

M23D - MHC-II Model Database

Josef Laimer and Peter Lackner

University of Salzburg, Department of Biosciences
Hellbrunnerstraße 34, 5020 Salzburg/Austria



Introduction

MHC class II molecules are essential for initiating immune responses. Knowledge about their 3D structure is a potential key to understand their capabilities of binding antigen peptides or their interaction with T-cell receptors. M23D is a database of 3D models of MHC-II molecules, currently for all HLA-DR alleles listed in the IPD-IMGT/HLA database [1]. Its final version will contain data for all HLA class II alleles. M23D is indented for applications in antigen peptide binding studies or the analysis of fold stability.

Result

M23D provides experimentally determined structures, derived from the PDB database [2], as well as predicted models utilizing MODELLER [3]. All data are accessible via an easy to use web interface and a REST web service.

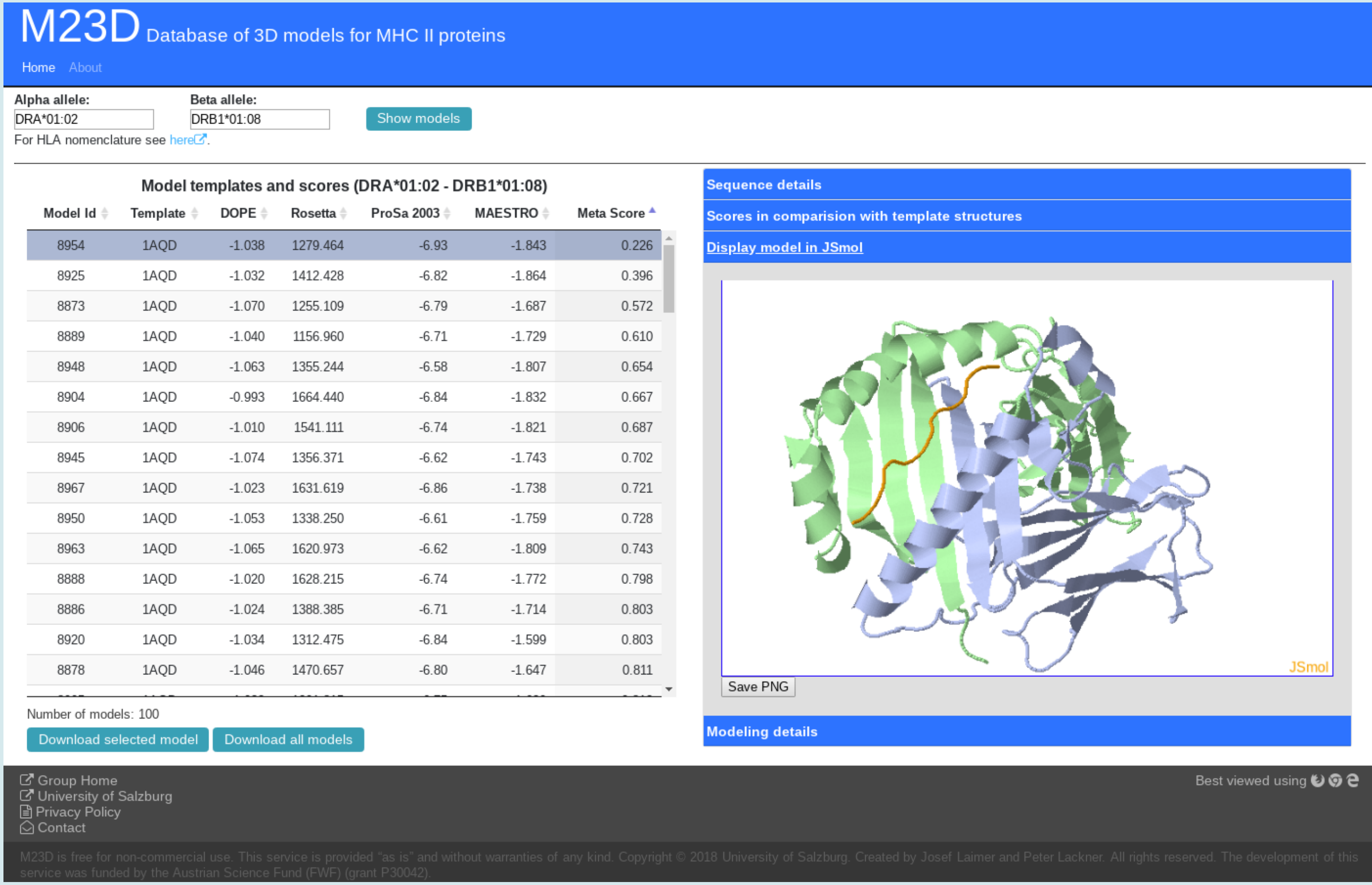


Figure 1: Screenshot of the M23D web interface.

After selecting an alpha and a beta allele, a list of at least 100 models is provided. More models are available if multiple templates were found in the modeling process (see below). All models are scored with a set of four widely used evaluation tools: DOPE [4], Rosetta [5], ProSa2003 [6], and MAESTRO [7]. In addition, a meta score is computed based on Stouffer's z-score method, which combines the aforementioned scores. An analysis of this meta score in comparison with the others is presented in figure 2.

Besides a list of models and their scores, search results include sequence details, including links to UniProt [8] and the IPD-IMGT/HLA database, a comparison of the model scores with those of template structures, modeling details including alignments between the allele and template sequences utilizing JSAV [9], and a 3D view of a selected model using JSmol [10].

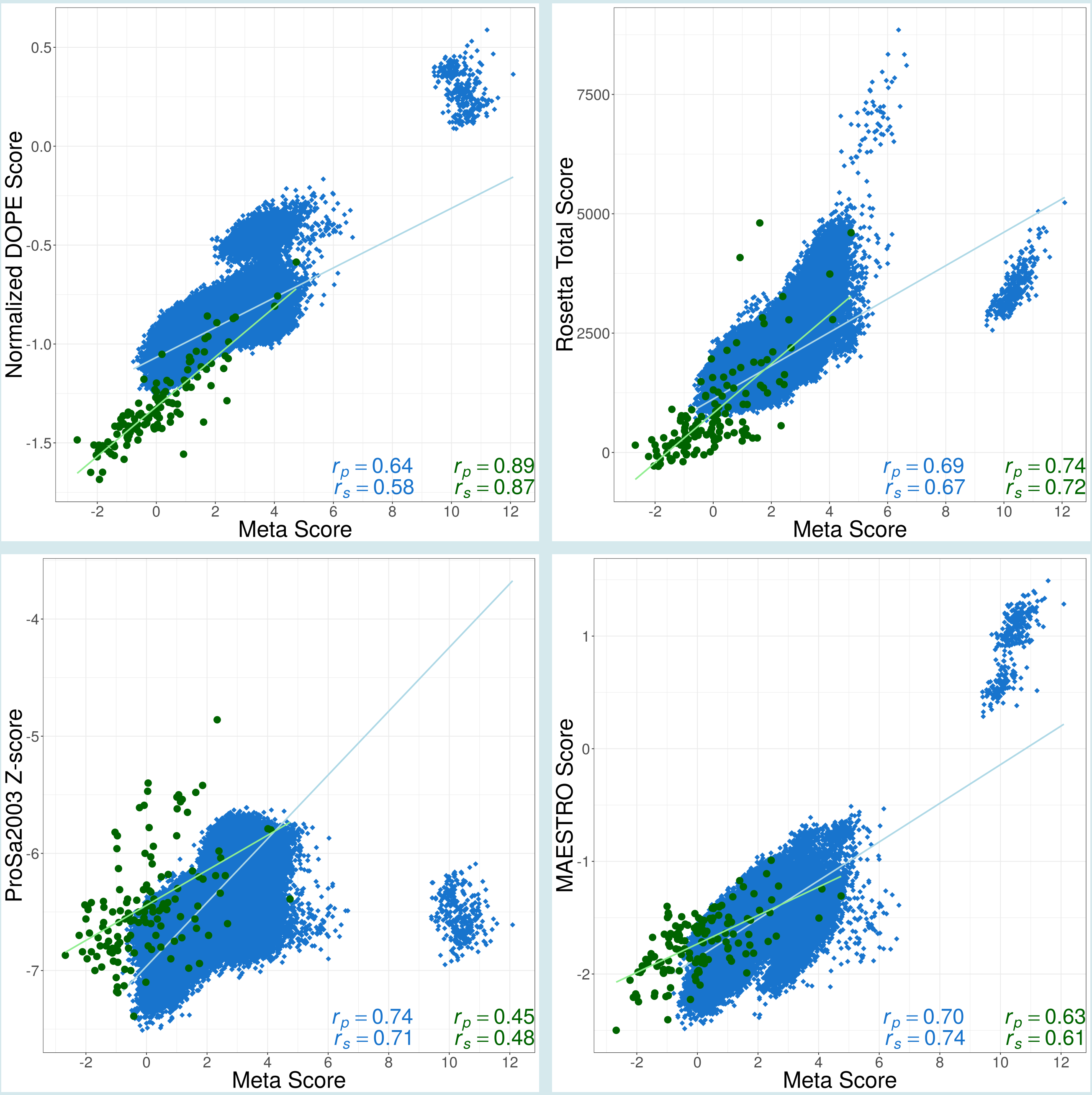


Figure 2: Correlations (r_p ... Pearson correlation coefficient, r_s ... Spearman's rank correlation coefficient) between scores computed by DOPE, Rosetta, ProSa2003, and MAESTRO and the combined meta score. Results for template structures are shown in green, those for modeled structures in blue.

Implementation

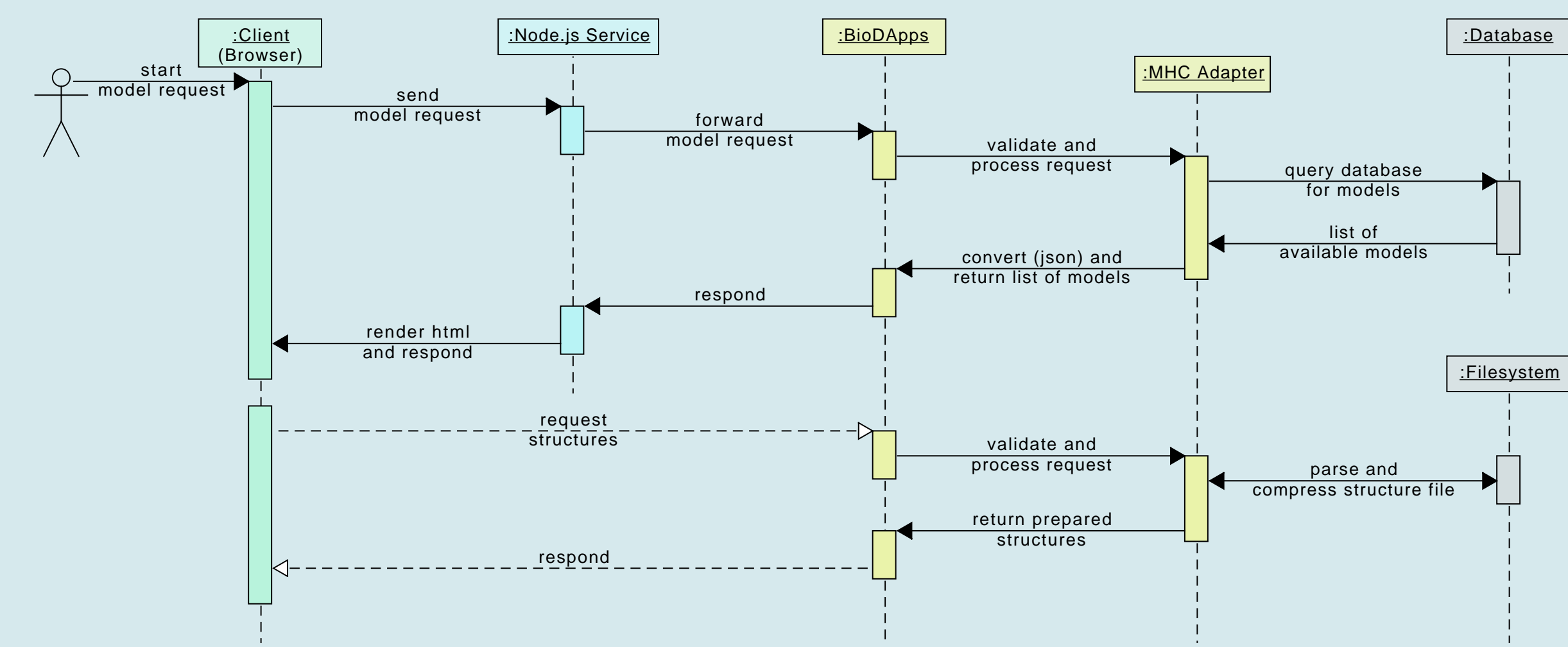


Figure 3: Module interaction.

M23D was realized with a combination of server-side (Node.js) and client-side JavaScript and utilizes our REST service BioDApps [11], a Python-based platform for accessing distinct biomedical tools and web resources. All data, except for model files, are stored in a relational database (MariaDB).

The actual web interface is realized with the Node.js framework, an asynchronous event-driven JavaScript runtime. The M23D instance is utilizing the EJS engine to render websites.

Modeling

MHC-II models are generated utilizing MODELLER, based on a set of 161 experimental determined template structures.

Allele templates are primarily selected based on their sequence identity. Further, the number of unresolved residues (gaps) in the possible template structure, missing atoms, and the resolution are taken into account. Templates are selected independently for each chain. If multiple, equally suitable, templates are found, models are generated based on each of the templates.

Conclusion and Outlook

M23D is a database for MHC-II models, currently including over 700000 models for all HLA-DR alleles listed in the IPD-IMGT/HLA database. Models for the remaining HLA class II alleles will follow soon.

All models include an alanine nonamer binding peptide as a placeholder. Future versions will provide the possibility to generate models with user-specified binding peptide sequences for selected allotypes. The final version will integrate a structure based prediction of potential binding sequences, which is the ultimate project aim and major purpose of the development of M23D.

M23D is free for non-commercial usage and can be accessed via an easy to use web interface as well as a REST web service.

References

- [1] **The IPD and IMGT/HLA database: allele variant databases.** Robinson *et al.*, 2015. <https://www.ebi.ac.uk/ipd/imgt/hla/>
- [2] **The Protein Data Bank.** Berman *et al.*, 2000. <https://www.rcsb.org/>
- [3] **Comparative protein modelling by satisfaction of spatial restraints.** Sali and Blundell, 1993. <https://salilab.org/modeller/>
- [4] **Statistical potential for assessment and prediction of protein structures.** Shen and Sali, 2006.
- [5] **The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design.** Alford *et al.*, 2017. <https://www.rosettacommons.org/>
- [6] **Recognition of errors in three-dimensional structures of proteins.** Sippl, 1993. <https://www.came.sbg.ac.at/>
- [7] **MAESTRO -multi agent stability prediction upon point mutations.** Laimer *et al.*, 2015. <https://biwww.che.sbg.ac.at/maestro>
- [8] **UniProt: the universal protein knowledgebase.** The UniProt Consortium, 2016. <https://www.uniprot.org/>
- [9] **Viewing multiple sequence alignments with the JavaScript Sequence Alignment Viewer (JSAV).** Martin, 2014. <http://www.bioinf.org.uk/software/jsav/>
- [10] **JSmol and the next-generation web-based representation of 3D molecular structure as applied to proteopedia.** Hanson *et al.*, 2013. <http://jmol.sourceforge.net/project/>
- [11] **BioDApps - Distributed Bioinformatics Applications.** Bernwinkler *et al.*, 2014. https://biwww.che.sbg.ac.at/wp-content/uploads/2016/05/biodapps_oegmbt2014.pdf