Integration of heterogeneous biological data in multiscale mechanistic model calibration

J-L. Palgen[1], A. Perrillat-Mercerot[1], N. Ceres[1], E. Peyronnet[1], M. Coudron[1], J. Bosley[1], A. L'Hostis[1], C. Monteiro[1]

[1] Novadiscovery, Lyon, France

Contact: jean-louis.palgen@novadiscovery.com

CALIBRATION CHALLENGE

Knowledge-based mechanistic models are built using knowledge as primary information source, with well-established biological interactions and physical laws determining the causal relationships within the model. Once the causal structure of the knowledge-based model is established with parameters and equations encompassing each of these relationships, parameter values or distributions must be estimated so that the simulation accurately reproduce experimental data, throughout a process called calibration.

This calibration process can be particularly challenging in the context of knowledge-based models of pathophysiologies, especially with heterogeneous data. Indeed, data scarcity and format can make it difficult to integrate all the information gathered in multiple heterogeneous datasets and experiments. We suggest a dedicated procedure to tackle this challenge and describe it hereafter.

DEDICATED STRATEGY

To make use of heterogeneous data with inconsistent format, we suggest a calibration strategy based on a list of successive steps, each step relying on the previous ones, and having as objective to match a set of biological behaviors -translated into computational constraints- observed in literature with computational model outputs. The steps should be made from the lowest levels of granularity of the model to the highest, and start with:

1. Biological phenomena that are well-detailed in the literature, in order to introduce as few uncertainties as possible in the first calibration steps
2. Biological phenomena whose internal dynamics are the most detailed in the model so that the model displays a consistent behavior
3. Biological phenomena that are less connected to others -phenomena with a lower number of inputs/outputs between one another

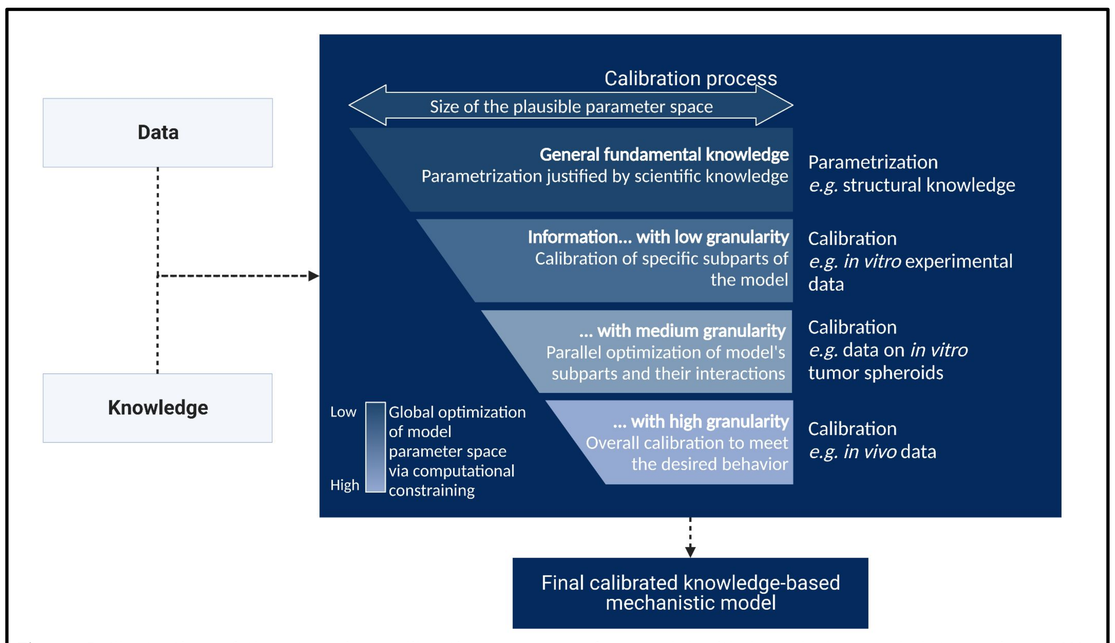

Figure 1. Illustration of the stepwise calibration process of a model with knowledge and data as input. We propose an iterative calibration process where the size of the plausible parameter space decreases at each step, and the effect of the computational constraints applied to the model on the efficiency of the calibration process increases. *Figure created with Biorender.com.*

We define two kinds of scores evaluating how well a constraint is met:

- Knowledge source with an associated binary score that takes value 0 or 1
- Data source with a score that continuously varies in [0, 1]

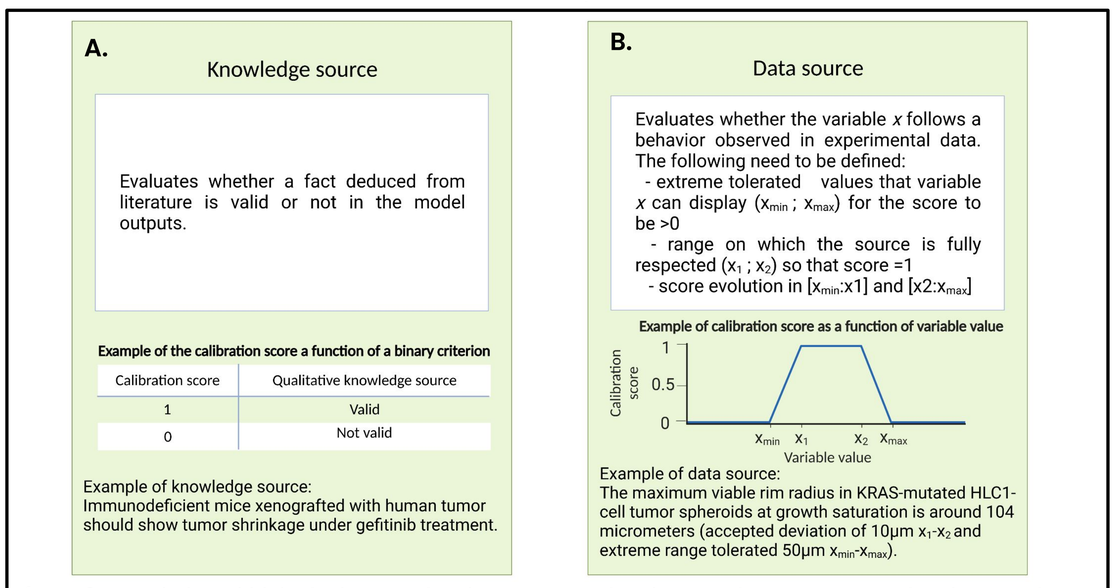

Figure 2. Translation of biological constraints into computational constraints. A: Computational constraint based on knowledge, B: Computational constraint based on experimental data.
Figure created with Biorender.com

An objective function to maximize is then defined as the weighted mean of all scores:

$$f(\theta) = \frac{\sum_k \omega_k s_k(y(\theta))}{\sum_k \omega_k}$$

where θ is a given simulation, y(θ) the simulation output, s_k the k^{th} sub-score and ω_k its associated weight (representing the confidence in the given scoring).

APPLICATION EXAMPLE: LUNG ADENOCARCINOMA

Context and modeling

Non-small-cell lung cancer is the most common type of lung cancer comprising notably lung adenocarcinomas (LUAD). The model used here, referred to as the *in silico* EGFR+ (Epidermal Growth Factor Receptor) Lung Adenocarcinoma (ISELA) model, relies on a mechanistic representation of tumor evolution, from specific mutations to tumor size evolution, in the context of patients with advanced EGFR mutated LUAD, treated with standard of care first-generation tyrosine kinase inhibitors.

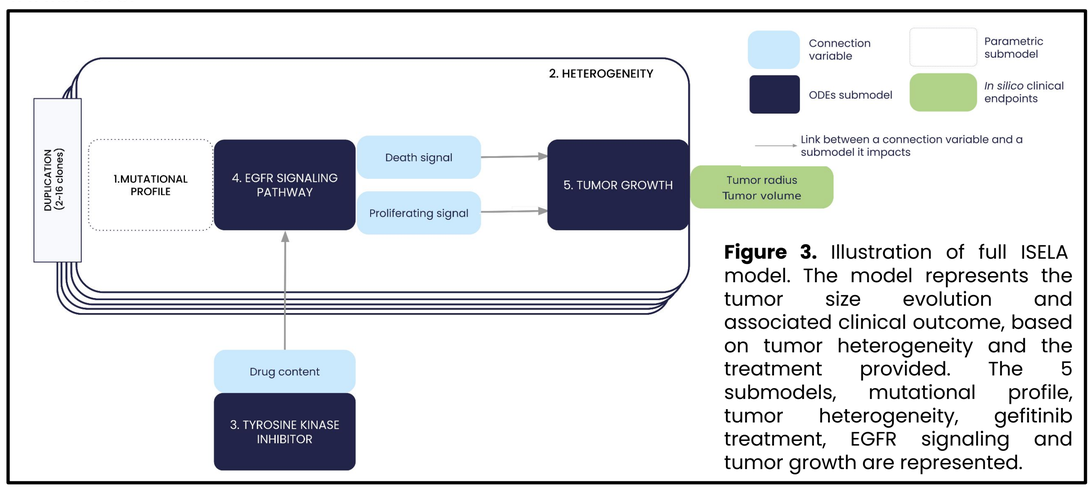

Figure 3. Illustration of full ISELA model. The model represents the tumor size evolution and associated clinical outcome, based on tumor heterogeneity and the treatment provided. The 5 submodels, mutational profile, tumor heterogeneity, gefitinib treatment, EGFR signaling and tumor growth are represented.

Calibration protocol

The experimental data found in literature to characterize EGFR+ LUAD tumor growth are either obtained from *in vitro* spheroids or from xenografted mice. As a consequence and in order to reproduce the EGFR+ LUAD cell growth observed experimentally, we defined the following calibration strategy for the ISELA model, with two sequential steps (that followed previous calibrations of EGFR signaling pathways and treatment pharmacokinetics):

1. Reproduce the tumor radius evolution and the proportion of viable cells of *in vitro* spheroids. This addresses *in vitro* scale, focusing on tumor growth in the absence of extracellular signals related to neo-angiogenesis, immune system or treatment.
2. Reproduce the tumor volume evolution of human EGFR+ LUAD tumor cells transplanted into immunocompromised mice. This addresses *in vivo* scale; focusing on tumor growth in the presence of neo-angiogenesis, weakened immune system and treatment.

Calibration results

Following the calibration, the ISELA model successfully reproduces the evolution of tumor size, both *in vitro* and in xenografted mice.

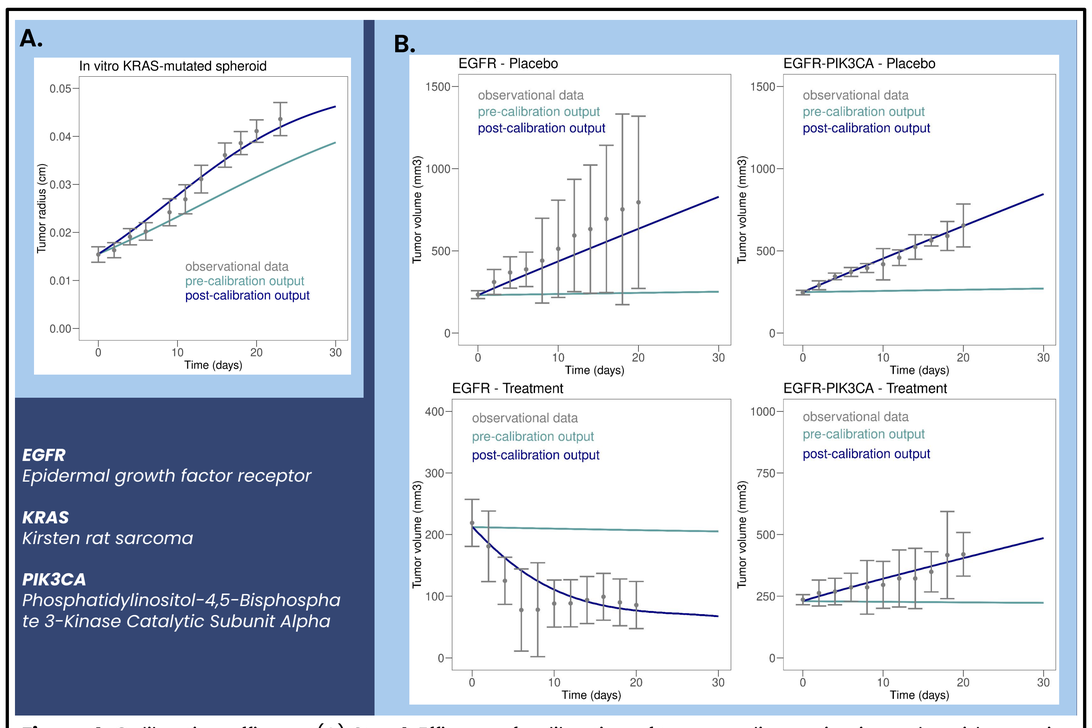

Figure 4. Calibration efficacy. (A) Step 1. Efficacy of calibration of tumor radius on *in vitro* spheroids carrying KRAS mutation. Experimental data are shown in gray (B) Step 2. Efficacy of calibration of tumor volume with xenograft data. (A-B) The type of mutations carried by the tumor is indicated on top of each plot. Plots were generated using R Core Team (2022). Experimental data are shown in gray (error bars represent 2 standard deviations from the experimental data reported by *Jagiella et al. (2016)* and *Kang et al. (2018)*), simulated data prior to calibration in light blue and simulated data after calibration in dark blue.

DISCUSSION & CONCLUSION

- Despite data heterogeneity constraints in scale -molecular pathways, pharmacokinetics and tumor size evolution- experimental conditions -*in vitro, in vivo*-, and format -qualitative and quantitative-, our approach led to a successful calibration. We deemed this approach the most adequate here, as no alternative method was available to our knowledge.
- The calibration strategy can be applied to a wide range of applied mathematical models to support their creation and drug development.

ACKNOWLEDGMENTS

We would like to thank Janssen-Cilag France for their support.

REFERENCES

Palgen, J. L., Perrillat-Mercerot, A., Ceres, N., Peyronnet, E., Coudron, M., Tixier, E., Illigens, B., Bosley, J., L'Hostis, A., & Monteiro, C. (2022). Integration of Heterogeneous Biological Data in Multiscale Mechanistic Model Calibration: Application to Lung Adenocarcinoma. Acta biotheoretica, 70(3), 19.

Jagiella N, Müller B, Müller M, et al. (2016) Inferring growth control mechanisms in growing multi-cellular spheroids of NSCLC cells from spatial-temporal image data. PLOS Computational Biology 12(2):e1004,412. 10.1371/journal.pcbi.1004412

Kang HN, Choi JW, Shim HS, et al. Establishment of a platform of non-small-cell lung cancer patient-derived xenografts with clinical and genomic annotation. Lung Cancer. 2018;124:168–178. doi: 10.1016/j.lungcan.2018.08.008.