

# EE4RPP: Project Overview

The purpose of the project is to make an intuitive yet powerful bioinformatics search engine which provides online access to a large dataset of protein isoelectric points which has been compiled by Aston University researchers and students over the course of several years.

## Background

Bioinformatics is a multidisciplinary field which uses computational methods to aid in biological research by creating systems for storing, organising and analysing complex biological data. Within this field there are many online databases categorising biological information at the molecular level, and one such purpose of these is for storing the functional and physical properties of proteins. Currently, no such database exists for one of the most widely-used, important, and useful properties of proteins: the isoelectric point (pI). An isoelectric point is the acidity (pH) at which a molecule carries no net charge; below the isoelectric point, proteins have a net positive charge, above it a net negative charge. Additionally, proteins are at their lowest solubility at their isoelectric point, and this makes the isoelectric point a vitally important property when both characterising and purifying proteins.

The dataset which has been compiled is a collection of entries stored as a non-relational table, and for each entry it records the name of the protein, its identity, origin, experimental conditions, its isoelectric point, and other pertinent data. There are also links to a heterogeneous collection of databases containing associated data, such as amino acid sequence, function, etc. A web-accessible database that warehouses this data and offers a robust and adaptable GUI for searching, viewing and downloading results would greatly increase the accessibility of the dataset.

## Objectives

1. To build a free (as in freedom) web application for viewing protein isoelectric points.
2. To produce a bioinformatics tool with real world value for future research.
3. The application should provide intuitive but powerful searching facilities.
4. The application should provide a convenient means for a certified user to edit and upload additional data.
5. The application should present information in a usable and efficient form.
6. Users should be allowed to download generated results for offline use.
7. Adequate security precautions should be taken to minimise the risk of data being sabotaged or stolen.
8. The implementation should use a clean model view controller architecture.
9. Comprehensive test coverage of the API and common use cases should be automated.
10. The application should be fully scalable for much larger datasets.