Rapid, qualitative prediction of antimicrobial resistance by alchemical free energy methods

Philip W Fowler[1,2]

[1]Nuffield Department of Medicine, John Radcliffe Hospital, University of Oxford, Headley Way, Oxford, OX3 9DU, UK

[2]National Institute of Health Research Oxford Biomedical Research Centre, John Radcliffe Hospital, Headley Way, Oxford, OX3 9DU, UK

The evolution of resistance to antibiotics was predicted by Fleming in his Nobel Prize speech and is now accepted as posing a threat to modern medicine requiring urgent and concerted action. Helping clinicians make appropriate treatment decisions by improving the coverage, portability, speed, accuracy and cost of species identification and drug susceptibility testing will be an important part of the solution. A promising approach is to sequence the genome of any infecting pathogen(s) found in a clinical sample and, by looking up genetic variants found in genes known to confer resistance to the action of antibiotics, return a prediction of the effectiveness, or otherwise, of a panel of antibiotics to the clinician. The exemplar for this approach is tuberculosis, partly because its growth rate is so slow that culture-based clinical microbiology can take up to two months to return a result to the clinician, and partly because its genetics is simpler than other pathogens and therefore the current second-generation sequencing technologies work well. Genetics clinical microbiology has been shown to be cheaper, faster and probably more accurate than traditional culture-based clinical microbiology for the drug susceptibility testing of tuberculosis[1] and, in addition, facilitates the rapid identification of epidemiological clusters. Although catalogues relating genetic variants to phenotype have been carefully and extensively developed, a potential weakness remains: such an approach is fundamentally inferential and so cannot make a prediction when it encounters a genetic variant not present in the catalogue, such as is the case for rare genetic mutations.

Alchemical free energy methods, a simulation method derived from classical statistical mechanics, can be employed to calculate the effect of individual amino acid mutations on the action of an antibiotic. Our test cases are the action of trimethoprim on *S. aureus* and rifampicin on *M. tuberculosis*. Trimethoprim is a competitive inhibitor of dihydrofolate reductase (DHFR), an enzyme in the essential folic acid pathway encoded by the

chromosomal gene *dfrB*. The most common mutation that confers resistance to trimethoprim is F99Y (Fig. 1A), a comparatively small mutation and therefore a stringent test of any predictive method. In previous work we demonstrated that conventional alchemical free energy methods (replica exchange thermodynamic integration with separate decharging, decoupling and recharging steps) was able to successfully and reproducibly distinguish mutations that confer resistance to trimethoprim from mutations that do not[2]. Some additional data using a fewer number of repeats but with longer λ simulations is shown in Fig. 1B-D. Rifampicin binds to the RNA polymerase, preventing the extension of the RNA and the most common resistance-conferring mutation is S450L in the B subunit of the complex (Fig. 1E), which is encoded by the *rpoB* gene.

For such a method to be deployed clinically it must be both fast and consume as little computational resource as possible, whilst still meeting the thresholds for accuracy and reproducibility as laid out by the existing international standards for new drug susceptibility testing methods. In the field of alchemical free energy calculations much attention has understandably been focussed on demonstrating that free energies can be calculated that agree with experimental data to a high degree of precision. By contrast, the dominant paradigm in clinical microbiology is qualitative: is the infection resistant (or not) to this antibiotic? A high degree of precision and accuracy is therefore spurious in this application; instead we wish to know whether the mutation will decrease the binding of the antibiotic sufficiently for it to be clinically considered resistant and that the process must be reliable (i.e. reproducible).

Here I shall examine how varying the computational resource allocated to the calculations affects the qualitative prediction and its reproducibility and thereby answer the question: "just how quickly can we reliably predict the effect of a mutation in *dfrB*/*rpoB* on the action of trimethoprim/rifampicin?". The answer to this question will guide whether it is yet feasible to deploy this kind of approach clinically, or whether we will need to wait for improvements in computational speed. It may also potentially enable the systematic prospective screening of mutations which could allow a pathogen to escape the action or an antibiotic. This information could, we hope, guide the selection and optimisation of lead compounds with antimicrobial activity and the approach more generally could be deployed early on in the drug development process, displacing less accurate methods.

One can think of this inversion of the usual requirement for high accuracy and precision as a sign that the field of alchemical free energy calculations is maturing. In the 1980s when the first protein calculations were performed[3], the duration of the simulations was extremely short, electrostatic forces were only calculated between pairs of atoms within a specified distance and it was unthinkable to repeat calculations to aid in the estimation of error and to test reproducibility. Despite this, the first studies reported good agreement with experiment, however, as CPU speeds increased allowing

longer simulations, precision appeared to suffer[4], suggesting that the first calculations underestimated their errors which the additional sampling was belatedly correcting. With the advent of computational grids and then cloud computing, the required simulations could be run in parallel, reducing the time to solution, however, it has only been in the last decade with the advent of GPUs that running repeats has become common place. Repeats simplify the estimation of error and naturally allow reproducibility to be tested[5]. By insisting on repeats, but testing how short the simulations can be and still correctly qualitatively predict the effect of a protein mutation on the action of an antibiotic we are, in some sense, turning back the clock but keeping several of the important computational and methodological developments.

References

1. Pankhurst LJ, del Ojo Elias C, Votintseva AA et al. (2016) *Lancet Resp Med* 4:49–58; Walker TM, Kohl TA, Omar SV et al. (2015) *Lancet Infec Disease* 15:1193–202; The CRyPTIC Consortium, 100000 Genomes Project (2018) *New England Journal of Medicine* 379:1403–1415.

2. Fowler PW, Cole K, Gordon NC, Kearns AM, Llewelyn MJ, Peto TE, Crook DW, Walker AS (2018) *Cell Chemical Biology* 25:339–349.e4.

3. Jorgensen WL, Ravimohan C (1985) *J Chem Phys* 83:3050; Bash P, Singh U, Brown F, Langridge R, Kollman P (1987) *Science* 235:574.

4. Chipot C, Pearlman D (2002) *Mol Sim* 28:1–12.

5. Bhati AP, Wan S, Hu Y, Sherborne B, Coveney PV (2018) *Journal of Chemical Theory and Computation* 14:2867–2880.

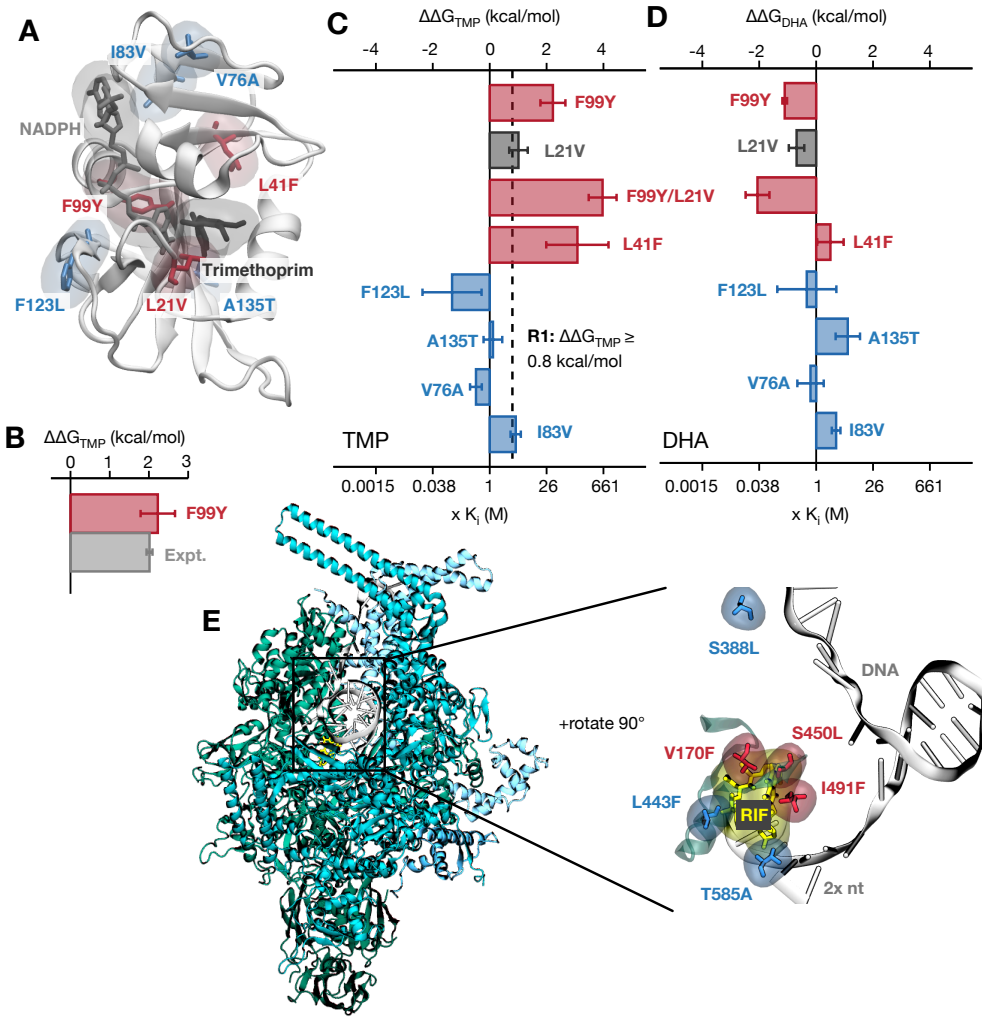

Figure 1: The test cases and some preliminary data. (**A**) The six studied mutations in *S. aureus* DHFR. Three confer resistance (red) to trimethoprim and three do not (blue). (**B**) In the previous work by Fowler *et. al*[2], 32 values of $\Delta\Delta G$ were calculated using λ simulations 0.25 ns in duration but did not agree with published isothermal titration calorimetry data. Here we show that calculating only 3 values of $\Delta\Delta G$ using λ simulations 10x longer leads to agreement with experiment, but arguable only by decreasing precision. (**C**) Using fewer, longer simulations does not alter the classification of the mutations made in Fig. 3 of the earlier work. (**D**) It does, however, reduce the predicted effect of the L41F DHFR mutation which was previously predicted to significantly reduce how well the natural substrate binds to DHFR. (**E**) The RNA polymerase of *M. tuberculosis* is an order of magnitude larger than DHFR. A pull out shows the mutations studied, again coloured according to whether they are clinically associated with resistance (red) or not (blue) to rifampicin.