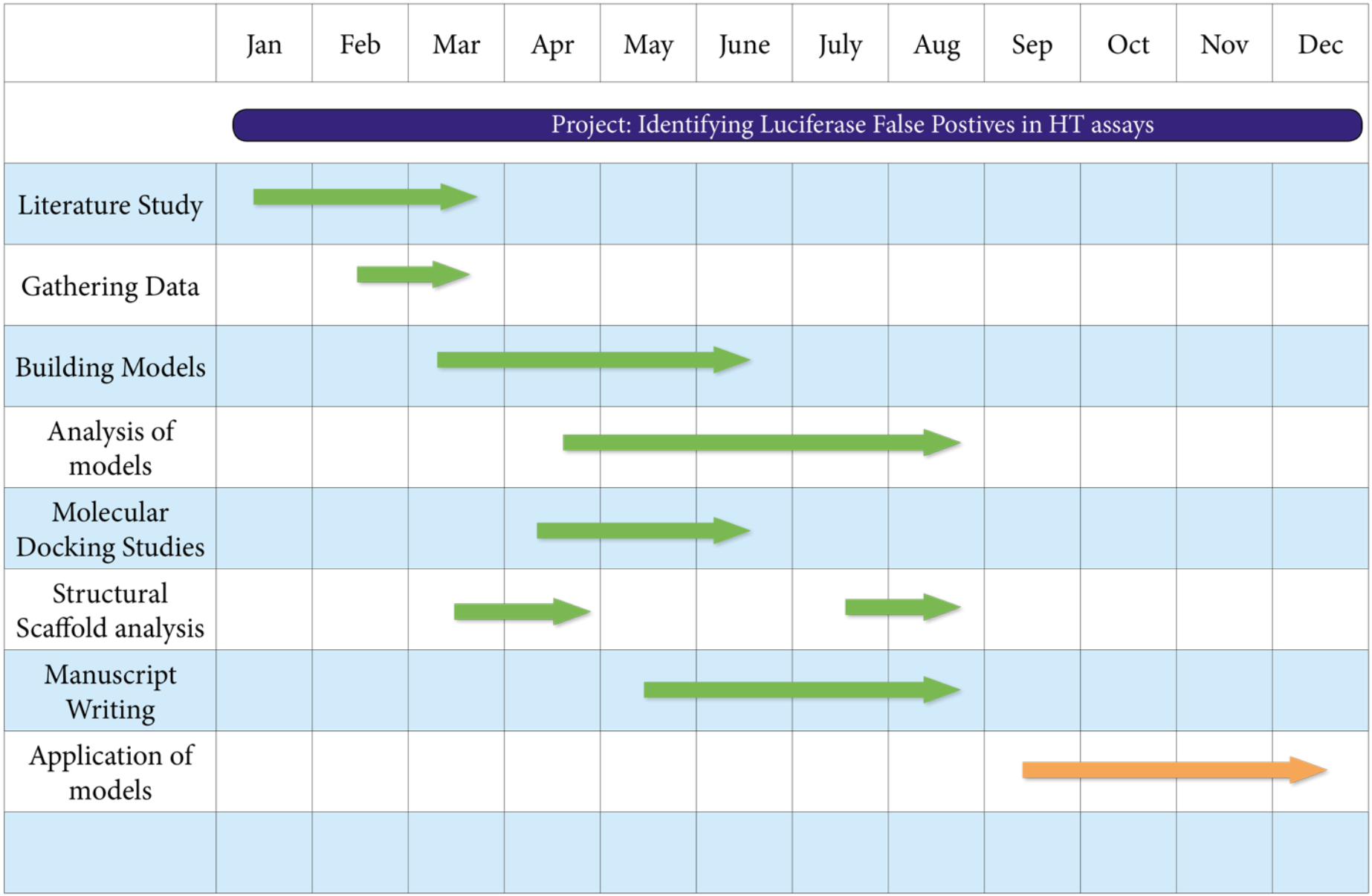
Primary objective of this project is to Development of frequent hitters filters for HTS screening.

Modelling False Positives in Luciferase HT Assay

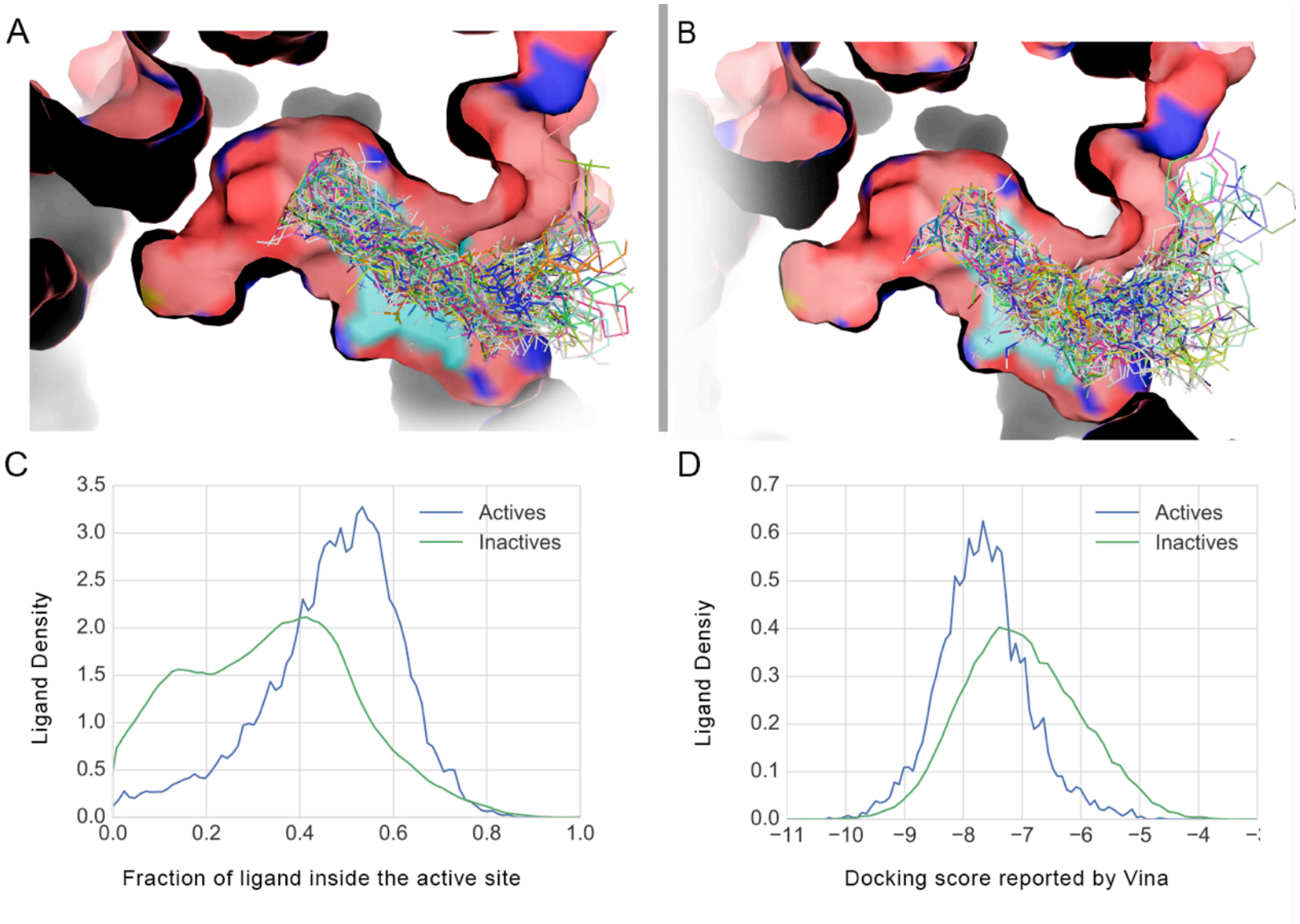


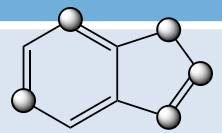
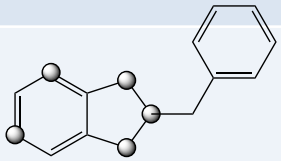
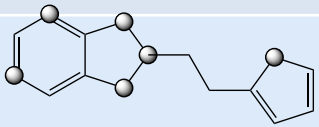
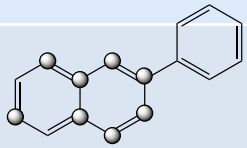
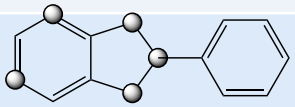
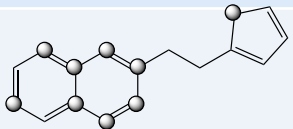
Work plan


3.1
Machine Learning
3.2
Structural Analysis
3.3
Molecular Docking

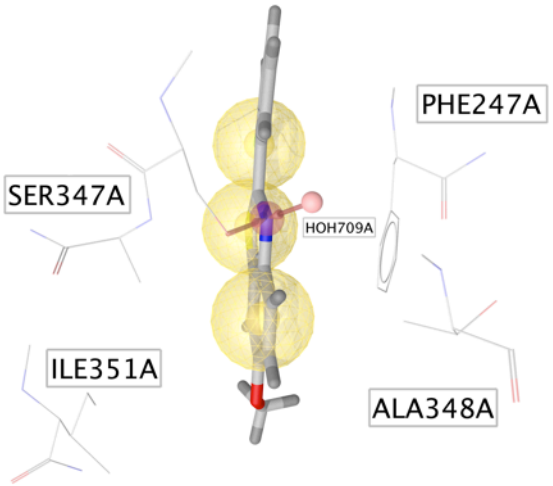

Jan
Feb
Mar
Apr
May
June
July
Aug
Sep
Oct
Nov
Dec
Project: Identifying Luciferase False Postives in HT assays
Literature Study
Gathering Data
Building Models
Analysis of models
Molecular Docking Studies
Structural Scaffold analysis
Manuscript Writing
Application of models

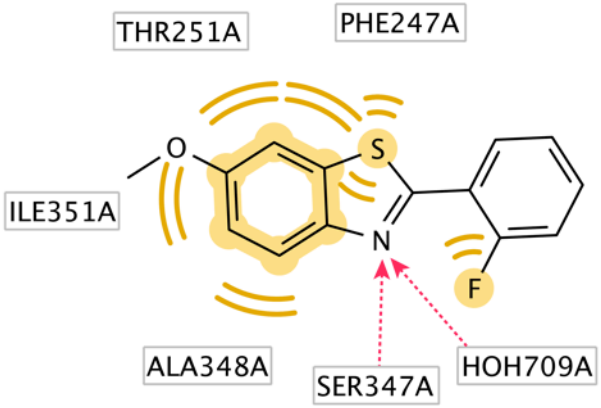
Docking Studies

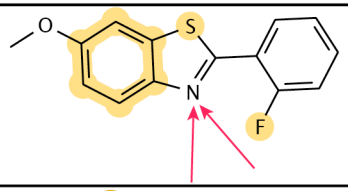
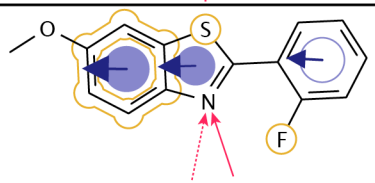
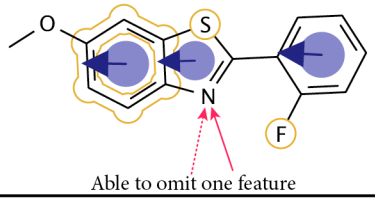


SMARTS Representation	Active Molecules Percentage Coverage	Inactive Molecules Percentage Coverage	Enrichment factor
	21.7	12.9	1.7
	4.5	1.1	4.2
	8.8	6.2	1.4
	0.7	0.1	10.1
	2.9	0.9	3.1
	6.8	4.1	1.7
	6	0.8	7.6
	0.2	0.1	1.9

Scaffold Analysis





Pharmacophore Representation†	Actives (%)	Inactives (%)
	8.2	3.5
	33.5	15.6
 Able to omit one feature	74.2	24.7

Models

Descriptor	Balanced Accuracy (%)		
	Set 1	Set2	Set3
CDK	83.5 ± 0.9	84.3 ± 0.3	88.0 ± 0.1
Dragon6 (blocks: 1-29)	83.7 ± 0.8	83.6 ± 0.3	88.1 ± 0.1
ALogPS, OEstate	81.3 ± 0.9	81.5 ± 0.3	86.6 ± 0.2
Fragmentor (Length 2 - 4)	81.3 ± 0.8	82.7 ± 0.4	87.7 ± 0.1
GSFrag	79.5 ± 0.9	80.7 ± 0.4	85.8 ± 0.2
Mera, Mersy	82.1 ± 0.8	81.8 ± 0.4	84.3 ± 0.2
Chemaxon Descriptors (7.4)	81.2 ± 0.8	81.8 ± 0.3	85.3 ± 0.2
Inductive Descriptors	78.1 ± 0.9	78.8 ± 0.4	80.7 ± 0.2
Adriana	85.1 ± 0.8	83.4 ± 0.3	86.7 ± 0.2
Spectrophores	78.1 ± 0.9	77.4 ± 0.4	78.4 ± 0.2
QNPR	79.3 ± 0.9	80.2 ± 0.4	85.4 ± 0.2
Structural Alerts	73.0 ± 1.0	72.7 ± 0.4	79.1 ± 0.2
SIRMS	78.1 ± 0.9	81.1 ± 0.4	85.3 ± 0.2
Consensus Model	86.2 ± 0.7	86.4 ± 0.3	89.7 ± 0.1

Balanced Accuracy = 0.5 * (TP / (TP + FN) + TN / (TN + FP))

Model profile

Statistical parameters, tables, charts - all the information related to the model.

Overview | Applicability domain

Model name: Consensus Luciferase Inhibition [rename] , published in High-accuracy models to predict false positives hits in Luciferase HTS Assays
Public ID is 697

Predicted property: **Luciferase_Inhibitory_Activity** modeled in CLASS
Training method: Consensus

4 individual models:
ModelID: 696
ModelID: 693
ModelID: 695
ModelID: 694
AVERAGE

No validation

Size: 1 Kb

Data Set	#	Accuracy	Balanced Accuracy	MCC	AUC
Training set: CombinedSet3Base	323435 records	88.21% ± 0.06	89.7% ± 0.1	-0.246 ± 0.005	0.957 ± 0.01

Show ROC curves

Real↓/Predicted→	Active	Inactive	Hit rate
Active	9822	933	0.913
Inactive	37202	275478	0.881
Precision	0.209	0.9966	
Training (Original)			

Applicability domain analysis

Model profile
Statistical parameters, tables, charts - all the information related to the model.

Overview | Applicability domain

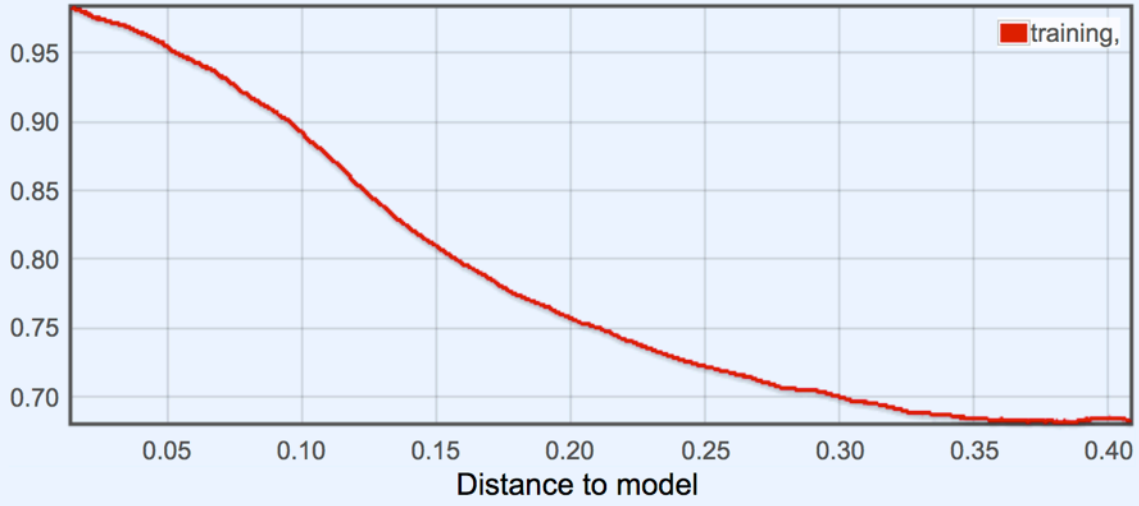
Williams plot with *CONSENSUS-STD* used as a distance to model.

Distance to model
CONSENSUS-STD

Averaging type
Default

Window size:
10% of the set

X axis
CONSENSUS-STD



APPLY THE MODEL TO NEW COMPOUNDS

Future developments

- AlphaScreen FHs
- FH for fragments

- Secondments:
  - AZ – January 2018
  - BI – Spring 2018

Acknowledgement

Dr. Igor Tetko

Barbara Gasset

Dr. Ekaterina Ratkova

Michael Withnall

Dr. Pavel Carpov
Zhonghua Xia

Dr. Uwe Koch
Dr. Kamiar Hadian