AMR-TB Methodology

Isoniazid–catalase molecular complex structure (WP1) ***Lead***: SPbRI.

The process of isoniazid (INH) ligand interaction with catalase protein is the key and most important mechanism that defines sensitivity of mycobacterial strains to INH. We will, therefore, devote a whole workpackage for the molecular genetic manipulations of *Escherichia coli* to recombinantly express wildtype and mutant *Mycobacterium tuberculosis* KatG orthologues that can be purified, complexed with INH and explored by structural biology techniques (including X-ray crystallography). This will provide a deeper mechanistic understanding of how resistance-conferring mutations in KatG lead to INH antibiotic resistance.

We will first, isolate the genomic DNA from samples of drug-resistant TB held in the unique collection (more than 100 clinical strains) at SPbRI (St Petersburg). The KatG genes of these isolates will be amplified by polymerase chain reaction (PCR) and cloned into a pET expression system for recombinant protein production in *E. coli.* The KatG proteins will be purified by immobilised metal affinity chromatography (IMAC) exploiting a polyhistidine tag. This will be performed by a team of biochemists from KBI (Kiev) during a series of secondments to SPbRI.

After preparing the samples of INH-KatG complexes we will perform X-ray measurements to deduce the molecular structure of the complexes. This will be done in the crystallography research centre at SPbU (St Petersburg) who possess state of the art X-ray crystallography equipment and highly experienced research team.

Originality: the collection of drug-resistant TB samples is unique, nobody else in the world has access to such set of mutants; the database of molecular structures from this collection, that will be built in the project, will put EU research in AMR to the leading positions in the area of TB related diseases.

Molecular Dynamics modelling of ligand-protein binding in isoniazid–catalase complex (WP2) ***Lead***: Aston.

We will conduct high performance Molecular Dynamics simulations using UK resources (ARCHER national supercomputer) as well as Japanese partner of the project RIKEN for simulating the interaction between INH and bacterial catalase. Using the measured atomistic structures of the INH-KatG complexes as initial approximation we will perform a series of simulations to investigate the molecular mechanisms of INH interaction with KatG. Qualitative data will be obtained for calculating the affinity constant of the INH ligand.

The main difficulty in computer modelling of biomolecular processes involving several molecules (the catalase and the ligand in our case) is the number of atoms in the model, which is very large. This is because the surrounding water should be included in the system explicitly as water molecules, (not as a continuum), and the number of water molecules can reach 95% of the number of atoms in the whole system, especially if processes of diffusion and reciprocal movements of the molecules before the actual interaction takes place are important. It has been recognised that the water surrounding biomolecules plays the major role in controlling the dynamics, emphasising certain motions whilst restricting others. It needs to be considered as an integral part of the molecular system [4,5]. Therefore, effective ways of reducing the number of water molecules, while preserving the quality of the biomolecular model need to be developed. Recently a number of methods that address this problem have been suggested, including hybrid approaches that seamlessly combine atomistic (MD) with hydrodynamic (Computational Fluid Dynamics, CFD) simulations. In Dmitry Nerukh (Aston) group we developed an original methodology in this framework that allows significant reduction of the number of atoms in the system without sacrificing the atomistic details in critical regions of the system (Hybrid Hydrodynamics Molecular Dynamics, HHMD). We have formulated the theory, implemented it in the popular Molecular Dynamics package GROMACS, and demonstrated its effectiveness for various systems, including a small virus PCV2 in water [6-11].

Originality: extensive MD modelling of the structures obtained in WP1 will elevate the structural biology information to the next level by supplying the artificially “frozen” experimental structures with realistic dynamical information at physiological conditions, that will ultimately allow much more accurate estimation of INH affinity; there are very few examples of such modelling in the literature to date.

Analysis of the dynamical couplings between catalase residues with elastic network models (WP2) ***Lead***: Madrid.

An attractive complement to Molecular Dynamics is represented by structure-based models that adopt simplified force fields only grounded on the experimentally known native structure. In particular, Elastic Network Models (ENM) allow for predicting the native dynamics of proteins without the need of lengthy simulations and detailed force fields, and it has been shown through a long series of works that their low-frequency normal modes describe collective movements strongly related with the functional dynamics of proteins. The Madrid group has recently developed an ENM in torsion angle space (TNM) whose normal modes strongly correlate with observed conformational changes, with structural ensembles observed in NMR experiments, and with crystallographic fluctuations. Through this model, we shall predict the dynamical couplings between residues involved in the interaction with INH, identifying key residues whose coordinated motion allows for the binding of the ligand and whose perturbation through mutation can destabilize the complex, and we shall compare our predictions with the properties of the experimentally studied mutants [12-15].

Originality: relating functional dynamics of proteins to INH binding is new and very promising.

Physical chemistry of isoniazid–catalase complex (WP2) ***Lead***: Messina.

The relationship among activity, flexibility and stability of isoniazid–catalase complex will be carried out by the synergistic use of complementary spectroscopic techniques. The following activities are proposed in the framework of the project:

- Characterization of the conformational changes as a response of the thermal and osmotic stress by Photon Correlation Spectroscopy (PCS) and Small Angle Neutron Scattering (SANS).

- Study of the correlation between dynamical properties and flexibility. For this study several spectroscopic techniques will be employed, such as Elastic Incoherent Neutron Scattering (EINS), which allows to determine the mean square displacement inside biomolecules, which is related to flexibility, and the macromolecular “resilience”, corresponding to an effective force constant associated to the structural stability and to the forces that maintain the folded structure and protein activity, and Quasi Elastic Neutron Scattering (QENS), which gives access to the correlation times of diffusion motions.

- Determination of the intermolecular and intramolecular hydrogen bonded network under stress conditions by IR absorption, Raman spectroscopy and Inelastic Neutron Scattering (INS), which allows to characterise the vibrational modes and the strengthening/weakening of the hydrogen bonded network. Such results will be integrated with the data obtained on flexibility.

- Characterization of biomolecular activity as a function of salt concentration and temperature, particularly at very low and very high temperature values.

Originality: an extensive experimental investigation of the unique collection of mutant isoniazid–catalase complexes will validate the results of structural measurements and modelling, the task has not been accomplished before for such an extensive collection of samples.

Biochemistry of KatG (WP3) ***Lead***: KBI.

The biochemistry of KatG will be studied in Kiev (KBI) and it will include: purification of the KatG proteins by affinity chromatography with catalase monoclonal antibodies and KatG specific inhibitors with the high affinity to enzyme as possible ligands; Proteins characteristics of recombinat KatG: PAAJE (poly acryl amid gel electrophoresis), w.blot and mass-spectrometry analysis for defining homogeneity; spectroscopy analysis for NADPH contamination detection determination of catalytic activity KatG (i) first order kinetic constant of reaction H_2O_2 decomposition and Km; (ii) AC 50% for KatG enzymes inhibition by isoniazid and other KatG-inhibitors; Crystallization of KatG proteins.

Isoniazid (INH) is the main drug used for TB treatment, because it interacts with the bacterial catalase that leads to the bacterial death. It is generally accepted that INH activates KatG resulting in a cascade of reactions that produces isonicotinic acid, which attacks harmful species, nevertheless the detailed molecular mechanism of stacking of INH and KatG is unclear as well as the accurate pathway of metabolic transformation of INH to isonicotinic acid.

We plan to attack this problem as follows: i) to perform metabolomics investigation of M. tuberculosis cells exposed to isoniazid (Surrey); ii) to evaluate the fitting of kinetic constants using the available experimental data for the time evolution of concentrations using the methodology developed by the participants of the project [16]; iii) to supply this networks by the genetic and structural information taking into account possible staking sites of KatG and its mutation that should result in the description suitable for the modelling using Boolean networks approach, and evaluate the corresponding simulations.

As a result, the constructed combined ODE-Boolean system should make it possible to reproduce the dynamics of the target process in silico and be optimized to reach conditions overcoming AMR.

Originality: the quantitative biochemistry of KatG is the key for understanding the mechanisms of INH treatment which, in combination with the extensive knowledge of various mutants, will provide a unique opportunity for suggesting new effective TB drugs overcoming AMR.

Modelling of drug metabolism (WP3) ***Lead***: SPbU.

Mathematical modelling (including detailed enzyme kinetics) of the drug metabolism will be conducted in Saint-Petersburg Research Institute Phthisiopulmonology, Saint-Petersburg State University and Kursk State University and at the University of Surrey using an in-house code written in conventional languages of technical computing and based on metabolic models of M. tuberculosis already constructed at Surrey [17]. One of the goals of such approach is to provide to users, involved into the practical laboratory investigations, a tool that can work on standard desktops and realise on-the-flight analysis. The mechanism of action of INH is very complex and involves several different concepts. The pro-drug INH is activated by catalase-peroxidase (KatG) to get active INH products that leads to the activation of following enzymes cascade and bacillus death. Also the mutation in the KatG gene is the major cause of INH resistance, therefore mathematical modelling will be focused on the detailed kinetics mechanism of involved enzymes in susceptible and specific resistant strains (obtained from patients in SPbRI). To describe real KatG kinetics we will use biochemical data obtained in Kiev (KBI) to compare kinetics of KatG in susceptible and resistant strains.

Our task is also to study how the full metabolism of Mycobacteria is changed in different strains, for this we plan to use the Boolean networks methodology. Its advantage is based on the principal advantage of this approach: the fast binary output, which depends on the topology of links in a metabolic network and character of interactions (activation/inhibition) between their nodes. Thus, it allows fast estimation of cascades actually acting in different strains. In addition to this coarse level of approximation, these conclusions will provide a background for reducing dimensionality of the further detailed kinetic ODE-based models aimed at adjusting the drug influence on the specific controlling parameters.

The complex modelling will allow us to elucidate the biochemical mechanisms operating behind INH action and would enable better detection of INH resistance. This information will aid novel drug design strategies.

Originality: the application of powerful mathematical modelling techniques to TB treatment with ultimate goal of providing end users with effective software for studying the bacterial metabolism.

Synergy between drugs (WP3) ***Lead***: Surrey.

TB is normally treated, yet resistance emerges, with synergistic drug combinations that should prevent resistance as the mutation rate to resistance to multiple drugs should be extraordinarily low. To understand how resistance develops from multi-drug regimes, we have developed a model at Surrey, with the non-pathogenic relative of M. tuberculosis, Mycobacterium smegmatis (MSm), exposed to sub-MIC levels of isoniazid + rifampicin in which mutants arise are resistant to both antibiotics at sub-MIC concentrations. We propose this scenario may represent the first step in development of multi-resistance, particularly in patients who are poorly compliant or in lesions in which antibiotics fail to penetrate effectively. However, as patients continue treatment or a strain is transmitted to another patient, then further mutations will inevitably arise leading to full clinical multi-resistance. We propose to investigate this scenario by sequencing the multi-resistant MSm strains we have already generated and monitoring their subsequent acquisition of additional resistance mutations when exposed to higher levels of drugs. The experiment may be repeated with Mtb, if time permits.

Originality: a unique set of results of synergetic action of TB drugs directly leading to practical recommendations.

Modelling of fitness landscape (WP4) ***Lead***: Madrid.

The library of isoniazid resistant strains can be analysed to identify the relevant loci of mutations, e.g. on KatG; and associated fitness measures for individual strains will be used to construct a digital fitness landscape. In parallel, population dynamics approaches of random mutations on the KatG gene will be used to model the appearance of resistance and multi-resistance, and to predict other mutations that possibly confer resistance (and inversely, to predict sequence with low probability of resistance) or recover fitness. This will be done at Aston University, the University of Surrey and in Quito. For this, a model for the genotype-phenotype map will be developed, relating sequence information with catalase structure. Evolutionary simulations will also be performed in Madrid to investigate how folding stability influences the site-specific evolutionary rates. This genotype-phenotype map will be incorporated into the experimental fitness landscape. The analysis of the library, the construction of the fitness landscape, and the population dynamics approach are performed at Aston University, Surrey, in Quito and in Madrid.

The predicted substitution rates will be compared to site-specific substitution rates deduced from multiple sequence alignments through maximum-likelihood (ML) phylogenetic inference. In this way, by comparing observed and predicted substitution rates, we will identify positions that are subject to stronger-than-expected constraints because of functional reasons not taken into account in the evolutionary model, and we can compare them with the positions that are predicted to be key for the native dynamics of catalase described above.

ML inference, coupled with stability-aware molecular evolution models, also allows for reconstructing ancestral sequences in such a way that the bias in ancestral protein stability inherent in ML methods is minimized. The ancestral sequences that we shall reconstruct in this way are expected to provide a precious information, since it has been proposed that targeting with drugs ancestral sequences at the origin of a large phylogenetic group may help developing drugs with broader scope.

This genotype-phenotype map is incorporated into the experimental fitness landscape. The analysis of the library, the construction of the fitness landscape, and the population dynamics approach are performed at Aston University, in Quito and in Madrid [18-25].

Originality: the relationship between an extensive set of KatG mutations and fitness landscape will be done for the first time for TB bacteria.