Computer-Assisted Deductive Integration method

The analytical procedure implemented is entitled Computer-Assisted Deductive Integration (CADI) and associates algorithms and heuristics. The logic behind this model-building approach (Figure 1) does not assume functional linearity within biological systems and the components of a model do not incorporate solely what is known. Indeed, since this approach relies upon strict and systematic implementation of negative selection of hypotheses, models arising from this procedure contain elements that have never been described but cannot be refuted by current knowledge and/or available biological data, thereby generating novel understanding.

The system generates working hypotheses, which have been directly constructed from datasets and the literature, and which have resisted all destruction attempts (“go boards”). These are then merged to produce interaction maps describing the pathways/mechanisms that have become functional and those that have become forbidden in response to local conditions imposed by the activation of defined biological mechanisms. These maps are then merged to produce hypothetical physiological mechanisms. During each phase, “undetected” biological events are revealed while novel working hypotheses are generated (dotted arrows). These novel hypotheses are then subjected to the iterative negative (destructive) selection procedure. Hence, this model-building process involves multiple levels of internal cross-check procedures designed to eliminate any hypothesis that is not directly or indirectly supported by multiple data intersects.

Once a biologically plausible hypothetical model has been generated, experiments can be designed to directly challenge/validate the model. The results of these experiments can then be incorporated into the iterative analytical process, allowing rapid and efficient correction of the model. Thus, the system first provides for rapid development of a clear and factual understanding of the biological processes under investigation, and then provides a directly exploitable representation of the biological reality addressed by any therapeutic intervention.

In practice, the procedure is initiated from a query-building interface linked to an initially empty database (DB DH) the purpose of which is threefold. First, to record all queries sent to external databases. Second, to harbor both the queries and retrieved information attached to working hypotheses demonstrated as incorrect. Third, to avoid unnecessary redundancies by filtering all new queries. Following this filtration step, the queries are then dispatched to small machines linked to public databases via a web information retrieval interface. The information retrieved in
answer to a query, largely under the form of published literature and images, is then processed to determine whether this information could support the hypothesis attached to the query, invalidate it or neither support nor invalidate the hypothesis but provide material for a new formulation of this hypothesis.

If the information supports the working hypothesis, the information and the query are directed to a dedicated database (DB H1). The fact that a working hypothesis finds support in the published literature does not mean that the hypothesis is correct. It merely means that it does not contradict publicly accessible information. If data from the literature is at variance with the hypothesis, the retrieved information and the query are directed to the DB DH database. If the hypothesis is neither invalidated nor supported, the retrieved information and the query are directed to the DB H2 database. This complex procedure is carried out by specialized biologists assisted by proprietary software. This first level of iterative query procedures is ended when most new queries lead to material directed to DB H2, the contents of this database growing over five times faster than those of the other two databases.

At this stage, it can be considered that most available “medium-sized” pieces of the puzzle have been obtained and the model-building process itself can now be implemented Figure 1). The indices of the databases DB H1 and H2 are visualized to generate “meta-hypotheses” from the merging of either already supported hypotheses (DB H1) or supported hypotheses associated with neither supported nor ruled out hypotheses (DB H1 + DB H2). Meta-hypotheses are in turn subjected to the testing mechanism described above. Meta-hypotheses finding support in the literature enter the model-building module, while those proved incorrect enter the DB DH database and those neither supported nor ruled out enter a new sector in DB H2. Once again, the process is ended when the contents of the new sector in DB DH2 grows much faster than those of either DB DH or of the model-building module.

At this stage, most of the “large pieces” of the puzzle that can be reconstructed using published information have been obtained. But numerous gaps and uncertainties still remain. Thus, during the model-building phase, numerous questions do arise (dotted arrows in Figure 1) and these are in turn processed to the query interface in order to find supported solutions or propose possible answers, i.e. hypotheses that are not in contradiction with publicly accessible information. Model-building ends when the query process mostly generates uncertainties.
This model-building approach has repeatedly proven its efficacy in the discovery of i) hitherto unsuspected biological mechanisms, pathways and interactions directly associated with phenotypic transitions in vivo (be they pathological or developmental),\textsuperscript{1-3} ii) patent-protected novel therapeutic approaches in fields ranging from oncology to neurodegenerative and infectious diseases,\textsuperscript{1,4-6} and iii) the development of novel and patent-protected technologies.\textsuperscript{7}

Figure 1: The logic of the model-building approach.

References

1. Iris F. Psychiatric systems medicine: closer at hand than anticipated but not with the expected portrait. Pharmacopsychiatry. 2012;45 Suppl 1:S12-21.
2. Iris F, Gea M, Lampe PH, Querleux B. Heuristic modelling applied to epidermal homeostasis. In: Querleux B, ed. Computational biophysics of the skin. New York, NY: Pan Stanford. 2015.
3. Iris F, Gea M, Lampe PH, Santamaria P. Production and implementation of predictive biological models. Med Sci (Paris). 2009;25:608-616.
4. Turck CW, Iris F. Proteome-based pathway modelling of psychiatric disorders. Pharmacopsychiatry. 2011;44 Suppl 1:S54-61.
5. Gadal F, Bozic C, Pillot-Brochet C, et al. Integrated transcriptome analysis of the cellular mechanisms associated with Ha-ras-dependent malignant transformation of the human breast epithelial MCF7 cell line. Nucleic Acids Res. 2003;31:5789-5804.
6. Gadal F, Starzec A, Bozic C, et al. Integrative analysis of gene expression patterns predicts specific modulations of defined cell functions by estrogen and tamoxifen in MCF7 breast cancer cells. *J Mol Endocrinol.* 2005;34:61-75.

7. Pouillot F, Blois H, Iris F. Genetically engineered virulent phage banks in the detection and control of emergent pathogenic bacteria. *Biosecur Bioterror.* 2010;8:155-169.