ScattBO Benchmark - Bayesian optimisation for materials discovery

Samuel A. Appiah[1], Giuseppe Fisicaro[2], Gabriel L. Graves[3] and Andy S. Anker[*4, 5]

1: Ghana National Gas Company

2: CNR Institute for Microelectronics and Microsystems, Catania, Italy

3: Georgia Institute of Technology

4: Department of Energy, Danish Technical University, Denmark

5: Department of Chemistry, University of Oxford, England

*** ansoan@dtu.dk / andy.anker@chem.ox.ac.uk**

Team #24

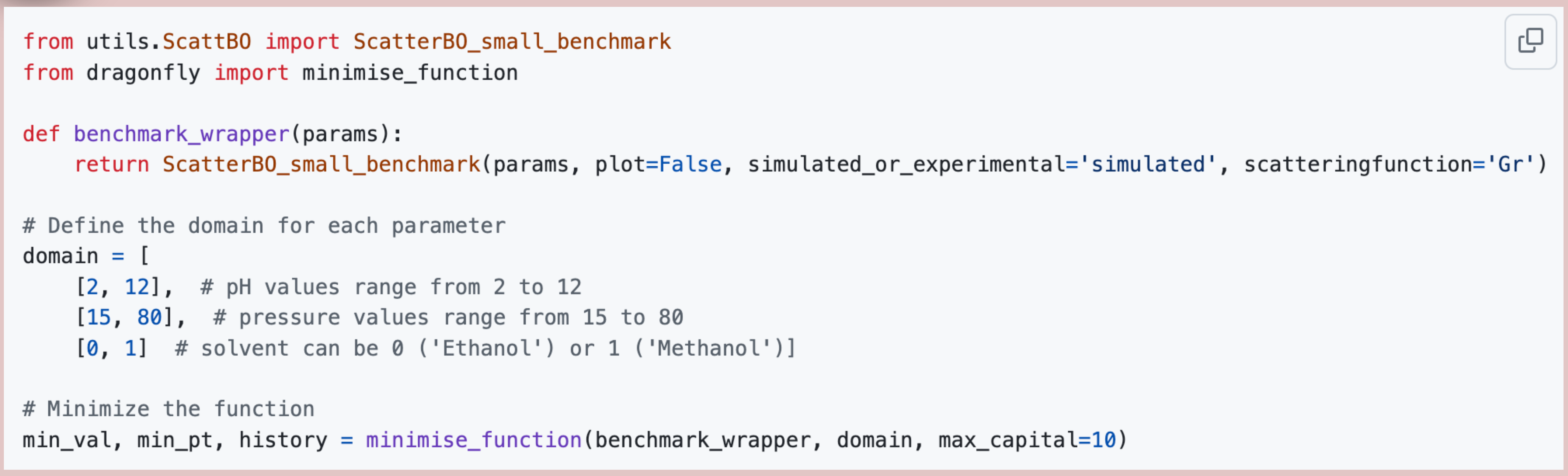
1 Introduction

A self-driving laboratory (SDL) is an autonomous platform that conducts machine learning (ML) selected experiments to achieve a user-defined objective. An objective can be to synthesise a specific material.[1] Such an SDL will synthesise a material, evaluate if this is the target material and if necessary optimise the synthesis parameters for the next synthesis. One way to evaluate if the material is the target material is by measuring scattering data and comparing that to the scattering pattern of the target material.

However, these types of SDLs can be expensive to run, which means that intelligent experimental planning is essential. At the same time, only a few people have access to an SDL for materials discovery. Therefore, it is currently challenging to benchmark Bayesian optimisation algorithms for experimental planning tasks in SDLs.

Here, we present a Python-based benchmark (ScattBO) that is an in silico simulation of an SDL for materials discovery. Based on a set of synthesis parameters, the benchmark 'synthesises' a structure, calculates the scattering pattern[2] and compares this to the scattering pattern of the target structure. Note: Scattering data may not be enough to conclusively validate that the target material has been synthesised.[3] The benchmark can include other types of data as long they can be simulated.

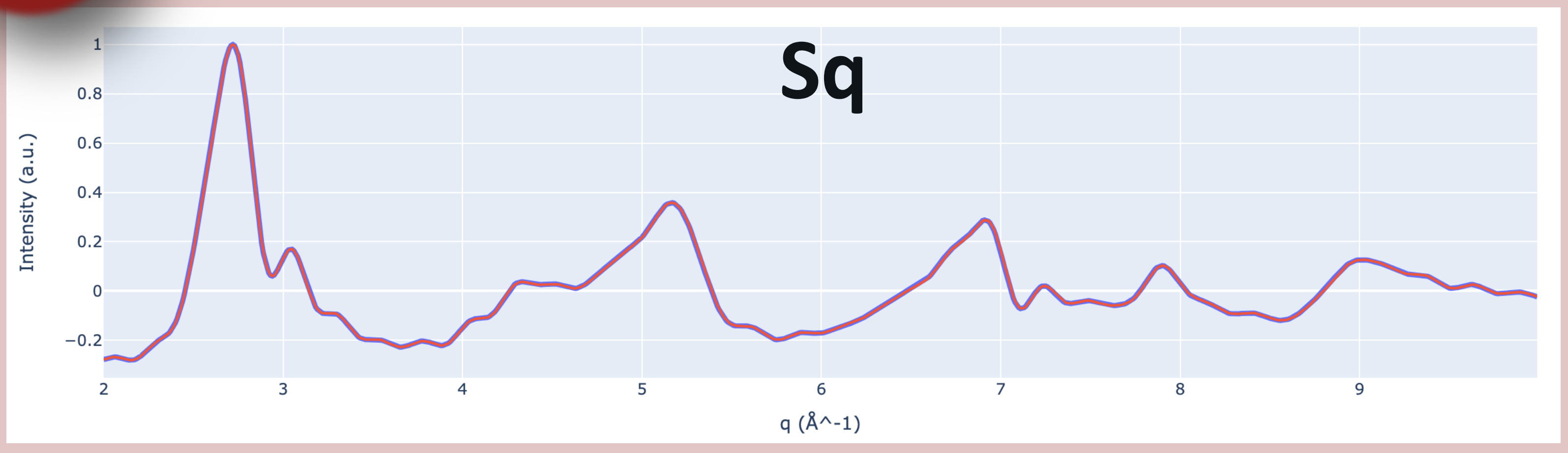
2 Minimal example using dragonfly[4]

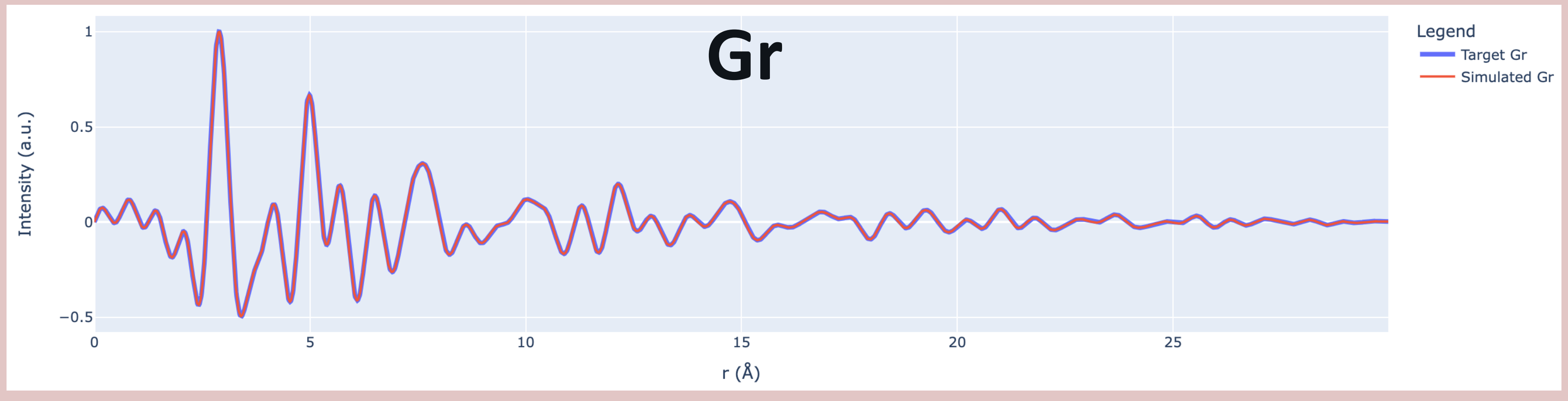


```python
from utils.ScattBO import ScatterBO_small_benchmark
from dragonfly import minimise_function

def benchmark_wrapper(params):
    return ScatterBO_small_benchmark(params, plot=False, simulated_or_experimental='simulated', scatteringfunction='Gr')

# Define the domain for each parameter
domain = [
    [2, 12],  # pH values range from 2 to 12
    [15, 80],  # pressure values range from 15 to 80
    [0, 1]  # solvent can be 0 ('Ethanol') or 1 ('Methanol')]

# Minimize the function
min_val, min_pt, history = minimise_function(benchmark_wrapper, domain, max_capital=10)
```

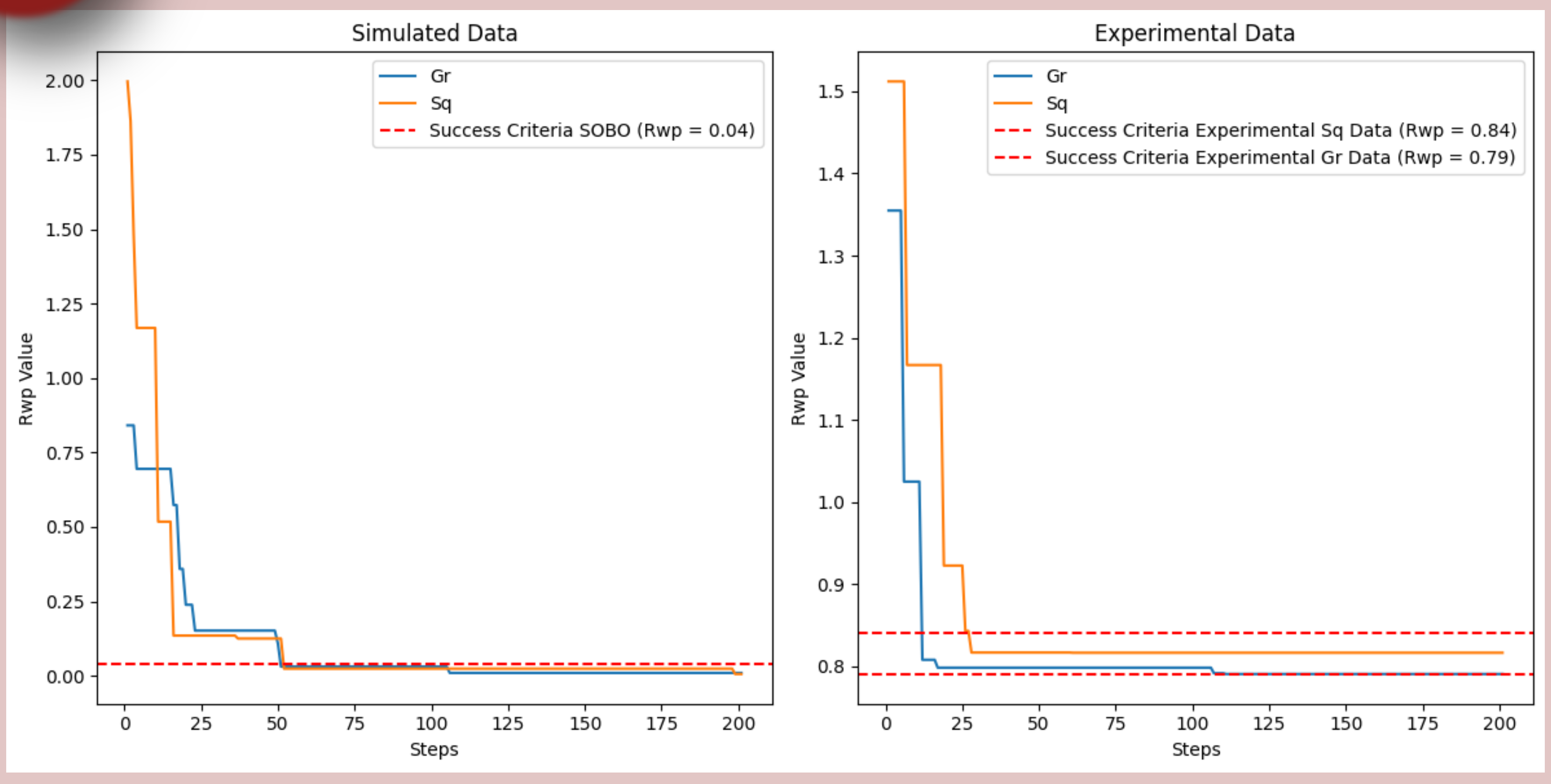
https://github.com/AndySAnker/ScattBO/edit/main/README.md

3 Scattering data

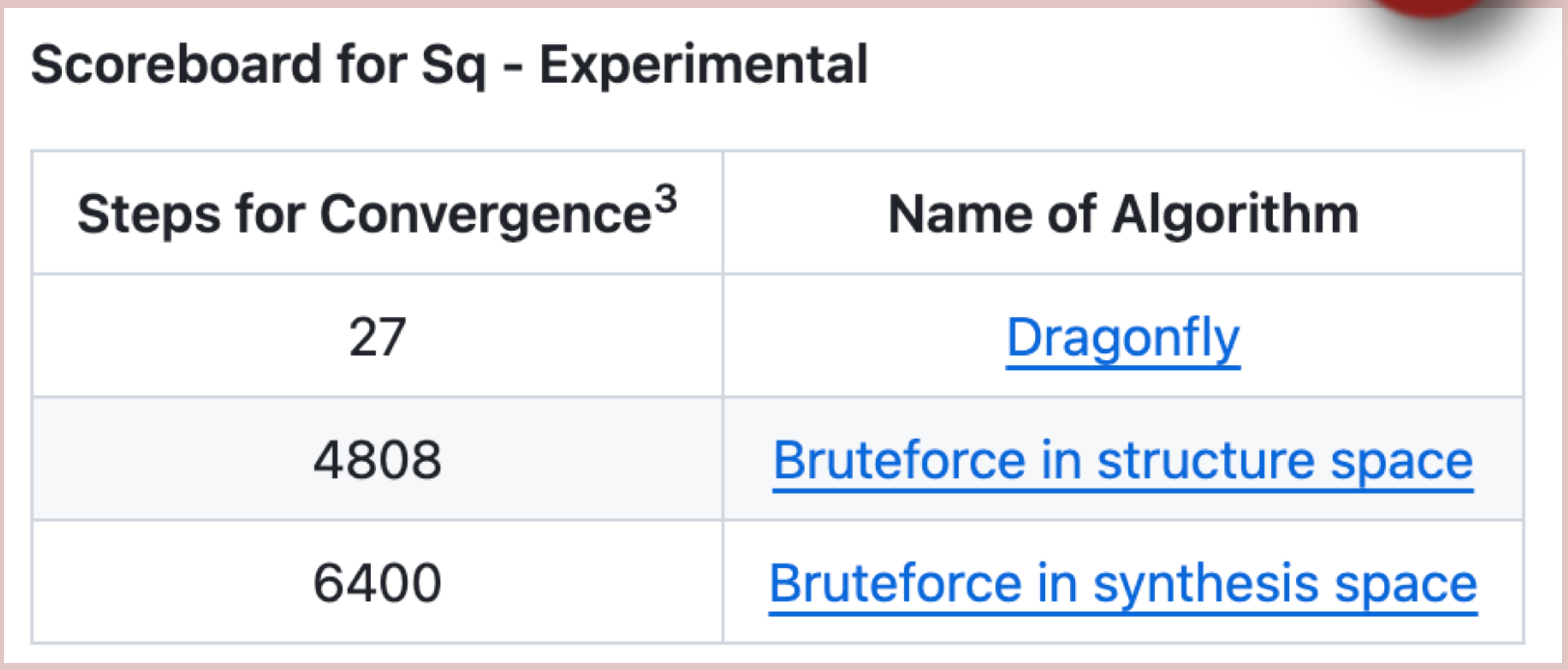




4 Convergence plots



Scoreboard 5



Scoreboard for Sq - Experimental

Steps for Convergence[3]	Name of Algorithm
27	Dragonfly
4808	Bruteforce in structure space
6400	Bruteforce in synthesis space

We have constructed two different baseline algorithms which find the optimal synthesis parameters through a bruteforce approach. These will use 4808 and 6400 steps to find the optimal synthesis parameters. Using Bayesian optimisation we can improve this to tens of experiments. We have during the BO hackathon been working on benchmarking the efficiency of various Bayesian optimisation algorithms for these tasks (*Bayes_opt*[5], *skopt*[6] and *BoTorch*[7]) but currently only have results for Dragonfly.

References

[1] Szymanski, Nathan J., et al., An autonomous laboratory for the accelerated synthesis of novel materials, Nature, 624(7990), 86-91 (2023)
[2] Frederik L. Johansen & Andy S. Anker, et al., A GPU-Accelerated Open-Source Python Package for Calculating Powder Diffraction, Small-Angle-, and Total Scattering with the Debye Scattering Equation, Journal of Open Source Software, 9(94), 6024 (2024)
[3] Leeman, Josh, et al., Challenges in High-Throughput Inorganic Materials Prediction and Autonomous Synthesis, PRX Energy, 3(1), 011002 (2024)
[4] https://github.com/dragonfly/dragonfly/tree/master [5] https://github.com/bayesian-optimization/BayesianOptimization [6] https://scikit-optimize.github.io/stable/auto_examples/bayesian-optimization.html [7] https://botorch.org/

Acknowledgements

This work is part of a project that has received funding from the Novo Nordisk Foundation (NFF) under the Postdoctoral Fellowships for MIT, Stanford Bio-X, University of Oxford and Weizmann Institute of Science – 2023 Programme (grant agreement No. NNF23OC0081359).