Report on the clinical adaptation and validation procedure of the Oncosimulator and its integration into the ACGT architecture

Project Number: FP6-2005-IST-026996

Deliverable ID: D8.4

"Report on the clinical adaptation and validation procedure of the Oncosimulator and its integration into the ACGT architecture"

Submission Date: 17/09/2010

Information Society Technologies

COVER AND CONTROL PAGE OF DOCUMENT	
Project Acronym:	ACGT
Project Full Name:	Advancing Clinico-Genomic Clinical Trials on Cancer: Open Grid Services for Improving Medical Knowledge Discovery
Document id:	D8.4
Document name:	Report on the clinical adaptation and validation procedure of the Oncosimulator and its integration into the ACGT architecture
Document type (PU, INT, RE)	PU
Version:	Final
Submission date:	17/09/2010
Editor: Organisation: Email:	Georgios S. Stamatakos ICCS-NTUA gestam@central.ntua.gr

Document type PU = public, INT = internal, RE = restricted

ABSTRACT:

The present document aims at both recapitulating the main features and modules of the ACGT Oncosimulator and providing a fairly comprehensive account of the latest work that concerns mainly the procedures of its clinical adaptation and validation. The integration of the Oncosimulator into the ACGT architecture is also dealt with. Both the scientific and the technological aspects of the endeavour are addressed. The document concludes with a critical evaluation of the Oncosimulator development and the process of its envisaged clinical translation.

KEYWORD LIST: *in silico* oncology, Oncosimulator, cancer modelling, tumour growth, clinical trial, oncology, clinical adaptation, clinical validation, ACGT, SIOP 2001/GPOH clinical trial, TOP trial, multiscale cancer modeling, nephroblastoma, Wilms tumour, breast cancer, individualized treatment optimization, cancer systems biology, grid, virtual reality, subjunctive interface, tumour segmentation, immune system, visualization

MODIFICATION CONTROL			
Version	Date	Status	Author
1.0	01/07/10	Initial Draft	Georgios S. Stamatakos, ICCS-NTUA
2.0	20/07/10	Revised Draft	Georgios S. Stamatakos, ICCS-NTUA incorporating the contributions of the involved partners
3.0	01/08/10	Revised Draft	Dimitra Dionysiou, ICCS-NTUA
4.0	10/09/10	Revised Draft	Georgios S. Stamatakos, ICCS-NTUA
5.0	12/09/10	Revised Draft	Dimitra Dionysiou, ICCS-NTUA
6.0	15/09/10	Revised Draft	Norbert Graf, USAAR
7.0	17/09/10	Final	Georgios S. Stamatakos, ICCS-NTUA

List of (Additional) Contributors

- E. Kolokotroni, ICCS-NTUA
- E. Georgiadi, ICCS-NTUA
- S. Giatili, ICCS-NTUA
- C. Desmedt, IJB
- M. Erdt, FhG
- J. Puckacki, PSNC
- R. Belleman, UvA
- A. Lunzer, UHok
- S. Rueping, FHG
- A. d' Onofrio, IEO
- K.Marias, FORTH
- M. Tsiknakis, FORTH

It is noted that several other ACGT participants have supported the Oncosimulator endeavour

CONTENTS

(Page numbers)

CONTENTS.....4

1. EXECUTIVE SUMMARY.....5

2. INTRODUCTION AND A BRIEF OVERVIEW *(CODE LETTER: I)*.....6

3. THE ACGT ONCOSIMULATOR FROM THE CLINICAL PERSPECTIVE – THE PARADIGM OF WILMS TUMOUR *(CODE LETTER : C)*.....11

4. THE BASIC SIMULATION MODULE OF THE ACGT ONCOSIMULATOR: AN ADVANCED DISCRETE STATE / DISCRETE EVENT MULTISCALE SIMULATION MODEL OF THE RESPONSE OF A SOLID TUMOUR TO CHEMOTHERAPY. MIMICKING A CLINICAL STUDY *(CODE LETTER : S)* 30

5. UPDATE ON THE UTILIZATION OF WILMS TUMOUR DATA FOR CLINICAL ADAPTATION AND VALIDATION OF THE ACGT ONCOSIMULATOR *(CODE LETTER : W)* 68

6. UPDATE ON THE UTILIZATION OF BREAST CANCER DATA FOR CLINICAL ADAPTATION AND VALIDATION OF THE ACGT ONCOSIMULATOR *(CODE LETTER : B)* 80

7. IMAGE PROCESSING OF THE DEDICAL DATA (*CODE LETTER : P)* .,.....91

8. SUPPORT FOR THE EXECUTION OF THE ACGT ONCOSIMULATOR CODE IN THE GRID ENVIRONMENT *(CODE LETTER : G)* 98

9. EXPLORATORY VISUALIZATION OF ONCOSIMULATOR RESULTS THROUGH LIGHTWEIGHT INTERACTIVE VISUALIZATION SERVICES *(CODE LETTER : V)* 110

10. A FRONT-END INTERFACE FOR THE ACGT ONCOSIMULATOR VALIDATION (*CODE LETTER: F)* 123

11. THE ONCOSIMULATOR AS AN ACGT-COMPLIANT GRID SERVICE (*CODE LETTER: A)* 146

12. A SCENARIO FOR THE FUTURE IMMUNOLOGICAL EXTENSION OF THE ONCOSIMULATOR *(CODE LETTER: M)* 149

13. THE ACGT ONCOSIMULATOR TEAM IN THE 4th INTERNATIONAL ADVANCED RESEARCH WORKSHOP ON *IN SILICO* ONCOLOGY AND CANCER INVESTIGATION (*CODE LETTER: O)* 158

14. CONCLUSION *(CODE LETER: N)* 159

15. APPENDIX 1 - ABBREVIATIONS AND ACRONYMS *(CODE LETTER: R)* 160

IMPORTANT NOTICE: Bibliographical references pertaining to each chapter are to be found at its end.

1. Executive Summary

The present document aims at both recapitulating the main features and modules of the ACGT Oncosimulator (OS) and providing a fairly comprehensive account of the latest work that concerns mainly the procedures of its clinical adaptation and validation. The integration of the Oncosimulator into the ACGT architecture is also dealt with. Both the scientific and the technological aspects of the endeavour are addressed. The document concludes with a critical evaluation of the Oncosimulator development and the process of its envisaged clinical translation. The document begins with an overview of the ACGT Oncosimulator at the conceptual level. Subsequently, the clinical positioning of the system is provided through the paradigm of nephroblastoma (Wilms tumour) treated with chemotherapy. A detailed mathematical and algorithmic description of an advanced version of the core simulation model along with a sensitivity analysis and a proposed clinical adaptation procedure are delineated. The latest advances in the technological modules of the systems are presented in relative detail. These include image processing, technological support for the execution of the OS code in the grid environment, exploratory visualization of of OS results through lightweight interactive visualization series, a front end interface for the OS validation (termed the OncoRecipeSheet) and the integration of the OS into the ACGT platform so as to function as an ACGT-compliant grid service. As a possible future extension of the OS a high level description of the integration of an immune system model-module into the OS is also presented. A brief account of the 4th International Advanced Research Workshop on *In Silico* Oncology and Cancer Investigation in which the general architecture as well as several aspects of the ACGT Oncosimulator were presented has also been also included in the document. The conclusions of the overall exposition complete the deliverable.

2. Introduction and a brief overview

(Code letter: I)

2.1 The ACGT Oncosimulator: from conceptualization to development via multiscale cancer modeling

The aim of this chapter is to briefly outline the modules and the developmental and translational stages of the *Oncosimulator* developed within the framework of the European Commission and Japan co-funded ACGT integrated project (www.eu-acgt.org, http://eu-acgt.org/acgt-for-you/researchers/in-silico-oncology/oncosimulator.html) [I1-I20]. The *Oncosimulator* is an integrated software system simulating *in vivo* tumour response to therapeutic modalities in the clinical trial context. It aims at supporting patient individualized optimization of cancer treatment. The four dimensional simulation module embedded in the *Oncosimulator* is based on the multiscale, *top-down*, discrete entity - discrete event cancer simulation approach developed by the *In Silico* Oncology Group, ICCS, National Technical University of Athens [www.in-silico-oncology.iccs.ntua.gr]. It has been specified for the cases of nephroblastoma and breast cancer being chemotherapeutically treated in the neoadjuvant setting according to protocols included in the SIOP 2001/GPOH and the TOP clinical trial respectively. The main difference between the two specific models (apart from the different values of the respective model parameters) lies in the different pharmacokinetics and pharmacodynamics submodules for the drugs / drug combinations administered.

Cellular automata, the generic Monte Carlo technique, novel dedicated algorithms and pharmacokinetic differential equations constitute the mathematical basis of the simulation. A discretizing mesh covers the anatomic area of interest. A system of biological cell clusters included within each geometrical cell of the discretizing mesh lies at the heart of the simulation approach. Critical mechanisms such as tumour expansion or shrinkage and the effects of particular drugs on the tumour under consideration have been incorporated into the model. In order to integrate biological mechanisms acting on different biocomplexity scales (levels) the *summarize and jump* strategy [I9] has been adopted.

The following processes-modules constitute the core of the integrated scientific and technological construct of the *Oncosimulator*: (i) anonymized / pseudonymized multiscale data collection (including imaging, histopathological, molecular, clinical and treatment data), (ii) image processing and preprocessing of the rest of data, (iii) description of previous or candidate treatment schemes / schedules, (iv) model and simulation code development, (v) invocation of code execution via an intelligent portal (*RecipeSheet*), (vi) execution of the code on either a cluster or the grid, (vii) collection and visualization of the predictions and numerical analysis (including sensitivity analysis) of the simulation model, (viii) evaluation of the predictions and system optimization, (ix) clinical validation of the system within the clinical trial context, (x) optimization of the system, (xi) eventual future translation of the *Oncosimulator* into the clinical practice to serve as a patient individualized treatment optimization support system.

Since the model reflects the natural instability of cancer regarding small changes in a limited number of critical parameters, *virtual cancer patient prototypes* are currently being developed in order to take this behaviour into account. Virtual cancer patient prototypes are constructed by combining representative plausible values of critical model parameters by taking into account the molecular, histological and imaging characteristics of a given tumour. Thus each patient prototype is in fact a cluster of combinations of plausible parameter values. By running the simulation code for all candidate treatment schemes for a virtual cancer patient prototype the user can rank the schemes based on the tumour response (e.g. overall cell kill or tumour shrinkage) they produce *in silico*. Thus the treatment scheme/schedule that induces the best tumour response in the majority of parameter value combinations for a given tumour is tentatively assumed to be the optimal one, provided that restrictions imposed by normal tissue side effects allows its clinical application. It is pointed out that the ACGT *Oncosimulator* constitutes a *global novelty*.

2.2 The medical problems addressed

In the case of nephroblastoma, the simulation algorithms address the cases of preoperative chemotherapy for unilateral stage I-III tumours treated in the framework of the clinical trials addressed by ACGT (SIOP 2001/GPOH) with a combination of actinomycin-D and vincristine (Fig. I1).

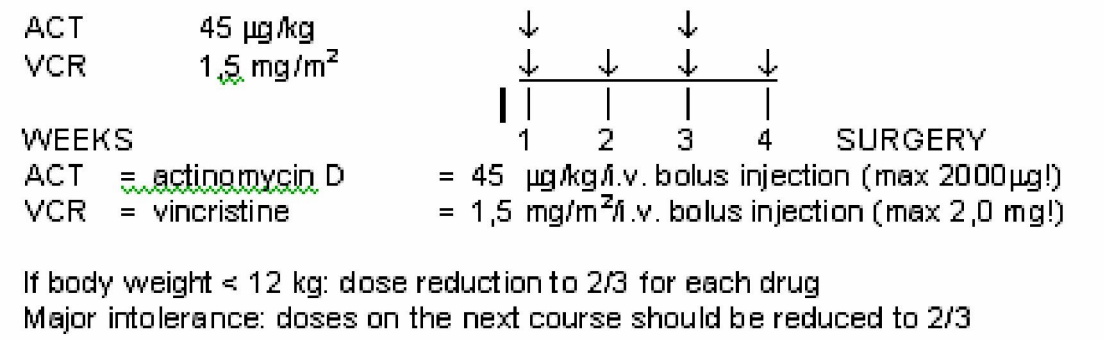


Fig. I1. The simulated treatment protocol for Wilms' tumours.

In the case of breast cancer, the simulation algorithms address the cases of primary chemotherapy ("neo-adjuvant" chemotherapy) with single-agent epirubicin (100 mg/m^2 i.v. once every 3 weeks for 4 consecutive cycles) for early breast cancer patients in the framework of the ACGT clinical trials (TOP trial) (Fig.I2).

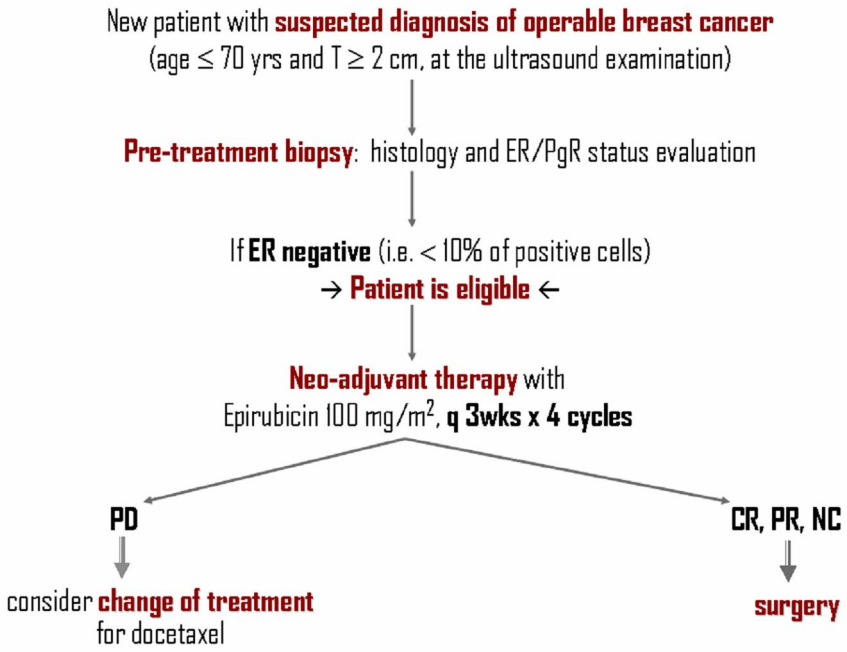


Fig. I2. The simulated treatment protocol for early breast cancer tumours.

2.3 Indicative simulation results

Fig. I3 Simulated time course of the limited mitotic potential (Limp) cell population before, during and after the chemotherapeutic treatment in the case of a typical real nephroblastoma tumour.

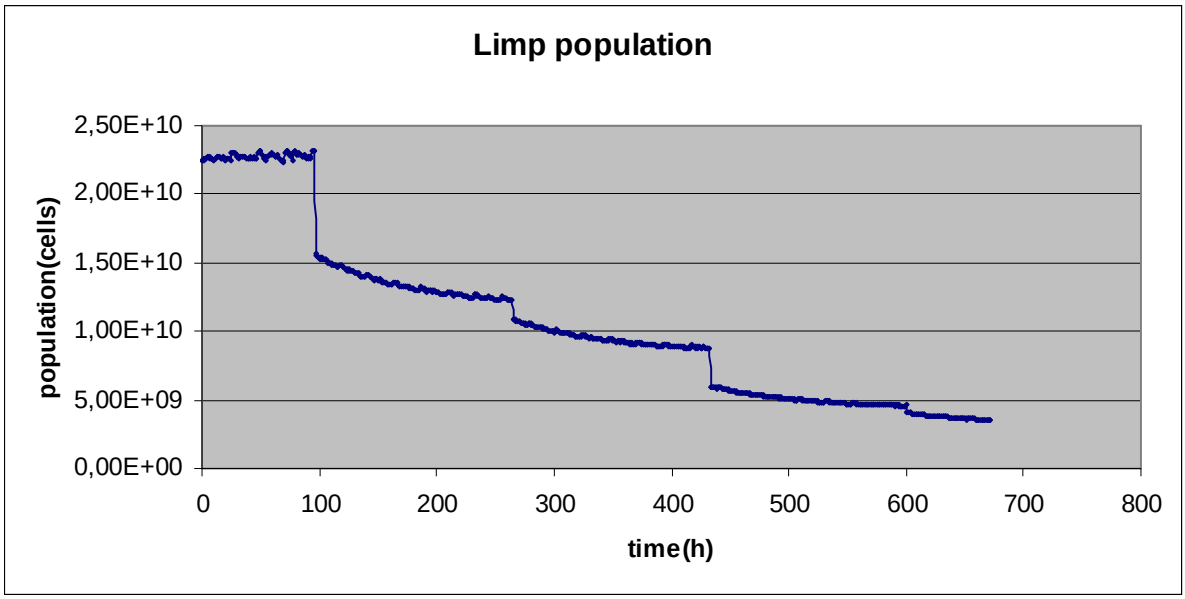


Fig.I3 shows a typical simulation prediction regarding nephroblastoma (Wilms tumour).

Exploratory analyses have revealed the importance of critical model parameters such as the symmetric division probability (i.e. the probability of a stem cell to give birth to two stem cells instead of a stem and a progenitor cell), the dormancy probability of a newborn tumour cell, the cell cycle duration and the cell kill probability for the response of a solid tumour to chemotherapeutic treatment. Furthermore, plausible parameter values have been adapted to real multiscale clinical trial data (e.g. 20 multiscale data sets in the case of nephroblastoma) in a consistent way thus supporting the predictive potential of the *Oncosimulator*. Clinical adaptation and validation of the ACGT *Oncosimulator* are in progress. An eventually successful completion of this time demanding phase is expected to lead to the clinical translation of the system [I12]. In parallel the *Oncosimulator* is currently expanded in order to provide a platform for the study of the immune system response to cancer and related phenomena and/or treatment techniques.

2.4 Summary

The clinically oriented character of the ACGT *Oncosimulator*, its flexible *top-down*, discrete entity-discrete event simulation philosophy as well as its preliminary clinical adaptation outcome support its potential for future clinical translation. Following the completion of a thorough numerical exploration and a strict clinical validation process, the *Oncosimulator* is expected to serve as treatment optimization system in the patient individualized context. Before reaching that stage utilization of the *Oncosimulator* as a multiscale cancer research tool is an obvious possibility.

2.5 References

[I1] Stamatakos G.S., D.D. Dionysiou, E.I. Zacharaki, N.A. Mouravliansky, K.S. Nikita, N.K. Uzunoglu 2002. *In silico* radiation oncology: combining novel simulation algorithms with current visualization techniques. Proc IEEE. Special Issue on Bioinformatics: Advances and Chalenges 90(11):1764-1777.

[I2] Dionysiou D.D., G.S. Stamatakos, N.K.Uzunoglu, K.S.Nikita, A. Marioli , A Four Dimensional In Vivo Model of Tumour Response to Radiotherapy: Parametric Validation Considering Radiosensitivity, Genetic Profile and Fractionation , J. theor. Biol. , 230 , 1-20 , 2004

[I3] G.S.Stamatakos, D.D.Dionysiou, N.M.Graf, N.A.Sofra, C.Desmedt, A.Hoppe, N.Uzunoglu, M.Tsiknakis , The Oncosimulator: a multilevel, clinically oriented simulation system of tumour growth and organism response to therapeutic schemes. Towards the clinical evaluation of in silico oncology , Proceedings of the 29th Annual International Conference of the IEEE EMBS Cite Internationale, August 23-26, SuB07.1: 6628-6631 , Lyon, France , 2007

[I4] Dionysiou, D.D., G.S. Stamatakos, D. Gintides, N. Uzunoglu K. Kyriaki. 2008. Critical parameters determining standard radiotherapy treatment outcome for glioblastoma multiforme: a computer simulation. Open Biomed Eng J 2: 43-51.

[I5] Graf, N., A. Hoppe , E. Georgiadi, R. Belleman, C. Desmedt, D. Dionysiou, M. Erdt , J. Jacques, E. Kolokotroni, A. Lunzer, M. Tsiknakis and G. Stamatakos. 2009. "*In silico* oncology" for clinical decision making in the context of nephroblastoma. Klin Paediatr 221: 141-149.

[I6] Kolokotroni E.A.,, G. S. Stamatakos, D. D. Dionysiou, E. Ch. Georgiadi, Ch. Desmedt, N. M. Graf, "Translating Multiscale Cancer Models into Clinical Trials: Simulating Breast Cancer Tumour Dynamics within the Framework of the "Trial of Principle" Clinical Trial and the ACGT Project.," Proc. 8th IEEE International Conference on Bioinformatics and Bioengineering (BIBE 2008), Athens, Greece, 8-10 Oct. 2008. IEEE Catalog Number: CFP08266, ISBN: 978-1-4244-2845-8, Library of Congress: 2008907441, Paper No. BE-2.1.1, length: 8 pages (in electronic format). 2008

[I7] Georgiadi E. Ch., G. S. Stamatakos, N. M. Graf, E. A. Kolokotroni, D. D. Dionysiou, A. Hoppe, N. K. Uzunoglu, "Multilevel Cancer Modeling in the Clinical Environment: Simulating the Behaviour of Wilms Tumour in the Context of the SIOP 2001/GPOH Clinical Trial and the ACGT Project," Proc. 8th IEEE International Conference on Bioinformatics and Bioengineering (BIBE 2008), Athens, Greece, 8-10 Oct. 2008. IEEE Catalog Number: CFP08266, ISBN: 978-1-4244-2845-8, Library of Congress: 2008907441, Paper No. BE-2.1.2, length: 8 pages (in electronic format). 2008

[I8] Giatili S.G., G. S. Stamatakos, D. D. Dionysiou, E. A. Kolokotroni, E. Ch. Georgiadi, Geometrical and Mechanical Aspects of Tumour Growth and Response to Chemotherapeutic Schemes in the Context of the ACGT Oncosimulator," Proceedings of the 3rd International Advanced Research Workshop on In Silico Oncology, Istanbul, Turkey, Sept. 23-24, 2008. Edited by G. Stamatakos and D. Dionysiou pp.35-37, 2008

[I9] Stamatakos G. and D. Dionysiou. 2009. Introduction of hypermatrix and operator notation into a discrete mathematics simulation model of malignant tumour response to therapeutic schemes *in vivo*. Some operator properties. Cancer Informatics 2009:7 239–251.

[I10] Stamatakos, G.S., E. A. Kolokotroni, D. D Dionysiou, E. C Georgiadi, C. Desmedt, An advanced discrete state - discrete event multiscale simulation model of the response of a solid tumour to chemotherapy. Mimicking a clinical study. J.Theor. Biol. 266 (2010) 124-139.

[I11] Stamatakos G. 2010. *In Silico* Oncology: PART I Clinically oriented cancer multilevel modeling based on discrete event simulation in T. S. Deisboeck and G. S. Stamatakos Eds, Multiscale Cancer Modeling. Chapman & Hall/CRC, Boca Raton, Florida, USA, 2010 (*in press*) ISBN: 9781439814406.

[I12] N.Graf, *In Silico* Oncology Part II: Clinical Requirements. in T. S. Deisboeck and G. S. Stamatakos Eds, Multiscale Cancer Modeling. Chapman & Hall/CRC, Boca Raton, Florida, USA, 2010 (*in press*) ISBN: 9781439814406.

[I13] G.Stamatakos, 2010. In Silico Oncology: A Hypermatrix – Operator Formulation of a Top - Down Multiscale Simulation Model of Tumour Response to Treatment. The Oncosimulator Concept. In Proc. 4th International Advanced Research Workshop on *In Silico* Oncology and Cancer Investigation, Athens, Greece Sept. 8-9, 2010, Edited by G.Stamatakos and D. Dionysiou. To appear in www.4th-iarwisoci.iccs.ntua.gr

[I14] D. Dionysiou, 2010. The ISOG, NTUA Tumour Response to Treatment Discrete Simulation Models: A Review of Basic Concepts and Algorithms. In Proc. 4th International Advanced Research Workshop on *In Silico* Oncology

and Cancer Investigation, Athens, Greece Sept. 8-9, 2010, Edited by G.Stamatakos and D. Dionysiou. To appear in www.4th-iarwisoci.iccs.ntua.gr

[I15] E.Kolokotroni, D.Dionysiou, E.Georgiadi, N.Uzunoglu, G.Stamatakos, 2010. Breast Cancer Modeling in the Clinical Context: Parametric Studies. In Proc. 4th International Advanced Research Workshop on *In Silico* Oncology and Cancer Investigation, Athens, Greece Sept. 8-9, 2010, Edited by G.Stamatakos and D. Dionysiou. To appear in www.4th-iarwisoci.iccs.ntua.gr

[I16] E.Georgiadi, D.Dionysiou, E.Kolokotroni, N.Uzunoglu, N.Graf, G.Stamatakos, 2010. Discrete Event Based Modeling of Nephroblastoma. Sensitivity Considerations. In Proc. 4th International Advanced Research Workshop on *In Silico* Oncology and Cancer Investigation, Athens, Greece Sept. 8-9, 2010, Edited by G.Stamatakos and D. Dionysiou. To appear in www.4th-iarwisoci.iccs.ntua.gr

[I17] G.Stamatakos, D.Dionysiou, E.Georgiadi, E.Kolokotroni, S.Giatili, A.Hoppe, C.Desmedt, A.Lunzer, M.Erdt, J.Jacques, J.Pukacki, R.Belleman, P.Melis, A.d'Onofrio, F.Buffa, B.Claerhout, S.Rueping, K.Marias, M.Tsiknakis, N.Graf, 2010. The ACGT Oncosimulator: from Conceptualization to Development via Multiscale Cancer Modeling. In Proc. 4th International Advanced Research Workshop on *In Silico* Oncology and Cancer Investigation, Athens, Greece Sept. 8-9, 2010, Edited by G.Stamatakos and D. Dionysiou. To appear in www.4th-iarwisoci.iccs.ntua.gr

[I18] J.Zepp, N.Graf, E.Skounakis, R.Bohle, E.Meese, H.Stenzhorn, Y.-J. Kim, C.Farmaki, V.Sakkalis, W.Reith, G.Stamatakos, K.Marias, 2010. Tumour segmentation: The impact of standardized signal intensity histograms in glioblastoma. In Proc. 4th International Advanced Research Workshop on *In Silico* Oncology and Cancer Investigation, Athens, Greece Sept. 8-9, 2010, Edited by G.Stamatakos and D. Dionysiou. To appear in www.4th-iarwisoci.iccs.ntua.gr

[I19] A.Lunzer, R.Belleman, P.Melis, J.Pukacki, P.Spychała, G.Stamatakos, 2010. Validating the ACGT Oncosimulator with a Grid-Supported Visualisation Environment. In Proc. 4th International Advanced Research Workshop on *In Silico* Oncology and Cancer Investigation, Athens, Greece Sept. 8-9, 2010, Edited by G.Stamatakos and D. Dionysiou. To appear in www.4th-iarwisoci.iccs.ntua.gr

[I20] T. S. Deisboeck and G. S. Stamatakos Eds, "Multiscale Cancer Modeling." Chapman & Hall/CRC, Boca Raton, Florida, USA, 2010 (book in press)

http://www.crcpress.com/product/isbn/9781439814406;jsessionid=Btz7pVD%20gkQQFtWh%201W3ng**

.

3. The ACGT Oncosimulator from the clinical perspective – the paradigm of Wilms tumour

(Code letter : C)

3.1 Introduction

The goal of cancer research is to find better ways to treat patients with cancer. During the last decades basic research and clinical trials have gathered a lot of new insights in the molecular biology of cancer providing new drugs and treatment approaches. This can be clearly shown for Wilms Tumour [C1] . Nevertheless the outcome for many cancers is still dismal demanding new and better treatments for patients.

A clinical trial is one of the final stages of a long and careful cancer research process. The search for new treatments begins in the laboratory. Molecular biology did help to better understand carcinogenesis, thus finding new targets for interfering with agents that are able to reverse the process of developing cancer cells. New methods and technologies in molecular biology will result in an exponential increase of information in future that can be handled by the advances of high-computing and informatics. It is of paramount importance to gather this information with clinical data to gain new knowledge for developing better treatments for cancer patients. This approach will result in clinicogenomic trials, as ACGT is proposing and running. Such trials will lead to more individualized cancer treatments and providing better chances of cure for patients with fewer side effects.

With the help of *In Silico Oncology* it is expected that cancer growth and response to different treatments can be simulated. Such *in silico* experiments might help clinicians in future to find the best way of treating an individual patient by simulating different treatments in the computer before starting the treatment in reality. The *In Silico Oncology* Group, ICCS, NTUA has been engaged in developing such simulation models, by fully exploiting the insight gained by molecular biology and other disciplines as well as the individual patient's data [C2-C6]

As medicine relies on models to understand and predict the physiology and pathophysiology of biological systems these models are verified *in vitro* and *in vivo* but they can also be analyzed theoretically with '*in silico'* techniques [C7] . Advances in systems-biology-driven concepts in biomedicine enhanced by the increasing volume of molecular data and the decreasing costs of computational power have made it possible to run such large and clinically relevant simulations today. If clinicians could accurately predict which treatment will fail in a patient before it is applied, this could save lives, time and resources, and might ultimately lead to more targeted, personalized therapies.

From the mathematical point of view approaches in '*in silico*' cancer modeling include the application of game theory [C8-C10] scaling laws [C11-C12] fractals [C13-C15] and graph theory [C16] . These strategies and methods will help scientists to go beyond the classic domains of volumetric tumour growth dynamics [C17-C18] and vascularisation patterns [C19-C22] in cancer. It is possible to investigate genetic instability [C23] and mutagenesis [C24] for instance, as well as the complexity of tumour–immune system interaction [C25] .

The aim of '*in silico'* oncology is to develop patient specific computer simulation models of malignant tumours and normal tissues in order to optimize the planning of various therapeutic schemes. Ultimately, the aim is to contribute to the process of effectively treating cancer and to contribute to the understanding of the disease at the molecular, cellular, organ and body level.

From a clinical point of view it is expected that cancer growth and response to different treatments can be simulated. In simulating response to treatment in a given cancer this knowledge is significant for assessing better methods for treatment efficacy as the RECIST criteria [C26-C27] provide. It might be time to improve traditional measures of clinical response as trial end points and to evaluate the activity on rare and resistant cancer cells [C28]. If '*in silico'* experiments are to be of more help for a clinician than providing a prediction of changes in tumour volume and shape, the response of treatment of the small fraction of resistant cancer cells will be of utmost importance in the future.

Such '*in silico'* experiments might help clinicians in the future to find the best way of treating an individual patient by simulating different treatments in the computer before starting the treatment in reality. Two preconditions are of utmost importance before one can rely on '*in silico*' oncology models [C29]:

1. every '*in silico'* method has to be part of a clinico-genomic trial
2. every prediction of an '*in silico'* method has to be compared with the reality

After establishing the '*in silico'* model it is necessary to define the needed data in a first step, including data from the tumour (molecular biology, pathology, imaging), from the patient (clinical data) and from the possible treatment (pharmacokinetics of drugs that will be used, the treatment schema) as well as from literature and open-source databases. To make the simulation predictions precise and realistic it is crucial to get as much information as possible from each of the different categories. The amount of data will be restricted by the availability of tumour material, imaging data and clinical data. Therefore '*in silico'* oncology must always be integrated into or be part of a clinico-genomic trial, where data management including data security and anonymisation or pseudonymisation of data, along with tumour banking, are well established. In addition the trial is always reviewed by an ethical committee and fulfils all other GCP criteria to get approval by regulatory authorities [C30-C32].

The simulation prediction of each '*in silico'* model must always be compared with the reality, in other words, the actual treatment outcome. The actual outcome provides feedback for tuning the '*in silico'* model to get better predictions. Such a control loop, executed automatically, must be a component in any '*in silico*' model whose predictions are to be used for treatment. In that way '*in silico'* experiments should be considered and established as learning systems. Only if there are no or minimal deviations between the prediction and the reality the '*in silico*' method can be allowed to be used in a clinical setting. The clinician has to define what can be accepted as a minimal deviation between prediction and reality in a trial. This definition should always be included in the biometrics part of a clinico-genomic trial protocol. For the safety of patients a stopping rule has to be defined, if clinical decisions are based on '*in silico'* experiments.

For a clinician it is important that the '*in silico*' experiments can address and answer precisely for each patient the following questions:

1. What is the natural course of the tumour growth over time in size and shape?
2. When and to where is the tumour metastasizing?
3. Can the response of the local tumour and the metastases to a given treatment be predicted in size and shape over time?
4. What is the best treatment schedule regarding drugs, surgery, irradiation and their combination, dosage, time schedule and duration to achieve cure?
5. Is it possible to predict severe adverse events (SAE) of a treatment and to propose alternatives to avoid them without deteriorating the outcome?
6. Is it possible to predict a cancer before it occurs, and to recommend treatment options to prevent the occurrence or a recurrence?

Which question will be addressed is decided by the clinician and will influence the model.

3.2 A brief description of the module work done throughout the entire ACGT lifespan from the clinical perspective

Nephroblastoma is the most common malignant renal tumour in children. Today treatments are based on several multicentre trials and studies conducted by the SIOP in Europe and the COG in North America. Information from these nephroblastoma studies on both sides of the Atlantic allowed identification of prognostic indicators independent of whether patients were treated by immediate surgery (COG: Children's Oncology Group, North America) or surgery after preoperative chemotherapy (SIOP: International Society of Paediatric Oncology). In particular, identification of histological subtypes of Wilms' tumour in addition to stage classification is of prognostic value. In this way the SIOP trials and studies largely focus on the issue of preoperative therapy [C33-C35]. Response to treatment can be

measured individually by tumour volume reduction and percentage of therapy-induced necrosis or remaining vital blastema at the time of surgery in the histological specimen. In nephroblastoma the blastemal subtype after preoperative chemotherapy is recognised as an unfavourable entity. This gives an early individual prognostic parameter and is used for further stratifying and more individualizing postoperative treatment as in other paediatric cancers [C36-C37].

In about 1.6 % of cases preoperative chemotherapy is given erroneously [C38] as it is based solely on imaging studies. Despite the fact that for neuroblastoma - the most important differential diagnosis for nephroblastoma - such treatment is without influence on outcome [C39] , unnecessary chemotherapy must always be avoided in benign tumours. As long as imaging studies cannot rule out precisely tumours other than nephroblastoma, this risk has to be well balanced with the advantage of downstaging of the tumour and thus applying less therapy to the patient after surgery. Therefore knowing about '*in silico'* oncology, models of preoperative chemotherapy have to be developed in nephroblastoma. These models should be able to predict precisely the response to treatment in every case, to avoid unnecessary treatment in non-responding tumours so as to go primarily to surgery in these cases, while applying chemotherapy only to those patients that benefit most.

As treatment in Wilms Tumours starts in the SIOP trials without histological proven diagnosis the prediction of a correct diagnosis and the response to preoperative chemotherapy is of highest clinical relevance. By integrating all available data at the time of diagnosis in the so called 'Oncosimulator' the prediction of a correct diagnosis and the response to the preoperative chemotherapy will be given as a result. This result of the Oncosimulator can be validated with the real diagnosis and the achieved response to the given treatment in an individual patient. Input data at the time of diagnosis are:

- molecular biology data from serum

  autoantibodies against nephroblastoma
- imaging studies with data from tumour rendering
- predicted pharmacokinetic data of vincristine and actinomycin-D
- clinical data

A schema of the scenario is given in figure C1

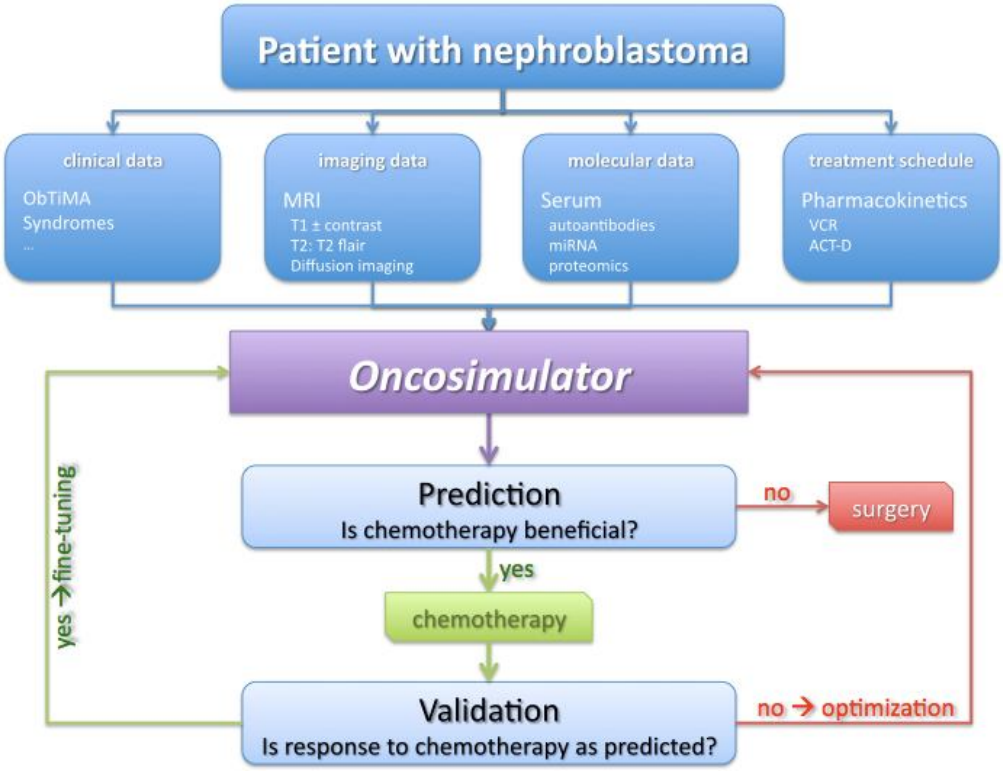


Figure C1: Schema of the Wilms Tumour VPH scenario including further developments of more molecular data.

The following molecular biological data will be provided for this scenario.

1. Autoantibodies against Wilms tumour.

It is well known that tumours develop autoantibodies against tumour specific antigens. Our group could show this for nephroblastoma. First results are already presented to the scientific community [C40-C41]. The analysis of these autoantibodies is established and will be part of the next SIOP nephroblastoma trial.

In future developments this dataset will be enlarged by the following data:

2. miRNA in serum and urine of patients with Wilms tumour

For many diseases miRNA serves as tumour specific markers. We are currently analyzing miRNA in serum of patients with Wilms tumour. First results lend strong support to the idea of using specific miRNA profiling of human blood as a diagnostic tool [C42].

3. Platform for clinical proteomics applied to blood and tissue profiling

A platform for identifying the disease-proteome signature will be established to define protein expression patterns that can identify specific phenotypes (diagnosis), establish a patient's specific outcome independent of treatment (prognosis) and predict a potential outcome from the effects of a specific therapy (prediction). For the benefit of a personalized medicine this platform requires the proteomic tools, the 'hardware', and the 'software', to extract meaningful statistical and biological information from samples, which are defined by hundreds or, thousands of measurements.

The following imaging studies will be provided for this scenario:

MRI (T1 with and without contrast enhancement, T2, T2 flair and diffusion weighted imaging) at the time of diagnosis and after 4 weeks of preoperative chemotherapy will serve as the input images. These data

needs pre-processing before entering the Oncosimulator to get information of tumour volume and morphology. DoctorEye will be used to segment the tumour.

The following clinical data and proposed pharmacokinetic data will be provided for this scenario [C43]:

These data will be available from ObTiMA and open source databases (pharmacokinetics)

All original and pre-processed data will be used after pseudonymisation and according to the legal and ethical framework of ACGT.

An outline of the version of the mathematical model underlying the ACGT simulator under consideration has been given by Georgiadi ECh et al [C44]. The simulation model application (ACGT Oncosimulator) can be used as a standalone tool but it can also benefit from the Grid environment that is provided by the ACGT infrastructure making the simulation process much faster. Grid computing is the application of several computers to a single problem at the same time. It depends on software to divide and apportion pieces of a program among several computers, sometimes up to many thousands. Grid computing can also be thought of as distributed and large-scale cluster computing, as well as a form of network-distributed parallel processing [C45-C46].

In the first version of the model a macroscopically homogeneous tumour is considered as an initial approximation to a real nephroblastoma. The model will therefore be applied only to such tumours that are found to be homogeneous in imaging studies. In future versions the model will be expanded to inhomogeneous tumours. The initial simulation conditions of the tumour of an individual patient are acquired using a segmentation tool to demarcate the tumour and get data on volume and the shape of the tumour for the model (Fig. C2). The tumour is discretized using a cubic mesh (Fig. C3). Each elementary cube of the mesh is called a geometrical cell (GC) [C47-C50] and is used as the unit for the description of the biological activity of an imageable tumour [C51]. Each GC of volume $1mm^3$ is assumed to contain 10^6 biological cells.

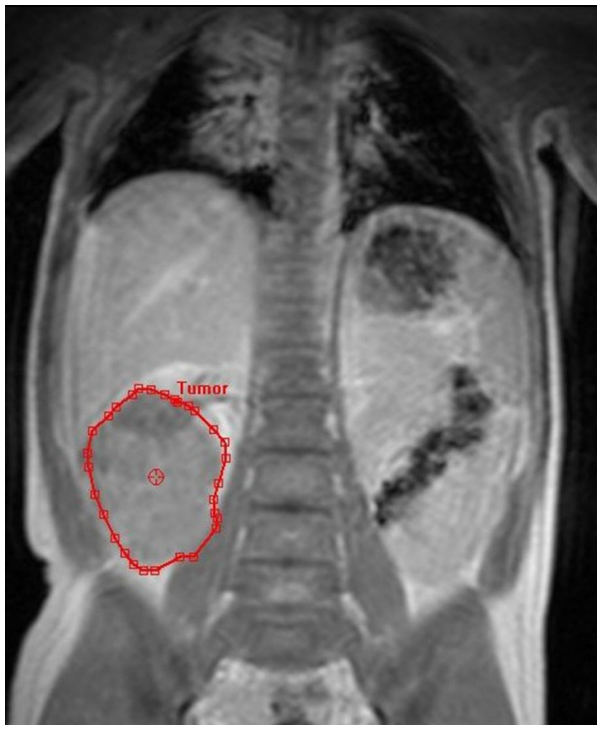


Fig. C2: Demarcating of a nephroblastoma by using a segmentation tool. The tool has been provided by the Fraunhofer Institute for Computer Graphics, Darmstadt, Germany.

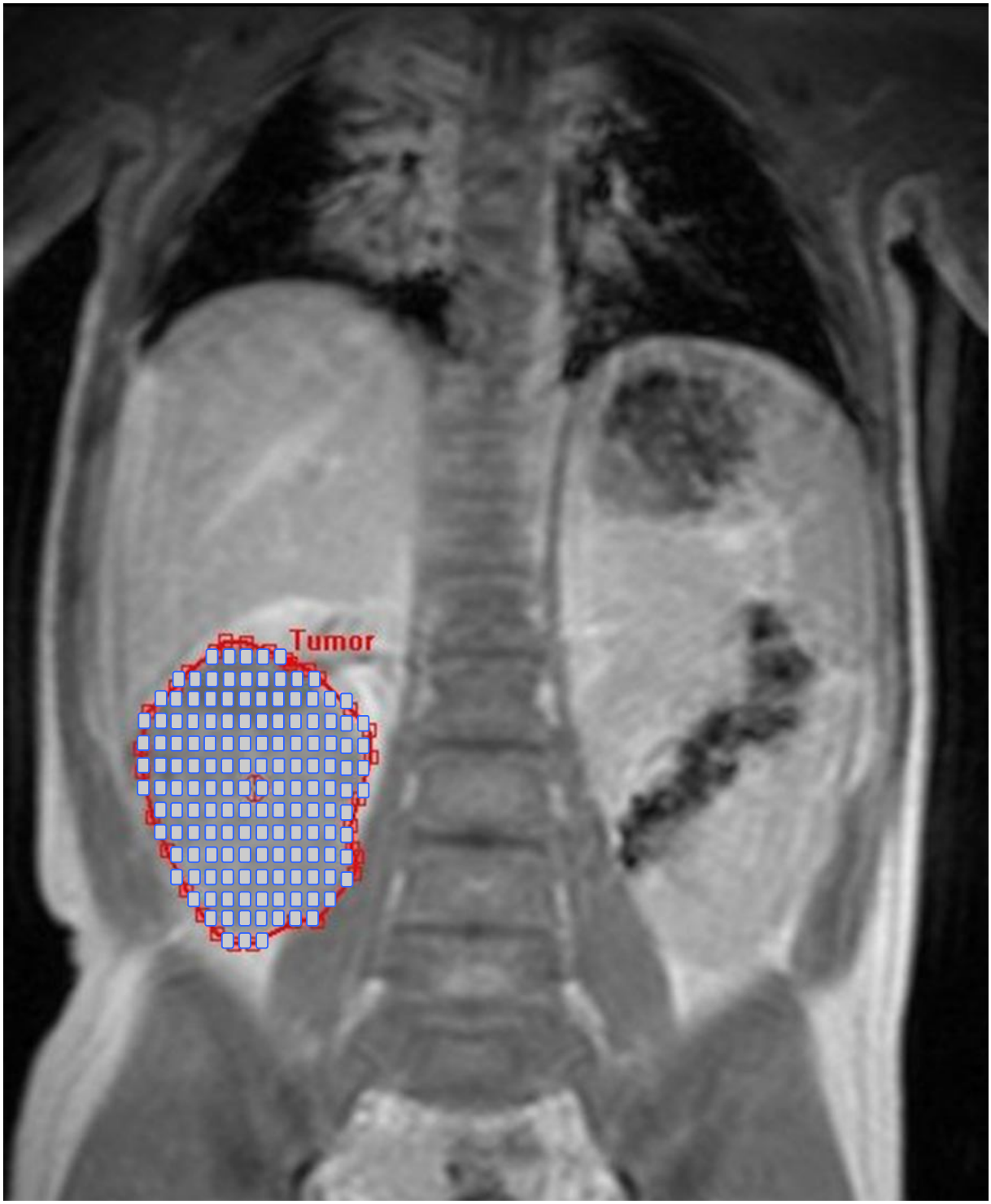


Fig. C3: Discretization of the tumour using a cubic mesh.

The following categories of tumour cells are considered:

- Stem cells
- Limited mitotic potential or progenitor cells (LIMP cells)
- Differentiated cells
- Dead cells

Proliferating stem cells and LIMP cells are further distinguished according to their cell cycle phase. Dead cells are divided into necrotic and apoptotic [C52]. The simulation algorithms developed so far address the preoperative chemotherapy with a combination of actinomycin-D and vincristine in unilateral stage I-III nephroblastoma treated according to the SIOP 2001/GPOH clinical trial. According to this protocol, i.v. vincristine bolus injection is directly followed by an i.v. bolus injection of actinomycin-D. The corresponding cell kill fractions computed according to the pharmacodynamics of each drug are added in order to acquire the total cell kill fraction. In a future version of the model the individual patient's serum immune response to specific tumour antigens will be added. If the autoantigen pattern correlates with tumour histology (blastemal, epithelial, stromal cell fractions) this will considerably affect chemotherapeutic responsiveness.

In order to effectively simulate tumour expansion or shrinkage a provisionally acceptable upper limit (NBC-upper) and a provisionally acceptable lower limit (NBC-lower) of the number of cells contained within each GC are defined. If the number of tumour cells within a given GC becomes less than NBC-lower during the simulation process a procedure starts that attempts to "unload" the remaining cells in the neighbouring GCs. If the given GC becomes empty it is "removed" from the tumour. An appropriate shift of a chain of GCs intended to fill in the "vacuum" leads to tumour shrinkage. Similarly differential tumour expansion is achieved by an appropriate shift of a chain of GCs towards the boundaries of the tumour.

To run the ACGT Oncosimulator for preoperative chemotherapy in nephroblastoma the data as described above are needed. A summary of the model parameters and their values considered is provided in table C1.

Table C1: Main clinical parameter values used for the prediction of the nephroblastoma tumour shrinkage shown in figures C4a, C4b and C5 following chemotherapy with vincristine and actinomycin-D. Technical parameters are not listed in this table.

PARAMETER NAME	PARAMETER VALUE
Cell cycle duration (h)	23.0
G0 phase duration (h) for stem cells	96
G0 phase duration (h) for limp cells	96
Spontaneous apoptosis rate [*fraction of cell number per hour*]	0.001
Spontaneous apoptosis rate for differentiated cells [*fraction of cell number per hour*]	0.003
Necrosis rate for differentiated cells [*fraction of cell number per hour*]	0.001
Fraction of dormant cancer stem cells re-entering the G1 phase per hour	0.01
Fraction of dormant cancer progenitor cells re-entering the G1 phase per hour	0.01

Fraction of cells in the necrotic layer that will enter G0 following mitosis	0.28
Fraction of cells in the proliferative layer that will enter G0 following mitosis	0.28
Fraction of the stem cells in the necrotic tumour layer that divide symmetrically	0.42
Fraction of the stem cells in the proliferative tumour layer that divide symmetrically	0.42
Number of mitoses that a progenitor cell undergoes before it becomes differentiated	3
Number of geometrical cells along the x axis	51
Number of geometrical cells along the y axis	52
Number of geometrical cells along the z axis	61
Number of cells contained within a geometrical cell (GC)	$8*10^6$
Which drug or drug combination is to be simulated? (1=vincristine, 2=actinomycin-D, 3=combination of vincristine and actinomycin-D)	3
Will geometric reconstruction be included in the simulations? (1: yes, 2:no)	1
Interval between two subsequent administrations of vincristine (h)	168
Time point of the first vincristine administration since initialization (h)	96
Number of consecutive vincristine sessions to be simulated	4
Cell kill ratio of stem cells for a specific vincristine dose	0.3
Cell kill ratio of limp cells for a specific vincristine dose	0.3
Interval between two subsequent administrations of actinomycin-D (h)	336
Time point of the first actinomycin-D administration since initialization (h)	96
Number of consecutive actinomycin-D sessions to be simulated	2
Cell kill ratio of stem cells for a specific Dactinomycin dose	0.2
Cell kill ratio of limp cells for a specific Dactinomycin dose	0.2
Time interval between last treatment session and post-treatment scan (h)	72

Some of the parameter values such as the cell cycle duration have been based on pertinent literature whereas others have been based on both qualitative data and logic. **Only values relevant to the particular type of tumour are used. In the case that different values are available for a given parameter, all are considered in the simulations.** Extensive use of pseudonymized actual SIOP 2001/GPOH clinical trial data is done to considerably refine the parameter value assignment. In order to simulate a realistic treatment scenario chemotherapy is assumed to start not later than 4 days after the pre-treatment imaging data are collected. Simulation continues up to 3 days after the last chemotherapy administration. This also represents a real SIOP 2001/GPOH case.

In this scenario the most important prediction of the ACGT Oncosimulator for the clinician is the amount of shrinkage of the given tumour in % of the initial volume. Provided that an eventually favourable clinical

validation of the ACGT Oncosimulator will have been accomplished in the future, one of the following two situations will arise based on the ACGT Oncosimulator prediction:

1. Prediction of reduction of tumour volume by more than 10%.

- In this case the clinician would judge that preoperative chemotherapy would be beneficial to the patient.

2. Prediction of reduction of tumour volume by less than 10% or no reduction of the tumour, or even increase in tumour volume.

- In this case the clinician would judge that preoperative chemotherapy would not be beneficial to the patient.

However, before completion of the clinical validation process the patient will always receive preoperative chemotherapy independent of the above judgement. The actual chemotherapy administration schedule is then registered. At the end of the preoperative treatment the reduction of the tumour size in reality is compared with the predicted reduction by the '*in silico'* experiment (Table C2).

Table C2: The initial and the final clinical and simulated volumetric data

CLINICAL VOLUMETRIC DATA	**Reality**	***In Silico* Experiment**
Initial volume of the tumour in mm^3	15236	15236
Final volume of the tumour in mm^3	4184	3917
Volume shrinkage percentage	72.54%	74.29%

A "perfect correlation" between the ACGT Oncosimulator and the in vivo situation for the nephroblastoma case is defined by getting the same result of more or less than 10% tumour volume reduction by both methods in every patient and by predicting the correct histological subtype (Table C3).

Table C3: Correlation of the tumour volume reduction between the in vivo situation and the in silico experiment

		In vivo	
		< 10 %	*> 10 %*
Oncosimulator	*< 10 %*	*good correlation / no preoperative chemotherapy is indicated*	*bad correlation / Oncosimulator has to be improved*
	> 10 %	*bad correlation / Oncosimulator has to be improved*	*good correlation / preoperative chemotherapy is indicated*

If there is no "perfect correlation" the ACGT Oncosimulator will be further tuned in a feedback loop as a learning system. Tumour volume response and histological outcome have to be evaluated independently. For the clinical situation the prediction of the correct tumour volume reduction before starting treatment is most important. This will help to choose the best treatment for an individual patient in advance. Before applying the prediction of the ACGT Oncosimulator in the clinical setting a prospective trial has to prove the correctness of the predictions in every patient.

The ACGT Oncosimulator combines tumour information obtained from medical imaging techniques (CT, MRI, PET and ultrasound) with mathematical models that predict the growth of tumours and the response to chemotherapy in the preoperative phase of nephroblastoma. The ACGT Oncosimulator produces predictions of the composition and shape of the tumour over the course of time. These predictions provide clinicians with valuable information on the most effective treatment out of several alternatives, as well as detailed parameters on the optimal composition of a treatment scheme, including the total treatment period, the type of drug(s), dose, and interval between treatments.

In a feedback, learning loop by comparing the prediction with the actual tumour response the ContraCancrum platform will be optimized. A nine-step scenario can be generalized to any cancer type, as shown in Table C4:

Table C4 Generalized clinical scenario for optimizing in Silico studies

	Action
1	Obtain patient's specific data
2	Preprocess patient's data
3	Describe a number of candidate therapeutic schemes and/or schedules
4	Run the simulations either sequentially or in parallel (e.g. on a grid platform)
5	Visualize and report the predictions
6	Evaluate the predictions and decide on the optimal scheme to be applied
7	Apply the optimal therapeutic scheme to individual patients

8 Compare the prediction of the simulation with the achieved result in th individual patients

9 Further optimize the ContraCancrum integrated simulator by tuning the mode with the post-treatment patient data

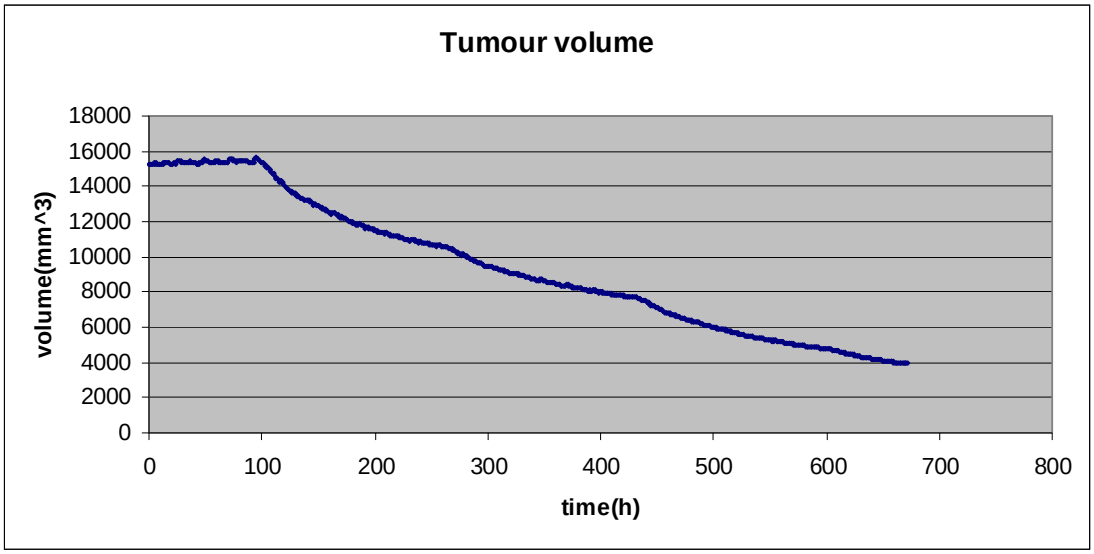


Fig. C4a: Nephroblastoma tumour volume reduction as a function of time for the preoperative chemotherapy with vincristine and actinomycin-D for 4 weeks based on one real patient. The parameter values of table 1 are included in the experiment.

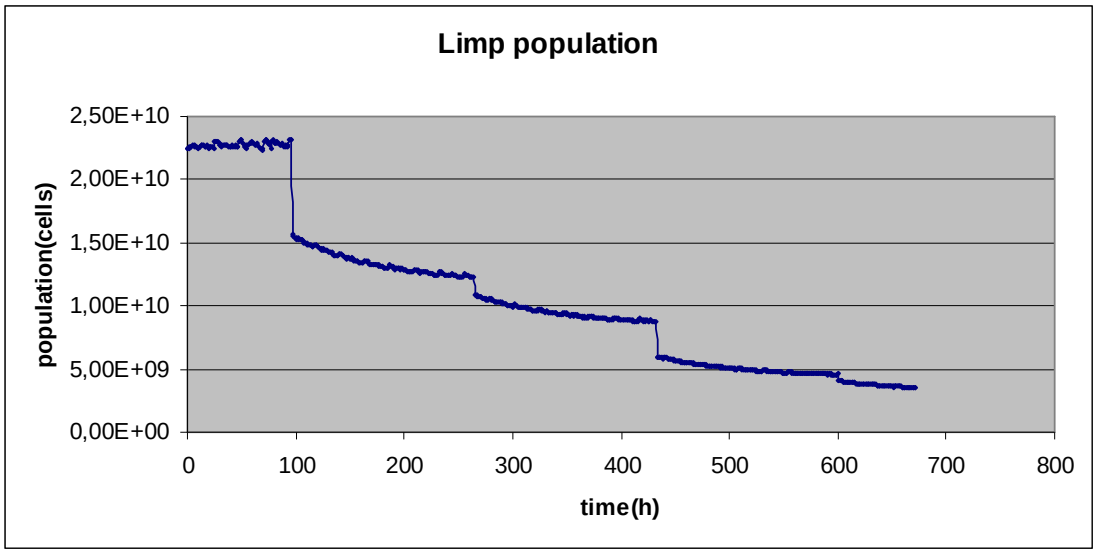


Fig. C4b: Time course of the limited mitotic potential cells (limp or progenitor cells) as an example of the various cell categories addressed for the tumour considered in figure 4a based on one real patient.

Figures C4a, C4b and C5 show the results of one '*in silico'* experiment in a single concrete patient with a nephroblastoma. Further details focusing on the model itself rather on its applicability are provided in chapter 4. The output of a simulator run is a volume for each hour of simulated time. Other ACGT Oncosimulator outputs include counts of the different cell categories used internally by the simulator. This cell count is best represented in 2D time-versus-count plots (fig C4b). The volume output is represented by 3D isosurfaces or volume rendering (Fig. C5). Different visualization methods for the ACGT Oncosimulator are possible including e.g. line graphs showing the predicted change in overall tumour volume or shape or even the predicted composition of the tumour.

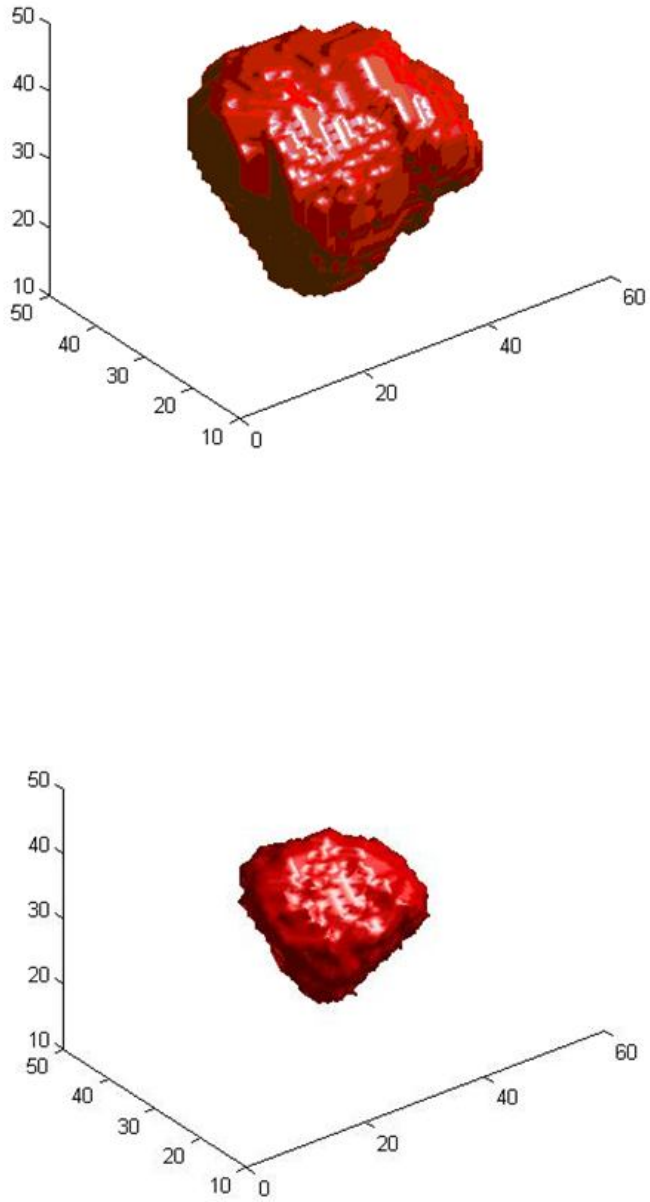


Fig. C5: 3D tumour visualization based on one real patient. Upper panel: tumour four days before treatment initiation, lower panel: predicted tumour three days after completion of the treatment course.

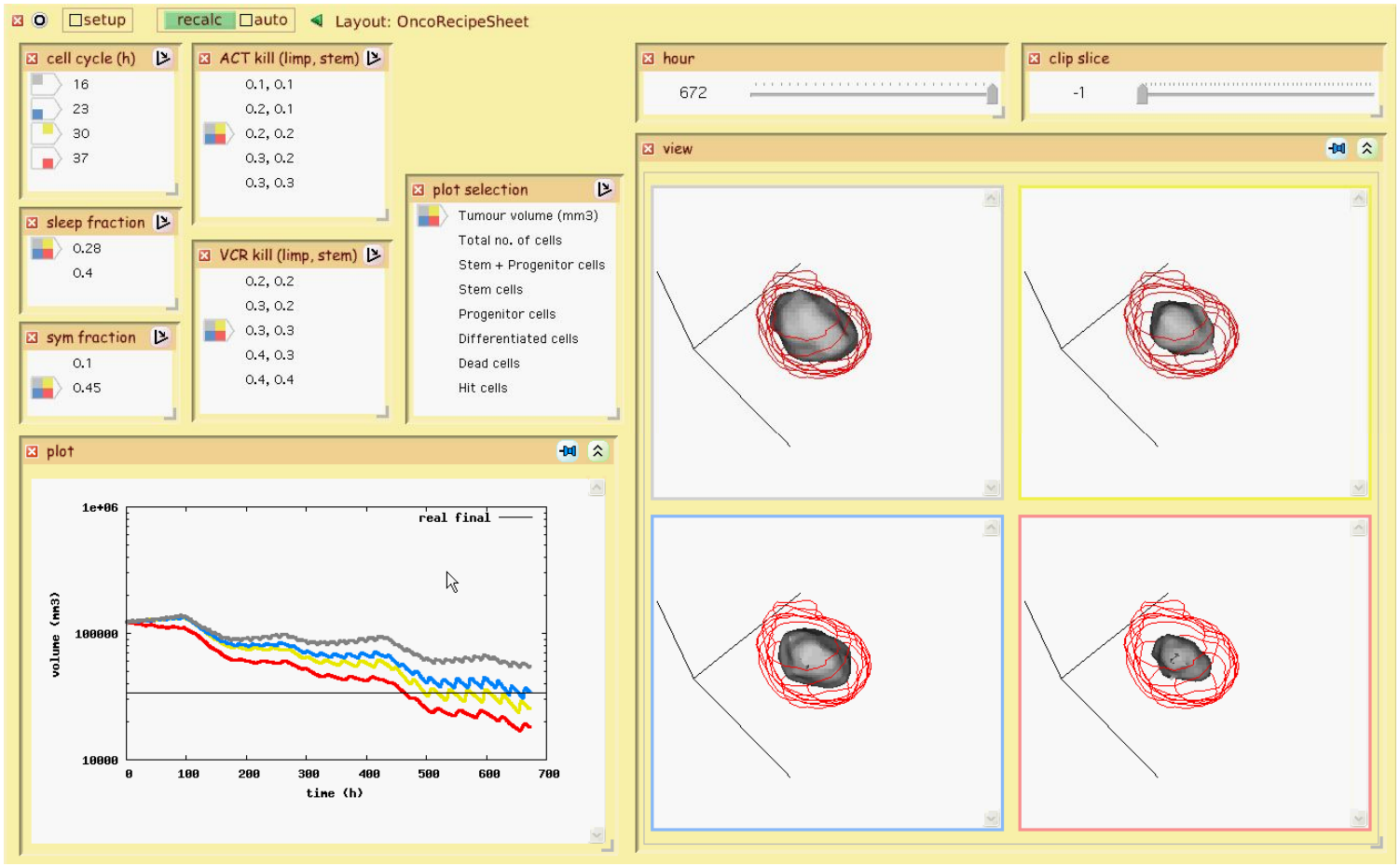


Fig. C6: Screenshot of the visualization services for one real patient used in the RecipeSheet. The 2D graph in the lower left and the four 3D views on the right are generated by visualization services. In this specific run the duration of the cell cycle has been put as a variable.

To support interactive exploration of treatment scenarios by using different values for the parameters given in Table C1 the simulation can be invoked and rerun through a specialized application built on a spreadsheet-inspired programming environment (RecipeSheet) [C53]. . Through such a sheet one or more simulator jobs can be set up and started. When each of the simulator runs completes, its output is transferred to the RecipeSheet [C54]. Figure C6 shows the visualization services integrated with the RecipeSheet. In this specific case of one real patient the output is given for four simulations based on alternative values for the duration of the cell cycle. The RecipeSheet merges the 2D line graphs from each simulation, and also maintains synchronization among the four 3D tumour views on the right of the sheet.

It is noted that the starting parameter values of the model are based on pertinent available literature (providing average parameter values and/or value intervals) or reasonable quantitative assumptions concerning those parameters for which literature provides qualitative rather than exact quantitative data. Known restrictions regarding the interrelation of several parameters (e.g. higher resistance of stem cells to treatment in relation to progenitor cells) are taken into account. As more and more sets of medical data are exploited new parameter values based on the previously exploited data and literature can arise. This will be amenable to further tuning the Oncosimulator.

3.3 Description of the latest Oncosimulator developments from the clinical perspective

Data of more patients with Wilms tumour will be provided for the Oncosimulator. Segmentation will be done using DoctorEye in the future [C55] . Histograms of the signal Intensities of MRI images before and after preoperative chemotherapy will be analyzed and correlated to histological findings to get better input data for the Oncosimulator. Is has to be taken into consideration that there is a great variability inter- and intra-individual when dealing with histologies as shown by Bueno-de-Mesquita et al. for breast cancer [C56]. Molecular data will be enhanced by miRNA and proteomics.

3.4 Clinical views, comments and suggestions on the Oncosimulator integration into the ACGT architecture

At the time a new patient is diagnosed with Wilms tumour, the treating physician will send images to the trial centre, where segmentation of the tumour is done and automatically send to the Oncosimulator. Blood will be analyzed in real time for the molecular parameters and results will be automatically entered in the Oncosimulator. Clinical data are provided by ObTiMA or any other data management system to which the Oncosimulator has a direct connection. All data will be automatically pseudonymized before used in the Oncosimulator. As soon as all necessary data are available in the Oncosimulator the model will be executed and results will be presented to the treating physician and the trial centre. After preoperative chemotherapy the prediction of the Oncosimulator will be compared with the reality and a learning loop will be started for optimization of the Oncosimulator. For the treating physician a user friendly portal to the Oncosimulator via ACGT is needed. To optimize the system there should be only one access point for a treating physician to the system, including a DICOM Server, the data management system and the Oncosimulator.

3.5 Clinical views, comments, and suggestions on the validation and future clinical translation of the Oncosimulator

For a part of Wilms tumour patients the Oncosimulator should be integrated in the next clinical trial for nephroblastoma. In that case the Oncosimulator should be tested prospectively. For that purpose ObTiMA will be enhanced with a DICOM Server and a link to DoctoryEye. Histograms of segmented Wilms tumours will be analyzed and their added value tested. A concrete plan for the learning loop needs to be developed. Wilms Tumour can be seen as a test case for other tumours.

3.6 Conclusions

The question "What is the best treatment for a given tumour in an individual patient?" leads to a high level of individualization. To attain this goal it is of utmost importance that the result of the '*in silico'* experiment is available in a short timeframe after diagnosis. This implies that all data that are necessary for running the '*in silico'* experiment have to be available in a timely manner. This is especially important for molecular biologists, radiologists and clinicians, who have to produce reliable data very fast. The experiment run by the ACGT Oncosimulator itself is not a time-consuming experiment [C57].

The combined chemotherapy simulation model is validated and optimized using pseudonymized data from the SIOP 2001/GPOH nephroblastoma clinical trial as a proof of principal. The model presented has been shown to be able to reproduce important aspects of the clinical reality and practice and generally produce reasonable predictions. Nevertheless, in order to enter routine clinical practice as a decision making tool, an exhaustive validation, adaptation, and optimization procedure has to take place. Furthermore, molecular methods of extraction of the crucial histological constitution of the tumour will be tested and integrated into the model. Following completion of this testing procedure the simulation model is expected to support clinicians' decisions concerning various candidate cancer treatment schemes and thus facilitating individualized treatment optimization, help suggest new therapeutic strategies based on '*in silico'* experiments, and help train or inform doctors, life scientists, researchers or interested patients by demonstrations of the likely tumour response to different therapeutic schemes.

The successful performance of the combined ACGT Oncosimulator platform as reviewed in this chapter is seen as a strongly encouraging step towards the clinical translation of such integrated systems. The Oncosimulator can be adjusted to any other cancer type using data as described in this review [C58-C60]. It is always necessary to go through a learning loop until the prediction of the simulator is corresponding to the reality of the tumour response in an individual patient. The ACGT Oncosimulator described here is the first of its kind worldwide.

3.7 References

[C1] Graf N, Tournade MF, deKraker J: The role of preoperative chemotherapy in the management of Wilms' Tumour. Urologic Clin North Am 27:443-454, 2000

[C2] Sofra N, Stamatakos G, Graf N, Uzunoglu N: A four dimensional simulation model of the in vivo response of nephroblastoma to vincristine. Edts.: Kostas M, Stamatakos G: Proceedings: 2nd International Advanced Research Workshop on In Silico Oncology (IARWISO), Kolympari, Chania, Greece , 25th and 26th September 2006

[C3] Stamatakos GS, Dionysiou DD, Graf N, Sofra NA, Desmedt C, Hoppe A, Uzunoglu NK, Tsiknakis M: "The “Oncosimulator”: a multilevel, clinically oriented simulation system of tumour growth and response to therapeutic schemes. Towards clinical evaluation of in silico oncology." Proceedings of the 29th Annual International Conference of the IEEE EMBS, Cité Internationale, Lyon, France, August 23-26, 2007, SuB07.1, S. 6628-6631; Conf Proc IEEE Eng Med Biol Soc. 2007;1:6628-16631

[C4] Kolokotroni EA, Stamatakos GS, Dionysiou D, Georgiadi ECh, Desmedt C, Graf N: Translating Multiscale Cancer Models into Clinical Trials: Simulating Breast Cancer Tumour Dynamics within the Framework of the “Trial of Principle” Clinical Trial and the ACGT Project. Proceedings of the 8th IEEE International Conference on Bioinformatics and Bioengineering (BIBE 2008), Athens, Greece, 8-10 October 2008. IEEE Catalog Number: CFP08266, ISBN: 978-1-4244-2845-8, Library of Congress: 2008907441, Paper No. BE-2.1.1, length: 8 pages (in electronic format)

[C5] Georgiadi ECh, Kolokotroni EA, Dionysiou DD, Graf N, Hoppe A, Uzunoglu NK, Stamatakos GS: Simulating the response of Nephroblastoma Tumour to chemotherapy in the clinical context. Proceedings of the 3rd International Advanced Research Workshop on In Silico Oncology, Istanbul, Turkey. Sept. 23-24, 2008: pp 27-30

[C6] Georgiadi ECh, Stamatakos GS, Graf N, Kolokotroni EA, Dionysiou DD, Hoppe A, Uzunoglu NK: Multilevel Cancer Modeling in the Clinical Environment: Simulating the Behaviour of Wilms Tumour in the Context of the SIOP 2001/GPOH Clinical Trial and the ACGT Project. Proceedings of the 8th IEEE International Conference on Bioinformatics and Bioengineering (BIBE 2008), Athens, Greece, 8-10 October 2008. IEEE Catalog Number: CFP08266, ISBN: 978-1-4244-2845-8, Library of Congress: 2008907441, Paper No. BE-2.1.2, length: 8 pages (in electronic format)

[C7] Deisboeck TS et al. In silico cancer modeling: is it ready for prime time? Nat Clin Pract Oncol. 2009; 6:34-42

[C8] Axelrod R et al. Evolution of cooperation among tumour cells. Proc Natl Acad Sci USA 2006; 103:13474–13479

[C9] Gatenby RA,Vincent TL. An evolutionary model of carcinogenesis. Cancer Res 2003; 63:6212–6220

[C10] Mansury Y et al. Evolutionary game theory in an agent-based brain tumour model: exploring the ‘Genotype-Phenotype’ link. J Theor Biol 2006; 238:146–156

[C11] Guiot C et al. Does tumour growth follow a “universal law”? J Theor Biol 2003; 225:147–151

[C12] Guiot C et al. The dynamic evolution of the power exponent in a universal growth model of tumours. J Theor Biol 2006; 240:459–463

[C13] Baish JW, Jain RK. Fractals and cancer. Cancer Res 2000; 60:3683–3688

[C14] Cross SS. Fractals in pathology. J Pathol 1997; 182: 1–8

[C15] Norton L. Conceptual and practical implications of breast tissue geometry: toward a more effective, less toxic therapy. Oncologist 2005; 10:370–381

[C16] Goh KI et al. The human disease network. Proc Natl Acad Sci USA 2007; 104:8685–8690

[C17] Marusic M et al. Analysis of growth of multicellular tumour spheroids by mathematical models. Cell Prolif 1994; 27:73–94

[C18] Vaidya VG, Alexandro FJ Jr. Evaluation of some mathematical models for tumour growth. Int J Biomed Comput 1982; 13:19–36

[C19] Alarcon T et al. A cellular automaton model for tumour growth in inhomogeneous environment. J Theor Biol 2003; 225:257–274

[C20] Bauer AL et al. A cell-based model exhibiting branching and anastomosis during tumour-induced angiogenesis. Biophys J 2007; 92:3105–3121

[C21] McDougall SR et al. Mathematical modelling of dynamic adaptive tumour-induced angiogenesis: clinical implications and therapeutic targeting strategies. J Theor Biol 2006; 241:564–589

[C22] Picci P et al: Prognostic significance of histopathologic response to chemotherapy in nonmetastatic Ewing's sarcoma of the extremities. J Clin Oncol 1993; 11:1763-1769

[C23] Sole RV, Deisboeck TS. An error catastrophe in cancer? J Theor Biol 2004; 228:47–54

[C24] Spencer SL et al. Modelling somatic evolution in tumourigenesis. PLoS Comput Biol 2006; 2:e108

[C25] Castiglione F et al. Computational modeling of the immune response to tumour antigens. J Theor Biol 2005; 237:390–400

[C26] Bogaerts J et al. Individual patient data analysis to assess modifications to the RECIST criteria. EJC 2009; 45:248-260

[C27] Therasse P et al. New guidelines to evaluate the response to treatment in solid tumours. J Natl Cancer Inst 2000; 92:205-16

[C28] Huff CA et al. The paradox of response and survival in cancer therapeutics. Blood 2006; 107:431-434

[C29] Graf N, Hoppe A. What are the expectations of a Clinician from In Silico Oncology? In: Marias K, Stamatakos G (Hrsg). Proc. 2nd International Advanced Research Workshop on In Silico Oncology, Kolympari, Chania, Greece, 25-26 Sept. 2006, pp. 36-38 (http://www.ics.forth.gr/bmi/2nd-iarwiso/)

[C30] Göbel U, Jürgens H. Translation der kliniknahen Grundlagenforschung in die pädiatrische Onkologie. Klin Pädiatr 2003; 215:289-290

[C31] Göbel U, Witt O. Das Dilemma der klinischen Register in der pädiatrischen Onkologie und Hämatologie. Klin Pädiatr 2008; 220:129-133

[C32] Graf N, Göbel U: Therapieoptimierungsstudien der Gesellschaft für Pädiatrische Onkologie und Hämatologie (GPOH) und 12. Novelle des Arzneimittelgesetzes zur Umsetzung der EU-Richtlinie. Klin Pädiatr 2004; 216:129-131

[C33] DeKraker J et al. Reduction of postoperative chemotherapy in children with stage I intermediate-risk and anaplastic Wilms' tumour (SIOP 93-01 trial): a randomised controlled trial. Lancet 2004; 364:1229-1235

[C34] Graf N et al. The role of preoperative chemotherapy in the management of Wilms Tumour – The SIOP Studies. Urol Clin North Am 2000; 27:443-454

[C35] Reinhard H et al. Results of the SIOP 93-01/GPOH trial and study for the treatment of patients with unilateral nonmetastatic Wilms Tumour. Klin Pädiatr 2004; 216:132-140

[C36] Rosen G: Neoadjuvant chemotherapy for osteogenic sarcoma: a model for the treatment of other highly malignant neoplasms. Recent Results Cancer Res 1986; 103:148-157

[C37] Winkler K, Bielack SS, Delling G, et al: Treatment of osteosarcoma: experience of the Cooperative Osteosarcoma Study Group (COSS). Cancer Treat Res 1993; 62:269-277

[C38] Tournade MF et al: Results of the Sixth International Society of Pediatric Oncology Wilms´ Tumour Trial and Study: A Risk-Adapted Therapeutic Approach in Wilms´ Tumour. J Clin Oncol 1993; 11:1014-1023

[C39] Hero B et al. Neuroblastoma preoperatively treated as nephroblastoma: does inadequate therapy worsen the prognosis. Klin Pädiatr 2002; 214:157-161

[C40] Nourkami N, Fischer U, Leidinger P, Heisel S, Habel N, Hoppe A, Graf N, Meese E: Immune response pattern in Wilms Tumour patients: New biomarkers for early diagnosis of malignant childhood tumours. 7th International Meeting on the Biology of Childhood Renal Tumours. Banff; 1st – 3rd of March 2010

[C41] Heisel S, Habel NC, Hoppe A, Keller A, Nourkami N, Berthold F, Lenhof HP, Gessler M, Graf N, Meese E: Identification of serological markers and generation of autoantibody signatures to improve differential diagnosis of Wilms and Non-Wilms tumours. 7th International Meeting on the Biology of Childhood Renal Tumours. Banff; 1st – 3rd of March 2010

[C42] Keller A, Leidinger P, Bauer A, ElSharawi A, Haas J, Borries A, Wendschlag A, Giese N, Tjaden Ch, Nikolaus S, Ruprecht K, Huwer H, Huebers J, Jacobsen G, Rosenstiel P, Sina Ch, Wullich B, Graf N, Reichrath J, JagerSU, Staehler P, Staehler C, Beier M, Scheffler M, Buechler MW, Wischhusen J, Häusler S, Dietl J, Mueller-Quernheim J, Backes CH, Lenhof HP, Schreiber S, Katus HA, Rottbauer W, Meder B, Franke A, Hoheisel J, Meese EmiRNA signatures of human blood – promising biomarkers for human diseases. Submitted, 2010

[C43] Graf N, Hoppe A, Georgiadi E, Bellemann R, Desmedt C, Dionysiou D, Erdt M, Jacques J, Kolokotroni E, Lunzer A, Tsiknakis M, Stamatakos G: 'In Silico' oncology for clinical decision-making in the context of nephroblastoma. Klin Pädiatr 221:141-149, 2009

[C44] Georgiadi ECh et al. Multilevel Cancer Modeling in the Clinical Environment: Simulating the Behaviour of Wilms Tumour in the Context of the SIOP 2001/GPOH Clinical Trial and the ACGT Project. Proceedings of the 8th IEEE International Conference on Bioinformatics and Bioengineering (BIBE 2008), Athens, Greece, 8-10 October 2008. IEEE Catalog Number: CFP08266, ISBN: 978-1-4244-2845-8, Library of Congress: 2008907441, Paper No. BE-2.1.2, length: 8 pages (in electronic format)

[C45] Buyya, Rajkumar. Grid Computing: Making the Global Cyberinfrastructure for eScience a Reality. CSI Communications (Mumbai, India: Computer Society of India (CSI)) 2005; 29 (1). ISSN 0970-647X. http://www.gridbus.org/~raj/papers/CSICommunicationsJuly2005.pdf

[C46] GridCafe. http://www.gridcafe.org/version1/whatisgrid/whatis.html

[C47] Dionysiou DD et al. A four-dimensional simulation model of tumour response to radiotherapy in vivo: parametric validation considering radiosensitivity, genetic profile and fractionation. J Theor Biol 2004; 230:1–20

[C48] Stamatakos GS et al. In silico radiation oncology: combining novel simulation algorithms with current visualization techniques. Proceedings of IEEE: Special Issue on Bioinformatics: Advances and Challenges 2002; 90:1764-1777

[C49] Stamatakos GS et al. A spatiotemporal, patient individualized simulation model of solid tumour response to chemotherapy in vivo: the paradigm of glioblastoma multiforme treated by temozolomide. IEEE Transactions on Biomedical Engineering 2006; 53:1467-1477

[C50] Stamatakos GS et al. The Oncosimulator: a multilevel, clinically oriented simulation system of tumour growth and organism response to therapeutic schemes. Towards the clinical evaluation of in silico oncology. Proceedings of the 29th Annual International Conference of the IEEE EMBS, Lyon, France, 2007; SuB07.1:6628-6631

[C51] Lodish D et al. Molecular Cell Biology. New York: Scientific American Books, 1995; pp 1247–1294

[C52] Georgiadi ECh et al. Multilevel Cancer Modeling in the Clinical Environment: Simulating the Behaviour of Wilms Tumour in the Context of the SIOP 2001/GPOH Clinical Trial and the ACGT Project. Proceedings of the 8th IEEE International Conference on Bioinformatics and Bioengineering (BIBE 2008), Athens, Greece, 8-10 October 2008. IEEE

Catalog Number: CFP08266, ISBN: 978-1-4244-2845-8, Library of Congress: 2008907441, Paper No. BE-2.1.2, length: 8 pages (in electronic format)

[C53] Lunzer A et al. RecipeSheet: Creating, Combining and Controlling Information Processors. Proceedings of the 19th Annual ACM Symposium on User interface Software and Technology (UIST '06), Montreux, Switzerland, ACM Press, Oct 2006, 145-153

[C54] Belleman RG et al. Interactive Simulation and Visualization for Cancer Treatment Planning with Grid-based Technology, ERCIM News, special theme: The Digital Patient. 2007; 69:22-23

[C55] Emmanouil Skounakis, Vangelis Sakkalis, Kostas Marias, Konstantinos Banitsas, Graf N: DoctorEye: A Multifunctional Open Platform for Fast Annotation and Visualization of Tumours in Medical Images. Conference: 31st Annual International IEEE EMBS Conference Schedule Code: FrDPo04.44. Conf Proc IEEE Eng Med Biol Soc. 2009;1:3759-3762

[C56] Bueno-de-Mesquita JM et al.: The impact of inter-observer variation in pathological assessment of node-negative breast cancer on clinical risk assessment and patient selection for adjuvant systemic treatment. Ann Oncol 21:40-47, 2010

[C57] Lavenier D, Jacques J. Parallelizing the ACGT Oncosimulator. Proceedings of the 3rd International Advanced Research Workshop on In Silico Oncology, Zografeio Lyceum, Istanbul, Turkey, September 23/24, 2008; pp 38-40 http://www.3rd-iarwiso.iccs.ntua.gr/procs.pdf

[C58] Tsiknakis M, Rueping S, Martin L, Sfakianakis S, Bucur A, Sengstag T, Brochhausen M, Pucaski J, Graf N: Developing a European Grid infrastructure for cancer research: vision, architecture, and services. Ecancermedicalscience 1: DOI: 10.3332/eCMS.2007.56, 2007

[C59] Graf N, Desmedt C, Buffa F, Kafetzopoulos D, Forgó N, Kollek R, Hoppe A, Stamatakos G, Tsiknakis M: Post-genomic clinical trials – the perspective of ACGT. Ecancermedicalscience 1: DOI: 10.3332/eCMS.2007.66, 2008

[C60] Graf N: Clinical requirements regarding In Silico Oncology. Lecture from the European Union ICT BIO 2008 Conference (Oct. 23-24, 2008). Ecancermedicalscience; ecancer.tv 2009; Video: http://www.ecancermedicalscience.com/tv/?play=120

4. The basic simulation module of the ACGT Oncosimulator: an advanced discrete state / discrete event multiscale simulation model of the response of a solid tumour to chemotherapy. Mimicking a clinical study *(Code letter : S)*

4.1 Introduction

Over the last years it has become clear that in order to understand both the disease and the complex natural phenomenon of cancer in a quantitative way sophisticated multiscale models are needed. Furthermore, multiscale cancer models are expected to substantially support treatment optimization in the patient individualized context. To this end several research groups have been working on this field on the global level. The major modeling approaches can be distinguished into the following three categories: continuous mathematics based models, discrete mathematics based models and hybrid models.

Continuous mathematics based models rely primarily on the diffusion equation, which is applied in order to describe the diffusion of molecules such as oxygen and glucose and/or the invasion of tumour cells into the surrounding tissue(s) in the case of diffusive tumours (e.g. glioblastoma) (Murray, 2003; Swanson et al., 2002; Breward et al., 2003; Clatz et al., 2005; Frieboes et al., 2006; Guiot et al., 2006; Enderling et al., 2007). In many cases continuous equations are solved by finite difference methods such as the Crank-Nicholson method. Discrete mathematics based models rely primarily on the consideration of several discrete states in which cells may be found (e.g. cell cycle phases, stem cell state etc.) and the transitions between states (e.g. cell necrosis following a prolonged residence of a tumour cell in the G0 phase etc.) (Duechting and Vogelsaenger, 1981; Ginsberg et al., 1993; Stamatakos et al., 2002, 2006a, 2006b, 2007; Dionysiou et al., 2004, 2006 , Dionysiou and Stamatakos, 2006; Deisboeck et al., 2009). State transitions can be decided by making use of several possible “decision calculators” such as cytokinetic diagrams, agent-based techniques etc.

Hybrid methods exploit the potential of both continuous and discrete methods in order to address in more detail several tumour dynamics mechanisms. A clinically oriented example of the latter is the combination of pharmacokinetics differential equations with discrete state/event simulations in order to simulate the response of large imageable tumours to chemotherapeutic treatment (Stamatakos et al., 2006b). A further example refers to the modeling of cancer cell invasion to tissue (Ramis-Conde et al., 2008). As more and more biological mechanisms are being quantitatively understood most new models tend to fall in this category i.e. to make use of a combination of continuous and discrete mathematics. In that way the particular advantages of both approaches are exploited and therefore mechanisms of both predominantly continuous and predominantly discrete mechanisms can be addressed.

Hybrid models with a strong discrete character have been shown to be particularly adequate for the simulation of the response of large imageable clinical tumours to therapeutic interventions such as radiotherapy and chemotherapy (Stamatakos et al., 2002, 2006a, 2006b, 2007; Dionysiou et al., 2004, 2006). Since clinical

validation is a *sine qua non* necessity for clinically oriented multiscale cancer models, this chapter presents an actual clinical trial driven model concerning the response of early breast cancer to epirubicin. A branch of the Trial of Principle (TOP: Topoisomerase II Alpha Gene Amplification and Protein Overexpression Predicting Efficacy of Epirubicin) trial (TOP trial) has been addressed as a paradigm of a chemotherapy optimization targeted clinicogenomic trial.

Carcinoma of the breast is the most commonly diagnosed cancer in women. Even though a variety of effective treatments exists, their benefits and adverse effects vary considerably. Early results indicate that gene expression profiles may contribute to predicting response to breast cancer treatment (Potti et al., 2006). A number of clinical trials are running with the aim to identify gene expressions that correlate with chemotherapy results. The TOP trial, which is also part of the ACGT European project (ACGT), addresses a particular molecular subgroup of breast cancer patients and aims to determine the predictive rate of a single class of chemotherapeutic agent. More specifically, this study aims to evaluate the topoisomerase II alpha gene amplification and protein overexpression as markers predicting the efficacy of epirubicin monotherapy in the primary treatment of breast cancer patients. Inclusion criteria are non-metastatic, early breast cancer patients with estrogen receptor (ER) - negative tumours with a maximum dimension larger than 2 cm as defined by ultrasound. The patients are treated with single-agent epirubicin as neoadjuvant treatment, followed by surgery and adjuvant chemotherapy (Durbecq et al., 2004).

The model presented, although having its roots in previous work done and published by the *In Silico* Oncology Group, National Technical University of Athens (*In Silico* Oncology Group) is characterized by several crucial new features such as an advanced generic cytokinetic model for the response of a tumour cell to chemotherapeutic treatment, the explicit distinction of proliferating cells into stem cells and cells of limited mitotic potential and the adaptation of the relative populations of the various cell categories/phases to the cell category/phase transition rates (a novel tumour initialization strategy aiming at avoiding abnormal patterns of free tumour growth as described in subsequent paragraphs). Available imaging-based, histological, molecular and treatment data are exploited by the model in order to strengthen patient individualization modeling and in the future patient individualized treatment optimization. The expected use of the model following thorough clinical adaptation, optimization and validation is to simulate either several candidate treatment schemes for a particular patient and support the selection of the optimal one for him or her or to simulate the expected extent of tumour shrinkage for a given time instant and decide on the adequacy or not of the simulated scheme. The integrated simulation system has been termed “Oncosimulator” (Stamatakos et al., 2007). Two “oncosimulators” are currently being developed, clinically adapted using real clinical trial multiscale data and clinically validated within the framework of the EC funded projects ACGT (ACGT) and ContraCancrum (ContraCancrum).

The structure of the chapter is as follows. First the cytokinetic model proposed and adopted is presented. Subsequently the basic algorithms of the spatiotemporal discretization of the tumour are provided, followed by the tumour initialization strategy. Considerations on the pharmacokinetics and pharmacodynamics of epirubicin are presented. A description of the model parameters is provided. A sensitivity analysis of the model regarding the effect of critical parameters on the dynamics of the biological system is presented, followed by the description of an initial step towards the clinical adaptation and validation of the model. Simulations

for two real case studies from the TOP trial are presented and discussed. The chapter concludes with a discussion of the main points and the conclusions of the work (Stamatakos et al 2010).

4.2 The discrete state – discrete event cytokinetic model

The simulation model presented is based on the well documented assumption that tumour sustenance is due to the existence of stem cells i.e. cells with theoretically unlimited mitotic potential. Most cancers comprise a heterogeneous population of cells with differences in their proliferative potential. Cancer stem cells are generally thought to represent a small portion of the total tumour cell population (Stingl and Carlos, 2007). They possess the ability of self renewal i.e. the ability to give birth to more stem cells. Two types of stem cell division are possible: *symmetric* and *asymmetric*. An *asymmetric* stem cell division gives rise to one daughter cell with stem cell fate and one daughter cell (limited mitotic potential or committed progenitor cell) that follows the path towards differentiation. A *symmetric* stem cell division gives rise to two daughter cells both with a stem cell fate (Morisson and Kimble, 2006).

Specifically, the following five categories of cancer cells are considered in the model:

i. *Stem cells*

ii. *Limited mitotic potential (LIMP) or restricted/committed progenitor cells:* cells able to perform a limited number of divisions before becoming terminally differentiated

iii. *Differentiated cells*: terminally differentiated cells

iv. *Apoptotic cells*: cells that have died through apoptosis

v. *Necrotic cells*: cells that have died through necrosis

Stem, LIMP and differentiated cells constitute three categories with distinct mitotic potential.

Fig. S1 depicts the generic cytokinetic model proposed for the case of tumour growth and response to chemotherapy. A cytokinetic model limited to free tumour growth has been presented in a previous publication (Kolokotroni et al., 2008).

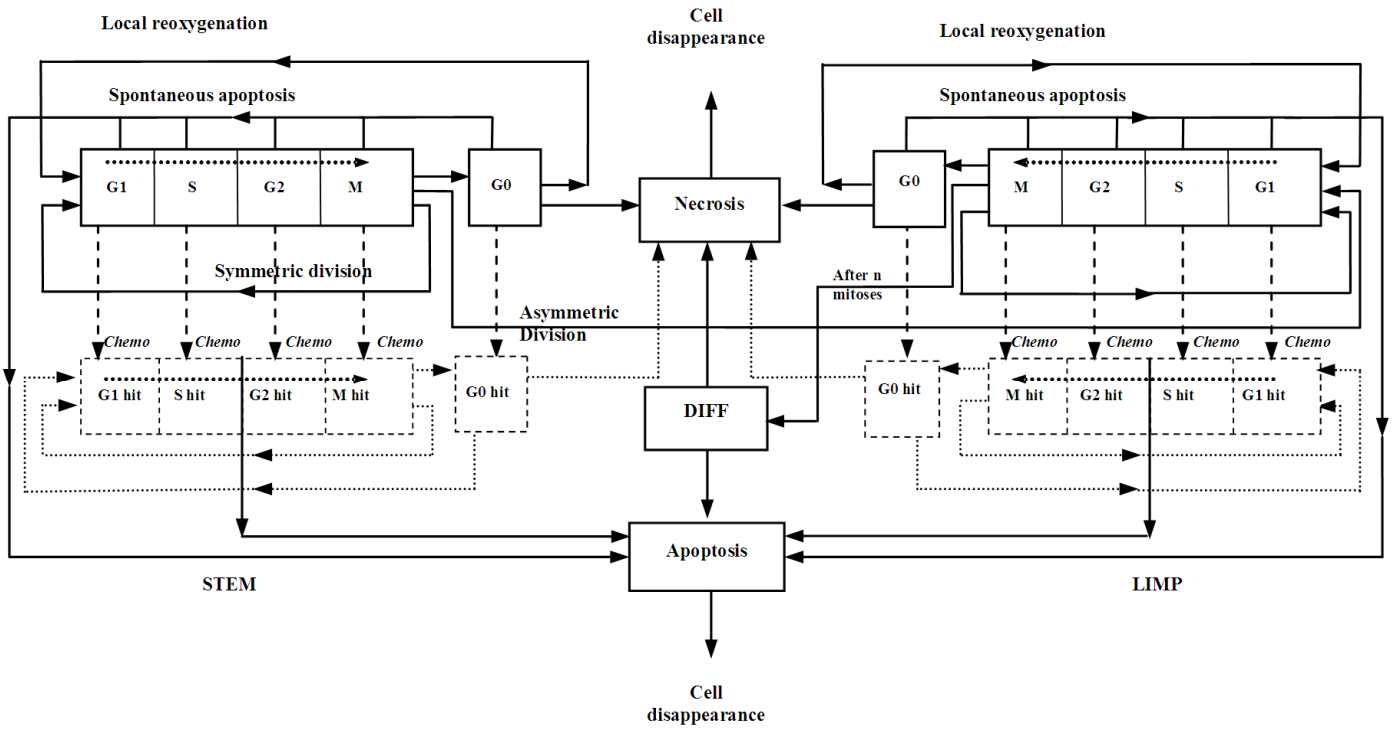


Fig S1. Generic cytokinetic model (cell category / phase transition diagram) for tumour response to chemotherapy. Abbreviations: STEM: stem cell. LIMP: LImited Mitotic Potential tumour cell (also called committed or restricted progenitor cell). DIFF: terminally DIFFerentiated tumour cell. G1: Gap 1 cell cycle phase. S: DNA synthesis phase. G2: Gap 2 phase. M: Mitosis phase. G0: dormant, resting phase. Chemo: chemotherapeutic treatment. Hit: cells lethally hit by the drug.

Proliferating cells, including stem and LIMP cells, go through the four phases of the cell cycle: gap 1 (G1) phase, DNA synthesis (S) phase, gap 2 (G2) phase and mitosis (M) phase. Depending on the nutrient and oxygen supply level of the local environment the daughter cells may re-enter the cell cycle at G1 phase or enter the dormant (resting) G0 phase following mitosis. Cells that are dormant due to insufficient nutrient supply and oxygenation can survive for a limited period of time. Subsequently, they die through necrosis unless in the meantime the local metabolic conditions regarding nutrition and oxygenation have become adequate. In the latter case dormant cells re-enter the G1 phase. Proliferating, dormant and differentiated cells may die due to aging and spontaneous apoptosis. Differentiated cells may also die through necrosis.

When a tumour is chemotherapeutically treated, a fraction of cancer cells are lethally hit by the drug or its metabolites. Lethally hit cycling tumour cells enter a rudimentary cell cycle that leads to apoptotic death via a specific phase dictated by the action mechanism of the chemotherapeutic agent used. Similarly, in the case of cell cycle non specific drugs, lethally hit dormant (G0) cells enter the G0hit phase. Marking of a cell as *hit* by the drug is assumed to take place at the instant of drug administration although its actual time of death is dictated by the specific pharmacokinetics and pharmacodynamics of the drug. For the special case of epirubicin treatment, the fraction of cells marked as *hit* by the chemotherapeutic

agent is considered the same for all cell cycle phases and the G0 phase since epirubicin is a cell cycle non specific drug as well as a cell cycle *phase* non specific drug (FDA, 1999). It is pointed out however that cell cycle phase specific drugs can be readily modeled by the cytokinetic model shown in Fig. S1 by appropriately selecting the “Chemo” induced exit from the normal cell cycle for both cases of stem and LIMP cells.

4.3 Spatiotemporal discretization of the tumour

Although macroscopically inhomogeneous tumours of generic shapes can be simulated by the full version of the model developed, in this chapter the special case of a macroscopically homogeneous tumour of generic shape is presented. The application examples, however, concern the even simpler case of a macroscopically homogeneous spherical tumour since this has been the best approximation to the imaging data available from the TOP trial (TOP trial). Such a spatially simple tumour model is a reasonable first approximation to the modeling of a wide range of breast cancer tumours. The latter is supported by both the rather homogeneous normal soft tissue biomechanics of the breast and the accumulated clinical experience regarding breast tumours. In fact the tumour size as reported in the Case Report Forms (CRFs) of the TOP trial is defined as the maximum diameter based on ultrasound examination. It should also be noted that macroscopic spatial homogeneity implies that the model parameter values used refer actually to their spatial average throughout the tumour.

In contrast with its spatial simplicity the model is characterized by a rather high complexity regarding the number of mitotic potential cell categories and cycling phases considered, as well as the corresponding transitions. A three dimensional (3-D) cubic discretizing mesh is superimposed upon the anatomical region of interest. The elementary cube of the mesh is called geometrical cell (GC) and in the cases considered in this chapter its size is 1 mm^3. Each GC occupied by the tumour is assumed to initially contain a *Number of Biological Cells* (*NBC*) (see also section 6(i)). The biological cells residing within each geometrical cell of the mesh are distributed into the five categories mentioned above i.e. the stem, LIMP, differentiated, apoptotic and necrotic categories. From the mathematical standpoint each cell category defines an equivalence class. Distribution of the cells into the five equivalence classes creates one level of biological cell population partitioning within each GC. At each given instant each stem or LIMP cell can be either proliferating or dormant. Proliferation or dormancy creates another level of cell population partitioning. Cell cycle phases (G1, S, G2, M) introduce a finer partitioning of proliferating cells (stem and LIMP) into subclasses. A further partitioner in the case of therapeutic intervention is treatment hitting i.e. a boolean variable denoting whether a biological cell has been hit by treatment. The relative population (expressed as the fraction of the total tumour cell population) of each equivalence class and its equivalence subclasses is initialized based on cell category and cell phase transition rates as described in section 4.4 and Appendix SA.

The initial distribution of the proliferating cells throughout the cell cycle phases is assumed analogous to the corresponding cell cycle phase durations (see also section S6(ii)). In order to tackle computing power limitations a number of carefully chosen quantizations are introduced. Time is discretized. The time unit in the practical cases considered in this chapter is taken one hour since this is approximately the duration of mitosis, the shortest cell cycle phase (Bast et al., 2000). For any given instant the biological cells belonging to the same cell

category and cell cycle phase within a given GC are assumed synchronized. However, biological cells belonging to different GCs or to different categories and cell cycle phases within the same GC are not assumed synchronized. From the computational standpoint a sufficient number of registers are used to describe the state of each GC occupied by the tumour. They include i.a. the number of biological cells residing in each equivalence class and subclass and the time spent at each subclass. Mean time spent by the biological cells of a given equivalence subclass in the same subclass is initialized using a random number generator (Monte Carlo technique). The time under consideration can vary between 0 and the maximum time of the corresponding phase.

At each time step i.e. every hour the discretizing mesh covering the anatomical region of interest is virtually scanned in order to apply the basic metabolic, cytokinetic, pharmacokinetic/pharmacodynamic and mechanical rules that govern the spatiotemporal evolution of the tumour system. For practical reasons each complete virtual scan can be viewed as consisting of two mesh scans. A brief outline of the scanning procedure is given below.

(i) First Scan

The first scan aims at updating the state of each GC according to the proposed and adopted cytokinetic model of Fig. S1. The time registers of the various cell subclasses within each geometrical cell are updated and the cytokinetic diagram is applied within each GC as follows. *Spontaneous apoptosis* induced cell loss from each non treatment perturbed cell cycle phase and the G0 phase is calculated for each cell category based on the spontaneous apoptotic rates assumed. Any necessary transitions between equivalence subclasses (G1→S, S→G2, G2→M, M→G1 or M→G0) take place for biological cells clustered in the same subclass. If the mean time that the clustered cells have spent in the corresponding phase has become equal to or larger than the phase duration then the cells enter a new phase and equivalence subclass.

In any one of the cases of dormant (including stem and LIMP), differentiated, necrotic and apoptotic cells a fraction of the corresponding subclass(es) population may be transferred to another subclass or disappear from the tumour at each time step according to the cytokinetic model (Fig.S1). Therefore the following transitions may take place. *For stem and LIMP cells*: G0→G1 or G0→Necrosis or G0→Apoptosis. *For differentiated cells*: Differentiated→Necrosis or Differentiated→Apoptosis. *For dead cells of any mitotic potential category:* Apoptosis→Cell disappearance, Necrosis→Cell disappearance. Most of the corresponding rates are parameters of the model (see Table SI and sections S6(iii),S6(iv)).

As far as chemotherapy is concerned, at any time instant corresponding to drug administration, the numbers of proliferating and dormant cells belonging to each one of the phases G1, S, G2, M or G0 and to each one of the stem or LIMP mitotic potential categories that are designated as *hit* by the drug are computed. The latter is achieved through the utilization of the *cell kill ratio* (*CKR*) parameter that corresponds to the drug and dose (per m^2 of the patient surface) considered. In terms of the simulation model's parameters the cell kill ratio is the percentage of LIMP and stem cells hit by the chemotherapeutic agent after each drug administration. The above mentioned cell numbers are added to the corresponding cell numbers of the drug affected equivalence subclasses, designated as

"phase"hit e.g. G1hit in the cytokinetic diagram of Fig.S1. A general strategy for the calculation of the *cell kill ratio* for the case of epirubicin is described in section 4.5.

(ii) Second Scan

The second scan aims to preserve a roughly uniform cell density throughout the tumour volume. To this end, adequately shaped morphological rules are introduced, which may lead to tumour expansion, as is the case in free tumour growth, or no change in tumour volume or tumour shrinkage as is usually the case after treatment administration.

More specifically, at any given time point the total cell population that can be accommodated in each GC is allowed to fluctuate between a minimum (0.9**NBC*) and a maximum (1.1**NBC*) value. If the total population exceeds the maximum value of 1.1**NBC* then a procedure is initiated that attempts to unload the *total GC population minus NBC* to neighboring GCs (26 GC neighborhood is considered) possessing empty space i.e. GCs with total cell population less than *NBC*. The procedure starts from the neighboring GC possessing the maximum free space. If two or more neighboring GCs possess the same free space then a random number generator is used so as to select the visiting order of the GCs. The procedure is repeated until all the excess cells have been transferred if possible. If the procedure fails to reduce the total population of the GC under consideration below the upper limit (maximum value) then an adjacent GC is freed from its contents which are moved outwards. The latter push the contents of a chain of geometrical cells outwards too. The excess contents of the GC under consideration are placed into the newly freed adjacent GC. The previous process leads to differential tumour expansion. The position of the GC to be freed from its contents relative to the GC with the excess contents is determined using a random number generator. The shifting of the chain of GCs mentioned above can take place along any randomly selected direction.

On the other hand if the GC's total cell population is below the minimum value then a similar procedure attempts to unload all cells to neighboring GCs possessing free space. If the GC becomes empty then a chain of GC contents is shifted towards the GC under consideration so as to fill the vacuum generated. The latter leads to differential tumour shrinkage. Shifting of the GC content chain takes place in a way analogous to the previously described one.

The above procedure, however, may give rise to the following "side effects". (a) *Premature extensive tumour fragmentation*: some GCs belonging to the tumour become artificially separated from the main tumour mass. (b) *Vacuum enclosures*: holes that correspond to empty GCs are created inside the tumour. The algorithm developed in order to avoid the occurrence of the above side effects is given in appendix SB. It should be noted that explicit information regarding tumour fragmentation induced by chemotherapy is rather scarce in the relevant literature. It is possible to expect that in some cases some degree of fragmentation will result, although pressure from adjacent normal tissue is one possible counter-acting mechanism. The available post-chemotherapy data from the TOP trial are in the form of tumours' maximum dimension, which led us to the assumption that at least the particular breast cancer tumours remained fairly cohesive.

The "second scan" algorithms introduced into the simulation model ensure uniform tumour cell density throughout the tumour volume (in the absence of any special information) and conformal tumour shrinkage, as depicted e.g. in (Perez and Brady, 1998). Our primary concern is to avoid artificial tumour fragmentation,

namely an extensive premature fragmentation caused by the very procedure of shifting chains of adjacent geometrical cells along randomly selected directions. There is still the possibility to have geometrical cells with less or more than the typical cell content (currently used margin=0.1*typical number of cells in a geometrical cell) and this flexibility in cell number actually “incorporates” the phenomenon of tumour “fragmentation”, in the sense that when chemotherapy is simulated geometrical cells become emptier over time. The above mentioned margin has been adjusted in order to achieve a rather uniform cell density throughout the tumour volume, but may be for example increased to permit a more extensive fragmentation in case that relevant information is available. These algorithms relate to geometrical aspects of the simulation and may be easily omitted from simulation runs in cases of macroscopically homogeneous tumours (such as those addressed in this chapter) as they have no effect on (biological) cell numbers. The latter is the most critical aspect of the simulation both in case of untreated tumour growth and in tumour shrinkage induced by chemotherapy treatment.

A simplified flowchart of the entire simulation procedure pertaining to a macroscopically homogeneous solid tumour of arbitrary shape is provided in Fig. S2.

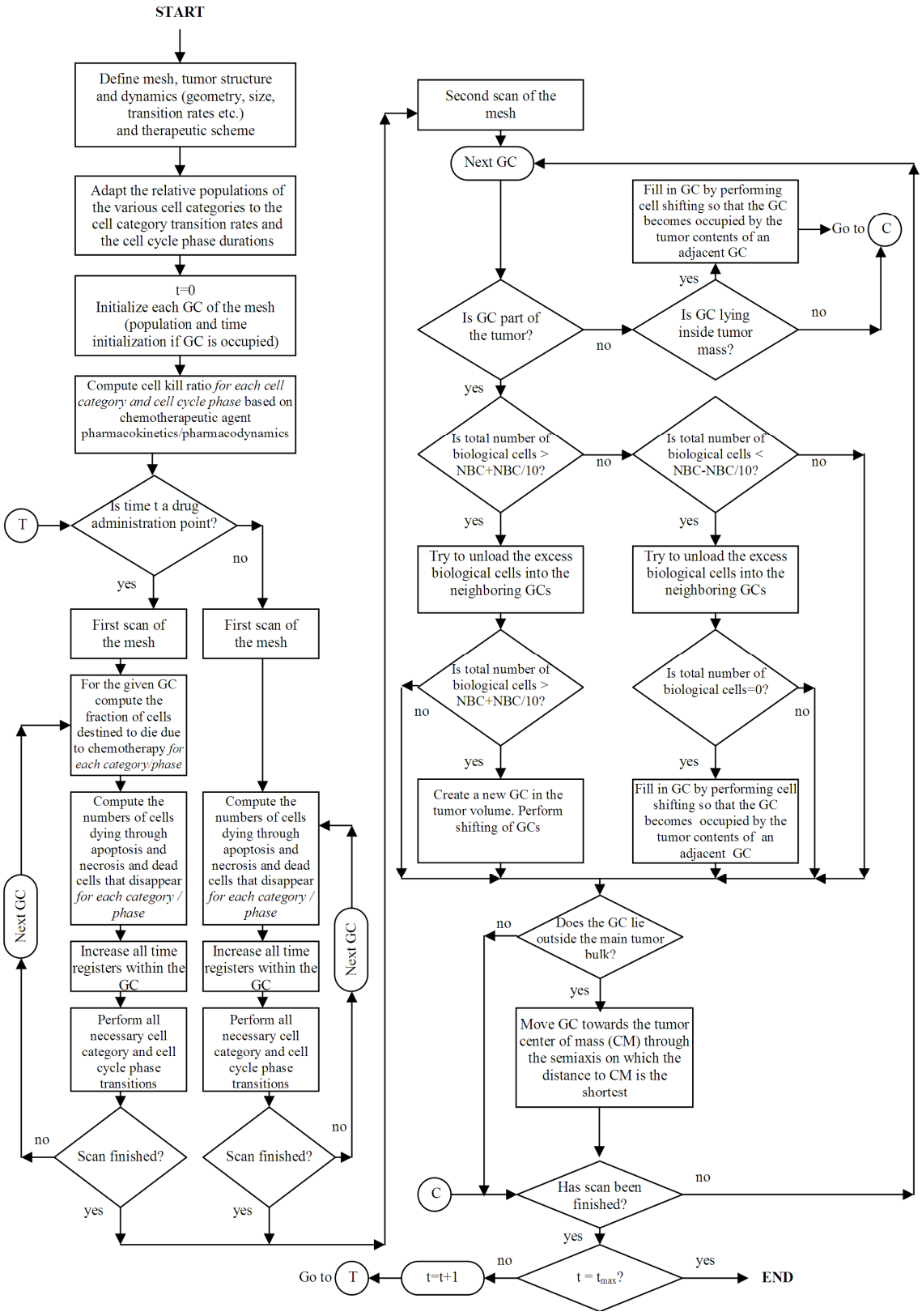


Fig.S2. Simplified flow chart of the simulation procedure for a macroscopically homogeneous solid tumour of arbitrary shape. Abbreviations: GC: geometrical cell. NBC: Number of Biological Cells normally contained within a geometrical cell of the mesh, t: simulation time step.

The core output of the simulation code is a number of matrices corresponding to the discretization mesh of the anatomic region of interest for a series of time points. These matrices contain all the necessary biological information to be used for the 3D metabolic, kinetic etc. reconstruction of the tumour at several time points or the creation of several diagrams describing tumour dynamics and response to therapy as a function of space and/or time.

It should be noted that the full model (not presented in this chapter) supports the division of tumour area into different metabolic regions (e.g. necrotic and proliferative) and the handling of each region separately. In this case, GCs are initially characterized as *necrotic* or *proliferative* based on pertinent imaging data available and a process of image segmentation, interpolation and 3-D reconstruction. Different values of specific model parameters can be assigned to each region. A special procedure has been devised in order to conformally expand or shrink any internal necrotic region(s) during tumour evolution.

4.4 Tumour constitution initialization

If tumour GCs are initialized with arbitrary values of tumour cell populations belonging to the various mitotic potential categories and phases (stem, LIMP, differentiated, dead, proliferating, dormant etc.) whereas at the same time specific category / phase transition rates are used, it is very likely to observe an abnormal free tumour growth behaviour. A decrease in tumour volume followed by a volume increase is a very common pattern (Kolokotroni et al., 2008). In order to avoid such artificial behaviour the concept of a nomogram linking cell category / phase relative populations with cell category /phase transition rates was initially introduced (Kolokotroni et al., 2008). However, due to its shortcomings the nomogram method has been replaced by a new, more efficient and flexible technique, described in appendix SA.

4.4.1. Condition for monotonic free growth

As far as untreated tumour growth simulations are concerned, depending on the values assigned to the code input parameters, a tumour that grows over time (which is normally the expected case) or a tumour that gradually diminishes (and therefore is usually biologically irrelevant) may result. The outcome strongly depends on the properties and the resulting behaviour of the stem proliferating cells. In order to create a growing tumour the number of stem cells must increase over time and the following inequality must hold:

$$(1-R_A)^{T_C}(1+P_{sym})(1-\mathrm{P}_{\mathrm{sleep}}) \geq 1 \qquad \text{(S1)}$$

where T_c is the cell cycle duration, R_A the apoptosis rate of cancer stem cells, P_{sym} the fraction of stem cells that perform symmetric division and P_{sleep} the fraction of newborn cells that enter G0 phase (Table SI). The rationale behind the above inequality, which holds true for the majority of cases, can be found in a previous publication (Kolokotroni et al., 2008). Further analysis concerning free tumour growth preconditions is underway.

4.5 Pharmacokinetics and pharmacodynamics of the chemotherapeutic agent

Epirubicin is an anthracycline chemotherapeutic agent, derivative of doxorubicin. It exerts its cytotoxic action through various mechanisms; the most established one is intercalation between bases of double stranded DNA, thereby inhibiting nucleic acid (DNA - RNA) and protein synthesis. It also interferes with DNA replication and transcription, blocks helicase activity, inhibits DNA cleavage by topoisomerase II, and forms toxic oxygen-free radicals causing DNA and cell damage (FDA 1999). In specific cases epirubicin is favored over other anthracycline drugs (doxorubicin), as it appears to cause fewer side-effects due to its less toxic nature at equivalent therapeutic doses. Epirubicin is considered a cell cycle non-specific as well as a cell cycle *phase* non specific drug (FDA, 1999).

In the simulation model tumour cells are assumed to absorb the drug (to be treatment *hit*) at all cycling phases as well as at G0, whereas apoptotic death of hit cells is assumed to occur at the end of S phase (i.e. between S and G2 phases). The latter is supported by the observation that the maximal cytotoxic effect of epirubicin has been observed on S and G2 phases (FDA, 1999, p.40).

Following intravenous administration, epirubicin is rapidly and widely distributed into tissues. Its higher concentration is observed in liver, spleen, kidney and small intestine. Epirubicin undergoes extensive hepatic metabolism (Danesi et al., 2002) and is also metabolized by other organs and cells, including red blood cells. It is eliminated mainly through biliary excretion and, to a lesser extent, by urinary excretion (FDA, 1999).

Epirubicin pharmacokinetics has been described in various studies by an open three-compartment model with elimination from the central compartment (Danesi et al., 2002; FDA, 1999). If a specific value of CKR was to be used in a simulation run, then based on the parameter values of such a model and making use of e.g. the SAAM pharmacokinetics software (SAAM II, 2009) various pharmacokinetic quantities of interest could be estimated such as the Area Under Curve (*AUC*) for a given drug dose.

Subsequently, a reasonably good initial approximation would be to compute the survival fraction for epirubicin treated tumour cells on the basis of experimental FDA data concerning the pharmacodynamics of epirubicin, and more specifically the *in vitro* cytotoxicity of epirubicin on HeLa cells (FDA, 1999). The survival fraction for any realistic dose for which no experimental data are available could be approximately calculated through linear interpolation. A lower *CKR* might be expected for the *in vivo* case due to imperfect vascularization, unless the particular genetic profile of a specific tumour somehow increases responsiveness to epirubicin treatment. It is noted that in cases of human breast cancer steep concentration gradients have been shown for the highly relevant drug doxorubicin in pertinent studies (Lankelma et al., 1999).

In the simulations presented in the following sections of the chapter, the suggested value of the CKR for each case study (the “apparent” CKR) is the CKR that produces good agreement between the evolution of the simulated tumour and that of the real tumour according to the clinical data. The individual patient’s genetic profile data, if available, could be used to perturb the population based mean cell survival fraction. In this way an appropriate molecular signature of the patient will be exploited and therefore further individualization of the treatment plan will be achieved.

4.6 Description of the model parameters

Table I presents the simulation model's input parameters and their range of values according to pertinent literature or based on logical assumptions supported by basic science or clinical experience in case of lack of literature data. In the following a summarizing description of the model's parameters and their adopted values is provided for reasons of clarity:

(i) NBC

As has been described previously, this parameter refers to the number of biological cells initially occupying each GC belonging to the tumour. In order to preserve a roughly constant mean cell density throughout the tumour volume, the *NBC* parameter is allowed to fluctuate between a minimum and a maximum value. Therefore, as described in previous paragraphs, accordingly shaped morphological rules governing the special evolution of the tumour have been introduced. A typical value of tumour cell density found in literature is 10^6 biological cells/mm^3 (Steel, 1997, p.9), which corresponds to 10^6 biological cells per occupied GC since the considered GCs' dimensions for the simulations that have been performed so far is 1mm x 1mm x1mm.

(ii) Cell cycle phases duration

According to literature the duration of the various cell cycle phases follows a normal distribution. Although this observation has been explicitly taken into account in an *in vitro* tumour growth and treatment response model by our group (Zacharaki et al., 2004), in the present model cell cycle phase durations are considered constant and equal to their literature based mean values in order to accelerate executions. It is pointed out however that algorithmically constructing a normal distribution of phase durations is a pretty simple task.

Based on literature, breast cancer cell cycle duration may vary from approximately 20h to 96h (Sterin et al., 2001; Cos et al., 1996; Descamps et al., 1998; Barnes et al., 2001; Meyer et al., 1984). The duration of mitosis is considered constant and equal to 1h (Bast et al., 2000). The rest of the cell cycle phases durations are computed based on Salmon et al. (2001), after having taken into consideration the above assumption regarding the constant duration of mitosis. More specifically the following equations are used: $T_{G1} = T_S = 0.41(T_c - T_M)$, $T_{G2} = 0.18(T_c - T_M)$, T_M = 1h.

(iii) Duration of G0, duration of apoptosis and necrosis

According to literature dormant cells resting in G0 phase can survive under hypoxic conditions for 4-10 days (Maseide and Rofstad, 2000). Tumour apoptotic cells are generally considered to be rapidly phagocytosed in vivo (Dewey et al., 1995), contrary to the time-consuming process of necrosis products removal.

(iv) Cell category/phase transition rates/fractions

The values of these parameters have been selected based on both qualitative or semi-quantitative information and the dictates of the accumulated basic science and clinical experience. Systematic use of TOP clinical trial data is expected to permit a quantitative refinement of the initial assumptions.

(v) Number of mitotic divisions that LIMP (committed progenitor) cells undergo before they become terminally differentiated (also called LIMP mitotic stages)

By varying the number of LIMP mitotic stages (assumed range 1 to 10) tumours characterized by different differentiation degrees can be simulated. To the best of our knowledge, there exists no specific information regarding possible values for this parameter in the pertinent literature. Nevertheless, the choice of its value is governed by the need to achieve specific relative percentages of stem and LIMP cells (which is also a reflection of a tumour's degree of differentiation). A larger number in this interval leads to lower percentages of stem cells.

4.7 Sensitivity analysis

The presented model is characterized by a considerable number of input parameters as presented in the previous section and summarized in Table SI. A number of parametric studies have been performed in order to study the model's behaviour in relation to the value of each input parameter for tumour free growth and chemotherapeutic response simulations and thereby detect the most critical parameters of the model. An extensive sensitivity analysis has been performed and reported in the ACGT project deliverable D8.3 'Report on the Refinement and Optimization of the Algorithms and Codes, and the Initial Clinical Validation and Adaptation of the "Oncosimulator". The following section presents indicative results. A detailed sensitivity analysis involving the role of several model parameters on tumour free growth and response to therapy will be the topic of a paper to appear shortly. A characteristic approach that will be presented in that separate paper is to adopt a ±10% variation around a reference value of each model parameter and the inspection of the variation in the output. Such analyses permit for example a sorting of the model's parameters according to the magnitude of their effect on the output.

Table II presents the values assigned to the model's input parameters for the indicative sensitivity analyses presented in this chapter.

4.7.1 Tumour free growth parametric studies

4.7.1.1 Doubling time

Different types of tumours with respect to the degree of aggressiveness can be simulated by assigning suitable values to the model's parameters. In the following, an indicative number of exploratory simulations are presented. These have been performed in order to study the effect of various input parameters on tumour growth rate. The model's parameters considered in these simulation studies are the fraction of cells entering the G0 phase following mitosis (sleep fraction, P_{sleep}), the fraction of stem cells that perform symmetric division (symmetric division fraction, P_{sym}) and the necrosis rate of differentiated tumour cells ($R_{NDiff,}$) (see Table SI).

Table SI. Model parameters.

Parameter symbol	Description	Value range and/or specific values considered	References
Cell phase durations			
T_c[class] class∈{stem, LIMP[a]}	Cell cycle duration[b]	20–96 h	Sterin et al. (2001), Cos et al. (1996), Descamps et al. (1998), Barnes et al. (2001), Meyer et al. (1984)
T_{G1}[class] class∈{stem, LIMP[a]}	G1 phase duration[b]	0.41 (T_c−T_M)	Salmon and Sartorelli (2001) *in conjunction* with Bast et al. (2000) (slight adaptation performed)
T_S[class] class∈{stem, LIMP[a]}	DNA synthesis phase (S) duration[b]	0.41 (T_c−T_M)	Salmon and Sartorelli (2001) *in conjunction* with Bast et al. (2000)
T_{G2}[class] class∈{stem, LIMP[a]}	G2 phase duration[b]	0.18 (T_c−T_M)	Salmon and Sartorelli (2001) *in conjunction* with Bast et al. (2000) (slight adaptation performed)
T_M[class] class∈{stem, LIMP[a]}	Mitosis (M) phase duration[b]	1 h	Bast et al. (2000)
T_{G0}[class] class∈{stem, LIMP[a]}	G0 (dormant phase) duration[b] i.e. time interval before a dormant cell enters necrosis	96–240 h	Maseide and Rofstad, 2000
T_N[region] region∈{proliferating, necrotic}	Time needed for both necrosis to be completed and its lysis products to be removed from the tumour[c]	20 h (0–100 h)	Duechting et al. (1992), Wein et al. (2000)
T_A[region] region∈{proliferating,	Time needed for both apoptosis to be completed and its products to be removed	6 h (0–25 h)	Ribba et al. (2006), Dewey et al. (1995)

Parameter symbol	Description	Value range and/or specific values considered	References
necrotic}	from the tumour[c]		
Cell category/phase transition rates			
R_A	Apoptosis rate of living stem and LIMP tumour cells (fraction of cells dying through apoptosis per h)	0.0–1.0 h^{-1}	
R_{NDiff}	Necrosis rate of differentiated tumour cells per hour	0.0–1.0 h^{-1}	
R_{ADiff}	Apoptosis rate of differentiated tumour cells per hour	0.0–1.0 h^{-1}	
P_{G0toG1}[region] region∈{proliferating, necrotic}	Fraction of stem and LIMP cells having just left the G0 compartment that re-enter the cell cycle	0.0–1.0 h^{-1}	
P_{sleep}[region] region∈{proliferating, necrotic}	Fraction of cells entering the G0 phase following mitosis[c]	0.0–0.5	
P_{sym}[region] region∈{proliferating, necrotic}	Fraction of stem cells that perform symmetric division[c]	0.0–1.0	
Miscellaneous parameters			
NBC	Number of biological cells normally contained within a geometrical cell of the mesh	10^6	Begg and Steel, 2002
N_{LIMP}	Number of mitoses performed by LIMP cells before becoming differentiated	1–10	
Margin_percent	Acceptable temporary over-loading or under-loading of each geometrical cell as a	0.0–0.5	

Parameter symbol	Description	Value range and/or specific values considered	References
	fraction of unity		
Color_criterion	Minimum percentage of tumour cells that should be dead in order to denote (“paint”) the corresponding geometrical cell as necrotic	0.9–0.999	
Distance_factor[region] region∈{proliferating, necrotic}	Factor adapting the cell killing probability as estimated by pharmacodynamics to each tumour region	0.0–1.0	
x_{dim}, y_{dim}, z_{dim}	Number of geometrical cells along the *x*, *y*, *z* axis respectively	Depends on tumour size and computing resources	
tumour_length, tumour_breadth, tumour_width	Dimensions of the three tumour axes in numbers of geometrical cells (GCs) in the case a triaxial ellipsoidal tumour is considered	Depends on tumour imageable characteristics	
necrotic_length, necrotic_breadth, necrotic_width	Dimensions of the necrotic region along the three axes in GCs in case a triaxial ellipsoidal tumour is considered	Depends on tumour imageable characteristics	
Epirubicin drug administration parameters			
D	Dose	100 mg/m^2	TOP trial[d]
$T_{1st,adm}$	Time point of the first drug administration since pretreatment scan (simulation initialization)	0–300 h	TOP trial[d]
$T_{2nd,adm}$	Interval between the 1st and	Around	TOP trial[d]

Parameter symbol	Description	Value range and/or specific values considered	References
	the 2nd administration of epirubicin	504 h (21 d)	
$T_{3rd,adm}$	Interval between the 2nd and the 3rd administration of epirubicin	Around 504 h (21 d)	TOP trial[d]
$T_{4th,adm}$	Interval between the 3rd and the 4th administration of epirubicin	Around 504 h (21 d)	TOP trial[d]
T_{stop}	Time interval between the last administration and the simulation completion time (h)	0–300 h	TOP trial[d]
CKR	Cell kill ratio	0.0–1.0	

[a] A LIMP tumour cell denotes a limited mitotic potential tumour cell (also referred to as LIMP or committed progenitor tumour cell). It leads to a terminally differentiated tumour cell.
[b] Phase durations can be defined separately for the stem and the LIMP tumour cell category.
[c] Defined separately for the proliferating and the necrotic region of the tumour, for spatially inhomogeneous tumour cases.

[d] http://clinicaltrials.gov/ct2/show/NCT00162812 last visited on 13 June 2009.

For the analyses presented below, homogeneous tumours have been considered, consisting of a single proliferating region with no distinguishable necrotic or other areas. This assumption has been dictated by the availability of only volumetric and not necrosis distribution data in the TOP trial. This implies that microscopic necrotic regions which are primarily due to imperfect neovascularization of the tumour are postulated, rather homogeneously distributed over the entire tumour.

The cell category transition rates are assumed to be constant. Such an approximation is considered applicable for a relatively short time interval compared to the tumour lifetime (for example during the chemotherapeutic treatment course), and the constant values are assumed to reflect the means of the actual cell category transition rates over the interval. The use of the means of several tumour dynamics parameters over a substantial time interval is quite customary in radiobiology (Begg, 1997, pp.14-16). Such a tumour would be characterized by a grossly exponential growth pattern which in fact approximates a segment of the Gompertzian curve (Steel, 1997, p.10; Zoubek et al., 1999; Graf and Hoppe, 2006). Therefore for a short time interval the population of the various cell categories evolves over time according to the equation:

$N(t)=N_o e^{at}$ (S2)

where N_o is the population at time 0 and a is the growth rate constant.

Only for theoretical purposes in certain simulation cases is the considered time interval deliberately extended beyond what would seem a short segment of the Gompertzian curve.

The doubling time, T_d, defined as the period of time required for the total tumour cell population to double, is a widely used parameter characteristic of the growth rate of a tumour. Under the previous restrictions the following equation holds:

$T_d = (\ln 2)/a$ (S3)

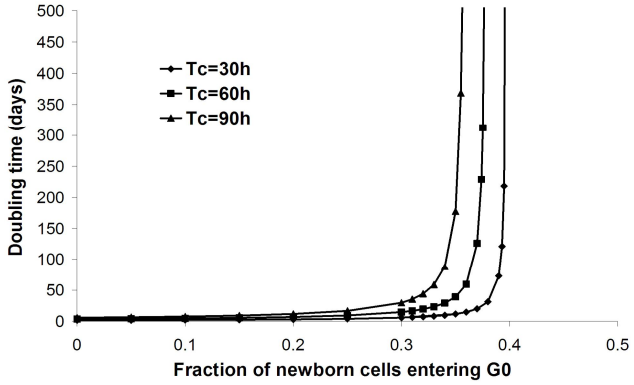
The doubling time is the same for all cell category populations and can be computed based on the following relationship:

$T_d = (\ln 2)*(t_2-t_1)/\ln(N_2/N_1)$ (S4)

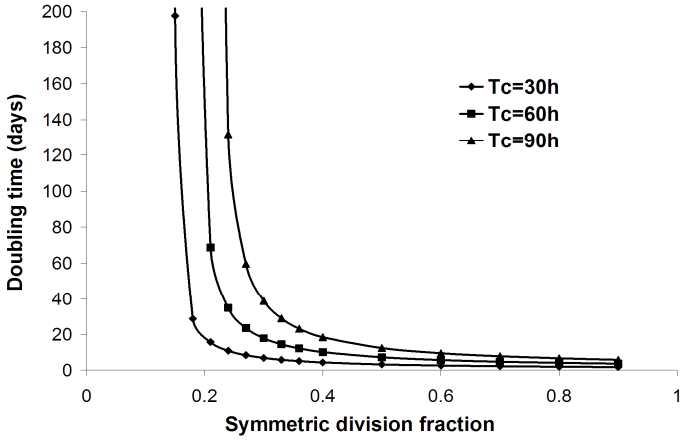
where N_1 and N_2 are the populations at times t_1 and t_2 respectively.

Table SII. Model parameter values used for the sensitivity analysis.

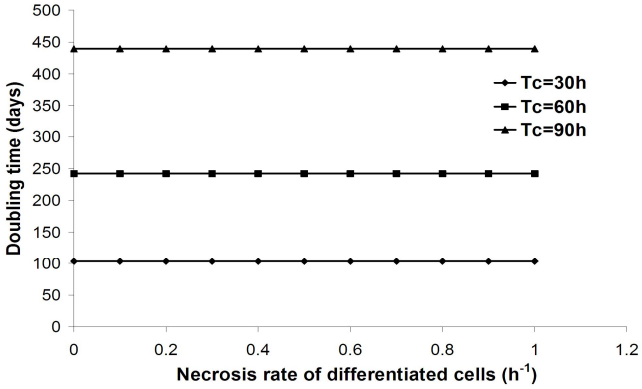
Parameter	Study of the effect of P_{sleep}	Study of the effect of P_{sym}	Study of the effect of R_{NDiff}	Study of the effect of CKR
T_c	30, 60, 90 h	30, 60, 90 h	30, 60, 90 h	30, 60, 90 h
T_{G1}	$0.41(T_c-T_M)$	$0.41(T_c-T_M)$	$0.41(T_c-T_M)$	$0.41(T_c-T_M)$
T_S	$0.41(T_c-T_M)$	$0.41(T_c-T_M)$	$0.41(T_c-T_M)$	$0.41(T_c-T_M)$
T_{G2}	$0.18(T_c-T_M)$	$0.18(T_c-T_M)$	$0.18(T_c-T_M)$	$0.18(T_c-T_M)$
T_M (h)	1	1	1	1
T_{G0} (h)	96	96	96	96
T_N (h)	20	20	20	20
T_A (h)	6	6	6	6
N_{LIMP}	7	7	7	7
R_A (h^{-1})	0.001	0.001	0.001	0.001
R_{NDiff}	0.01	0.01	0–1	0.01
R_{ADiff} (h^{-1})	0.001	0.001	0.001	0.001
P_{G0toG1} (h^{-1})	0.01	0.01	0.01	0.01
P_{sleep} (h^{-1})	0–0.5	0.1	0.1	0.1
P_{sym}	0.7	0–1	0.4	0.22
CKR	–	–	–	0.0–1.0



S3a



S3b



S3c

Fig. S3. The effects of (a) the fraction of cells entering the G0 phase following mitosis (P_{sleep}) , (b) the fraction of stem cells that perform symmetric division (P_{sym}), and (c) the necrosis rate of differentiated cells (R_{NDiff}) on the tumour doubling time for variable cell cycle durations (T_c = 30, 60, 90h). The values assigned to the rest of the code input parameters are reported in Table SII.

The effect of the parameters considered on tumour doubling time is revealed in Fig. S3. The following remarks can be made:

(a) *Fraction of cells entering the G0 phase after mitosis (Sleep fraction, P_{sleep}):* For low values of the sleep fraction, the simulated tumour is characterized by a small doubling time of a few days implying a rather aggressive tumour. As P_{sleep} increases, the doubling time increases as well, initially slowly for a wide range of values, and then rapidly towards infinity. There exists an upper limit of the sleep fraction that corresponds to the equality in (S1). This limit depends on the cell cycle duration and corresponds to a tumour that remains constant over time. As the cell cycle duration increases, the permitted value range of the sleep fraction narrows, since the upper limit decreases. Values of P_{sleep} larger than the upper limit result in a tumour that diminishes over time (Fig. S3a).

(b) *Fraction of stem cells that perform symmetric division (Symmetric division fraction, P_{sym})*: In contrast to the sleep fraction, in the case of the symmetric division fraction there exists a lower limit that corresponds to the equality in (S1) and increases as the cell cycle duration increases. As P_{sym} decreases, the doubling time approaches this limit asymptotically (Fig. S3b).

(c) *Necrosis rate of differentiated cells* (R_{NDiff}): According to the simulations that have been performed, this parameter appears to have no noticeable effect on tumour doubling time. There exists no upper or lower limit for its values, as the necrosis rate of differentiated cells does not appear in (S1) (Fig. S3c).

(d) *Cell cycle duration* (T_c): Higher values of T_c are associated with slowly evolving tumours and higher values of doubling time, on the condition that the rest of the model parameters are kept constant. However the same doubling time can result from different values of cell cycle duration after proper adjustment of the rest of the model parameters.

4.7.1.2 Relative cell category populations

As far as the effect of the model's parameters on the relative populations of cells of various categories is concerned, P_{sleep} and T_c have been chosen as characteristic parameters for such a study. A presentation of a more detailed analysis covering all the model's parameters is out of the scope of the present chapter. Based on Fig.S4 the following observations can be made:

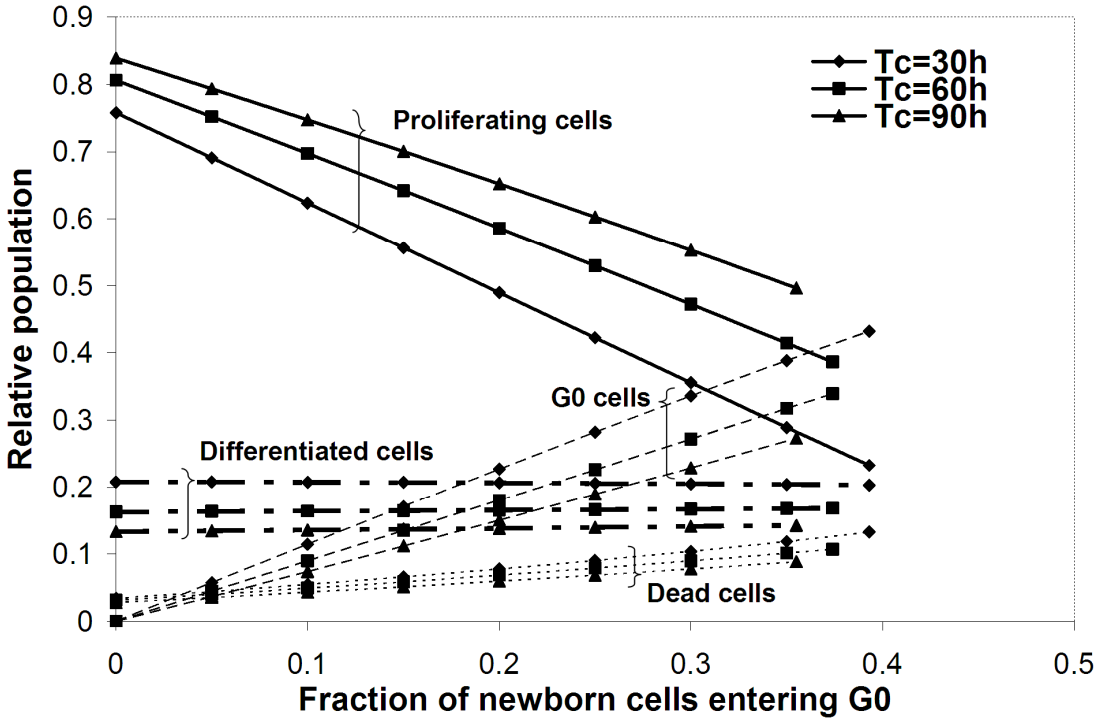


Fig. S4. The effect of the fraction of cells entering the G0 phase following mitosis (P_{sleep}) on the relative populations of the proliferating, dormant (G0), terminally differentiated and dead cells for several cell cycle durations (T_c = 30 h, 60 h, 90 h). The values assigned to the rest of the code input parameters are included in Table SII.

(i) It is expected that variations in the value of the fraction of cells that enter G0 phase after mitosis would mainly result in a redistribution of LIMP and stem cells in the dormant and the proliferative cell cycle phases. Indeed, according to our simulation results, as P_{sleep} increases the percentage of proliferating cells (proliferating LIMP and stem cells residing in the various proliferative cell cycle phases) decreases whereas the percentage of cells resting in G0 phase (dormant LIMP and stem cells) increases.

(ii) Increasing cell cycle duration T_c results in an increase of proliferating cells (the sum of LIMP proliferating and stem proliferating cells) at the expense of mainly the G0 and dead cell populations.

(iii) The effect of P_{sleep} on differentiated cells relative population appears to be insignificant. For higher values of the cell cycle duration there is a slight increase in the population, whereas for low values of the cell cycle duration the differentiated cells relative population slightly decreases with the increase of P_{sleep} (in the order of 10^{-3} for the whole range of P_{sleep}).

(iv) For high values of the sleep fraction the population of dead cells increases. This is due to the increase of necrotic cells, a class that is populated by cells formerly residing in the G0 phase.

(v) The relative population (fraction) of stem to LIMP cells remains unaffected by the parameter P_{sleep} (results not shown).

(vi) The sum of LIMP and stem cells relative populations decreases slightly as P_{sleep} increases (in the order of 10^{-2}) to account for the increase of necrotic cells relative population (results not shown).

4.7.2 Response to Chemotherapy Parametric Studies

The effect of the cell kill ratio (i.e. the percentage of LIMP and stem cells hit by the chemotherapeutic agent after each drug administration) on the tumour's volume and diameter is shown in Fig. S5 for various cell cycle durations. The simulation algorithms address the case of primary chemotherapy ("neo-adjuvant" chemotherapy) with single-agent epirubicin (administration once every 3 weeks for 4 consecutive cycles) for early breast cancer patients, in accordance with the TOP trial. The simulated tumour grows freely for 3 weeks after the end of chemotherapy. The percentage decrease in tumour volume/diameter is calculated based on the following equation:

$$\text{Percentage decrease} = ((X_{initial} - X_{final}) \times 100\%)/X_{initial} \quad \text{(S5)}$$

where X refers to tumour volume or diameter.

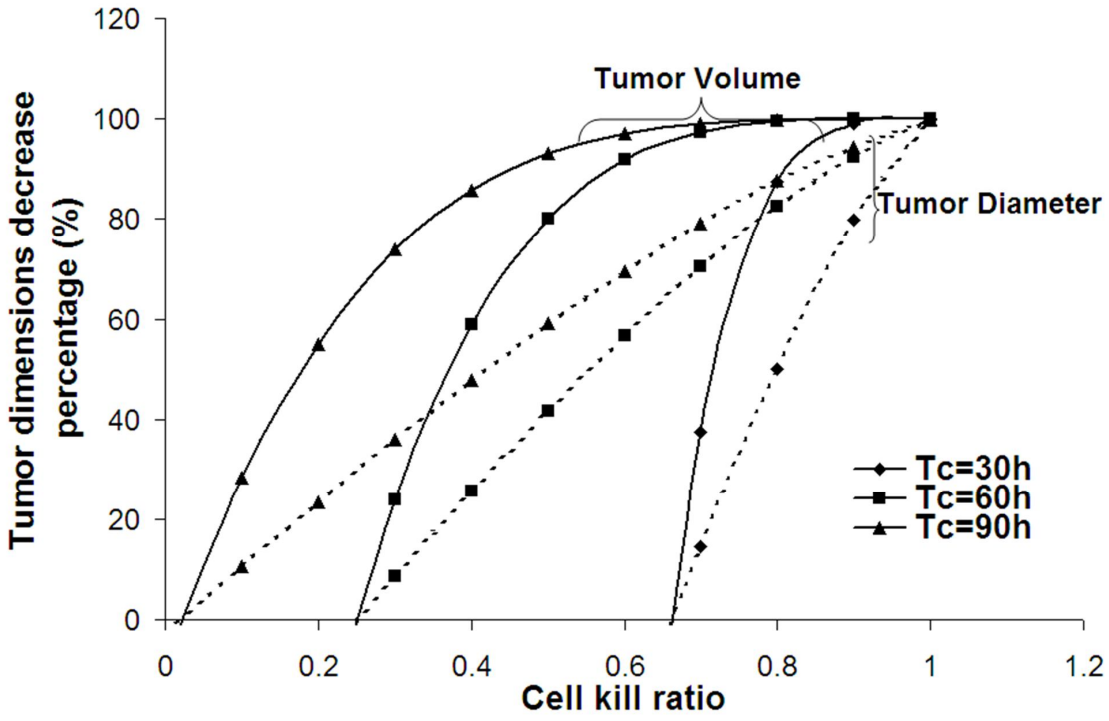


Fig. S5 The effect of cell kill ratio (CKR) on tumour volume and tumour diameter. Three cell cycle durations (T_c) have been considered. The values assigned to the model parameters are included in Table SII. Higher values of cell cycle duration correspond to more slowly evolving tumours, hence tumours with larger doubling time.

The following observations can be noted:

(i) Obviously the outcome of a chemotherapeutic scheme strongly depends on the pharmacodynamics of the chemotherapeutic agent used. The model successfully simulates the shrinkage of a tumour to a higher degree as the cell kill ratio increases. The value of the cell kill ratio could be thought of as summarizing important genetic determinants influencing a particular patient's tumour response to chemotherapy.

(ii) The free growth volume doubling time of a tumour is expected to influence the effectiveness of a chemotherapeutic fractionation scheme. In general, the time for tumour volume to double (volume doubling time) is determined by three main factors: the cell cycle duration, T_c, the growth fraction, *GF*, which is the percentage of actively proliferating cells, and the rate of cell loss (e.g. through necrosis or apoptosis) (Steel et al., 1997). Our simulation studies involved the cell cycle duration, as a determinant of tumour doubling time. By increasing the value of the cell cycle duration, while keeping the rest of the model parameters constant, a slower evolving tumour can be simulated. The effectiveness of the chemotherapeutic scheme is pronounced for tumours with higher doubling times, due to the restrained repopulation of the tumour between two therapeutic sessions.

(iii) There exists a limit in the cell kill ratio values, below which the specific therapeutic scheme fails to shrink the tumour. This limit depends on the tumour's doubling time. As doubling time increases (due to an increase in the cell cycle duration) this limit decreases.

4.8 Mimicking a branch of the Trial of Principle (TOP) trial. Towards clinical adaptation of the simulation model.

As presented in previous paragraphs, the model's behaviour substantiates its potential to serve as the basis of a treatment optimization system following a successful completion of the clinical adaptation and validation process which will rely on the use of anonymized real data before and after chemotherapeutic treatment for the case of the TOP breast cancer clinical trial. The TOP trial aims to evaluate the topoisomerase II alpha gene amplification and protein overexpression as markers predicting the efficacy of epirubicin in the primary treatment of breast cancer patients. Regarding the correlation of topoisomerase II alpha (TOP2A) amplification/ expression with response to anthracyclines, and epirubicin in particular, controversial results have been reported by several investigators during the last decade (e.g. Knoop et al., 2005; Gennari et al., 2008; Pritchard et al., 2008).

The presented simulation model is characterized by the ability to incorporate actual molecular data in order to predict the outcome of a chemotherapeutic treatment in a patient-individualized context. The effect of a patient's molecular profile on its tumour's response to chemotherapy can be modeled through adequate adaptation of the parameters related to the pharmacodynamics of the chemotherapeutic agent. Depending on the status and expression of critical genes such as topo II α, p53 HER-2/neu, etc., the drug pharmacodynamics is expected to vary around its population based mean value.

In the present work, two real clinical cases from the TOP trial with variable molecular profile (including topo II α amplification status) have been simulated. More specifically, the patient specific data that have been provided and utilized by the model are the following:

(i) Maximum dimension of the initial tumour before the beginning of the therapy as measured by ultrasound examination and the date that the above examination was performed.

(ii) Maximum dimension of the tumour during or after the completion of chemotherapy (depending on the available data) as measured by ultrasound examination and the date that the above examination was performed.

(iii) Chemotherapeutic scheme (number of therapeutic sessions, time and dose of epirubicin administration)

(iv) Molecular profile information, including topo II α gene amplification status as measured by FISH (fluorescent in situ hybridization). Amplification of the gene was defined as a relative copy number ratio ≥2.

(v) Tumour histological grade (determining the degree of tumour's differentiation). A tumour is described in TOP trial's CRF forms as poorly, moderately or well differentiated according to the the Elston-Ellis grade system (Elston and Ellis, 1991).

Fig. S6 describes each case study in terms of the above presented data. The date of the initial tumour measurement is considered as the time point reference (time t=0) and all time data refer to elapsed time from this reference time point.

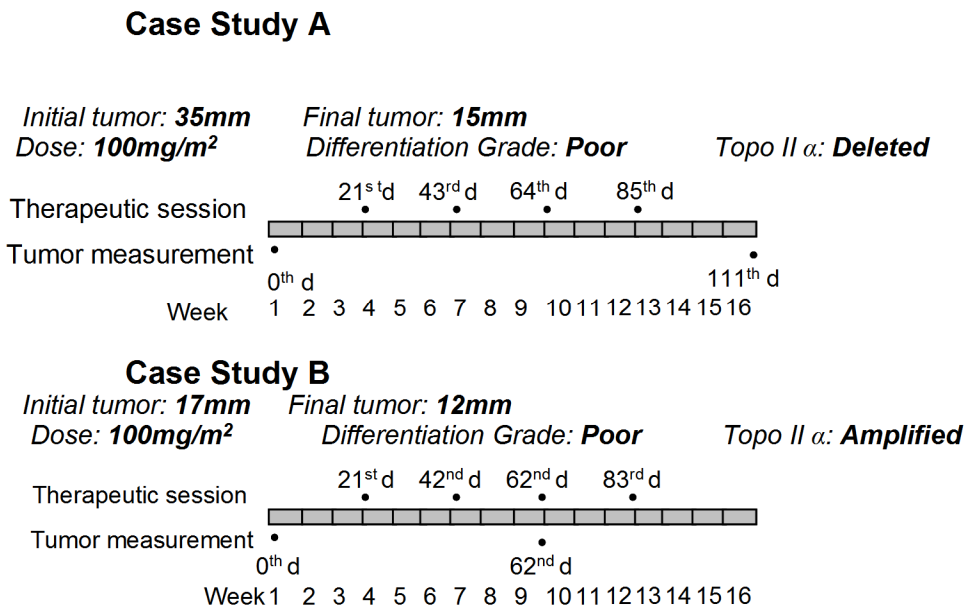


Fig.S6. The epirubicin administration schedules considered in case studies A and B.

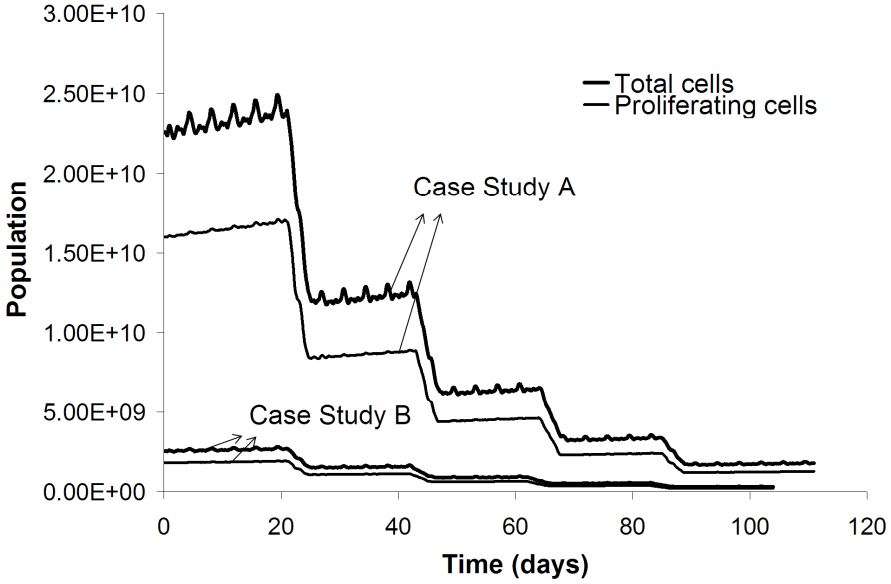
Due to non availability of all desirable clinical details at this early phase of the clinical validation process, specific assumptions had to be made regarding tumour shape and size. The data available (till the completion of the preparation of the present chapter) included the maximum dimension of the tumour as measured by ultrasound examination. No MRI data have been provided as yet. Therefore a refined 3D reconstruction of the tumour or the consideration of internal metabolic regions could not be made at this stage. Based on the above, a spatially homogeneous tumour of spherical shape has been assumed with a diameter of the given size. Therefore, various model parameters, which would otherwise change their values according to the metabolic region considered, now take an identical value throughout the entire tumour. The consideration of spherical tumours is a first approximation based on accumulated clinical experience regarding the shape of breast cancer tumours. More general shapes and metabolically inhomogeneous tumour structures will be addressed in subsequent versions of the model.

Table SIII. Parameter values used for the clinical cases' simulations and resulting tumour characteristics.

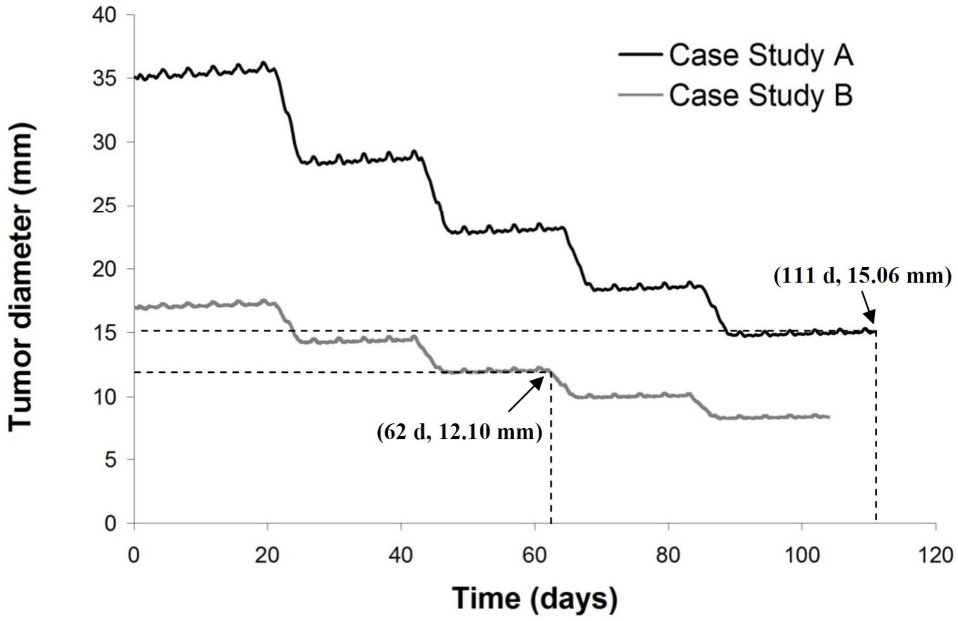
Model parameter	**Value**
T_c (h)	90
T_{G1} (h)	36
T_S (h)	36
T_{G2} (h)	17
T_M (h)	1
T_{G0} (h)	96
T_N (h)	20
T_A (h)	6
N_{LIMP}	7
R_A (h^{-1})	0.001
R_{NDiff} (h^{-1})	0.01
R_{ADiff} (h^{-1})	0.03
P_{G0toG1}	0.01
P_{sleep}	0.1
P_{sym}	0.228
Tumour characteristic	**Value**
Doubling time	232 days
Relative population of tumour stem cells	0.026
Relative population of tumour LIMP (committed progenitor) Cells	0.761
Relative population of terminally differentiated tumour cells	0.145
Relative population of dead cells	0.068
Growth fraction	0.71
Dormant cells fraction	0.078

In our simulations different tumour differentiation degrees are reflected in the relative percentage of proliferating (stem and LIMP) cells. For both case studies A and B, the tumour is characterized as poorly differentiated, which according to the Elston-Ellis system (Elston and Ellis, 1991) can be grossly translated into a tumour with percentage of proliferating cells larger than 75%, whereas a moderately differentiated tumour would have a percentage of proliferating cells between 10 and 75%. Well differentiated tumours are assumed to possess a growth fraction less than 10%. Based on preceding sensitivity analyses the model parameters cell cycle duration (T_c), symmetric division fraction (P_{sym}) and apoptotic rate of differentiated cells (R_{ADiff}) have been appropriately adjusted to achieve the above population percentages. LIMP cells have been considered undergoing seven mitoses before becoming terminally differentiated, in order to acquire relative percentages of stem and LIMP cells supported by the currently available literature.

The values assigned to the code input parameters and the resulting tumour characteristics are presented in Table III. Critical properties of the resulting tumour, such as doubling time (232 days) and percentage of proliferating cells (2.6%) are in accordance with breast cancer literature data (e.g. Spratt et al., 1981; Al-Hajj et al., 2003; Liu et al., 2005; Stingl and Carlos, 2007).



S7a



S7b

Fig.S7. Simulated time course of the population of proliferating cells and total cell population (a) and tumour diameter (b) for the case studies A and B. The simulated tumour's diameter at the time point for which the second set of imaging data is available is indicated for each clinical case.

Table SIV.Comparison of maximum tumour dimension between clinical trial data and simulation results (two case studies)

	Day	Clinical data (mm)	Simulation (mm)
Initial tumour maximum dimension			
Case study A	0	35	35.07
Case study B	0	17	16.96
Final tumour maximum dimension			
Case study A	111	15	15.06
Case study B	62	12	12.10

For each case study one possible combination of the model input parameter values that successfully simulates the clinical case i.e. predicts the actual tumour volume shrinkage and satisfies several biological boundary conditions is reported below. Fig. S7 shows the time course of the following quantities of interest for the two clinical cases: population of proliferating cells (stem and LIMP) and total cell population (Fig. S7a), and tumour diameter (Fig.S7b). The above populations include both intact cells and cells affected by the drug (and therefore destined to die) but not yet actually dead. Qualitatively a fairly expected and reasonable tumour dynamics behaviour can be easily noticed. Shrinkage of the tumour after each chemotherapeutic session and tumour repopulation are successfully demonstrated. Most importantly though, in quantitative terms, a reduction in tumour size in agreement with the clinical data has been achieved in both cases. According to the clinical data the tumours in case A and B have a maximum diameter of 12mm in day 62 and 15 mm in day 111 respectively, which is clearly reproduced by the simulation results (Fig S7b and Table SIV).

The values assigned to the cell kill ratio are shown in Table SV. The value of the CKR that produced good agreement with the clinical data is reported as the suggested “apparent CKR” for each particular clinical case, which can be thought of as summarizing important genetic determinants influencing the tumour’s response to epirubicin monotherapy. Table V includes topo II a amplification status for demonstration purposes (since topo II a amplification status has a particular role in the TOP trial), but the apparent CKR value incorporates the effect of all critical genetic factors constituting the tumour’s molecular profile. It is noted that no statistical conclusions can be drawn concerning specifically the effect of topo II α amplification status on the therapeutic outcome, a study that is out of the scope of the present work. In conclusion, by properly adjusting the pharmacodynamics of epirubicin as reflected in the cell kill ratio value, the patient’s molecular profile data (including topo II a amplification status) can be taken into account. The relevant methodology is to be refined according to the details provided by the TOP trial clinical cases.

4.9. Discussion

The model presented in this chapter , having its roots in previous work of the *In Silico* Oncology Group, National Technical University of Athens (*In Silico* Oncology Group) is an actual clinical trial driven model concerning the response of early

breast cancer to epirubicin monotherapy. Available imaging-based, histological, molecular and treatment data are exploited in order to strengthen patient individualization modeling and in the future patient individualized treatment optimization. Clinical validation constitutes an undeviating prerequisite for clinically oriented multiscale cancer models. A branch of the Trial of Principle (TOP) trial has been addressed as a paradigm of a chemotherapy optimization targeted clinicogenomic trial.

A basic feature of the simulation model is the explicit consideration of cells of distinct mitotic potential, i.e. stem cells of infinite proliferative potential, LIMP (or committed progenitor) cells of limited proliferative activity and terminally differentiated cells. It is well documented in the relevant literature that tumour sustenance is due to the existence of stem cells, which generally represent a small but potentially critical portion of the total tumour cell population (Dean et al., 2005). Cancer stem cells may be naturally resistant to chemotherapy; indeed considerable evidence exists for many types of tumours, including breast cancer (Dean et al., 2005; Diehn et al., 2007; Fillmore and Kuperwasser, 2008). A major advantage of the present model is the fact that such a diversification of chemotherapeutic resistance is facilitated by its discrete nature and the characteristics of the particular cytokinetic model adopted.

A further prominent feature of the model is the introduction of an improved tumour initialization technique, which dispenses shortcomings of previously used methods and is now incorporated into the hard code, thereby offering high flexibility to the user of the model. This is of particular importance since the clinically oriented nature of the model implies that the evolution of already fully developed clinical tumours is to be simulated. Proper tumour initialization excludes the possibility of artificial tumour behaviour which could interfere with the interpretation of the simulation results.

A number of parametric studies have been performed in order to study the model's behaviour in relation to the value of each chosen input parameter with the primary aim to deepen and advance quantification of our understanding of tumour response to chemotherapeutic treatment in the breast cancer and more specifically the TOP trial context. Similar studies are to be performed for all code parameters in order to thoroughly analyze the sensitivity of the model's behaviour to its parameters variations. An extensive use of the clinical trial data is expected to crucially support this procedure, while pertinent optimization techniques such as artificial neural networks, genetic algorithms etc. have been planned to be used in this context.

By using real medical data in conjunction with plausible values for the model parameters a reasonable prediction of the actual tumour volume shrinkage for two clinical cases of the TOP trial has been made possible. The value of the cell kill ratio parameter that produced good agreement with the clinical data is suggested as the "apparent CKR" for each particular clinical case, which can be thought of as summarizing important genetic determinants influencing the tumour's response to epirubicin monotherapy. The whole system can be seen as a paradigmal case of hybridizing clinicogenomics trials with multiscale computer modeling primarily focused on the cell and higher levels.

For each clinical case one possible combination of the model input parameter values that successfully predicts the actual tumour volume shrinkage has been reported. It should be pointed out that the large number of inherent biological boundary conditions (e.g. monotonic increase of all tumour cell class populations

for an imageable freely and smoothly growing tumour) dramatically limits the number of possible solutions (i.e. sets of parameter values that are able to predict real tumour shrinkage following treatment). This is particularly important since such limitations drastically facilitate the approach to the solution best representing clinical reality for each given medical data case.

The model has been developed to support and incorporate individualized clinical data such as imaging data (e.g., CT, MRI, PET slices, possibly fused) (e.g. Marias et al 2007), including the definition of the tumour contour and internal tumour regions (proliferating, necrotic), histopathologic (e.g., type of tumour) and the genetic data (e.g., p53 status, if available). The use of anonymized real data before and after chemotherapeutic treatment for the case of the TOP breast cancer clinical trials constitute the basis of the clinical adaptation and validation process. Obviously as more and more sets of medical data are exploited the reliability of the model "tuning" is expected to increase.

Finally, a detailed description of various technical issues involved in the present work (such as image processing, grid execution and visualization techniques) will be the topic of a dedicated paper.

4.10 Conclusions

An advanced multiscale model of clinical tumour response to chemotherapy has been presented. The paradigm of early breast cancer treated by epirubicin according to a branch of the TOP (Topoisomerase II Alpha Gene Amplification and Protein Overexpression Predicting Efficacy of Epirubicin) trial has been addressed. The model, stemming from previous work of our research group, is characterized by several crucial new features such as an advanced generic cytokinetic model for the response of a tumour cell to chemotherapeutic treatment, the explicit distinction of proliferating cells into stem cells and committed progenitor cells of limited proliferative capacity and the adaptation of the relative populations of the various cell categories/phases to the cell category/phase transition rates for free tumour growth, which constitutes an important novel tumour initialization strategy.

A sensitivity analysis regarding critical parameters of the model has revealed their effect on the behaviour of the biological system. In particular, the joint effect of the cell cycle duration and the fraction of newborn tumour cells that enter the G0 phase on the relative cell category populations, the joint effect of the cell cycle duration and the fraction of cells that perform symmetric division on the doubling time and the joint effect of the cell cycle duration and the necrosis rate of differentiated cells on the tumour doubling time have been studied. Furthermore, the joint effect of the cell kill ratio and the cell cycle duration on the therapeutic outcome (volume and diameter decrease percentages) has been also presented. Finally, a realistic time course of the tumour diameter for two clinical cases of the TOP trial has been achieved by selection of reasonable model parameter values, thus demonstrating a possible adaptation process of the model to real clinical trial data. The above studies and their results support the potential of the model to serve as both a theoretical investigation and a patient specific treatment optimization tool following completion of an ongoing clinical adaptation, optimization and validation process.

4.11 References

ACGT Advancing Clinicogenomic Trials on Cancer: Open Grid Services for Improving Medical Knowledge Discovery. EC and Japan funded R&D project. (FP6-2005-IST-026996) http://eu-acgt.org/acgt-for-you/researchers/in-silico-oncology/oncosimulator.html and http://www.eu-acgt.org/ (last visited on 22 Aug. 2009)

Al-Hajj M.A., Wischa M.S., Benito-Hernandez A., Morrison S.J., Clarke M.F., 2003. Prospective indentification of tumourigenic breast cancer cells. PNAS 100(7), 3983-3988.

Barnes, J. A., Dix, D. J., Collins, B. W., Luft, C., Allen J. W., 2001. Expression of inducible Hsp70 enhances the proliferation of MCF-7 breast cancer cells and protects against the cytotoxic effects of hyperthermia. Cell Stress & Chaperones 6 (4), 316–325.

Bast, R.C., Kufe, D.W., Pollock, R.E., Weichelbaum ,R.R., Holland, J.F., Frei E., Eds. 2000. Cancer Medicine 5th ed., B.C. Decker Inc, Hamilton, Ontario, Canada

Begg, A.C., 1997. Cell proliferation in tumours. In Steel, G.G., Ed. Basic Clinical Radiobiology. 2nd edition. Arnold, London,UK.

Breward, C.J., Byrne, H.M., Lewis, C.E., 2003. A multiphase model describing vascular tumour growth. Bull Math. Biol. 65(4), 609-40.

Cos, S., Recio, J., Sanchez-Barcelo, E. J., 1996. Modulation of the lenght of cell cycle time of MCF-7 human breast cancer cells by melatonin. Life Sci. 58, 811-816.

Clatz, O., Sermesant, M., Bondiau, P.Y., Delingette, H., Warfield, S.K., Malandain, G., Ayache, N., 2005. Realistic simulation of the 3-D growth of brain tumours in MR images coupling diffusion with biomechanical deformation. IEEE Trans. Med. Imaging. 24 (10), 1334-46.

ContraCancrum: Clinically Oriented Translational Cancer Multilevel Modelling (FP7-ICT-2007-2- 223979) http://contracancrum.eu/?q=node/1 (last visited on 22 Aug. 2009)

Danesi, R., Innocenti, F., Fogli, S., Gennari, A., Baldini, E., Paolo, A.D., Salvadori, B., Bocci, G., Conte, P.F., Del Tacca, M., 2002. Pharmacokinetics and pharmacodynamics of combination chemotherapy with paclitaxel and epirubicin in breast cancer patients. J. Clin. Pharmacol. 53, 508-518.

Dean M., Fojo T., Bates S., 2005. Tumour stem cells and drug resistance. Nat. Rev. Cancer 5, 275-284.

Deisboeck, T.S., Zhang, L., Yoon, J., Costa, J., 2009. In silico cancer modeling: is it ready for prime time?. Nat. Clin. Pract. Oncol. 6 (1), 34-42.

Descamps, S., Lebourhis, X., Delehedde, M., Boilly, B., Hondermarck H., 1998. Nerve Growth Factor Is Mitogenic for Cancerous but Not Normal Human Breast Epithelial Cells. J. Biol. Chem. 273 (27), 16659–16662.

Dewey, W.C., Ling, C.C., Meyn, R.E., 1995. Radiation-induced apoptosis: relevance to radiotherapy. Int. J. Radiat. Oncol. Biol. Phys. 33 (4), 781–796.

Diehn M., Cho R., Dorie M., Kulp A., Weissman I., Brown M., Clarke M., 2007. Analyzing the Sensitivity of Breast Cancer Stem Cells to Ionizing Radiation and Chemotherapy. *Int. J. Radiat. Oncol. Biol. Phys.* 69(3), S38-S39.

Dionysiou, D. D., Stamatakos, G. S., Uzunoglu, N.K., Nikita, K. S., Marioli, A., 2004. A four-dimensional simulation model of tumour response to radiotherapy in vivo: parametric validation considering radiosensitivity, genetic profile and fractionation. J. theor. Biol. 230, 1–20.

Dionysiou, D. D., Stamatakos, G. S., Uzunoglu, N.K., Nikita, K. S., 2006. A computer simulation of in vivo tumour growth and response to radiotherapy: New algorithms and parametric results. Comp. Biol. Med. 36, 448–464.

Dionysiou, D.D., Stamatakos, G.S., 2006. Applying a 4D multiscale in vivo tumour growth model to the exploration of radiotherapy scheduling: the effects of weekend treatment gaps and p53 gene status on the response of fast growing solid tumours. Cancer Informatics 2, 113-121.

Duechting, W., Vogelsaenger, T., 1981. Three-dimensional pattern generation applied to spheroidal tumour growth in a nutrient medium. Int. J. Biomed. Comput. 12 (5), 377–392.

Durbecq, V., Paesmans, M., Cardoso, F., Desmedt, C., Di Leo, A., Chan, S., Friedrichs, K., Pinter, T., Van Belle, S., Murray, E., Bodrogi, I., Walpole, E., Lesperance, B., Korec, S., Crown, J., Simmonds, P., Perren, T. J., Leroy, J.-Y., Rouas, G., Sotiriou, C., Piccart, M., Larsimont, D., 2004. Topoisomerase-IIA expression as a predictive marker in a population of advanced breast cancer patients randomly treated either with single-agent doxorubicin or single-agent docetaxel. Mol. Cancer Ther. 3, 1207-1214.

Elston, C.W., Ellis, I.O., 1991. Pathologic prognostic factors in breast caner. I. The valued of histological grade in breast cancer: experience from a large study with long-term follow-up. Histopathol. 19, 403-410.

Enderling, H., Chaplain, M.A.J., Anderson, A.R.A, Vaidya, J.S., 2007. A mathematical model of breast cancer development, local treatment and recurrence. J. theor. Biol. 246, 245–259.

FDA 1999, Food and Drug Administration. Division of Oncology Products, HDF-150. Review and Evaluation of Pharmacology and Toxicology Data. Original Review. Keywords: Epirubicin, breast cancer, anthracycline. NDA: 21-010 and 50-778. Type: NDA. CDR date: 12/15/98. Date Review Completed: June 14,1999.

Fillmore C.M., Kuperwasser C., 2008. Human breast cancer cell lines contain stem-like cells that self-renew, give rise to phenotypically diverse progeny and survive chemotherapy. Breast Cancer Research 10(2): R25.

Frieboes, H.B., Zheng, X., Sun, C.H., Tromberg, B., Gatenby, R., Cristini, V., 2006. An integrated computational/experimental model of tumour invasion. Cancer Res. 66 (3), 1597-604.

Gennari, A., Sormani, M.P., Pronzato, P., Puntoni, M., Colozza, M., Pfeffer, U., Bruzzi, P., 2008. HER2 Status and Efficacy of Adjuvant Anthracyclines in Early Breast Cancer: A Pooled Analysis of Randomized Trials. J. Natl. Cancer Inst. 100 (1), 14-20.

Ginsberg, T., Ulmer, W., Duechting, W., 1993. Computer simulation of fractionated radiotherapy: further results and their relevance to percutaneous irradiation and brachytherapy. Strahlenther. Onkol. 169, 304-310.

Graf, N., Hoppe, A., 2006. What are the expectations of a Clinician from In Silico Oncology ? Proceedings of the 2nd International Advanced Research Workshop on In Silico Oncology, Kolympari, Chania, Greece, Sept. 25-26, 2006, pp. 36-38.

Guiot, C., Delsanto, P.P., Carpinteri, A., Pugno, N., Mansury, Y., Deisboeck, T.S., 2006. The dynamic evolution of the power exponent in a universal growth model of tumours. J Theor Biol. 240 (3), 459-63.

In Silico Oncology Group, Institute of Communications and Computer Systems, National Technical University of Athens www.in-silico-oncology.iccs.ntua.gr (last visited on 22 Aug. 2009).

Knoop, A. S., Knudsen, H., Balslev, E., Rasmussen, B.B, Overgaard, J., Nielsen, K.V., Schonau, A., Gunnarsdo´ttir, K., Olsen, K.E., Mouridsen, H., Ejlertsen, B., 2005. Retrospective Analysis of Topoisomerase IIa Amplifications and Deletions As Predictive Markers in Primary Breast Cancer Patients Randomly Assigned to Cyclophosphamide, Methotrexate, and Fluorouracil or Cyclophosphamide, Epirubicin, and Fluorouracil: Danish Breast Cancer Cooperative Group. J. Clin. Oncol. 23 (30), 7483-7490.

Kolokotroni, E. A., Stamatakos, G. S., Dionysiou, D. D., Georgiadi, E. Ch., Desmedt, C., Graf, N.M., 2008. Translating Multiscale Cancer Models into Clinical Trials: Simulating Breast Cancer Tumour Dynamics within the Framework of the "Trial of Principle" Clinical Trial and the ACGT Project. Proc. 8th IEEE International Conference on Bioinformatics and Bioengineering (BIBE 2008), Athens, Greece, 8-10 Oct. 2008. IEEE Catalog Number: CFP08266, ISBN: 978-1-4244-2845-8, Library of Congress: 2008907441, Paper No. BE-2.1.1, length: 8 pages (in electronic format) 2008..

Lankelma, J., Dekker, H., Luque, R.F., Luykx, S., Hoekman, K., van der Valk P., van Diest P.J., Pinedo H.M., 1999. Doxorubicin gradients in human breast cancer. Clin. Cancer Res. 5, 1703-1707.

Liu S., Dontu G., Wischa M.S., 2005. Mammary stem cells, self-renewal pathways and carcinogenesis. Breast Cancer Res. 7(3), 86-95.

Marias K., D.Dionysiou, G.S. Stamatakos, F.Zacharopoulou, E.Georgiadi, T.G.Maris and I.Tollis. 2007. Multi-level analysis and information extraction considerations for validating 4D models of human function. Lect Notes Comput Sci 4561: 703-709.

Maseide, K. Rofstad, E.K. 2000. Mathematical modeling of chronical hypoxia in tumours considering potential doubling time and hypoxic cell lifetime. Radiother. Oncol. 54, 171-177.

Meyer, J.S., McDivitt, R., Stone, K., Prey, M., Bauer, W., 1984. Practical breast carcinoma cell kinetics: review and update. Breast Cancer Res. Treat. 4, 79-88.

Morisson, S.J., Kimble J., 2006. Asymmetric and symmetric stem-cell divisions in development and cancer. Nature 441, 1068-1074.

Murray, J.D., 2003. Mathematical Biology II. Spatial Models and Biomedical Applications, third ed. Springer-Verlag: Heidelberg, pp. 543-546.

Perez, C.A., Brady L.W. (Eds.) 1998. Principles and Practice of Radiation Oncology. Lippincott-Raven, Philadelphia, p.10.

Potti, A., Dressman, H.K., Bild, A., Riedel, R.F., Chan, G., Sayer, R., Cragun, J., Cottrill, H., Kelley, M.J., Petersen, R., Harpole, D., Marks, J., Berchuck, A., Ginsburg, G.S., Febbo, P., Lancaster, J., Nevins, J.R., 2006. Genomic signatures to guide the use of chemotherapeutics. Nat. Med. 12(11), 1294-1300.

Pritchard, K.I., Messersmith, H., Elavathil, L., Trudeau, M., Malley, F.O'., Dhesy-Thind, B., 2008. HER-2 and Topoisomerase II As Predictors of Response to Chemotherapy. J. Clin. Oncol., 26 (5), 736-744.

Ramis-Conde, I., Chaplain, M.A.J., Anderson, A.R.A., 2008. Mathematical modelling of cancer cell invasion of tissue. Math. Comp. Mod. 47, 533-545.

SAAM II 2009 http://depts.washington.edu/saam2/ (last visited on 25 Sept. 2009)

Salmon, S.E., Sartorelli, A.C., 2001. Cancer Chemotherapy. In: Katzung, B.G. (Eds.), Basic & Clinical Pharmacology. Lange Medical Books/McGraw-Hill: International Edition, pp. 923-1044.

Spratt, J.S., Heuser, L., Kuhns, J.G., Reiman, H.M., Buchanan, J.B., Polk, H.C., Sandoz J., 1981. Association between the actual doubling times of primary breast cancer with histopathologic characteristics and Wolfe's parenchymal mammographic patterns. Cancer 47: 2265-2268.

Stamatakos, G.S., Dionysiou, D.D., Zacharaki, E.I., Mouravliansky, N.A., Nikita, K.S., Uzunoglu, N.K., 2002. In silico radiation oncology: combining novel simulation algorithms with current visualization techniques. Proceedings of the IEEE 90 (11) , 1764-1777.

Stamatakos, G.S., Antipas, V.P., Uzunoglu, N.K., Dale, R.G., 2006a. A four dimensional computer simulation model of the in vivo response to radiotherapy of glioblastoma multiforme: studies on the effect of clonogenic cell density. Br. J. Radiol. 79, 389-400.

Stamatakos, G.S., Antipas, V.P. Uzunoglu, N.K., 2006b. A Spatiotemporal, Patient Individualized Simulation Model of Solid Tumour Response to Chemotherapy in Vivo: The Paradigm of Glioblastoma Multiforme Treated by Temozolomide. IEEE Trans. Biomed. Eng. 53, 1467- 1477.

Stamatakos, G.S., Dionysiou, D.D., Graf, N.M., Sofra, N.A., Desmedt, C., Hoppe, A., Uzunoglu, N., Tsiknakis, M., 2007. The Oncosimulator: a multilevel, clinically oriented simulation system of tumour growth and organism response to therapeutic schemes. Towards the clinical evaluation of in silico oncology. Proceedings of the 29th Annual International Conference of the IEEE EMBS Cite Internationale, August 23-26, SuB07.1: 6628-6631 , Lyon, France , 2007

Stamatakos, G.S., E. A. Kolokotroni, D. D Dionysiou, E. C Georgiadi, C. Desmedt, 2010. An advanced discrete state - discrete event multiscale simulation model of the response of a solid tumour to chemotherapy. Mimicking a clinical study. J.Theor. Biol. 266 (2010) 124-139.

Steel, G.G. 1997. Clonogenic cells and the concept of cell survival. In: Steel, G.G. (Ed.), Basic Clinical Radiobiology, second ed. Arnold: London, UK.

Sterin, M., Cohen, J.S., Mardor, Y., Berman, E., Ringel, I., 2001. Levels of phospholipid metabolites in breast cancer cells treated with antimitotic drugs. Cancer Res. 61, 7536-7543.

Stingl J., Carlos C., 2007. Molecular heterogeneity of breast carcinomas and the cancer stem cell hypothesis. Nat. Rev. Cancer 7, 791-799.

Swanson, K.R., Alvord, E.C., Murray, J.D., 2002. Virtual brain tumours (gliomas) enhance the reality of medical imaging and highlight inadequacies of current therapy. Br. J. Cancer 86, 14-18.

TOP trial [http://clinicaltrials.gov/ct2/show/NCT00162812] (last visited on 22 Aug 2009).

Zacharaki, E.I., Stamatakos, G.S., Nikita, K.S., Uzunoglu, N.K., 2004. Simulating growth dynamics and radiation response of avascular tumour spheroid model validation in the case of an EMT6/Ro multicellular spheroid. Comput. Methods Programs Biomed. 76, 193-206.

Zoubek, A. , Slavc, I., Mann, G., Trittenwein, G., Gadner, H., 1999. Natural course of a Wilms' tumour. Lancet 354(9175), 344.

Appendix SA

The principle of the tumour constitution initialization technique is to start with a small number of stem cells and with specific cell category transition rates that are assumed to hold true for a relatively small time interval about the treatment baseline. Specific values are assigned to the phase durations and transition rates. Gradually, all cell categories and phases become populated and after sufficient time the relative cell categories populations tend to reach an equilibrium state. If in subsequent simulations the GCs are initialized using the cell category /phase relative population values corresponding to this equilibrium state, a mathematically monotonous and biologically normal free growth behaviour will be achieved. The challenge is to successfully locate the point beyond which equilibrium has been achieved and use the relative populations (or “fractions of populations”) after that point for the correct initialization of the tumour. Certain combinations of category / phase transition rates cannot sustain tumour growth (see section 4.4 for the condition for monotonic free growth). In such cases the method will *correctly* fail to create the initial tumour and a relevant warning message will be issued by the simulation system.

More specifically, the technique consists in the following: A limited number of geometrical cells N_{GCs} are considered. Each GC initially contains a small number of stem cells, e.g. 100, residing in the various cell cycle phases (G1, S, G2, M) and the G0 phase. Time initialization, i.e. the time already spent by clustered stem cells in the phase they reside, is assigned using a pseudo-random generator. Different random number sets are assigned to different GCs. The aim is to avoid artificial synchronizations which would result in the group of GCs considered behaving as one big GC. The system is left to evolve and produce all cell category populations (distributed to the various cell phases). The code execution has to continue until equilibrium is reached and the various cell categories population percentages have been stabilized. The equilibrium condition applied is described by the following inequality:

$$\left| \frac{\sum_{\text{timestep}\,i=n-N+1}^{n} \frac{\text{cell category relat. population at time step i}}{N} - \sum_{\text{timestep}\,j=(n-M)-N+1}^{n-M} \frac{\text{cell category rel. population at time step j}}{N}}{\sum_{\text{timestep}\,j=(n-M)-N+1}^{n-M} \frac{\text{cell category rel. population at time step j}}{N}} \right| \leq \varepsilon$$

for *k* consecutive averages

(SA.1)

In words, the average of *N* consecutive values of the relative population of cells clustered in a given mitotic potential category is calculated every *M* time steps (hours). The variation of the average is also calculated. If the variation of the average is less than ε, where ε is a small positive real number, for *k* consequent average values, equilibrium has been reached. The condition must be true for all cell category relative populations.

In order to check the efficiency of the equilibrium condition and determine the most effective values of the parameters N_{GCs}, *M*, *N*, ε and *k* for the achievement of convergence a number of exploratory executions have been performed. The conclusions reached for each parameter are given below:

(a) Number of geometrical cells (N_{GCs}) used for the initialization of the tumour constitution

Fig.SA.1 presents the relative population of proliferating stem cells calculated for various values of N_{GCs}. For N_{GCs} = 1 the technique produces the highest value of the relative population. As more and more GCs are considered, the relative population of stem cells decreases fast towards a steady value. Similar observations have been made for the rest of the mitotic potential cell categories for various proliferation statuses. Taking into consideration the dependence of the execution time on the number of GCs considered, the selection of N_{GCs} = 64 has shown to ensure good convergence, while at the same time keep the code execution time for the tumour constitution initialization reasonably low. The previous observations hold true at least for the breast cancer parameter value subspace.

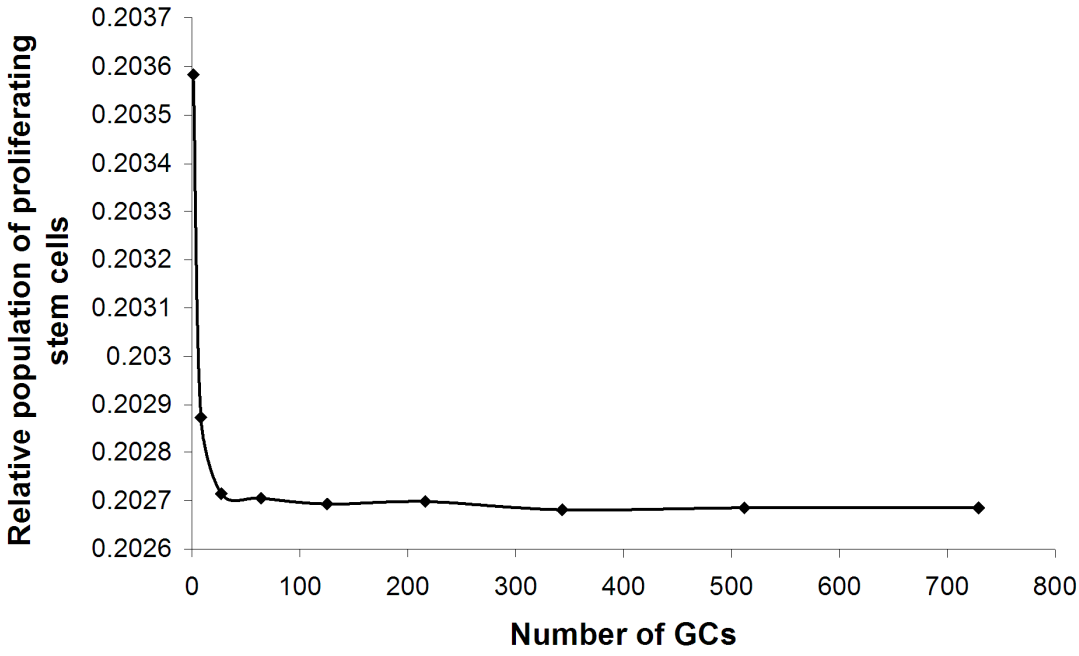


Fig. SA.1. Study of the model convergence. Relative population of proliferating stem cells as a function of the number of geometrical cells considered. The method described in section 4 has been applied . M=T_c, N=T_c, ε=10^{-5}, k=10.

(b) Number of N consecutive values of the relative population of cells clustered in a given mitotic potential category

It appears that the optimal value of *N* coincides with the cell cycle duration (in hours). This is shown in Fig. SA.2, which depicts the example of the relative population of proliferating stem cells along with its average value as a function of time for $N = T_c$ (in hours) = 96 and N =50 consecutive values. The relative population time course seems to follow a repeated pattern with a period equal to the cell cycle duration (T_c) of the tumour under consideration. On the other hand the average value over $N = T_c$ values of stem cell relative population follows a rather smooth course. Regarding the source of fluctuations the various quantizations of the model must play a major role. Nevertheless, the maximum fluctuation is about 2 % of the stabilized average value, which can be considered insignificant, when taking into account the uncertainties of the medical data to be used by the model.

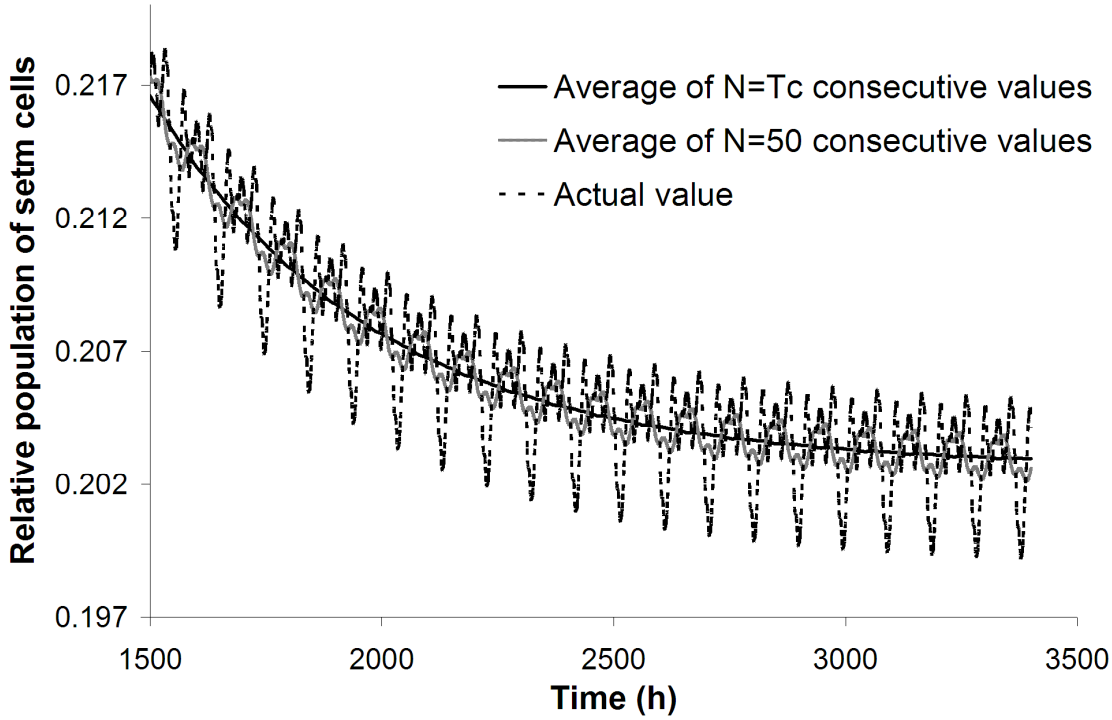


Fig. SA.2. Study of the model convergence. Fluctuation of the relative population of proliferating stem cells and its average value over time for N = T_c (in h) = 96 and N = 50 consecutive values. T_c denotes the duration of the cell cycle.

(c) Time interval M between consecutive calculations of the average values of the relative population of cells clustered in a given mitotic potential category

The average value of *N* consecutive values of the relative population of cells clustered in a given mitotic potential category is compared with its predecessor every *M* time steps (hours). It has been shown again that the most effective *M* value so that good convergence is achieved, is the value of the cell cycle duration in h (Fig. SA.3). Fig. SA.3 *triangles* shows the exemplary case of the relative variation of the average of $N = T_c$ (=96) consecutive instantaneous values of the relative population of stem cells, taken every $M = T_c$ (=96) time steps (hours) over time. The curve is fairly smooth and tends to zero. On the contrary, Fig. SA.3 *circles* depicts the same case in which *M* has not been set to T_c. A small fluctuation of the relative variation over time (about 2%) is apparent. Here again the fluctuation is still not significant.

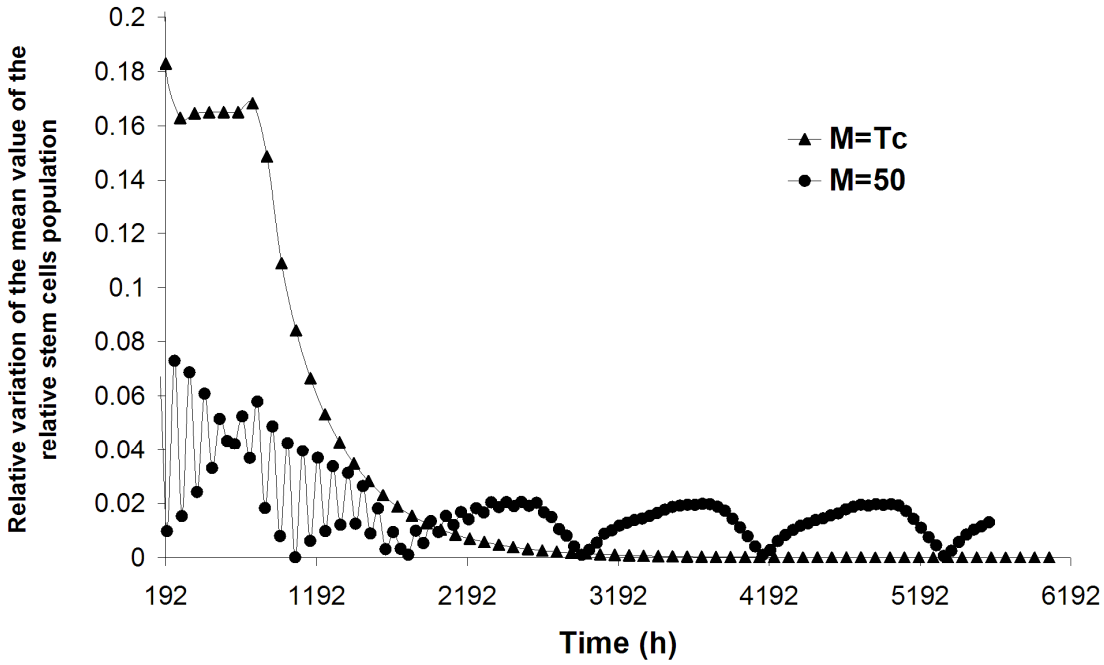


Fig. SA.3. Study of the model convergence. Relative variation of the average of N = Tc = 96 consecutive instantaneous values of the relative population of stem cells taken every M time steps (hours) over time. Triangles correspond to M equal to the numerical value of the cell cycle duration (96 h). Circles correspond to M = 50. Both curves correspond to the same tumour. The convergence of this parameter to zero is essential for the equilibrium condition described in appendix A to be applied.

(d) The small positive real number ε

Ideally, equilibrium of the relative cell populations has been reached when the relative variation of the average of *N* consecutive values of the relative population of each cell equivalence subclass over time has become zero. Since this may happen at infinity due to small quantization errors, the small real positive value of $\varepsilon = 0.00001 = 10^{-5}$ has been adopted. The latter has been shown to ensure convergence in the breast cancer parameter subspace considered.

(e) The number k of consecutive average values considered

Satisfaction of the equilibrium condition for $k = 10$ consecutive averages has been proved to ensure convergence in the parameter subspace considered.

In summary, the optimal parameter values for achieving convergence in the tumour constitution initialization procedure, for the specific parameter subspace considered in this chapter, are $N_{GCs} = 64$, $N = T_c$, $M = T_c$, $\varepsilon = 10^{-5}$ and $k = 10$. It is noted that the initialization code execution time ranges from a few fractions of a second to a few seconds on a standard laptop depending on the particular characteristics of the tumour (e.g. cell cycle duration, fraction of cells entering the G0 phase following mitosis etc.).

Appendix SB

The algorithm developed for the avoidance of premature extensive fragmentation and vacuum enclosures artifacts:

(I) detects tumour occupied GCs that are surrounded by empty GCs in a “6-GCs Neighborhood” and moves their contents (by one GC at each time step) towards the tumour’s center of mass. The direction of movement is chosen based on the minimum distance of the GC under consideration from the center of mass along the x, y, z coordinates. The corresponding quantity to be calculated each time is the following:

min{abs(GC.x-center.x), abs(GC.y-center.y), abs(GC.z-center.z)}

where abs() denotes absolute value, GC.x , GC.y and GC.z are the x , y and z coordinates of the GC respectively and center.x, center.y and center.z are the x, y and z coordinates of the tumour’s center of mass respectively.

If more than one direction is characterized by the same minimum distance then a random number generator is used for the selection of the movement direction.

(II) detects empty GCs that are surrounded by occupied GCs in a “6-GCs Neighborhood” and fills them with the contents of adjacent GCs by applying the tumour shrinkage procedure described above.

5. Update on the utilization of Wilms tumour data for clinical adaptation and validation of the ACGT Oncosimulator

(Code letter : W)

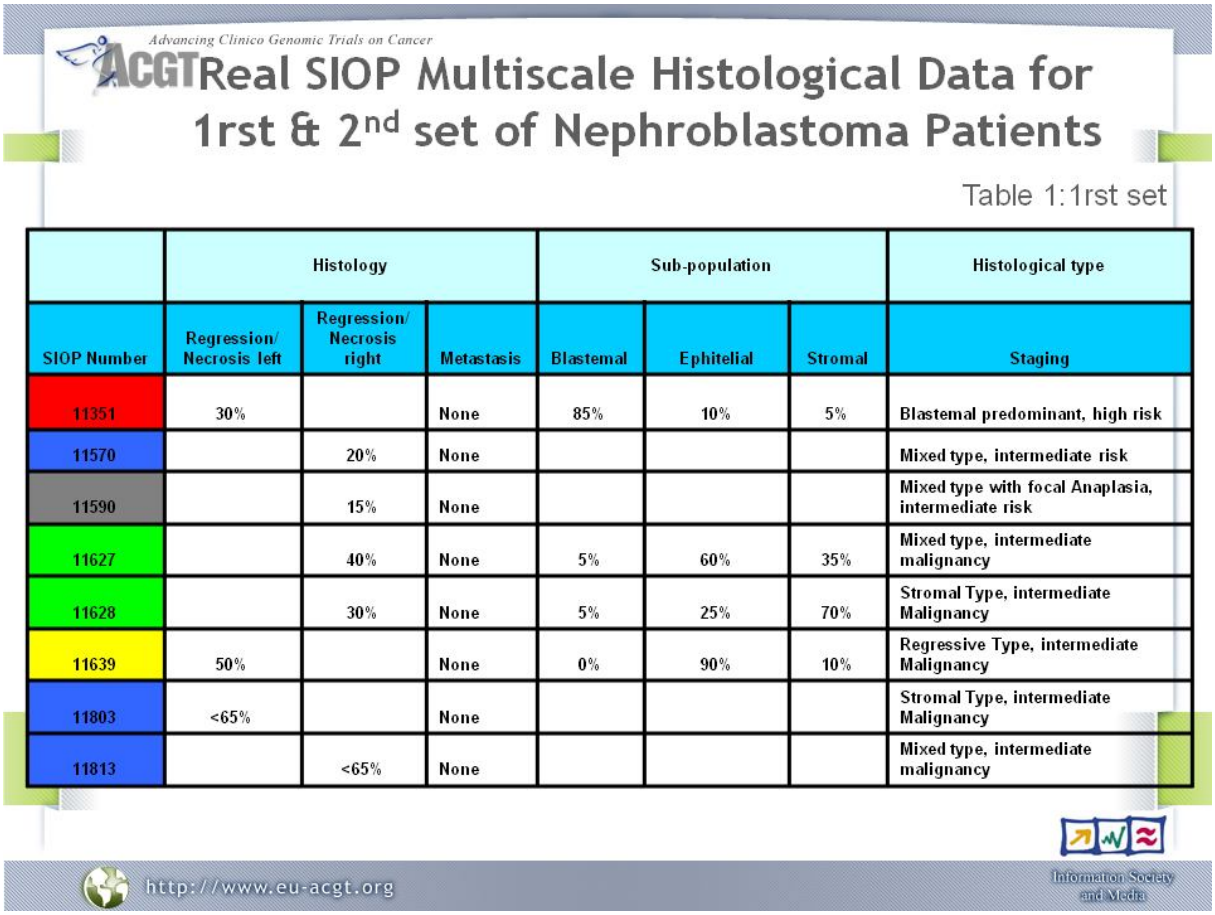
IMPORTANT NOTICE: In order to avoid serious waste of time a number of highly informative tables prepared for the review presentations of the related work have been included in this chapter. Short comments have been added on each figure/slide caption so as to render the tabular exposition easy to follow.

In this and the following chapter an account of the utilization of anonymized real Wilms tumour (nephroblastoma) and breast cancer data for the clinical adaptation of the ACGT Oncosimulator is provided. It is noted that clinical adaptation is a prerequisite for the clinical validation of the system and therefore it may be considered the first stage towards clinical validation in the strict sense of the term. It is pointed out that data explorations are still to be considered of a tentative nature. However, the high rate of "discovering" new highly specific and efficient ways to optimize the clinical adaptation procedures is expected to lead to the crystallization of several model parameter values and/or value ranges rather fast.

5.1 Real multiscale nephroblastoma data sets retrieved from the SIOP clinical trial

Up to now two sets of pertinent multiscale nephroblastoma data sets have been retrieved from the SIOP clinical trial repository. The first one consists of 8 anonymized cases whereas the second one of 9 cases (plus one case from the first set on which a more advanced image segmentation has been applied). This means that 17 patients have been considered in total.

Fig. W1 and W2 depict in a tabular form the real SIOP trial histological data for the nephroblastoma patients considered up to now.



Advancing Clinico Genomic Trials on Cancer

ACGT

Real SIOP Multiscale Histological Data for 1rst & 2nd set of Nephroblastoma Patients

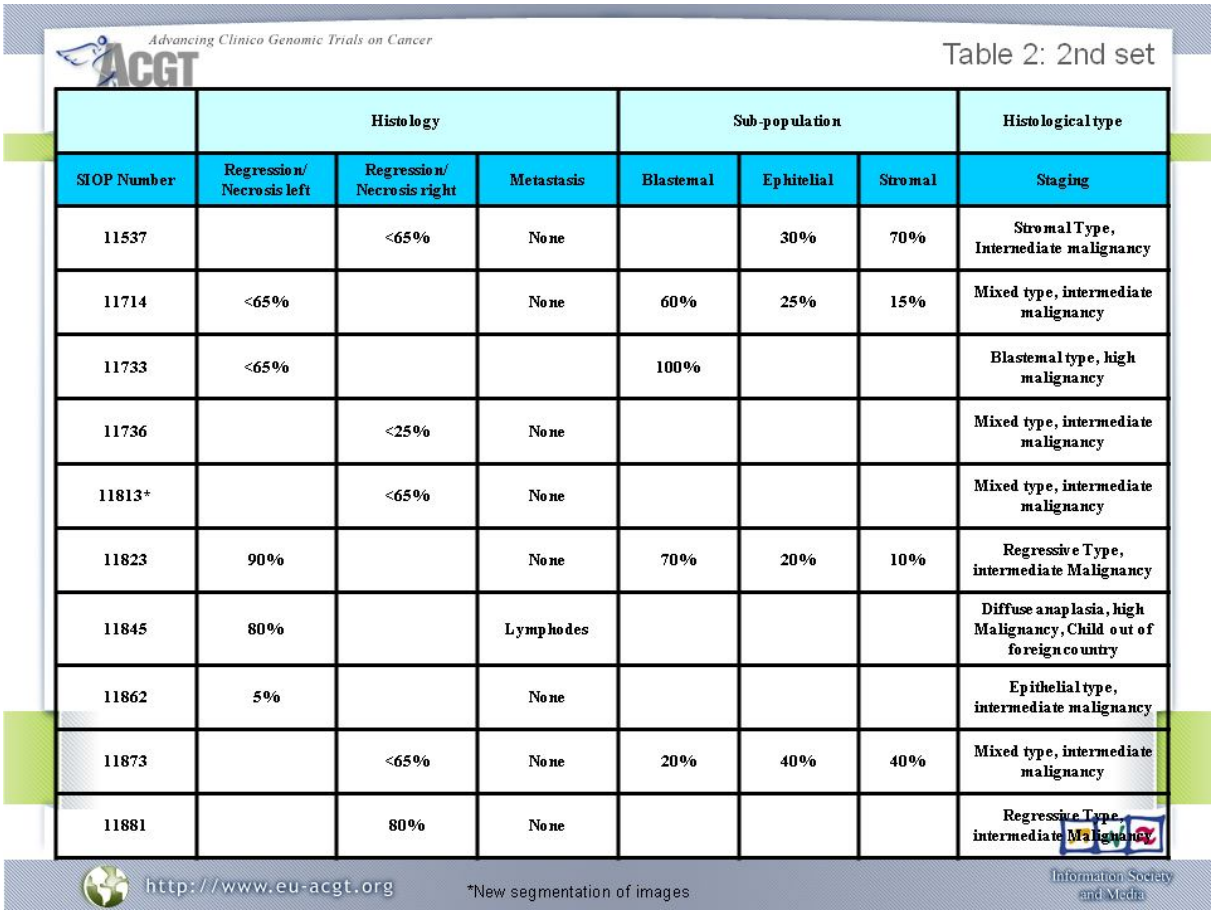
Table 1:1rst set

	Histology			Sub-population			Histological type
SIOP Number	Regression/ Necrosis left	Regression/ Necrosis right	Metastasis	Blastemal	Ephitelial	Stromal	Staging
11351	30%		None	85%	10%	5%	Blastemal predominant, high risk
11570		20%	None				Mixed type, intermediate risk
11590		15%	None				Mixed type with focal Anaplasia, intermediate risk
11627		40%	None	5%	60%	35%	Mixed type, intermediate malignancy
11628		30%	None	5%	25%	70%	Stromal Type, intermediate Malignancy
11639	50%		None	0%	90%	10%	Regressive Type, intermediate Malignancy
11803	<65%		None				Stromal Type, intermediate Malignancy
11813		<65%	None				Mixed type, intermediate malignancy

http://www.eu-acgt.org

Information Society and Media

Fig. W1 Real SIOP multiscale histological data for the first set of the anonymized nephroblastoma patients

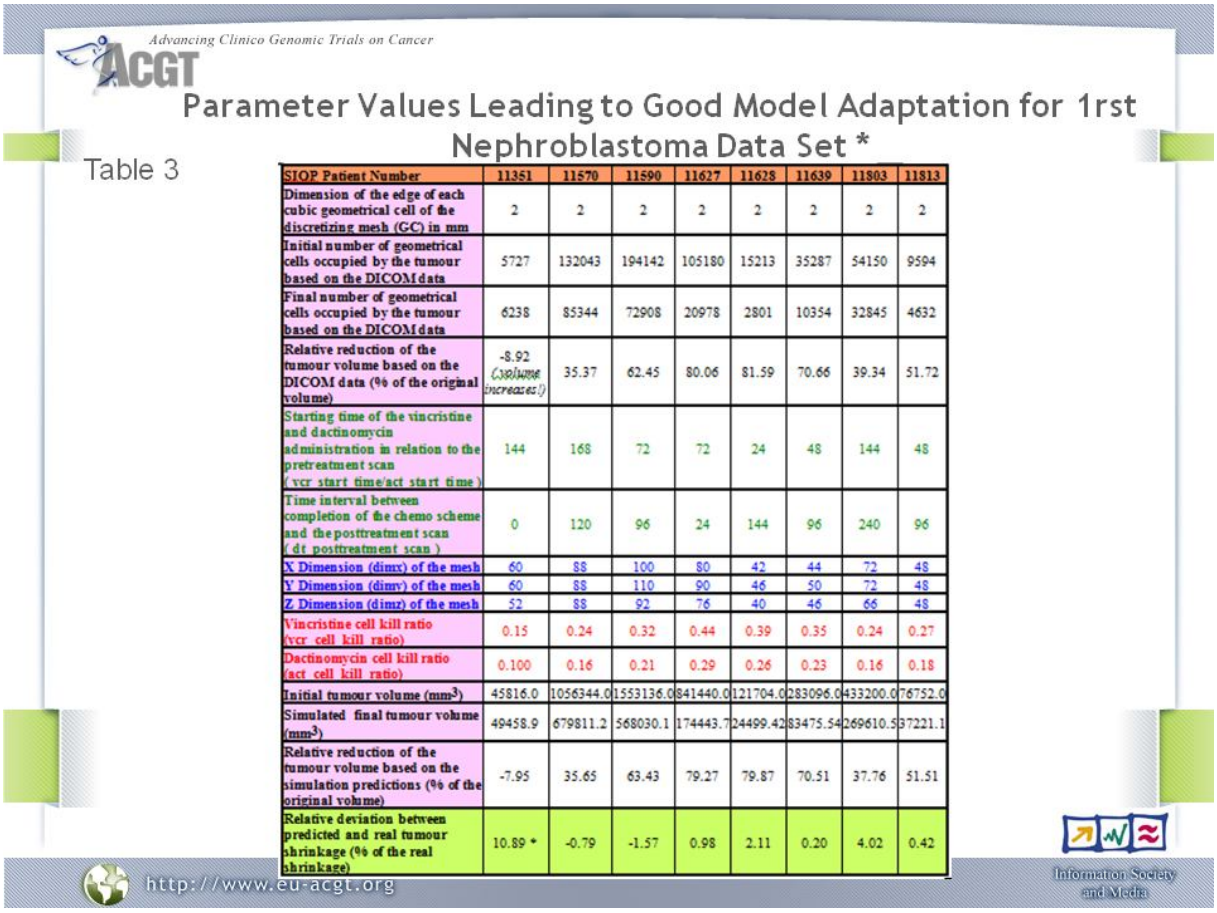


	Histology			Sub-population			Histological type
SIOP Number	Regression/ Necrosis left	Regression/ Necrosis right	Metastasis	Blastemal	Ephitelial	Stromal	Staging
11537		<65%	None		30%	70%	Stromal Type, Internediate malignancy
11714	<65%		None	60%	25%	15%	Mixed type, intermediate malignancy
11733	<65%			100%			Blastemal type, high malignancy
11736		<25%	None				Mixed type, intermediate malignancy
11813*		<65%	None				Mixed type, intermediate malignancy
11823	90%		None	70%	20%	10%	Regressive Type, intermediate Malignancy
11845	80%		Lymphodes				Diffuse anaplasia, high Malignancy, Child out of foreign country
11862	5%		None				Epithelial type, intermediate malignancy
11873		<65%	None	20%	40%	40%	Mixed type, intermediate malignancy
11881		80%	None				Regressive Type, intermediate Malignancy

Fig. W2. Real SIOP multiscale histological data for the second set of the anonymized nephroblastoma patients

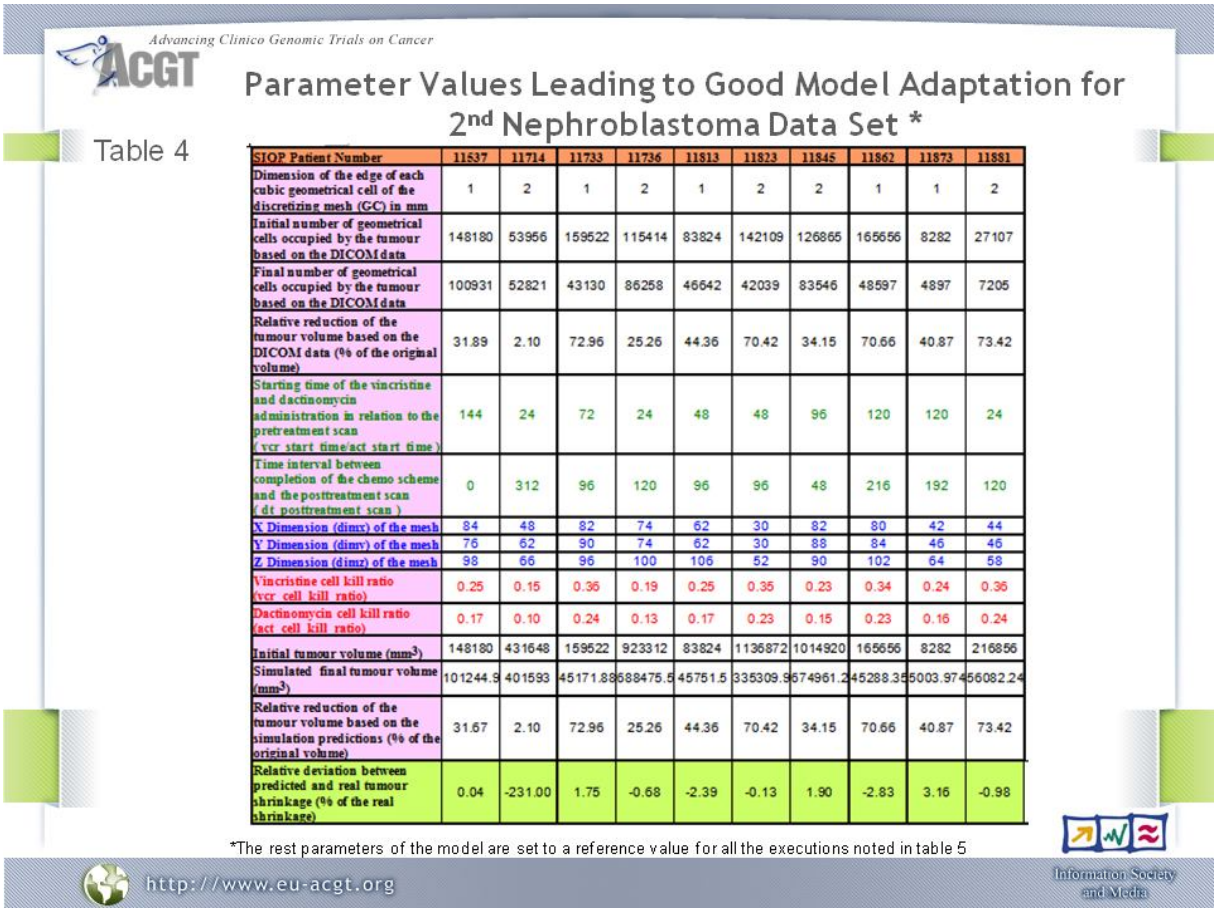
5.2 Indicative parameter value combinations leading to good model adaptation for the nephroblastoma datasets

Fig. W3, W4 and W5 provide sets of plausible parameter value combinations that lead to a good adaptation of the model to the clinical data. It is noted that the cell kill ratio (for both drugs administered) has been assumed to summarize the most critical tumour characteristics that determine its partial survival of kill following chemotherapeutic treatment.



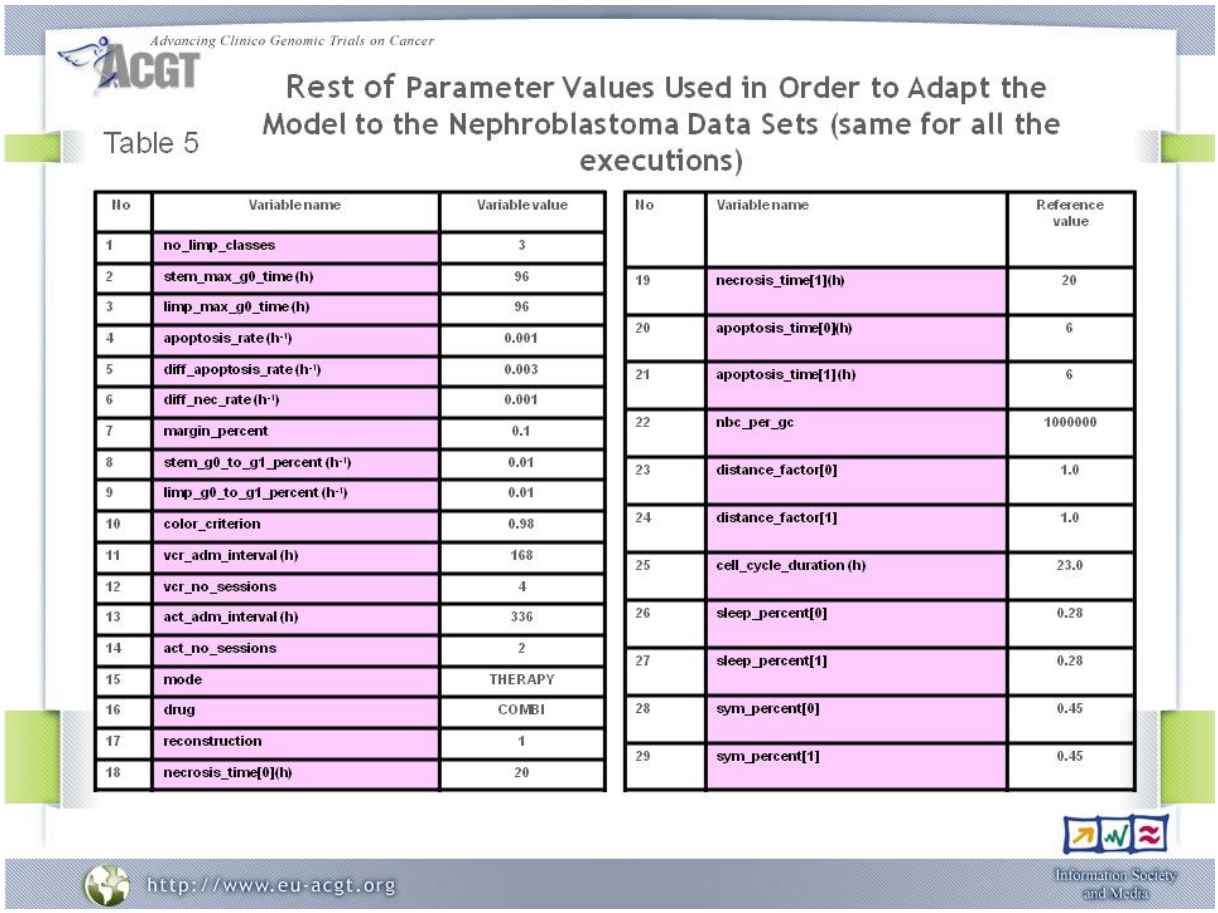
SIOP Patient Number	11351	11570	11590	11627	11628	11639	11803	11813
Dimension of the edge of each cubic geometrical cell of the discretizing mesh (GC) in mm	2	2	2	2	2	2	2	2
Initial number of geometrical cells occupied by the tumour based on the DICOM data	5727	132043	194142	105180	15213	35287	54150	9594
Final number of geometrical cells occupied by the tumour based on the DICOM data	6238	85344	72908	20978	2801	10354	32845	4632
Relative reduction of the tumour volume based on the DICOM data (% of the original volume)	-8.92 (volume increases!)	35.37	62.45	80.06	81.59	70.66	39.34	51.72
Starting time of the vincristine and dactinomycin administration in relation to the pretreatment scan (vcr_start_time/act_start_time)	144	168	72	72	24	48	144	48
Time interval between completion of the chemo scheme and the posttreatment scan (dt_posttreatment_scan)	0	120	96	24	144	96	240	96
X Dimension (dimx) of the mesh	60	88	100	80	42	44	72	48
Y Dimension (dimy) of the mesh	60	88	110	90	46	50	72	48
Z Dimension (dimz) of the mesh	52	88	92	76	40	46	66	48
Vincristine cell kill ratio (vcr_cell_kill_ratio)	0.15	0.24	0.32	0.44	0.39	0.35	0.24	0.27
Dactinomycin cell kill ratio (act_cell_kill_ratio)	0.100	0.16	0.21	0.29	0.26	0.23	0.16	0.18
Initial tumour volume (mm^3)	45816.0	1056344.0	1553136.0	841440.0	121704.0	283096.0	433200.0	76752.0
Simulated final tumour volume (mm^3)	49458.9	679811.2	568030.1	174443.7	24499.42	83475.54	269610.5	37221.1
Relative reduction of the tumour volume based on the simulation predictions (% of the original volume)	-7.95	35.65	63.43	79.27	79.87	70.51	37.76	51.51
Relative deviation between predicted and real tumour shrinkage (% of the real shrinkage)	10.89 *	-0.79	-1.57	0.98	2.11	0.20	4.02	0.42

Fig. W3 Paramerter value combinations that lead to a good model adaptation for the first nephroblastoma data set



SIOP Patient Number	11537	11714	11733	11736	11813	11823	11845	11862	11873	11881
Dimension of the edge of each cubic geometrical cell of the discretizing mesh (G-C) in mm	1	2	1	2	1	2	2	1	1	2
Initial number of geometrical cells occupied by the tumour based on the DICOM data	148180	53956	159522	115414	83824	142109	126865	165656	8282	27107
Final number of geometrical cells occupied by the tumour based on the DICOM data	100931	52821	43130	86258	46642	42039	83546	48597	4897	7205
Relative reduction of the tumour volume based on the DICOM data (% of the original volume)	31.89	2.10	72.96	25.26	44.36	70.42	34.15	70.66	40.87	73.42
Starting time of the vincristine and dactinomycin administration in relation to the pretreatment scan (vcr_start_time/act_start_time)	144	24	72	24	48	48	96	120	120	24
Time interval between completion of the chemo scheme and the posttreatment scan (dt_posttreatment_scan)	0	312	96	120	96	96	48	216	192	120
X Dimension (dimx) of the mesh	84	48	82	74	62	30	82	80	42	44
Y Dimension (dimy) of the mesh	76	62	90	74	62	30	88	84	46	46
Z Dimension (dimz) of the mesh	98	66	96	100	106	52	90	102	64	58
Vincristine cell kill ratio (vcr_cell_kill_ratio)	0.25	0.15	0.36	0.19	0.25	0.35	0.23	0.34	0.24	0.36
Dactinomycin cell kill ratio (act_cell_kill_ratio)	0.17	0.10	0.24	0.13	0.17	0.23	0.15	0.23	0.16	0.24
Initial tumour volume (mm^3)	148180	431648	159522	923312	83824	1136872	1014920	165656	8282	216856
Simulated final tumour volume (mm^3)	101244.9	401593	45171.88	688475.5	45751.5	335309.9	674961.2	45288.35	5003.97	456082.24
Relative reduction of the tumour volume based on the simulation predictions (% of the original volume)	31.67	2.10	72.96	25.26	44.36	70.42	34.15	70.66	40.87	73.42
Relative deviation between predicted and real tumour shrinkage (% of the real shrinkage)	0.04	-231.00	1.75	-0.68	-2.39	-0.13	1.90	-2.83	3.16	-0.98

Fig. W4 Paramerter value combinations that lead to a good model adaptation for the second nephroblastoma data set



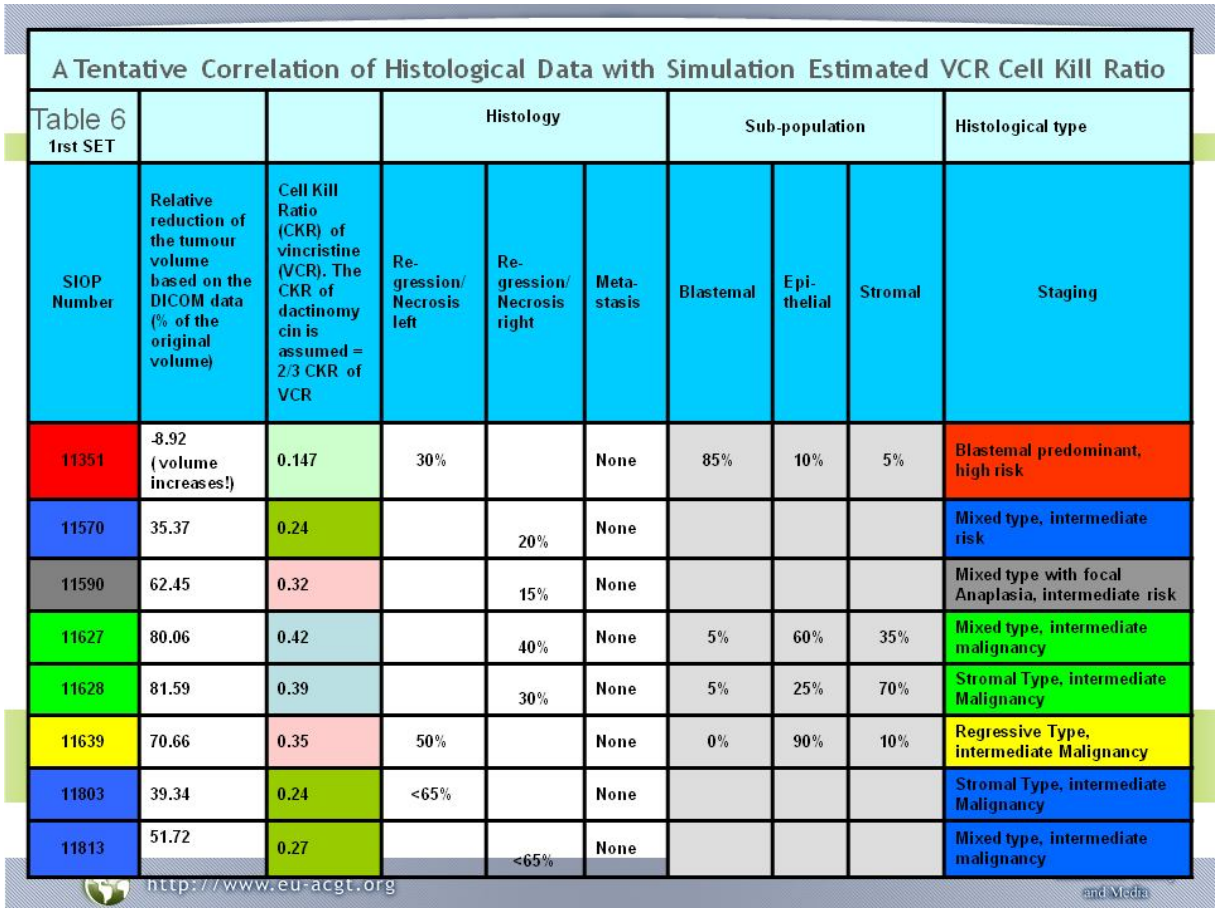
No	Variable name	Variable value
1	no_limp_classes	3
2	stem_max_g0_time (h)	96
3	limp_max_g0_time (h)	96
4	apoptosis_rate (h^{-1})	0.001
5	diff_apoptosis_rate (h^{-1})	0.003
6	diff_nec_rate (h^{-1})	0.001
7	margin_percent	0.1
8	stem_g0_to_g1_percent (h^{-1})	0.01
9	limp_g0_to_g1_percent (h^{-1})	0.01
10	color_criterion	0.98
11	vcr_adm_interval (h)	168
12	vcr_no_sessions	4
13	act_adm_interval (h)	336
14	act_no_sessions	2
15	mode	THERAPY
16	drug	COMBI
17	reconstruction	1
18	necrosis_time[0](h)	20

No	Variable name	Reference value
19	necrosis_time[1](h)	20
20	apoptosis_time[0](h)	6
21	apoptosis_time[1](h)	6
22	nbc_per_gc	1000000
23	distance_factor[0]	1.0
24	distance_factor[1]	1.0
25	cell_cycle_duration (h)	23.0
26	sleep_percent[0]	0.28
27	sleep_percent[1]	0.28
28	sym_percent[0]	0.45
29	sym_percent[1]	0.45

Fig. W5 Rest of the parameter values used in order to adapt the model to the nephroblastoma data sets

5.3 A tentative correlation of histological data with the vincristine cell kill ratio estimated through simulation

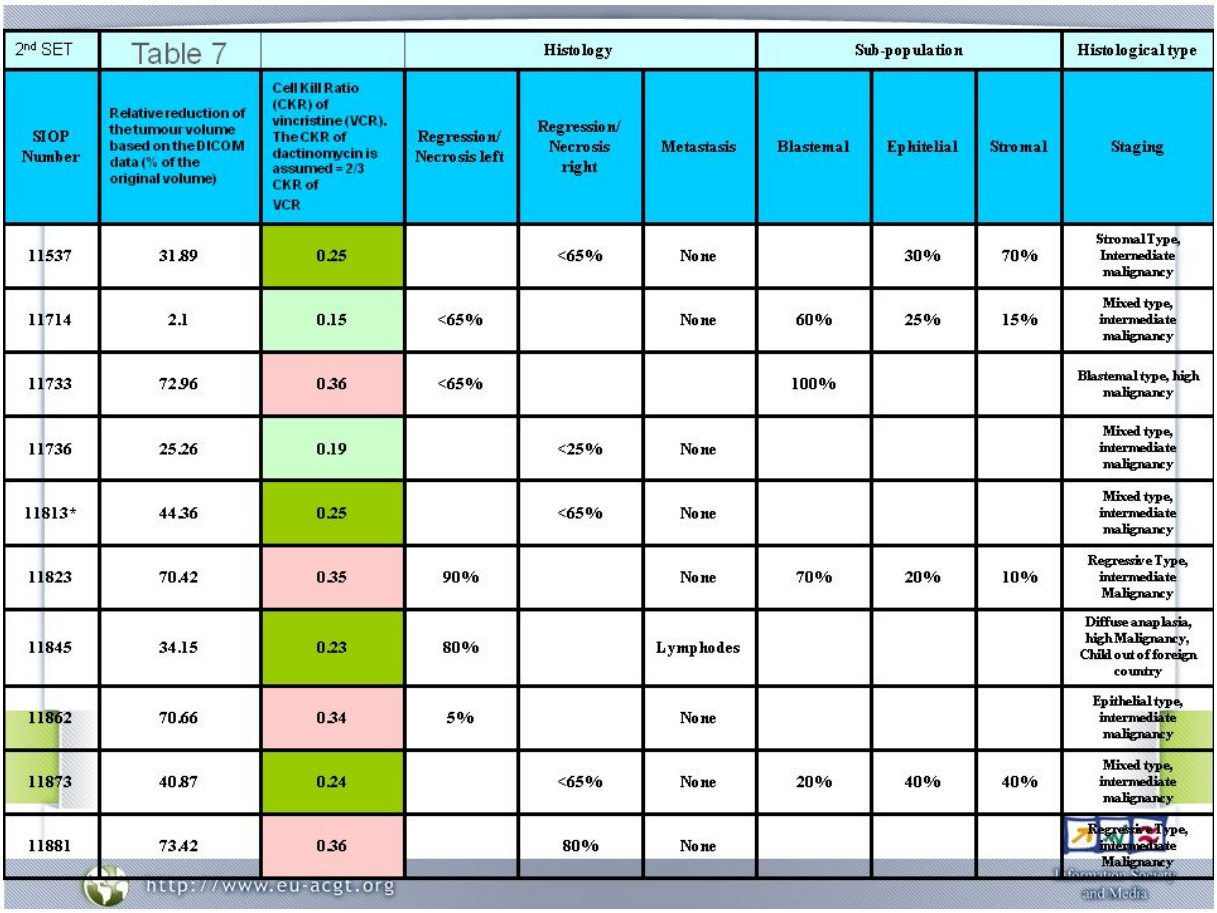
Figures W6 and W7 present efforts to correlate the histological data available with the vincristine cell kill ratio estimated through simulation.



A Tentative Correlation of Histological Data with Simulation Estimated VCR Cell Kill Ratio

Table 6 1rst SET			Histology			Sub-population			Histological type
SIOP Number	Relative reduction of the tumour volume based on the DICOM data (% of the original volume)	Cell Kill Ratio (CKR) of vincristine (VCR). The CKR of dactinomycin is assumed = 2/3 CKR of VCR	Re-gression/ Necrosis left	Re-gression/ Necrosis right	Meta-stasis	Blastemal	Epi-thelial	Stromal	Staging
11351	-8.92 (volume increases!)	0.147	30%		None	85%	10%	5%	Blastemal predominant, high risk
11570	35.37	0.24		20%	None				Mixed type, intermediate risk
11590	62.45	0.32		15%	None				Mixed type with focal Anaplasia, intermediate risk
11627	80.06	0.42		40%	None	5%	60%	35%	Mixed type, intermediate malignancy
11628	81.59	0.39		30%	None	5%	25%	70%	Stromal Type, intermediate Malignancy
11639	70.66	0.35	50%		None	0%	90%	10%	Regressive Type, intermediate Malignancy
11803	39.34	0.24	<65%		None				Stromal Type, intermediate Malignancy
11813	51.72	0.27		<65%	None				Mixed type, intermediate malignancy

Fig. W6 A tentative correlation of histological data available with the vincristine cell kill ratio estimated through simulation. Eight anonymized patients have been included in this table

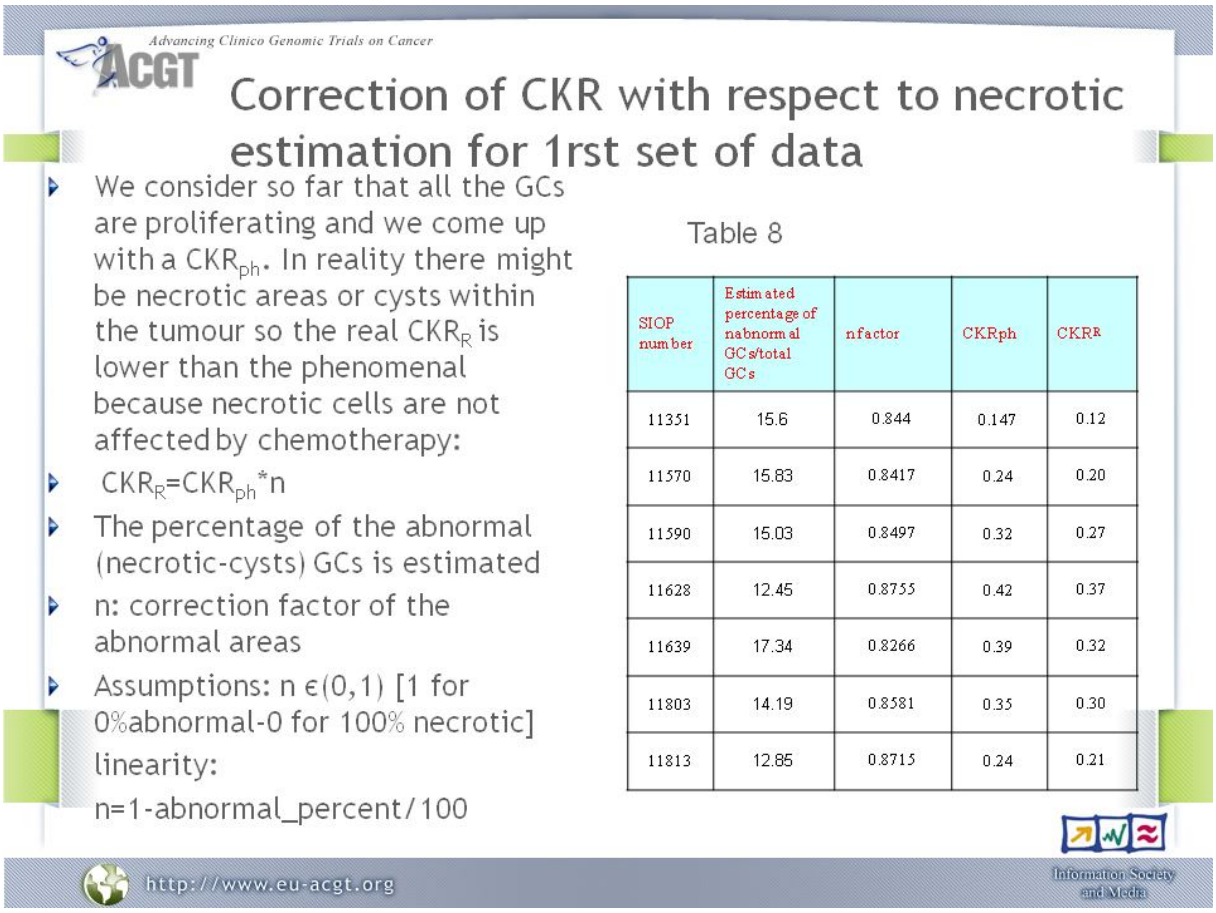


2nd SET	Table 7		Histology			Sub-population			Histological type
SIOP Number	Relative reduction of the tumour volume based on the DICOM data (% of the original volume)	Cell Kill Ratio (CKR) of vincristine (VCR). The CKR of dactinomycin is assumed = 2/3 CKR of VCR	Regression/ Necrosis left	Regression/ Necrosis right	Metastasis	Blastemal	Ephitelial	Stromal	Staging
11537	31.89	0.25		<65%	None		30%	70%	Stromal Type, Internediate malignancy
11714	2.1	0.15	<65%		None	60%	25%	15%	Mixed type, intermediate malignancy
11733	72.96	0.36	<65%			100%			Blastemal type, high malignancy
11736	25.26	0.19		<25%	None				Mixed type, intermediate malignancy
11813*	44.36	0.25		<65%	None				Mixed type, intermediate malignancy
11823	70.42	0.35	90%		None	70%	20%	10%	Regressive Type, intermediate Malignancy
11845	34.15	0.23	80%		Lymphodes				Diffuse anaplasia, high Malignancy, Child out of foreign country
11862	70.66	0.34	5%		None				Epithelial type, intermediate malignancy
11873	40.87	0.24		<65%	None	20%	40%	40%	Mixed type, intermediate malignancy
11881	73.42	0.36		80%	None				Regressive Type, intermediate Malignancy

Fig. W7 A tentative correlation of histological data available with the vincristine cell kill ratio estimated through simulation (cont.). Ten anonymized patients have been included in this table.

5.4 Correction of the estimated cell kill ratio based on the necrotic region percentage estimate

Fig. W8 depicts efforts to correlate kill ratio with respect to the estimated necrotic region percentage.

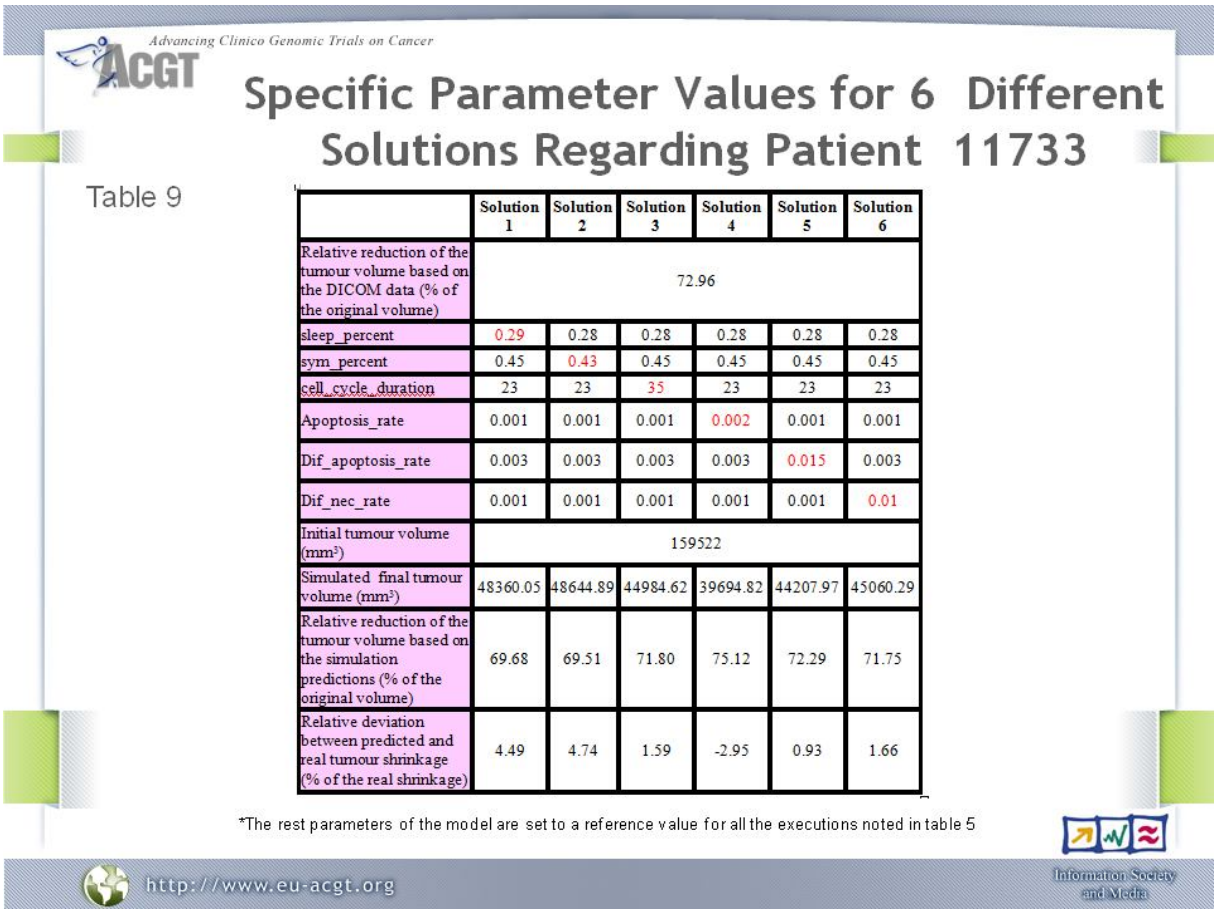


SIOP number	Estimated percentage of nabnormal GCs/total GCs	nfactor	CKRph	CKR^R
11351	15.6	0.844	0.147	0.12
11570	15.83	0.8417	0.24	0.20
11590	15.03	0.8497	0.32	0.27
11628	12.45	0.8755	0.42	0.37
11639	17.34	0.8266	0.39	0.32
11803	14.19	0.8581	0.35	0.30
11813	12.85	0.8715	0.24	0.21

Fig. W8. Correction of the cell kill ratio with respect to the estimated necrotic region percentage for the first data set.

5.5 Estimated parameter values for different clinical adaptation solutions regarding a single patient

Fig. W9 provides an exploration of the raher broad set of plausible parameter value combinations that can lead to clinical adaptation of the Oncosimulator. At a second stage by extracting further information from the multiscale medical data available as well as from literature, exlusion of several initially realistic looking parameter value combinations can take place thus leading to the narrowing of the possible solutions window and a better approximation of clinical truth.

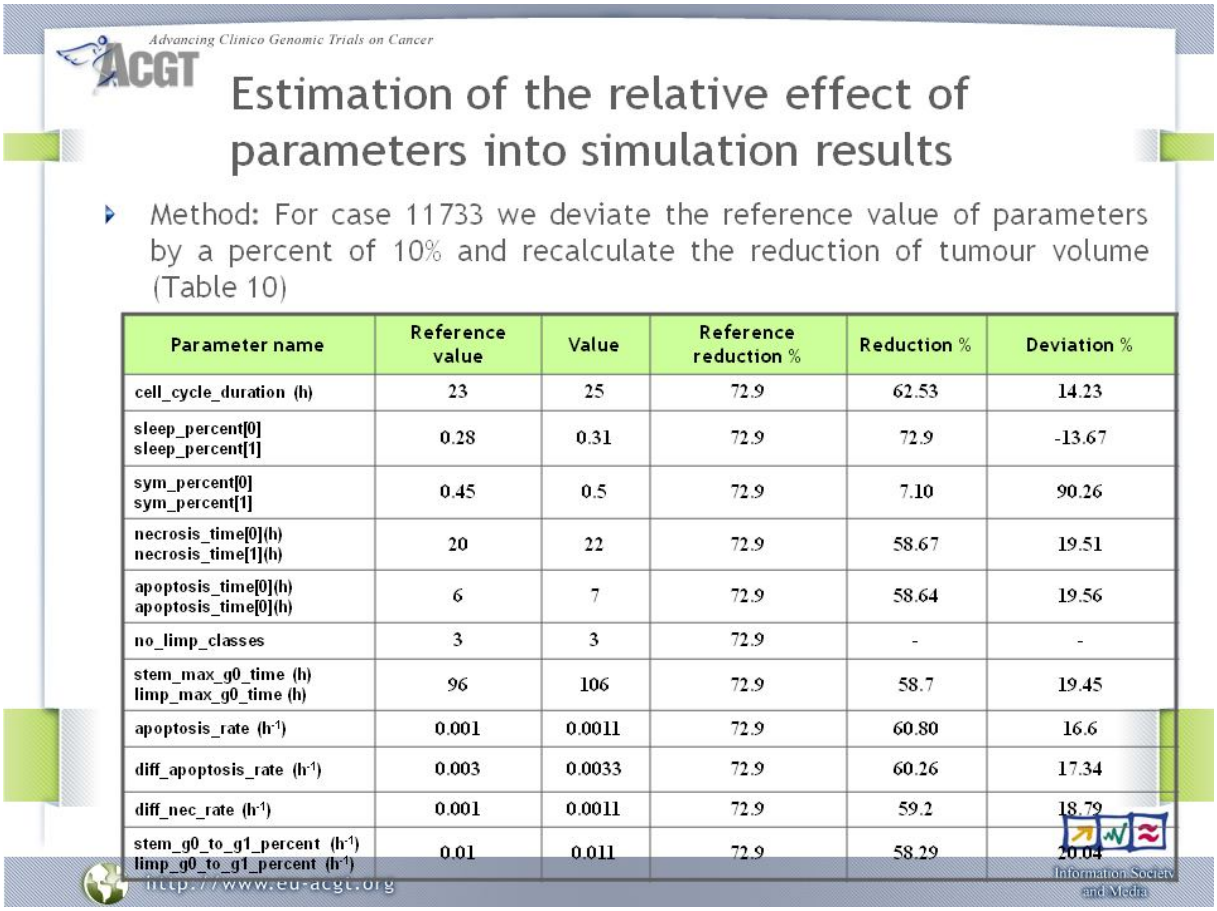


	Solution 1	Solution 2	Solution 3	Solution 4	Solution 5	Solution 6
Relative reduction of the tumour volume based on the DICOM data (% of the original volume)	72.96					
sleep_percent	0.29	0.28	0.28	0.28	0.28	0.28
sym_percent	0.45	0.43	0.45	0.45	0.45	0.45
cell_cycle_duration	23	23	35	23	23	23
Apoptosis_rate	0.001	0.001	0.001	0.002	0.001	0.001
Dif_apoptosis_rate	0.003	0.003	0.003	0.003	0.015	0.003
Dif_nec_rate	0.001	0.001	0.001	0.001	0.001	0.01
Initial tumour volume (mm^3)	159522					
Simulated final tumour volume (mm^3)	48360.05	48644.89	44984.62	39694.82	44207.97	45060.29
Relative reduction of the tumour volume based on the simulation predictions (% of the original volume)	69.68	69.51	71.80	75.12	72.29	71.75
Relative deviation between predicted and real tumour shrinkage (% of the real shrinkage)	4.49	4.74	1.59	-2.95	0.93	1.66

Fig. W9. Estimated parameter values for different clinical adaptation solutions regarding a single patient

5.6 Estimation of the relative effect of the parameters on the simulated outcome

Figure W10 provides an estimate of the relative effect of the parameters on the simulated results.



Parameter name	Reference value	Value	Reference reduction %	Reduction %	Deviation %
cell_cycle_duration (h)	23	25	72.9	62.53	14.23
sleep_percent[0] sleep_percent[1]	0.28	0.31	72.9	72.9	-13.67
sym_percent[0] sym_percent[1]	0.45	0.5	72.9	7.10	90.26
necrosis_time[0](h) necrosis_time[1](h)	20	22	72.9	58.67	19.51
apoptosis_time[0](h) apoptosis_time[0](h)	6	7	72.9	58.64	19.56
no_limp_classes	3	3	72.9	-	-
stem_max_g0_time (h) limp_max_g0_time (h)	96	106	72.9	58.7	19.45
apoptosis_rate (h^{-1})	0.001	0.0011	72.9	60.80	16.6
diff_apoptosis_rate (h^{-1})	0.003	0.0033	72.9	60.26	17.34
diff_nec_rate (h^{-1})	0.001	0.0011	72.9	59.2	18.79
stem_g0_to_g1_percent (h^{-1}) limp_g0_to_g1_percent (h^{-1})	0.01	0.011	72.9	58.29	20.04

Fig. W10. Estimation of the relative effect of the parameters on the simulated results

Figure W11 provides a sorting of the most critical model parameters.

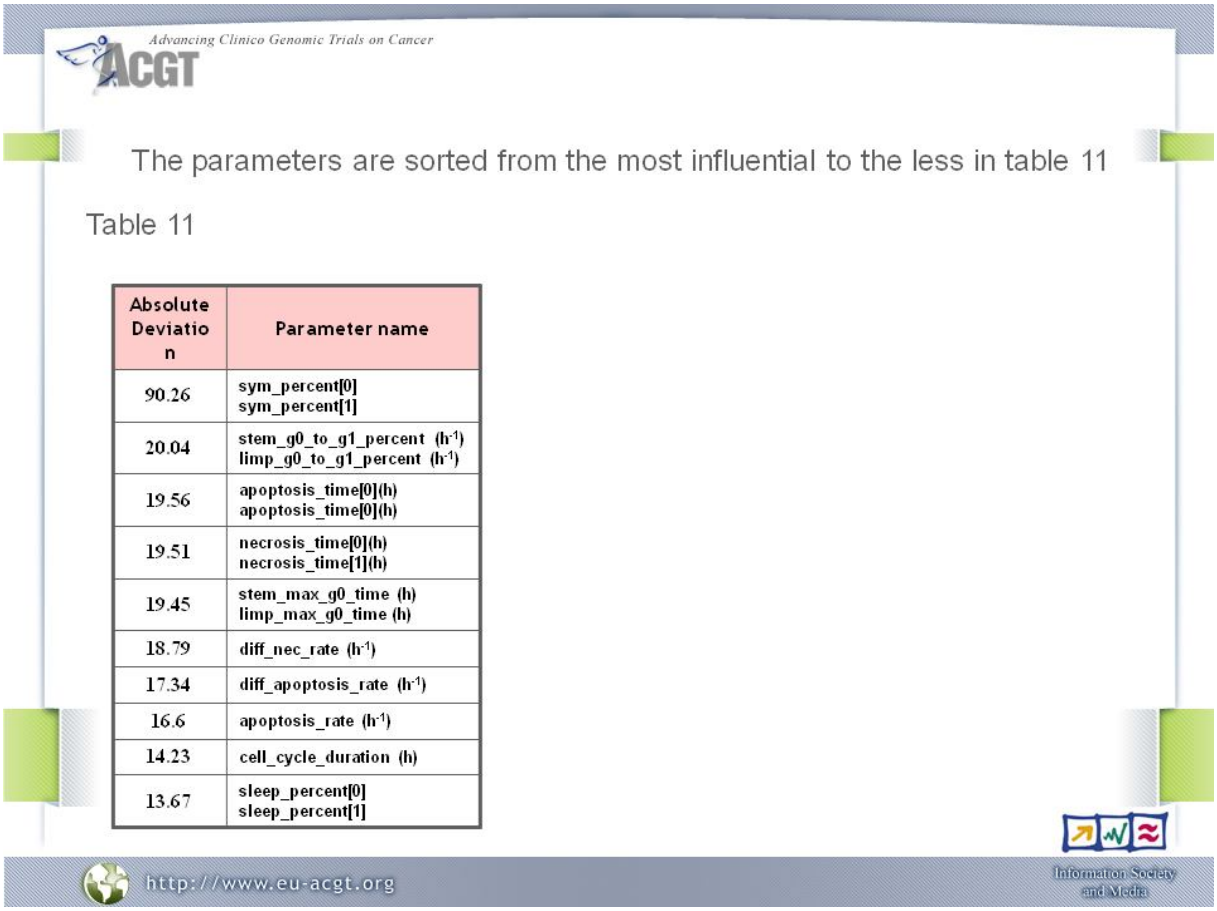


Fig. W11. Model parameter sorting from the more critical down to the less critical

6. Update on the utilization of breast cancer data for clinical adaptation and validation of the ACGT Oncosimulator

(Code letter : B)

IMPORTANT NOTE: In order to avoid serious waste of time a number of highly informative tables prepared for the review presentations of the related work have been included in this chapter. Short comments have been added on each figure/slide caption so as to render the tabular exposition easy to follow.

6.1 Breast cancer: TOP trial simulation

Twenty seven (27) real clinical cases from the TOP trial have been simulated. Only the maximum size of each tumour as measured using ultrasound has been provided. No MRI data has been available. As a firts approximation the tumour is considered a sphere of which the diameter is equal to the given size. The timing of both treatment (treatemnt schedule) and ultrasound scans has been made available. This chapter demonstrates how histopathological and moldeular data can be incorporated into the relevant Oncosimulator model.

6.2 Overview of the TOP trial clinical cases considered

Figures B1, B2 and B3 provide the detailed data of the TOP trial cases that have been considered.

Advancing Clinico Genomic Trials on Cancer

TOP trial clinical cases overview

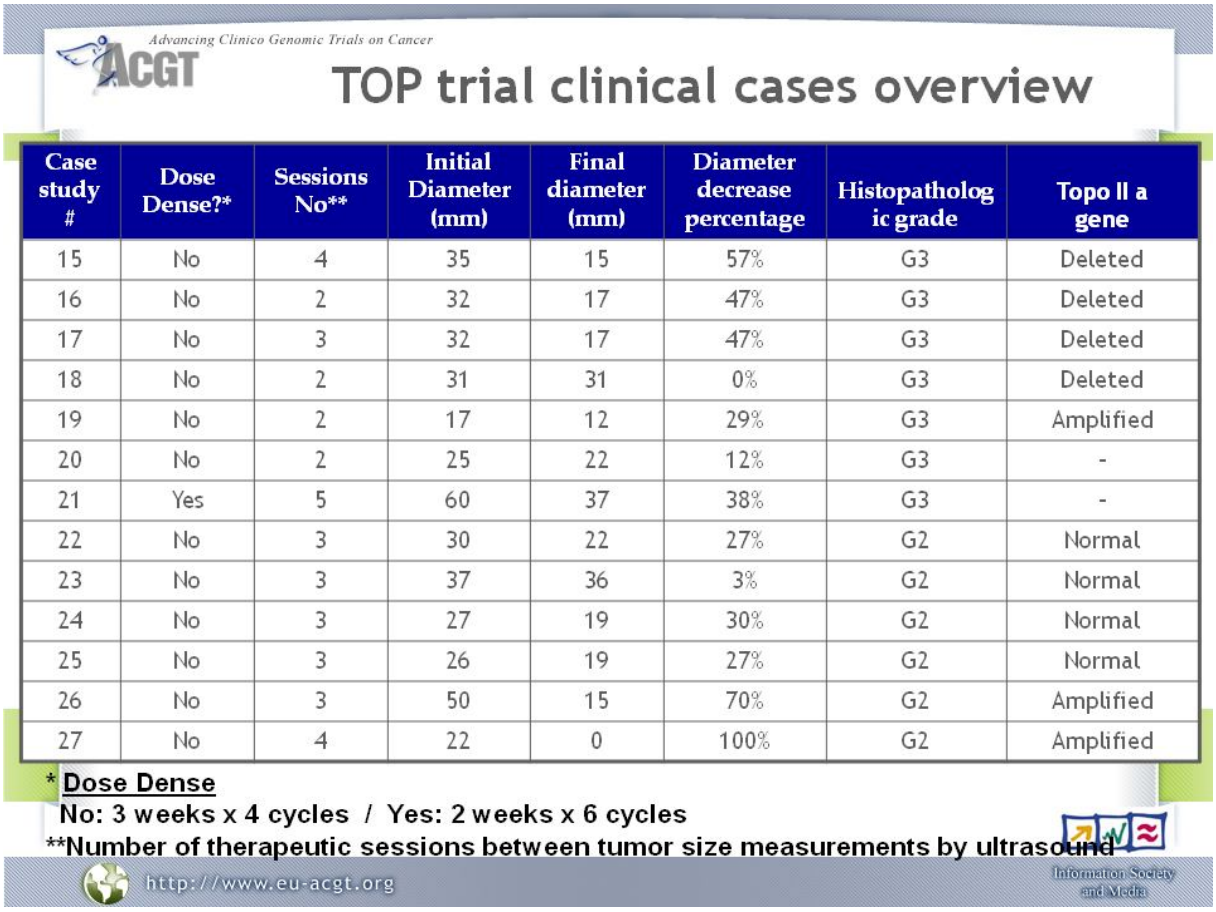
Case study #	Dose Dense?*	Sessions No**	Initial Diameter (mm)	Final diameter (mm)	Diameter decrease percentage	Histopathologic grade	Topo II a gene
1	No	3	50	30	40%	G3	Normal
2	No	3	60	70	-17%	G3	Normal
3	No	3	25	20	20%	G3	Normal
4	No	3	20	8	60%	G3	Normal
5	No	3	25	26	-4%	G3	Normal
6	Yes	3	32	0	100%	G3	Normal
7	Yes	5	20	15	25%	G3	Normal
8	Yes	5	34	25	26%	G3	Normal
9	Yes	5	20	18	10%	G3	Normal
10	Yes	5	22	18	18%	G3	Normal
11	Yes	6	13	0	100%	G3	Normal
12	Yes	2	50	33	34%	G3	Normal
13	Yes	5	27	13	52%	G3	Normal
14	Yes	5	50	12	76%	G3	Normal

*Dose Dense
No: 3 weeks x 4 cycles / Yes: 2 weeks x 6 cycles
**Number of therapeutic sessions between tumor size measurements by ultrasound

http://www.eu-acgt.org

Information Society and Media

Fig. B1. An overview of the TOP trial clinical cases considered (first part)



Case study #	Dose Dense?*	Sessions No**	Initial Diameter (mm)	Final diameter (mm)	Diameter decrease percentage	Histopathologic grade	Topo II a gene
15	No	4	35	15	57%	G3	Deleted
16	No	2	32	17	47%	G3	Deleted
17	No	3	32	17	47%	G3	Deleted
18	No	2	31	31	0%	G3	Deleted
19	No	2	17	12	29%	G3	Amplified
20	No	2	25	22	12%	G3	-
21	Yes	5	60	37	38%	G3	-
22	No	3	30	22	27%	G2	Normal
23	No	3	37	36	3%	G2	Normal
24	No	3	27	19	30%	G2	Normal
25	No	3	26	19	27%	G2	Normal
26	No	3	50	15	70%	G2	Amplified
27	No	4	22	0	100%	G2	Amplified

Fig. B2. An overview of the TOP trial clinical cases considered (second part)

The following list provides a brief molecular and histological overview of the TOP trial clinical cases considered:

21 poorly differentiated tumours have been identified out of which:

14 tumours with normal topo II a gene

4 tumours with deleted topo II a gene

1 tumour with amplified topo II a gene

2 tumours not defined in terms of the status of the topo II gene

6 moderately differentiated tumours have been identified out of which:

4 tumours with normal topo II a gene

2 tumours with amplified topo II a gene

6.3 Virtual implementation of tumour types

According to bibliography tumours with different histopathologic grade are characterized by:

Growth fraction - percentage of proliferating (stem + limp) cells.

They may may fall to one of the the following categories:

G3:Poorly differentiated:~25% (up to 75 %)

G2:Moderately differentiated: ~ 15%

Hypoxic fraction. They may fall to one of the following categories:

G3:~20%

G2: ~ 17%

Two tumour types have been implemented to reflect the variable degree of tumour differentiation that has characterized the case studies examined:

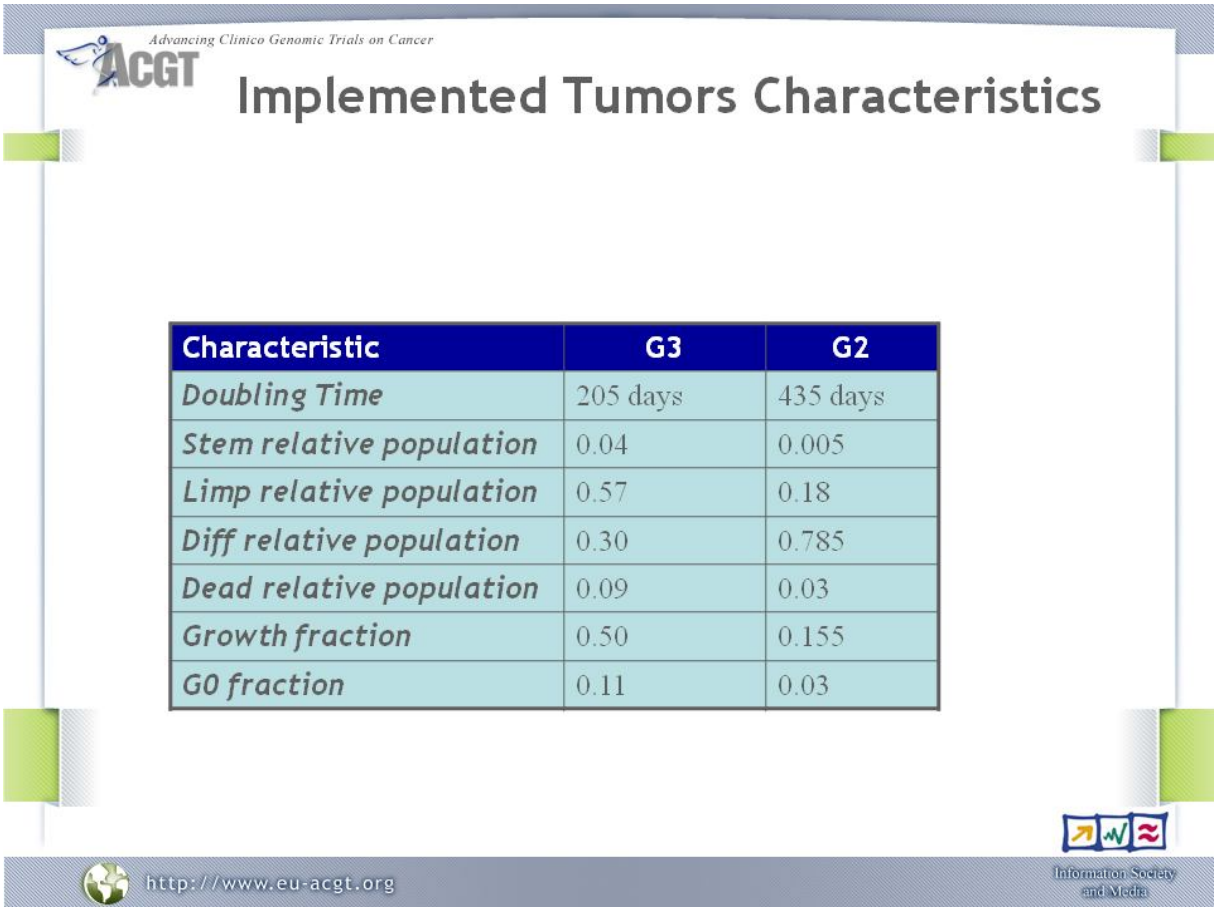
G3

and G2

The different degree of differentiation is implemented by assigning suitable values to the code input parameters in order to achieve the tumour characteristics found in bibliography.

6.4 Characteristics of the virtually implemented tumours

Fig. B3 provides the values adopted for certain critical characteristics of the tumours implemented.

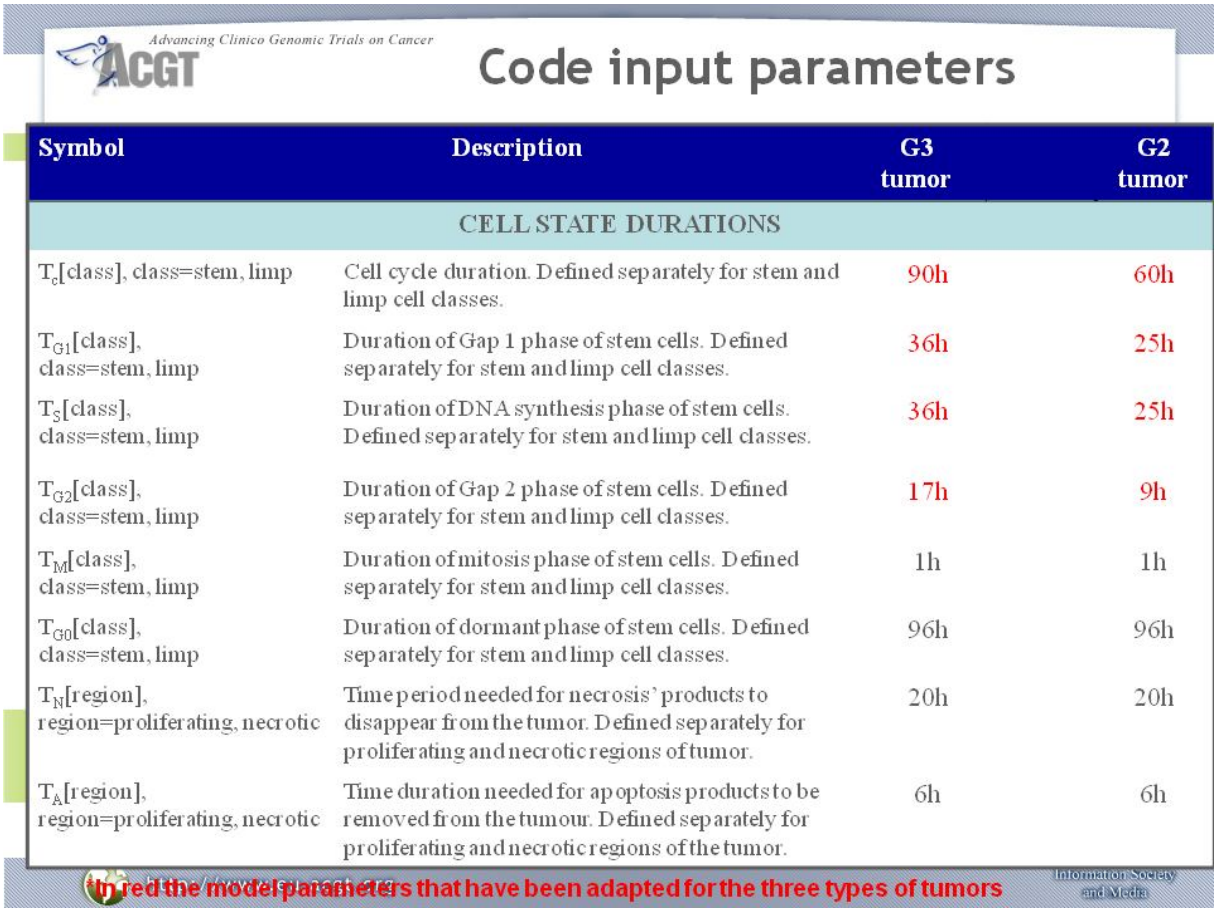


Characteristic	G3	G2
Doubling Time	205 days	435 days
Stem relative population	0.04	0.005
Limp relative population	0.57	0.18
Diff relative population	0.30	0.785
Dead relative population	0.09	0.03
Growth fraction	0.50	0.155
G0 fraction	0.11	0.03

Fig. B3. Characteristics of the tumours implemented

6.5 Simulation code input selected parameter values

Fig. B4 and B5 provide the input parameters of the code and their assumed values.



Symbol	Description	G3 tumor	G2 tumor
	CELL STATE DURATIONS		
T_c[class], class=stem, limp	Cell cycle duration. Defined separately for stem and limp cell classes.	90h	60h
T_{G1}[class], class=stem, limp	Duration of Gap 1 phase of stem cells. Defined separately for stem and limp cell classes.	36h	25h
T_S[class], class=stem, limp	Duration of DNA synthesis phase of stem cells. Defined separately for stem and limp cell classes.	36h	25h
T_{G2}[class], class=stem, limp	Duration of Gap 2 phase of stem cells. Defined separately for stem and limp cell classes.	17h	9h
T_M[class], class=stem, limp	Duration of mitosis phase of stem cells. Defined separately for stem and limp cell classes.	1h	1h
T_{G0}[class], class=stem, limp	Duration of dormant phase of stem cells. Defined separately for stem and limp cell classes.	96h	96h
T_N[region], region=proliferating, necrotic	Time period needed for necrosis' products to disappear from the tumor. Defined separately for proliferating and necrotic regions of tumor.	20h	20h
T_A[region], region=proliferating, necrotic	Time duration needed for apoptosis products to be removed from the tumour. Defined separately for proliferating and necrotic regions of the tumor.	6h	6h

Fig. B4. Input parameters of the code (first part)

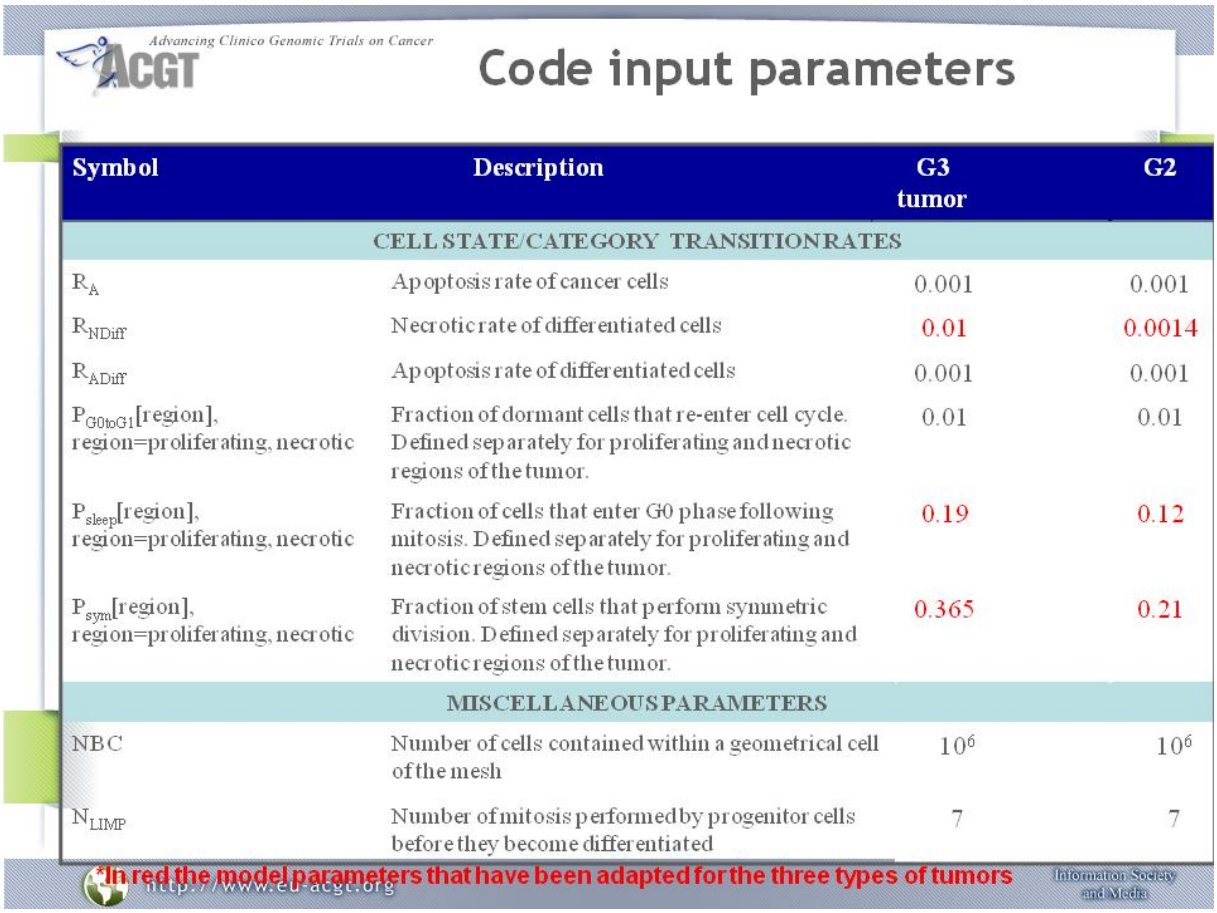


Fig. B5. Input parameters of the code (second part)

6.6 Results

Figures B6, B7, B8 and B9 depict the simulation results for poorly differentiated (G3) and moderately differentiated (G2) tumours. Pertinent comments and remarks are included within the slides.

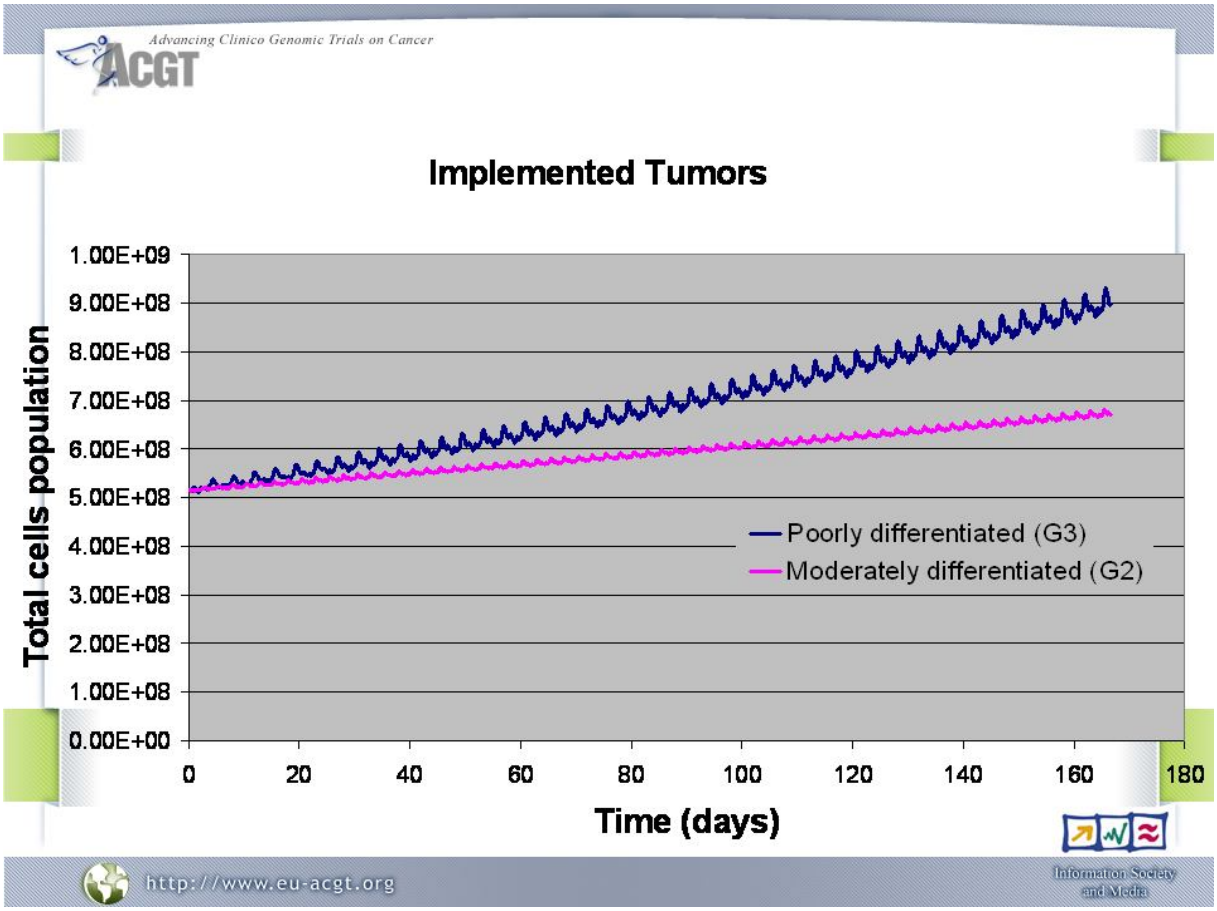


Fig. B6. Simulation results pertaining to poorly differentiated (G3) and moderately differentiated (G2) tumours. The waving of the curves is due to well known quantization errors which however do not affect the essential characteristics of the predictions. The latter are currently being considerably decreased using an improved quantization strategy.

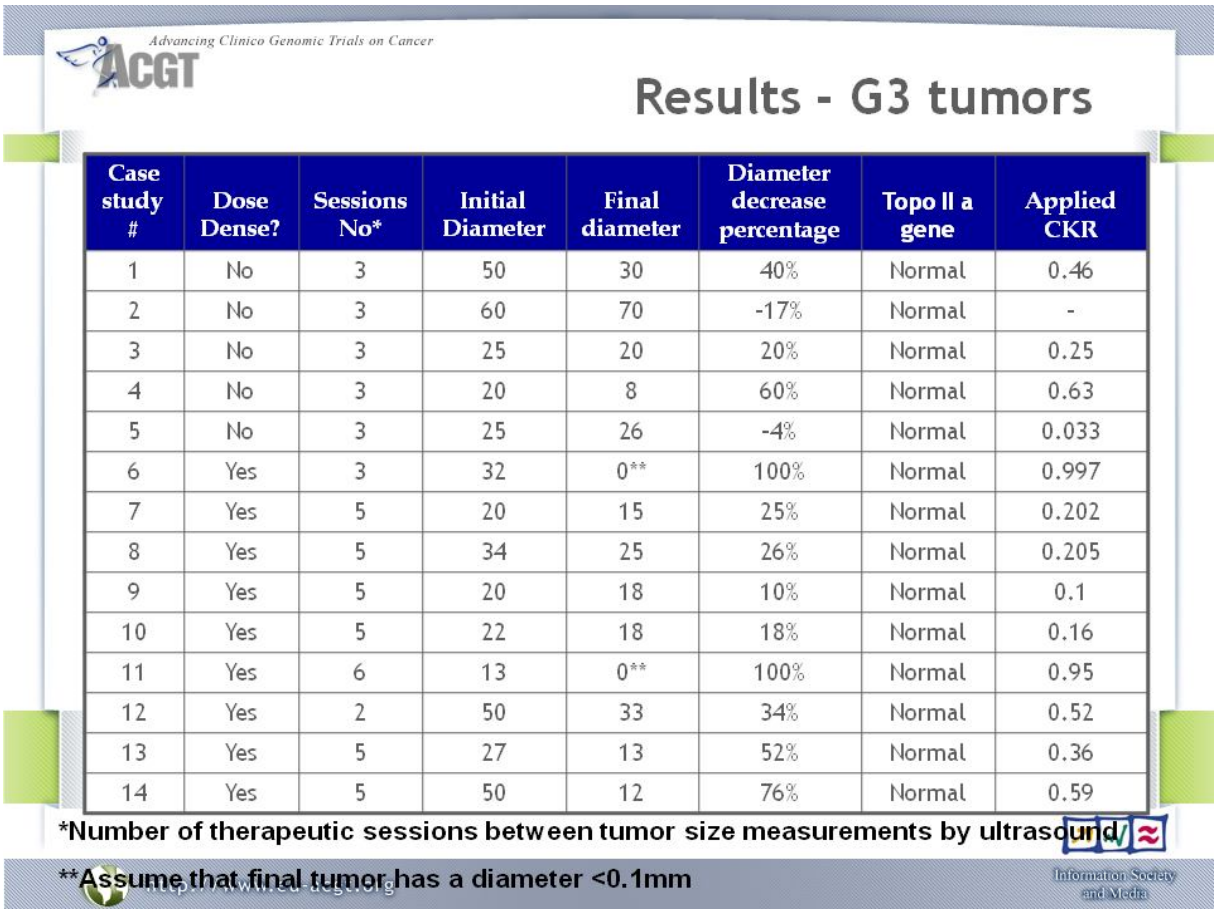


Fig. B7. Simulation results pertaining to poorly differentiated (G3) tumours (first part)

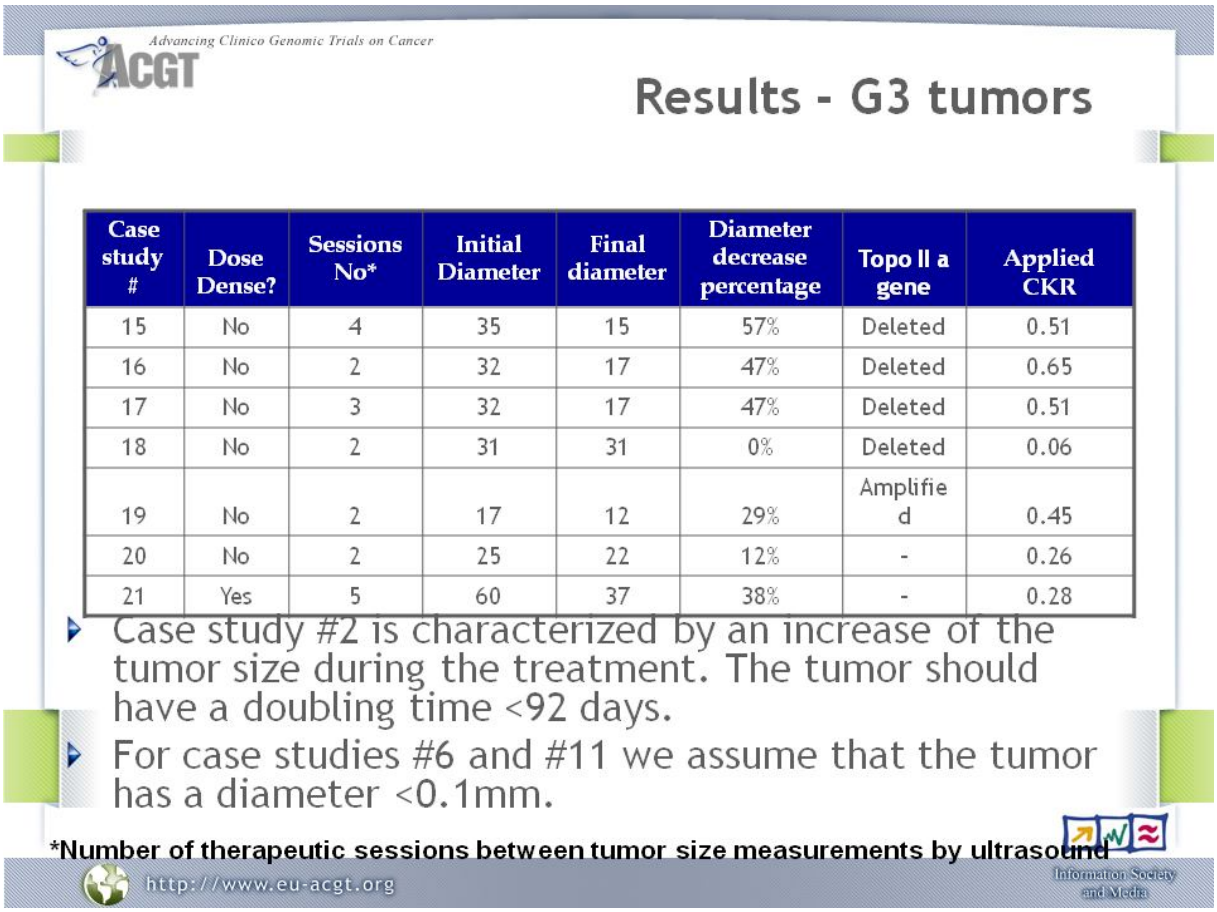


Fig. B8. Simulation results pertaining to poorly differentiated (G3) tumours (second part)

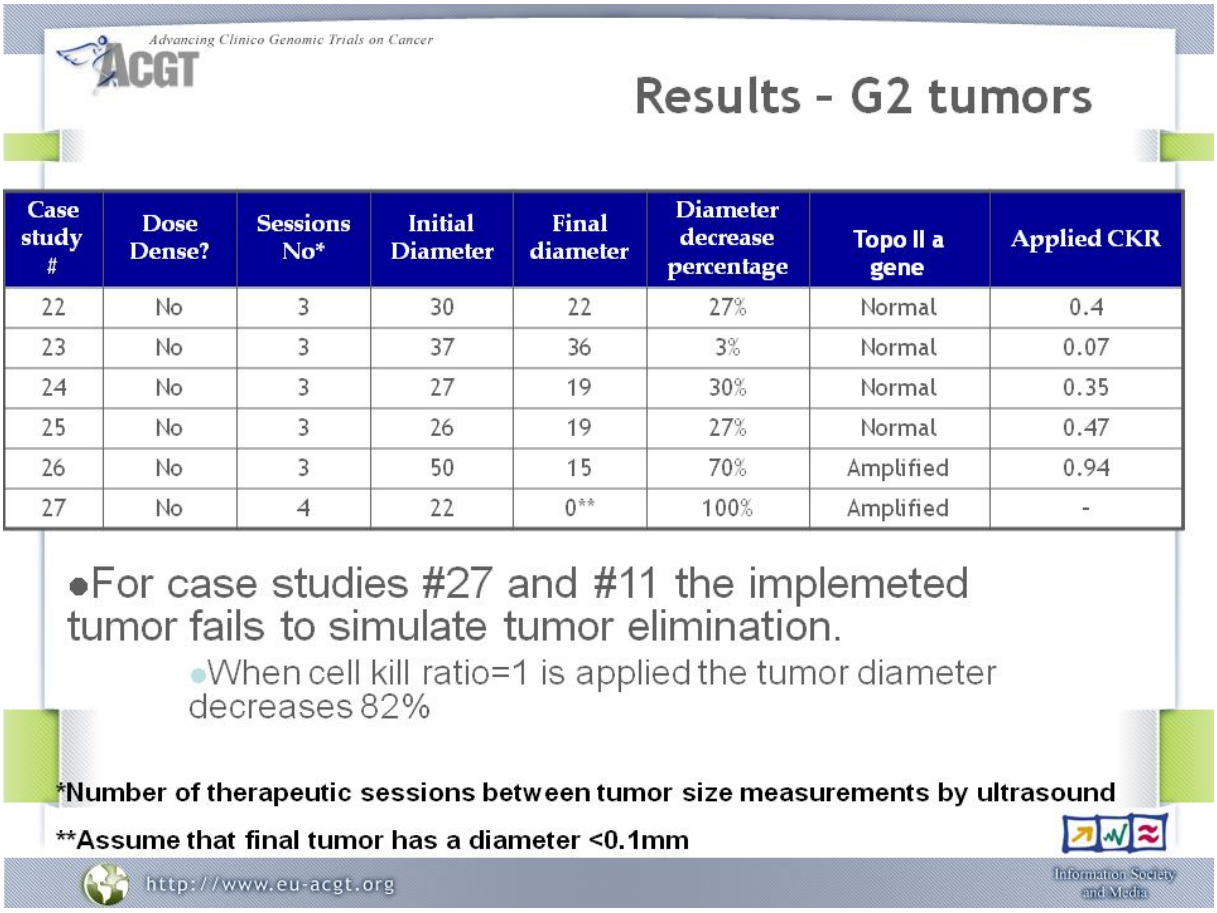


Fig. B9. Simulation results pertaining to moderately differentiated (G2) tumours

7. Image processing of the medical data *(Code letter : P)*

7.1 Introduction

In general, the term *segmentation* denotes the process of assigning sets of pixels or voxels to one or more distinct groups that are defined by the needs of the respective image processing task. The result of segmentation is a classification that labels every voxel to be part of a certain region. This is referred to as *binary segmentation* since a voxel either shares a property with its neighbours or not [P1].

Within ACGT, segmentation is needed to segment tumours from MRI data sets in order to provide ground truth delineations for both pre- and post-treatment images. In order to guarantee a high accuracy needed for clinical evaluation, segmentation is performed manually, i.e. slice by slice by experienced physicians. Simulated cancer growth is then compared with the ground truth growth calculated between pre- and post-treatment ground truth segmentations. In order to create an accurate model of the tumour, also the necrotic tissue inside the tumour needs to be considered. Such tissue may be finely spreaded inside the tumour making it hard to segment manually. Therefore, several automatic intensity based techniques are applied to estimate the percentage of necrotic tissue inside the tumour. In order to cope with the intensity bias of MRI, image normalization is performed [P2].

The simulation module of ACGT requires isotropic voxel dimensions in order to ease the computation. Since MRI slices are usually reconstructed containing highly non-isotropic voxels, interpolation of the binary segmentation volumes is performed [P3]. Furthermore, the volume may be resampled to a coarser resolution in order to speed up computation of tumour growth simulation. In order to give the simulation enough space for tumour growth, the extents of the volume are cropped and padded accordingly.

7.2 Description of the image processing module work

In ACGT, tumours had to be segmented from MRI data sets in order to simulate cancer growth and to compare the results with the real outcome of therapy. Therefore, a segmentation tool has been developed, that can be used by a physician to manually contour the tumours in every slice. Fig. P1 shows the result of such a slice-based segmentation. The contouring is done by defining a set of points placed on the borders of the tumour and by interpolating between the points using B-Spline interpolation, respectively. The results can also be viewed as a 3D rendering showing tumour and MRI volume visualization at the same time (cf. Fig. P2).

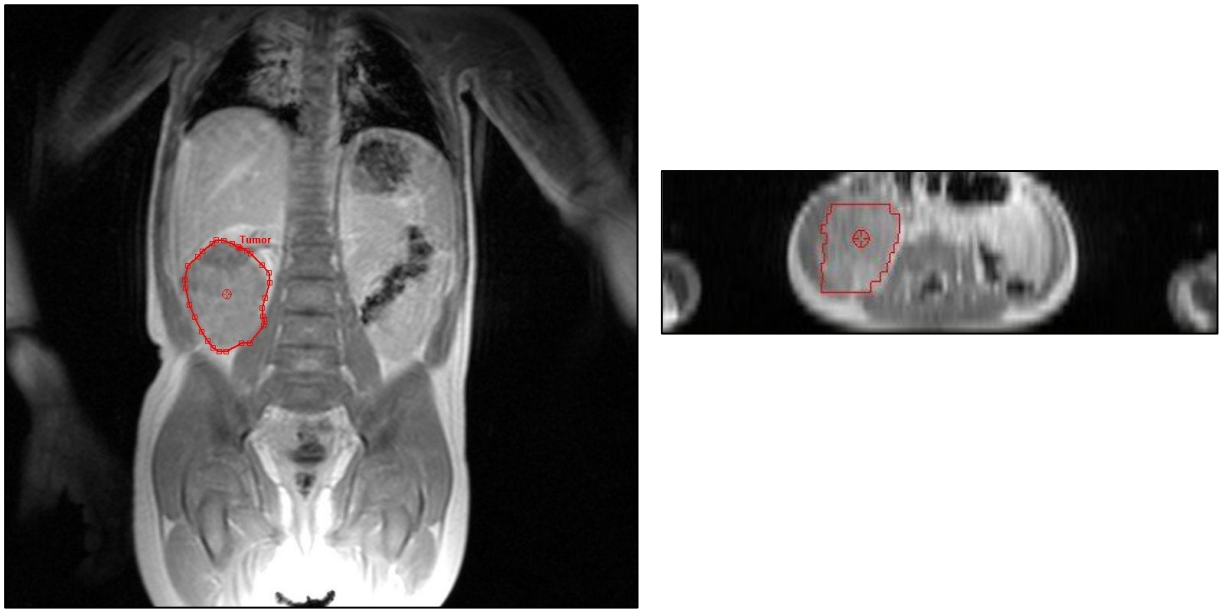


Fig P1: Contouring of tumours in MRI data sets.

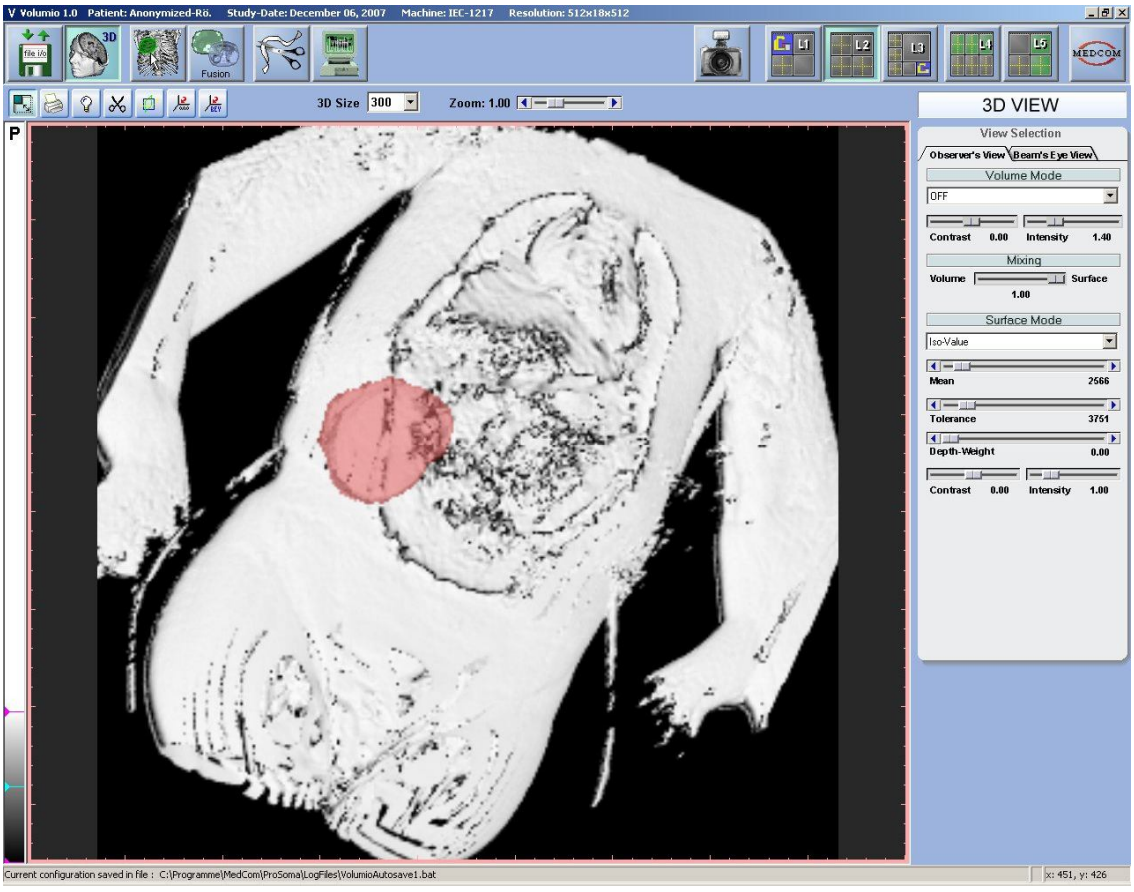


Fig.P2: 3D Visualization of tumour and thorax in a MR image.

7.3 Resampling and interpolation of binary segmentations

Usually voxels in MRI data sets are non-isotropic, i.e. their dimensions differ in x-,y- and z-direction. Therefore, binary labelled volumes are also non-isotropic. In order to allow for a proper simulation of tumour growth, the label volume has to be resampled. While resampling of the volume itself is not a difficult task, the resulting label should also be interpolated in order to have a smooth transition from slice to slice. Fig. P3 shows the principle of interpolating a coarse grained label to a voxel grid of higher resolution. Since we are not working on grey value data, there is not sufficient information available to interpolate between the slices. Therefore, we compute a distance transform for the binary label volume using level sets [P3]. The result is a grey level volume with the zero crossings denoting the zero level set. Fig. P4 gives an overview of the approach. One can think of the zero level set as an estimate for a surrounding minimal surface of the binary label data. By finding the zero-crossing in the volume, a new label volume can be defined which represents the resampling of the original data set.

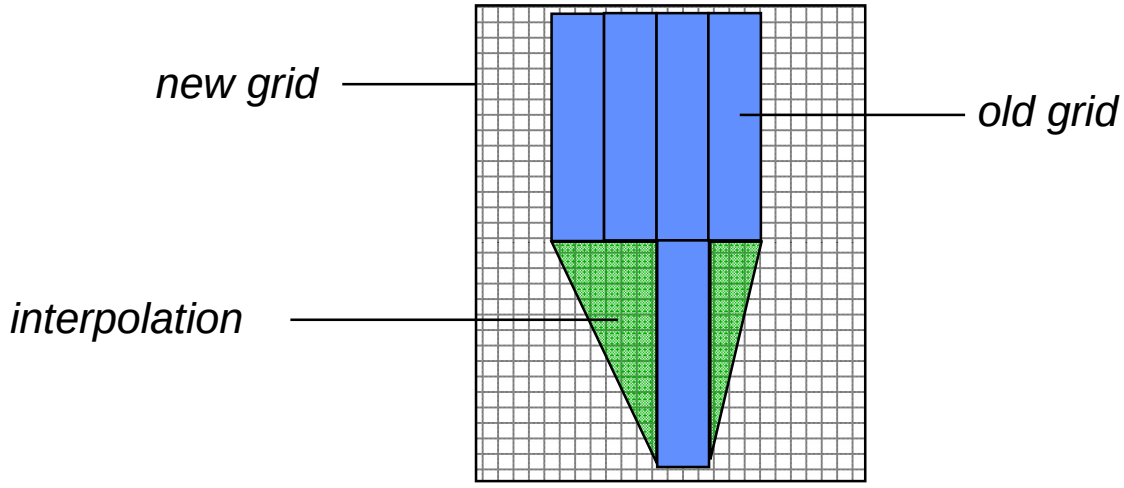


Fig. P3: Interpolation between two coarse grained slices to a voxel grid of higher resolution

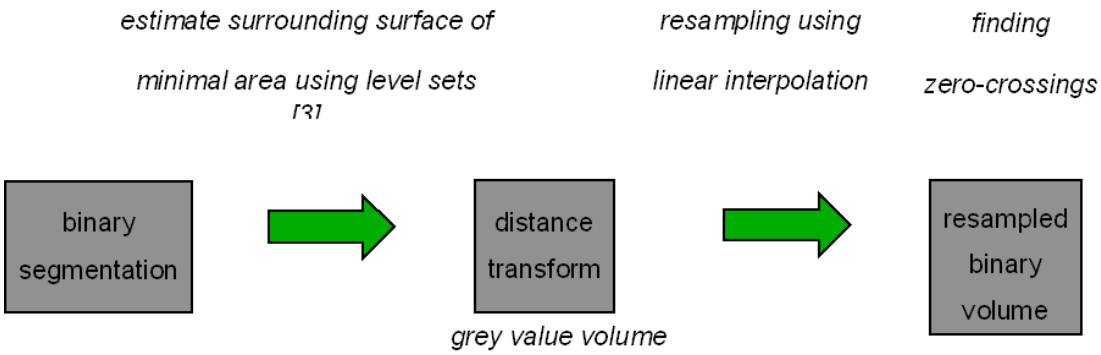


Fig. P4: Resampling of binary label volume using level sets

7.4 Latest Developments

Since deliverable 8.3, the image pre-processing pipeline has been extended by padding the tumour segmentation volume in order to provide enough space for the growth simulation. Analogously, regions far away from the tumour are cropped from the original volume in order to speed up the simulation execution (cf. Fig. P5). Furthermore, the pre-processing pipeline has been fully automated.

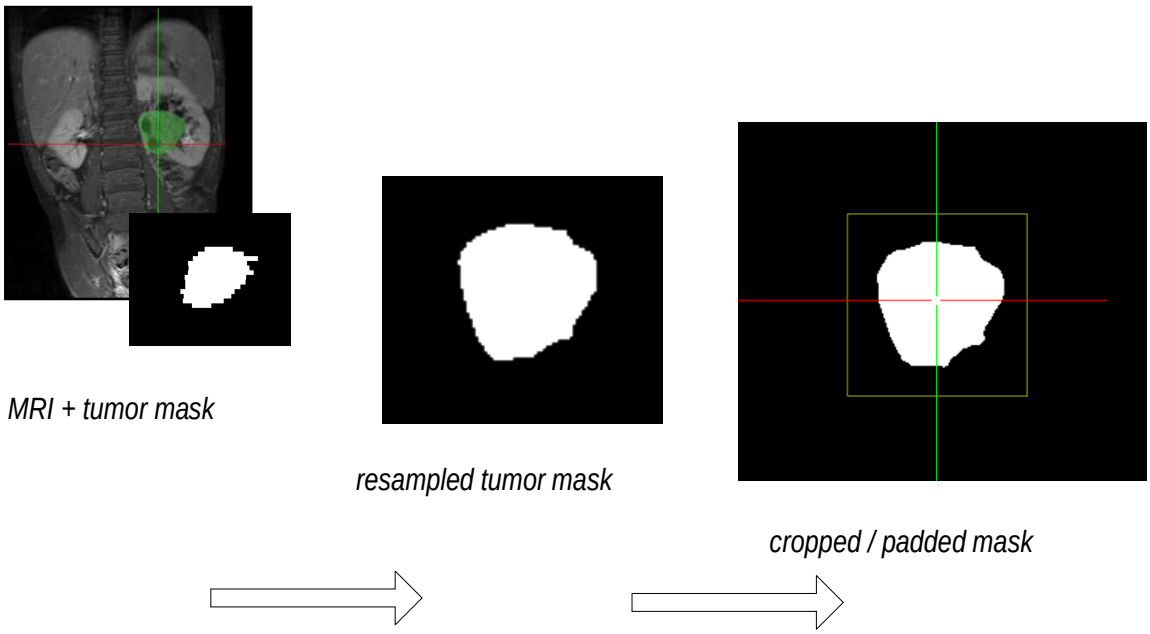


Fig.P5: Workflow of image data pre-processing

7.5 Estimation of abnormal tumour tissue

In order to create an accurate model of the tumour, also the non-living tissue inside the tumour needs to be considered. Such tissue may be finely spread inside the tumour making it hard to segment manually. Therefore, several semi-automatic and automatic intensity based techniques have been investigated to estimate the percentage of necrotic tissue inside the tumour.

In MRI images, intensity inhomogenities caused by magnetic settings, patients' position, and other factors are common. Therefore, intensity bias field estimation is performed. The method of Styner *et* al. is used [P2] since it does not need a pre-segmented image, but can be directly applied to the images (cf. Fig. P6).

Regarding the tumour growth simulation, only the amount of living tumour tissue is important. In the following we consider all tissue not containing living tumour tissue (e.g. vessels, cysts, necrotic tissue etc.) as “abnormal” tumour tissue. For estimation of this abnormal tissue, several methods have been investigated. The computation of mean and standard deviation of the whole tumour tissue independently on every slice followed by an automatic threshold turned out to provide the best results. Fig. P7 shows the results for an exemplary test scan.

Table P1 shows quantitative results for 16 scans from 8 patients (pre- and post-treatment). It can be seen that the magnitude of the estimated amount of abnormal tumour cells is in most cases comparable to the ground truth. However, the overlap of ground truth and estimated voxels is significantly less accurate. This is mainly due to partial volume effects at the border of the tumour.

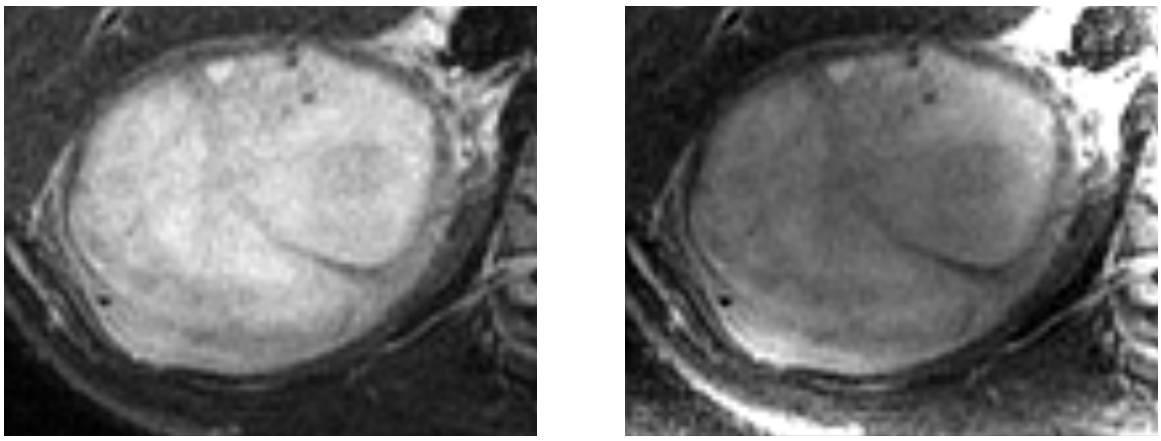


Fig. P6: uncorrected (left) and bias-corrected tumour tissue in a MRI image (right)

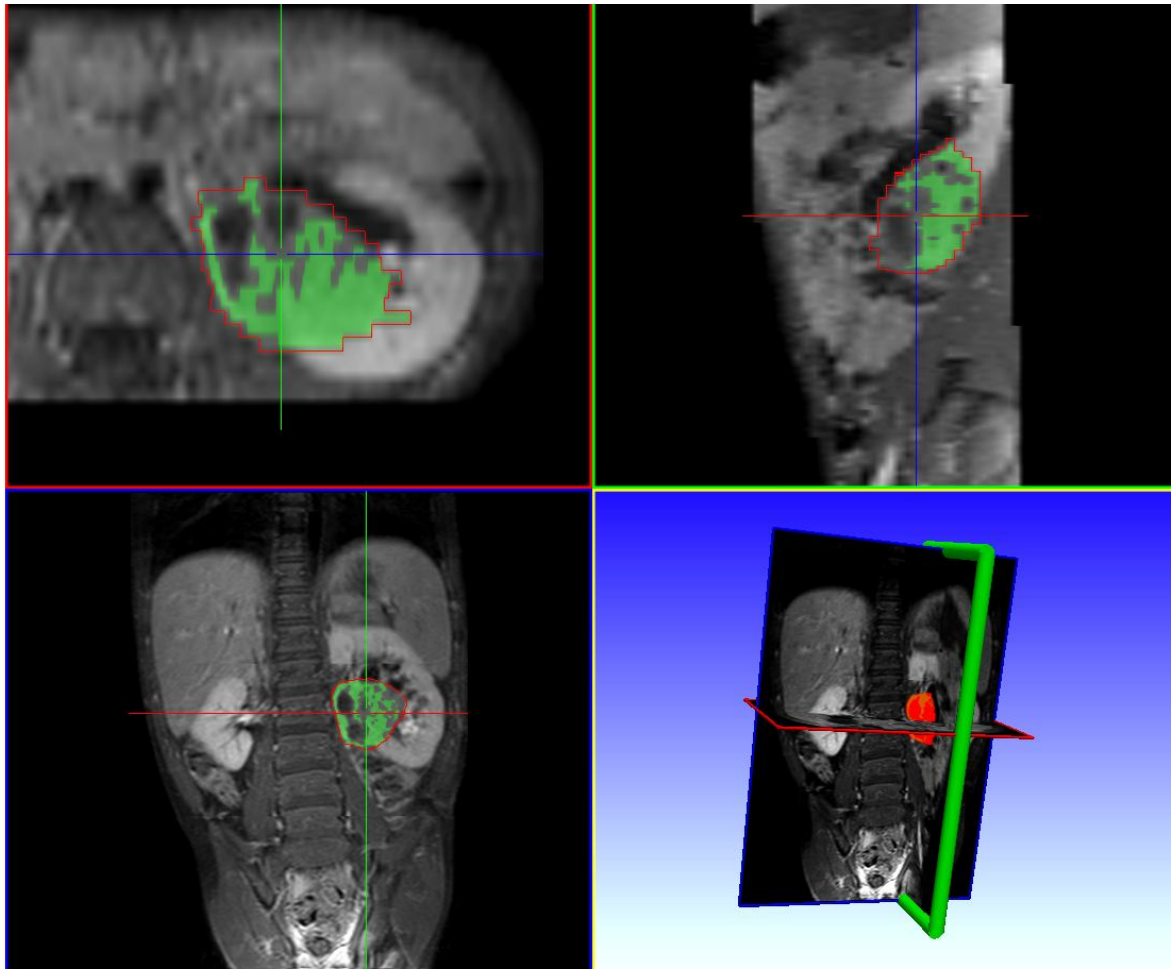


Fig. P7: Automated division of living (green) and abnormal tumour tissue in MRI using an intensity based estimation

Table P1: Estimation of abnormal tumour tissue (denoted as necrotic) for 16 test cases using an automatic estimation algorithm (X = no segmentation data available, x = no ground truth for necrotic tissue available)

	Number of necrotic ground truth cells	Volume of necrotic ground truth cells [mm³]	Number of estimated necrotic cells	Volume of estimated necrotic cells [mm³]	Number of total tumor cells	Volume of total tumor cells [mm³]	Percentage ground truth / total cells	Percentage estimated / total cells	Overlap necrotic ground truth with estimated ground truth (%)
1O	7166.00	8118.83	7569.00	8575.41	44797.00	50753.40	16.00	16.90	48.5832
1A	1505.00	3868.08	3083.00	7923.78	19757.00	50778.50	7.62	15.60	34.152
2O	70973.00	116960.00	71737.00	118219.00	418521.00	689701.00	16.96	17.14	40.7369
2A	97472.00	160629.00	102043.00	168162.00	644651.00	1062350.00	15.12	15.83	48.7763
3O	43653.00	611972.00	10798.00	151377.00	849100.00	1190360.00	51.41	12.72	22.4443
3A	6096.00	181384.00	7927.00	235865.00	52745.00	1569410.00	11.56	15.03	14.1752
4O	x	x	2100.00	27685.50	13063.00	172217.00	x	16.08	x
4A	X	X	X	X	X	X	X	X	X
5O	81.00	1635.18	187.00	3775.04	1276.00	25759.10	6.35	14.66	37.4359
5A	x	x	1160.00	16262.10	9315.00	130587.00	x	12.45	x
6O	x	x	11441.00	15013.50	65333.00	85733.40	x	17.51	x
6A	x	x	40838.00	50014.80	235561.00	288494.00	x	17.34	x
7O	14920.00	29450.20	22655.00	44718.20	134698.00	265877.00	11.08	16.82	27.9672
7A	x	x	6149.00	62252.30	43322.00	438591.00	x	14.19	x
8O	1902.00	10448.00	1222.00	6712.65	7098.00	38990.50	26.80	17.22	27.771
8A	639.00	3510.13	1842.00	10118.4	14335.00	78744.5	4.46	12.85	28.68

In order to ensure compatibility of the image pre-processing module with the tool “Doctor Eye” designed by FORTH for image annotation, a converter was developed. This module converts the exported Doctor Eye format into the internal format of the image processing module in case Doctor Eye was used for delineation.

7.6 Discussion and conclusion

Within ACGT an automated image pre-processing module has been developed that can interpolate, resample and crop/pad binary segmentations in order to provide adequate input for the cancer growth simulation module. Furthermore,

several methods for semi-automatic and automatic detection of non-living tumour tissue have been investigated and evaluated. In the tests, the developed algorithms yielded good results for estimation of the amount of non-living cells. However, the overlap of estimation and ground truth of those cells was not very accurate due to partial volume effects at the tumour borders. Furthermore, an intensity based differentiation between living and necrotic tissue using only MRI is often not possible. In order to improve the accuracy of non-living tumour tissue estimation, other imaging modalities --- particularly ultrasound (U/S) --- should be incorporated. In U/S, necrotic tissue can be better distinguished from living tissue as well as from vessels or cysts compared to MRI. A fusion of clinical MRI and U/S could lead to more precise results in necrotic tissue estimation and therefore to more accurate simulation outcome.

7.7 References

[P1] M. Erdt. Segmentation, Registration, and Fusion of Medical Images, In: Stergiopoulos, Stergios (Ed.): Advanced Signal Processing: Theory and Implementation for Sonar, Radar, and Non-Invasive Medical Diagnostic Systems. Boca Raton: CRC Press LLC, 2009, pp. 251-275.

[P2] M. Styner, G. Gerig, Christian Brechbuehler, and Gabor Szekely, Parametric estimate of intensity inhomogeneities applied to MRI, IEEE Transactions on Medical Imaging; 19(3), pp. 153-165, 2000.

[P3] R. Whitaker. Reducing Aliasing Artifacts in Iso-Surfaces of Binary Volumes. IEEE Volume Visualization and Graphics Symposium, October 2000, pp.23-32.

8. Support for the execution of the acgt oncosimulator code in the grid environment *(Code letter : G)*

8.1 Grid environment of ACGT

The most important components of the grid infrastructure in a context of Oncosimulator execution are GRMS - responsible for resource management, and DMS which is a grid storage system

8.1.1. GMRS

The component responsible for resource management within Gridge Toolkit is GRMS (Gridge Resource Management System). It is an open source meta-scheduling system, which allows developers to build and deploy resource management systems for large scale distributed computing infrastructures. GRMS is based on dynamic resource selection, mapping and an advanced scheduling methodology, combined with feedback control architecture, and deals with dynamic grid environment and resource management challenges, e.g., load-balancing among clusters, remote job control or file staging support. Therefore, the main goal of the GRMS is to manage the whole process of remote job submission to various batch queuing systems, clusters or resources. It has been designed as an independent core component for resource management processes which can take advantage of various low-level core services and existing technologies. Finally, GRMS can be considered as a robust system which provides abstraction of the complex grid infrastructure as well as a toolbox which helps to form and adapts to distributing computing environments.

GRMS is a central point for resource and job management activities and tightly cooperates with other services responsible for authorization, monitoring, and data management to fulfill the requirements of the applications. The main features of GRMS are job submission, job control (suspending, resuming, canceling), the ability to chose "the best" resource for the job execution using multi-criteria matching algorithm, support for job checkpointing and migration, support for file staging, storing information about the job execution, user notifications support, workflow jobs support etc.

GRMS has been designed as an independent set of components for the resource management processes which can take advantage of various low-level core services as, e.g., GRAM, GridFTP, and the Gridge Monitoring System, as well as various grid middleware services, e.g., the Gridge Authorization Service and the Gridge Data Management Service. All these services working together provide a consistent, adaptive and robust grid middleware layer which fits dynamically to many different distributing computing infrastructures. The GRMS implementation requires the Globus software [3] to be installed on the grid resources, and uses the following core Globus services deployed on the resources: GRAM, GridFTP, and MDS (optional). GRMS supports the Grid Security Infrastructure by providing GSI-enabled web service interfaces for all clients, e.g., portals or applications, and thus can be integrated with any other compliant grid middleware.

One of the main assumptions for GRMS is to perform remote job control and management in the way that satisfies users (job owners) and their applications requirements. All user requirements are expressed within an XML-based resource

specification document and sent to GRMS as SOAP requests over GSI transport layer connections. Simultaneously, resource administrators (resource owners) have full control over owned resources on which the jobs and operations will be performed by an appropriate GRMS setup and installation. GRMS together with the core services reduces operational and integration costs for administrators by enabling grid deployment across heterogeneous (and maybe previously incompatible) cluster and resources. Technically speaking, GRMS is a persistent service within a Tomcat/Axis container. It is written completely in Java so it can be deployed on various platforms.

8.1.2. Data Management

Data storage, management and access in the Gridge environment are supported by the Gridge Data Management Suite (DMS). This suite, composed of several specialized components, allows building a distributed system of services capable of delivering mechanisms for seamless management of large amounts of data. It is based on the pattern of autonomic agents using the accessible network infrastructure for mutual communication. From the external applications point of view DMS is a virtual file system keeping the data organized in a tree-like structure. The main units of this structure are meta-directories, which allow creating a hierarchy over other objects and metafiles. Metafiles represent a logical view of data regardless of their physical storage location.

Data Management System consists of three logical layers: the Data Broker, which serves as the access interface to the DMS system and implement the brokering of storage resources, the Metadata Repository that keeps information about the data managed by the system, and the Data Container, which is responsible for the physical storage of data. In addition, DMS contains modules which extend its functionality to fulfill the enterprise requirements. These include the fully functional web based administrator interface and a Proxy to external scientific databases. The Proxy provides a SOAP interface to the external databases, such as for example those provided by SRS (Sequence Retrieval System).

The Data Broker is designed as an access point to the data resources and data management services. A simple API of the Data Broker allows to easily access the functionality of the services and the stored data. The Data Broker acts as a mediator in the flow of all requests coming from external services, analyzes them and eventually passes to the relevant module. The DMS architecture assumes that multiple instances of the Data Broker can be deployed in the same environment, thus increasing the efficiency of data access from various points in the global Grid environment structure.

The Metadata Repository is the central element of the Gridge distributed data management solution. It is responsible for all metadata operations as well as their storage and maintenance. It manages metadata connected with the data files, their physical locations and transfer protocols that could be used to obtain them, with the access rights to the stored data and with the metadescriptions of the file contents. Currently each DMS installation must contain a single instance of the Metadata Repository, which acts as a central repository of the critical information about the metacatalogue structure, user data and security policy for the whole DMS installation.

The Data Container is a service specialized towards the management of physical data locations on the storage resources. The Data Container API is designed in a way to allow easy construction and participation in the distributed data

management environment of storage containers for different storage environments. The Data Containers currently available in the DMS suite include a generic file system Data Container, a relational database Data Container and a tape archiver Data Container. The data stored on the various storage resources can be accessed with one of the many available protocols including such as GASS, FTP and GridFTP.

The Proxy modules are services that join the functionality of the Metadata Repository allowing to list the available databanks, list their content, read the attached metadata attributes and to build and execute queries, and of the Data Container to provide the data using the selected data transfer protocol. Such Proxy container are highly customized towards the specific platform they are working with to allow building complex queries and executing operations on the found entries.

8.1.3 Security Services

The most important element of security infrastructure in Gridge is authorization service called GAS. The Gridge Authorization Service (GAS) is an authorization system which can be the standard decision point for all components of a system. Security policies for all system components are stored in GAS. Using these policies GAS can return an authorization decision upon the client request. GAS has been designed in such a way that it is easy to perform integration with external components and it is easy to manage security policies for complex systems. The possibility to integrate with the Globus Toolkit and many operating system components makes GAS an attractive solution for grid applications.

Generally, an authorization service can be used for returning an authorization decision upon the user request. The request has to be described by three attributes: user, object and operation. The requester simply asks if the specific user can perform the operation on the specific object. Obviously, the query to an authorization service can be more complex and the answer given by such service can be complicated, too. One of the services which can work in such scenario is the Gridge Authorization Service (GAS). GAS has been designed in a form which enables many possible applications. GAS can communicate in many ways with other components. By using the modular structure of GAS it is easy to write a completely new communication module. The GAS complex data structure can be used to model many abstract and real world objects and security policies for such objects. For example, GAS has been used for managing security policies: for many Virtual Organizations, for services (like Gridge Resource Management Service, Mobile Services and other) and for abstract objects like communicator conferences or computational centers. These and many other features give a possibility to integrate GAS with many existing solutions. Such integration can be very important, because it raises the security level of the existing solutions and makes it possible to use the newest security technologies.

The main goal of GAS is to provide a functionality that would be able to fulfill most authorization requirements of grid computing environments. GAS is designed as a trusted single logical point for defining security policy for complex grid infrastructures. As flexibility is the key requirement, it is to be able to implement various security scenarios, based on push or pull models, simultaneously. Secondly, GAS is considered as independent of specific technologies used at lower layers, and it should be fully usable in environments based on grid toolkits as well as other toolkits. The high level of flexibility is achieved mainly through the modular design of GAS and usage of a complex data structure which can model many scenarios and objects from the real world. It means that GAS can use many

different ways for communication with external components and systems, use many security data models and hold security policy on different types of storage systems. These features make GAS attractive for many applications and solutions (not only for those related with grids). GAS has to be the trusted component of each system in which it is used and it brings about that the implementation of GAS was written in ANSI C. This choice makes GAS a very fast and stable component which uses not much CPU power and little amount of memory. The main problem of many authorization systems is their management. It is not easy to work with a complex system in a user-friendly way. Based on many experiences and the end user comments together with GAS, the GAS administration portlet (web application) is provided, which makes management as easy as possible. Flexibility of this solution gives users a full possibility of presenting only these security policies which are important for them.

8.2 Grid execution of the application code

The first step of the scenario is devoted to preparation input data for the computation.

Clinician needs to prepare MRI sets of slices of a Wilms tumour, concerning the tumour before chemotherapeutic treatment (with vincristine and dactinomycin).

Using specialized tools he/she also needs to delineate tumour contours on provided slices.

The user uploads all required files from his/her machine to the Gridge Data Management System using web portal client. The files are then managed by DMS and can be accessed by the user or other services acting on behalf of the user.

To make the management of the submission process more flexible, also application code is stored in DMS. The main reason for it is a dynamic development process of Oncosimulator a new versions of the code appearing frequently so it would be impossible to manage a local deployment of it, on each grid node.

GRMS provides very useful workflow management mechanism that allows to design a sequence of actions that have to be done before or after simulation run. One of the actions like that, is downloading from DMS, and compiling the application code on a grid machine that was selected by GRMS to run the program. The described mechanism introduces some overheads - additional task managed by GRMS, compilation time - but on the other hand however provides a lot of flexibility and allows for fast testing of an updated code.

A full specification of the job managed by resource management system is described later.

The role of DMS is to store all the data that should be available for a grid execution. A final structure of the files stored there is presented in Fig. G1 .

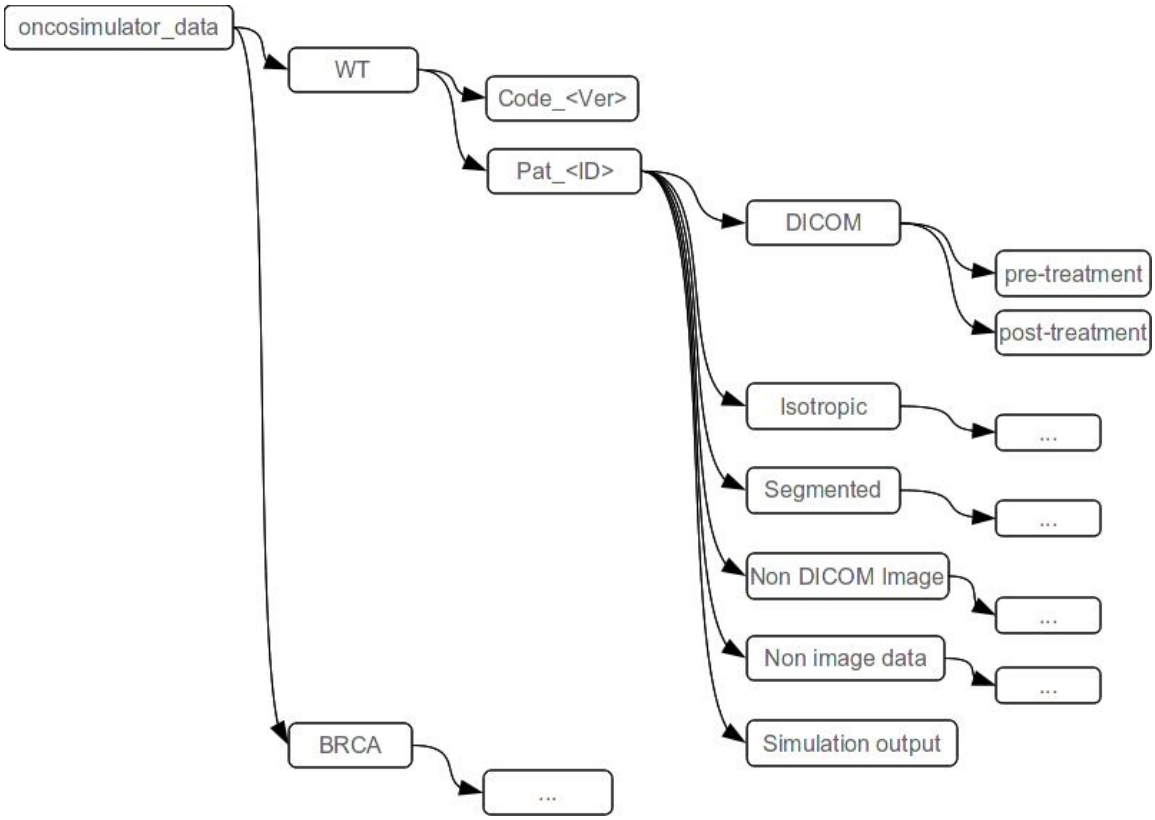


Fig. G1 Final structure of the files stored in DMS

To run the application in the grid environment of ACGT, the execution must be described in a proper way using XML document, and then submit to the resource management system of the grid - GRMS.

Job definition describes the job as a set of tasks with dependencies among them. Each task is defined with resource requirements and description of the executable: environment variables, parameters of the execution, input and output files (or directories), executable file, etc.

Resource requirements sections allows to narrow the search for available machine in the grid according to its parameters: total memory available, cpu speed, disk space, etc. For the Oncosimulator scenario in some cases it was very useful to use memory requirements specification capability because there were serious memory allocation problems for particular simulation runs.

It is not required for the end user to learn the language of GRMS - the client application called OncoRecipeSheet (ORS) is responsible for creating proper XML document and sending it to GRMS. ORS is generating the description based on forms that are presented to the user,that shows parameters of the application understandable to the user.

An example of the file generated by ORS is presented below:

```
<grmsJob appid="Oncosimulator">
  <task persistent="true" taskid="unpack_sim">
    <executable count="1" type="single">
      <execfile name="tar">
        <url>file:////bin/tar</url>
      </execfile>
      <arguments>
        <value>-xzf</value>
        <value>oncosim.tar.gz</value>
        <file name="oncosim.tar.gz" type="in">
          <logicalId>62839</logicalId>
        </file>
      </arguments>
    </executable>
  </task>
  <task taskid="failed_unpack">
    <executable count="1" type="single">
      <execfile name="echo1">
        <url>file:////bin/echo</url>
      </execfile>
      <arguments>
        <value>failed </value>
        <value>unpack</value>
      </arguments>
      <stdout>
        <logicalId>63334</logicalId>
      </stdout>
    </executable>
    <workflow>
      <parent triggerState="FAILED">unpack_sim</parent>
    </workflow>
  </task>
  <task extension="unpack_sim" persistent="true" taskid="make_sim">
    <executable count="1" type="single">
      <execfile name="make">
        <url>file:////usr/bin/make</url>
      </execfile>
      <arguments>
        <value>N_LIMP_CLASSES_MAX=8</value>
        <value>-f</value>
        <value>Makefile</value>
        <file name="wilmstumour" type="out">
          <reference>ONCOSIMULATOR_EXEC</reference>
        </file>
      </arguments>
    </executable>
    <workflow>
```

```
    <parent triggerState="FINISHED">unpack_sim</parent>
  </workflow>
</task>
<task taskid="failed_make">
  <executable count="1" type="single">
    <execfile name="echo2">
      <url>file:////bin/echo</url>
    </execfile>
    <arguments>
      <value>failed </value>
      <value>make</value>
    </arguments>
    <stdout>
      <logicalId>63334</logicalId>
    </stdout>
  </executable>
  <workflow>
    <parent triggerState="FAILED">make_sim</parent>
  </workflow>
</task>
<task persistent="true" taskid="simulation">
  <executable count="1" type="single">
    <execfile name="oncosimulator">
      <reference>ONCOSIMULATOR_EXEC</reference>
    </execfile>
    <arguments>
      <value>23</value>
      <value>2</value>
      <value>1000000</value>
      <value>20</value>
      <value>22</value>
      <value>32</value>
      <value>0.3</value>
      <value>0.3&amp;gt;</value>
      <value>0.2</value>
      <value>0.2</value>
      <value>0.28</value>
      <value>0.28</value>
      <value>0.45</value>
      <value>0.45</value>
      <value>20</value>
      <value>20</value>
      <value>6</value>
      <value>6</value>
      <value>1.0</value>
      <value>1.0</value>
      <value>8</value>
```

```
      <value>96</value>
      <value>96</value>
      <value>0.001</value>
      <value>0.003</value>
      <value>0.001</value>
      <value>0.1</value>
      <value>0.01</value>
      <value>0.01</value>
      <value>0.98</value>
      <value>168</value>
      <value>120</value>
      <value>4</value>
      <value>336</value>
      <value>120</value>
      <value>2</value>
      <value>192</value>
      <value>2</value>
      <value>2</value>
      <value>1</value>
      <value>input.raw</value>
      <value>outdata</value>
      <file name="input.raw" type="in">
        <logicalId>61417</logicalId>
      </file>
      <file name="./outdata/oncosim-parameters.log" type="out">
        <logicalId>63331</logicalId>
      </file>
      <file name="./outdata/tum_evol_file.dat" type="out">
        <logicalId>63332</logicalId>
      </file>
    </arguments>
    <stdout>
      <url>${TASK_DIR}/outdata/simulator.stdout</url>
    </stdout>
    <stderr>
      <url>${TASK_DIR}/outdata/simulator.stderr</url>
    </stderr>
  </executable>
  <workflow>
    <parent runSameHost="true" triggerState="FINISHED">make_sim</parent>
  </workflow>
</task>
<task extension="simulation" persistent="true" taskid="packed_results">
  <executable count="1" type="single">
    <execfile name="tar">
      <url>file:////bin/tar</url>
    </execfile>
```

```
    <arguments>
      <value>-czvf</value>
      <value>results.tgz</value>
      <value>outdata</value>
      <file name="results.tgz" type="out">
        <logicalId>63333</logicalId>
      </file>
    </arguments>
  </executable>
  <workflow>
    <or>
      <parent triggerState="FINISHED">simulation</parent>
      <parent triggerState="FAILED">simulation</parent>
    </or>
  </workflow>
 </task>
 <task extension="packed_results" persistent="false" taskid="completion_status">
  <executable count="1" type="single">
    <execfile name="ls">
      <url>file:////bin/ls</url>
    </execfile>
    <arguments>
      <value>-la</value>
      <value>outdata</value>
    </arguments>
    <stdout>
      <logicalId>63334</logicalId>
    </stdout>
  </executable>
  <workflow>
    <or>
      <parent triggerState="FINISHED">packed_results</parent>
      <parent triggerState="FAILED">packed_results</parent>
    </or>
  </workflow>
 </task>
</grmsJob>
```

The example above describes single run of the simulation. But one job description is not limited to the single run - it is possible to join many single runs in a one description. GRMS can then process it sequential or parallel way.

Fig. G2 presents schematic view of the single workflow in a GRMS job description.

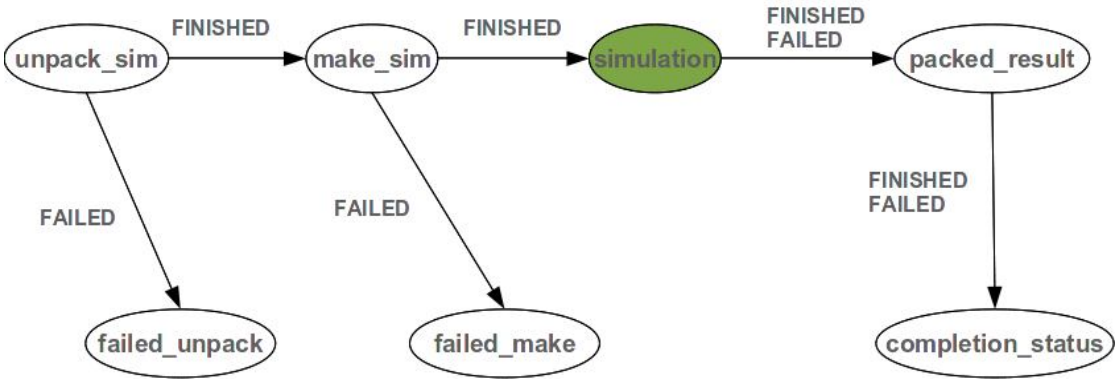


Fig. G2 A schematic view of the single workflow in a GRMS job description

The transitions between the tasks (components of the job) is based on task status changes. The first task of the workflow is called 'unpack_sim' and it is responsible for downloading the proper version of simulation code from DMS and unpacking it on the node chosen for Oncosimulator execution. In case of failure 'failed_unpack' task is launched - leaving the information on DMS about the failure. If unpacking is successful GRMS executes task called 'make_sim' - code compilation. After that step it is possible to run the simulation ('simulation' task). When the simulation is finished the results of it are compressed and send to DMS ('packed_result' task). The last task of the job is 'completion status' and it is responsible for leaving status of the execution on DMS.

Execution of the simulation in the context of ACGT architecture is presented in Fig. G3.

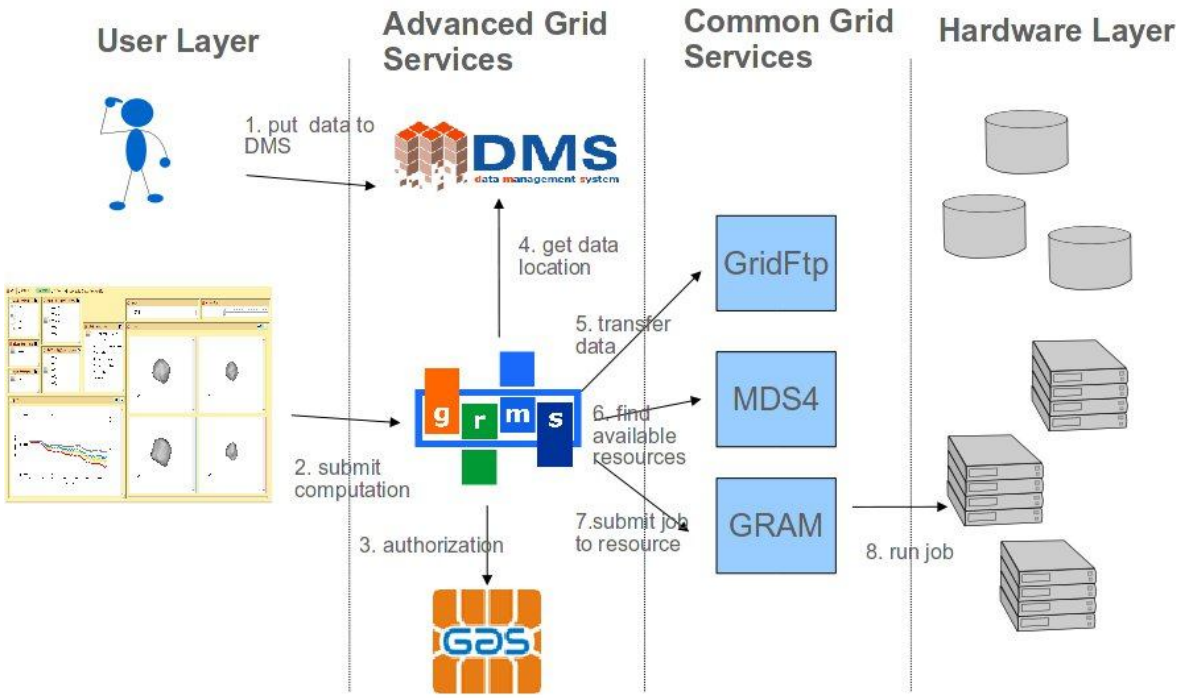


Fig. G3 Execution of the simulation in the context of ACGT architecture

8.3 The Oncosimulator as a service

The very important component of the ACGT infrastructure is a workflow environment. It is a set of services and tools that allows the users to create some scientific experiments by joining different services invocations in a one flow of actions. There are many services that are incorporated into environment: some of them are focused on accessing different databases while the others are able to process the data using statistical packages (e.g. R). Workflow environment lays on the top of the grid infrastructure of ACGT. Constructing workflow one is not using grid directly, but higher level services that are wrapping calls to grid layer.

Oncosimulator service was design and developed to take advantage of the workflow environment of ACGT and to provide means to integrate it with the rest of ACGT software.

The idea behind it, is to wrap grid execution of the Oncosimulator described in a previous chapter, with a web service compliant with the workflow environment

9. Exploratory visualization of Oncosimulator results through lightweight interactive visualization services

(Code letter : V)

9.1 Introduction

Interactive visualization is a powerful paradigm for the user-controlled exploration of scientific data. Visualization is used in situations where automated data analyses techniques fail or are non-existent. In these cases visualization algorithms are used to map scientific data to a graphical representation that allows the scientist to visually detect structure and patterns through his cognitive abilities, experience and expertise. Interaction methods that influence the visualization algorithms put the scientist in the driver seat of an interactive vehicle that allows him to explore his data.

This chapter describes the facilities that have been developed for the interactive visualization of Oncosimulator results and medical images. We have designed and built a lightweight framework that allows for the interactive visualization of data over a distributed architecture. Each visualization facility is wrapped into a service that allows integration into third-party applications including desktop applications, web portals and interactive Virtual Reality devices.

The structure of this chapter is as follows: in the next section we begin with a short explanation on why service-based interactive visualization is a necessary but challenging functionality for the ACGT project, the Oncosimulator in particular. We also provide an overview of related work on web- and grid-based distributed visualization. The visualization facilities implemented for to the ACGT Oncosimulator are described in section 9.3. Section 9.4 describes our implementation and outlines the structure of our visualization services. Next, we describe several applications that use the visualization services in section 9.5 and close with concluding remarks in section 9.6.

9.2 Service-based interactive visualization

One of the objectives of the ACGT project is to provide integrated access to the project's results through web based interfaces, often referred to as a "web portal". This has the advantage that all of ACGT's services are accessible to end-users through a standard web interface: a browser. The "ACGT Main Portal" is the main user interface that invokes the services required for the definition of clinical trials, data management and the definition of data analysis workflows. The decision to use a service-based design allows co-ordinated resource sharing and problem solving in a dynamic, multi-institutional, Pan-European virtual organization, including:

- precise control over access patterns to protect, for example, privacy-sensitive data and services, and
- controlled shared access to distributed computing resources.

Interactive visualization tools are used throughout the ACGT project, both by developers during the implementation of services and workflows but also by the end-users; the clinicians that define and manage a clinical trial. However, the

decision on a service-based design and the integration of these services into a web based portal poses a significant challenge in the development of interactive visualization services for the representation and exploration of data. In particular, the visualization services should be able to provide:

- integration - the services must be ready for integration by third-party applications, and shared web portals in particular;
- flexibility - the visualization needs in the ACGT project range from simple 2D graphs to interactively rendered 3D volumes;
- interaction - the visualization services must both provide interaction capabilities for navigation (i.e. pan, scale, orientation) as well as (re-)parameterization of visualization algorithms;
- responsiveness - a service based design that is implemented on a distributed architecture will add delays to response time as a result of communication overhead;
- scalability - the services must be able to withstand increases in the number of simultaneous service requests.

Non-interactive visualization methods for the graphical representation of information and scientific data are plentiful. Most of these can easily be integrated into third party applications, web portals included, and they often provide flexible 2D and 3D visualization methods. However, because of the lack of interaction, they do not support exploratory visualization. Unfortunately, for *interactive* visualization the above mentioned requirements do not go well together.

Supporting interaction with applications that run on a distributed architecture is far from trivial. Their distributed design implies that communication needs to take place between components which, in turn, negatively influences responsiveness. Furthermore, existing paradigms for the construction of service-oriented architectures (e.g. WSRF) impose a significant overhead on the performance of applications which has a negative effect on both responsiveness and scalability. This has led us to design and build a lightweight framework that allows for the implementation of third-party applications using network accessible interactive visualization facilities.

9.2.1 Related work

We have investigated the results of research projects that address distributed visualization and assessed its utility based on the requirements described in the previous section. We found that most are restricted to the visualization of specific data types or for particular technological domains, or they are limited in their support for interaction.

A good review of research areas and technologies for the support of distributed and collaborative visualization is provided by Brodlie et al [BDG 04]. For web-based visualization projects they make a useful distinction between *Full service* technologies where the service provider executes the visualization algorithms and provides graphical data to the client, *Software delivery* where the client executes the visualization algorithms, provided by the server, and *Local operation* where visualization software is assumed to be available at the client which is triggered by the download of data from the web.

Examples of the *Local operation* paradigm are servers that compute a graphical representation (such as VRML, X3D) of the data which is then transferred to the

client. The browser must support the rendering of this type of graphical data. Different plug-ins for VRML and X3D formats are available, but they differ in maturity and graphical capabilities. More importantly, the VRML format lacks support for non-polygonal types of visualization and has very limited support for interaction.

In *Software delivery* scenarios, the server provides visualization code and data for representation on the client. The responsibility of rendering graphical data lies with the client which has the advantage that local graphics hardware can be effectively used. Alternatives based on Java-based plug-in using Java3D [Jav] or JOGL [JOG] have the downside of having to create a "bridge" from the visualization server to the browser plug-in, and vice versa. The integration of visualization algorithms into either a browser plug-in or as a standalone application often implies a substantial amount of work to support. Transparently providing a plug-in installation that picks the right libraries for any given end-user platform and browser combination is possible, but cumbersome.

Probably the best known example of a *Full service* visualization service is VNC (Virtual Network Computing [VNC]). VNC allows the desktop of one machine (the server) to be displayed on a number of clients (the viewers) and features efficient image compression and input control between server and viewers. The main problem with VNC is that it is not trivial to integrate into third-party applications and that it does not perform well in 3D visualization applications. SGI VizServer [Viz] and Chromium [HHN*02] are similar in structure to VNC but designed to support remote high performance OpenGL rendering hardware.

The Grid Visualization Kernel (GVK [HK03]) was developed within the CrossGrid project and uses *glogin* [RK04] to provide a tunnel between a running grid job and the client as well as a facility that allows X11 port forwarding. This provides an interactive connection to running grid jobs, but this method is not suitable for integration into web based applications.

9.3 Visualization and the Oncosimulator

The Oncosimulator combines tumour information obtained from medical imaging techniques (CT, MRI, PET and ultrasound) with mathematical models that predict the growth of tumours and the response to chemotherapy or radiation therapy [SDG*07], specifically for the types of tumour under consideration: nephroblastoma and breast cancer. The Oncosimulator produces spatiotemporal predictions of the composition and morphology (shape) of the tumour over the course of time. These predictions provide clinicians with valuable information on the most effective treatment out of several alternatives, as well as detailed parameters on the optimal composition of a treatment scheme, including the total treatment period, the type of drug(s), dose, and interval between treatments.

A treatment is defined by several parameters, each of which has a range of possible values and is usually influenced by other parameters. Potential future clinical applications include scenarios where simulation models help to define the initial parameters that predict the treatment that is most likely to be most effective. Finding the optimal combination of parameters for a certain treatment is difficult. As no analytical method for finding this optimum exists, the ACGT Grid environment is used to perform a large number of simulations for combinations of parameters that are thought to be most successful. The architecture developed in ACGT consists of a Grid-based distributed computing and software framework that allows *in-silico* tumour simulation models, interactive visualization methods and

other data sources to be combined into an interactive visual exploration environment. This environment can be used to explore simulated predictions of tumour growth and treatment response.

The visualization needs for the Oncosimulator are varied. First: the desired visualization methods depend on the role of the end-user: the Oncosimulator model developer will be interested in visualization methods that help in tracing bugs, but a clinical end-user will want to see an immediate picture on tumour response given a range of different treatment. Second: the application context in which the visualization methods are used play a role: desktop applications will have other resource capabilities than a web portal or an interactive virtual reality device.

9.3.1 Visualization of Oncosimulator data

The Oncosimulator takes a bi-level 3D volume dataset as input. The two levels in the input denote which voxels (a volumetric pixel) of the volume are part of the tumour and which are not. The output of a simulator run is a volume (of the same dimensions as the input volume) for each hour of simulated time, containing an updated tumour region. Other Oncosimulator outputs include counts of the different cell categories used internally by the simulator. This cell count is best represented in 2D time-versus-count plots. The volume output is represented by 3D isosurfaces or volume rendering. The main visualization methods required for the Oncosimulator are categorized as follows:

1. Predicted change in overall tumour volume:

a. 2D line graph showing number of geometrical cells involved in the tumour, over time.

b. 2D line graphs showing number of cells of each type in the whole simulation, over time.

2. Predicted change in overall tumour shape:

a. Basic view: 3D rendering of the external shape of the tumour delineated by healthy tissue area or necrotic core.

b. Animated view: same as above, but with an animated camera view that exposes concealed geometry.

c. Animated growth: 3D rendering of the change in shape of the tumour over simulated time.

d. For comparison with after-treatment scans and segmentation: this will in general require registration of the after-treatment scan with the simulated tumour. This is non-trivial, but if done properly it offers overlaid presentation of the medical images with the simulated tumour.

3. Predicted composition of tumour:

a. Basic view: colour-coded geometric cells (healthy, mostly proliferating, mostly necrotic).

b. Internal view: 2D cut-through plane of a 3D rendering to reveal internal structure.

9.3.2 Initial steps: interactive visualization on the desktop

As a first step towards the visualization of Oncosimulator results we designed and implemented a desktop visualization environment called VTKFly2. This application was implemented using the Python programming language and the Visualization Toolkit (VTK) [SML06].

VTKFly2 provides a graphical user interface in which simulation datasets can be loaded and the evolution of the simulated tumour played back. The application allows multiple, colour-coded isosurfaces extracted from the simulation output to be displayed simultaneously to show the (change in) composition of a tumour (methods 2a, 2b, 2c and 3a listed in section 9.3.1). It can show simulation output together with volume rendered DICOM imaging data allowing visual comparison of pre-treatment and after-treatment tumour shapes (2c). An interactive cutting plane provides methods to slice through any data volume to reveal internal structure on a 2D sample plane (3b).

For the first ACGT Review meeting in Poznan in April 2007, VTKFly2 was integrated with a Grid-enabled version of the Oncosimulator. When a simulation run started, the sequential output files written by the Oncosimulator (one per hour of simulated time) are transferred “live” to the desktop system running VTKFly2. The application immediately appends new simulation results to the current visualization time line as they become available. This allows a user of VTKFly2 to inspect the simulation outcome *while* it is running on the remote system.

9.4 Visualization as a service

The application described in the previous section was successfully integrated within the initial ACGT infrastructure, but it was still a standalone desktop application. This limited its application within ACGT. To increase availability and flexibility we designed an architecture that provides visualization facilities through services. A service here resembles a resource that is accessible over a network and that produces one or more visualizations. The behaviour of a visualization facility is manipulated through “service parameters” that are defined and changed by the users of the service. A visualization service does not *have* to produce an image as output. Intermediate steps needed to produce that image, like isosurface extraction or the placement of glyphs, can also be provided by a service, if needed.

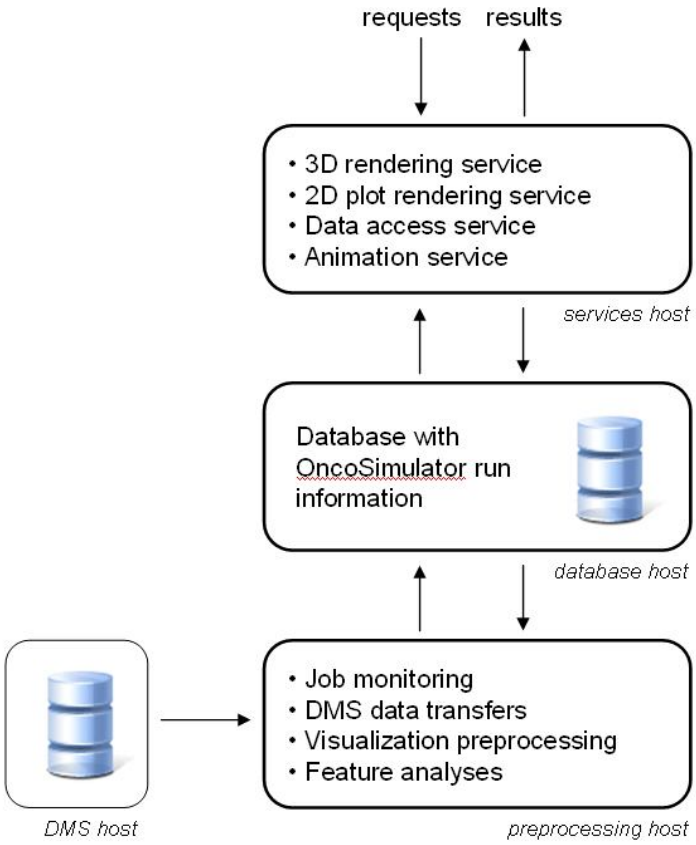


Figure V1. Schematic overview of the visualization architecture.

9.4.1 Architecture overview

A schematic overview of the implemented visualization architecture is shown in Figure V1. The top of this figure represents the perspective of the end-user: a collection of visualization services that return results in response to requests. The visualization services consult a database that contains a complete record of all Oncosimulator runs. This database contains information on (1) the input parameters used by each run, (2) the state of completion of each run, and (3) feature information on the output produced by each run (dimensions, number of timesteps, volume). These records are produced by a preprocessing host that (1) monitors the state of Oncosimulator runs on the grid, (2) transfers Oncosimulator output data from the DMS when runs have successfully completed, and (3) preprocesses Oncosimulator output data into a format that can be readily used by the rendering services.

9.4.2 Visualization service design

The visualization services are accessible through the standard HTTP web protocol. Each visualization service can be seen as a regular web server from which the end-user requests one or more images using conventional URLs. The visualization service returns the output in the form of an image (for the case of a rendering service), text (for the case of the data access service) or a byte stream (for the case of the animation service). All services are stateless. Parameterization of a

visualization service is done using HTTP GET and/or POST requests. As the visualization service communicates with the outside world using only the HTTP protocol, integration of the service with existing applications is straightforward. The users of a service do not necessarily have to be web-based applications, as will be described later in this chapter. Our service design does not require the installation of plug-ins in the browser. The design supports any kind of imaging output produced by the visualization algorithms and so is not limited by what is implemented on the browser side. The visualization services include support for interactive 2D and 3D renderings in a regular web browser with the only prerequisite that the browser supports Javascript. The architecture of this solution is presented next.

9.4.3 Service implementation

Each visualization service is implemented as a tiny application that consists of two parts: one part performs the actual visualization and rendering, the other part handles service requests via HTTP. Most visualization functionality is implemented using the classes provided by the Visualization Toolkit (VTK, [SML06]) and other libraries. Graphical rendering to images is performed through hardware-accelerated OpenGL libraries where available. To handle service requests, the application contains a minimal web server capable of servicing HTTP GET and POST requests [FGM99]. A tightly integrated system like this has multiple advantages over an architecture consisting of a regular visualization application coupled with a “full blown” web server, like Apache or Lighttpd. By merging both components into a single application we ensure the web server has direct access to the visualization output thereby greatly reducing overhead. Furthermore, incoming service requests can directly be applied to the visualization service without any intermediate processing or process communication induced by a web server.

Images sent from a visualization service to the browser are encoded in PNG format. For reduced image size and transmission time JPEG is available as an alternative choice. However, the compression algorithm used in JPEG is not lossless which may result in undesirable compression artefacts visible in the images.

9.4.4 Interaction with visualization services

We classify interaction with visualization in two categories: parameterization and navigation. Parameterization of scientific visualization services is done through HTTP GET and POST requests, as described in the previous sections. Navigation is the type of interaction that allows the user to change the view on a visualization, i.e. panning, rotation and scaling. To accomplish this type of interaction we use a form of AJAX (Asynchronous Javascript) programming. Whenever the user interacts with a 3D views, Javascript code in the HTML page sends updated view parameters to the instance of the visualization service that provides the image. The new view parameters are distributed by the service to the instances on the server to provide synchronized 3D views. After sending the updated parameters, the Javascript code in the HTML page invalidates all 3D view images which causes the browser to fetch updates images from the visualization instances. This method provides synchronized interactive 3D views in a web browser.

9.5 Applications

We have applied our visualization architecture to the following environments:

9.5.1 Web portals

Integration with web portals is illustrated through a working prototype, shown in **Error! Reference source not found.**. This figure shows a screenshot of a web browser in which the output of two visualization services are visible. Each service shows a different simulation dataset but at the same time step. The left dataset corresponds to untreated tumour growth, while the right one shows the results of treatment with the drug epirubicin.

Isosurfacing has been applied to the simulation output to extract the tumour shape. The use of clipping and colour coding reveals differing numbers of cells on the inside of the simulated tumour. The user changes simulation timesteps and location of the clipping plane using the buttons underneath the views. The user can manipulate the 3D views to pan, scale and rotate the tumour with the mouse. The two 3D views are kept synchronized on the server during interaction. For synchronization we use a tuple space model, similar to the LINDA system [Gel95].

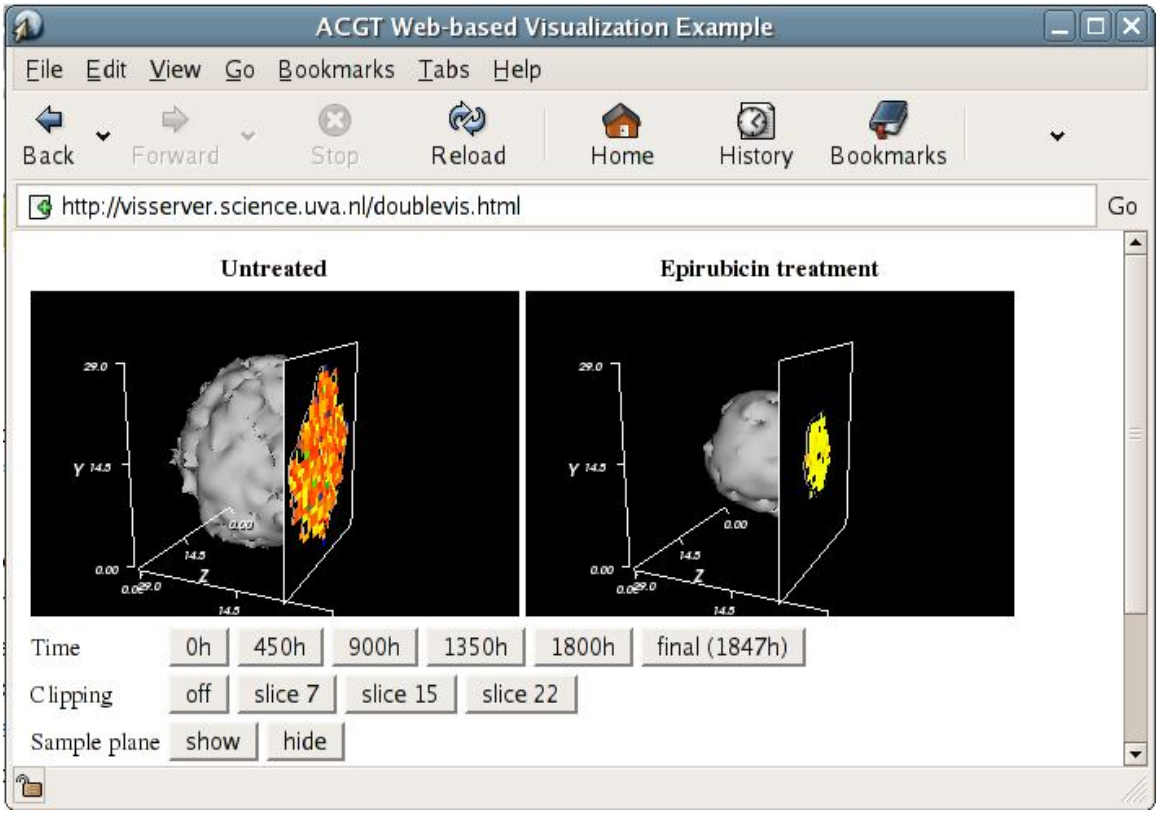


Figure V2. Visual comparison of two simulation datasets in a web portal. The left dataset corresponds to untreated tumour growth, the right shows the results of treatment with the drug epirubicin. The 3D views are interactive: the view can be panned, rotated and scaled interactively using the mouse

.

9.5.2 Test-case 2: The RecipeSheet

The RecipeSheet allows parameter spaces to be interactively navigated in an intuitive way by providing ways for different scenarios to be compared and easily altered [*The RecipeSheet is described in more detail elsewhere in chapter 10*]. It will show output for multiple scenarios so that the user can compare explore alternative treatment scenarios simultaneously. For this it provides special GUI components and interaction methods. The RecipeSheet is described in more detail elsewhere in this document.

To support interactive exploration of treatment scenarios, the RecipeSheet supports the execution of one or more simulator jobs. When each of the simulator runs completes, its output is transferred from the DMS to the visualization "preprocessing host", as described in the previous section.

Figure V3 shows the visualization services integrated with the RecipeSheet. The four views of the simulation output are not generated by the RecipeSheet. They are not even regular images. Instead, each image is an embedded web browser component that retrieves HTML pages from our visualization services. The four 3D views in these pages are interactive and synchronized, as described in Test-case 1. When a user changes any of the parameter values used for defining the current scenario, the RecipeSheet sends the updated values to the visualization service and reloads the image. The RecipeSheet does not know (and does not have to know) anything about how these visualizations are produced by the service, as the RecipeSheet is merely the service user. It only needs to know how to communicate with the service, what visualization elements the service provides and what parameters are available to influence these elements.

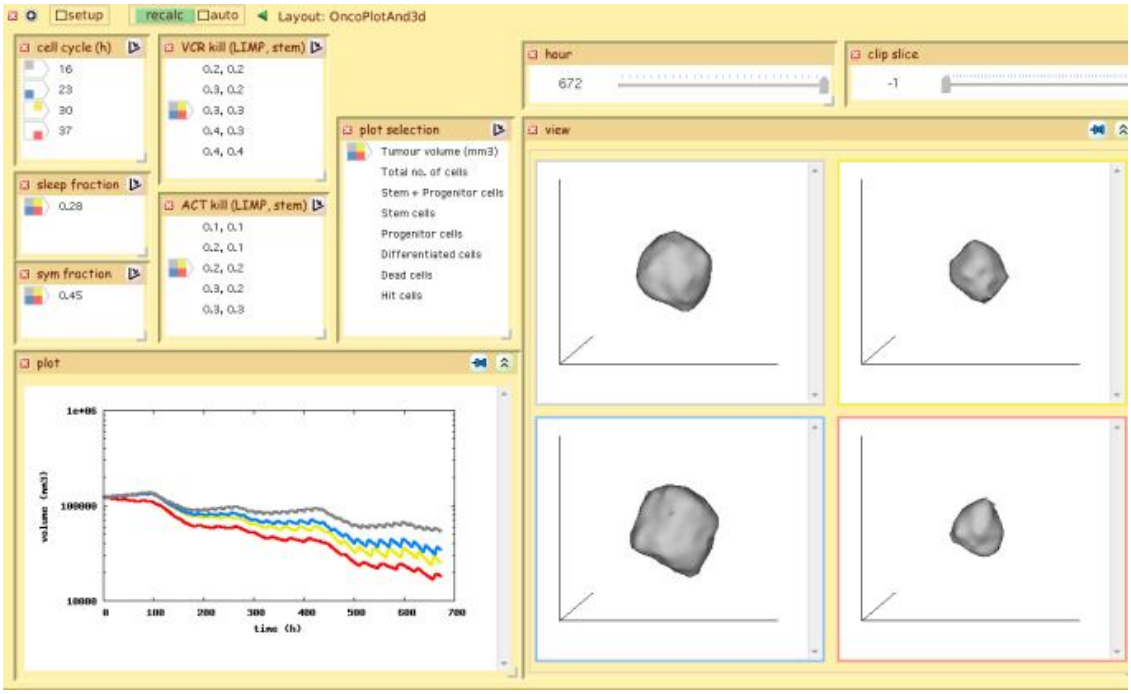


Figure V3. Screenshot of the visualization services used in the RecipeSheet. The 2D graph in the lower left and the four 3D views on the right are generated by visualization services.

9.5.3 Test-case 3: The Personal Space Station

We have constructed a third test-case for our visualization services in the shape of an application for a personal Virtual Reality device called the Personal Space Station (PSS, see also

Figure V4). The PSS supports stereoscopic 3D rendering on a mirror display and optical hand and object tracking to create an environment where the user can "reach in and touch their data". The optical tracking system tracks the user's head so that the orientation of the projected image can be altered to generate a motion parallax effect. This effect is an important depth cue used by the visual system to get a sense of distance and size of a 3D object. Object tracking adds the ability for a user to change the orientation and location of a 3D visualization by changing the orientation and location of a handheld object.

The combination of stereoscopic rendering together with the intuitive interaction methods provided by hand and head tracking creates the illusion that objects appear to float in front of the user. This provides the viewer with an almost instantaneous understanding of the morphology of 3D objects. Within the ACGT project, this type of visualization can provide information on the shape of a simulated tumour obtained from an Oncosimulator experiment within the context of the medical images obtained from the patient. Especially in cases where tumour resection will take place after chemotherapy or radiation therapy, the decision on when to terminate treatment often depends on the shape and size of the tumour with respect to the surrounding organs. An example of this type of visualization is shown in Figure V5.

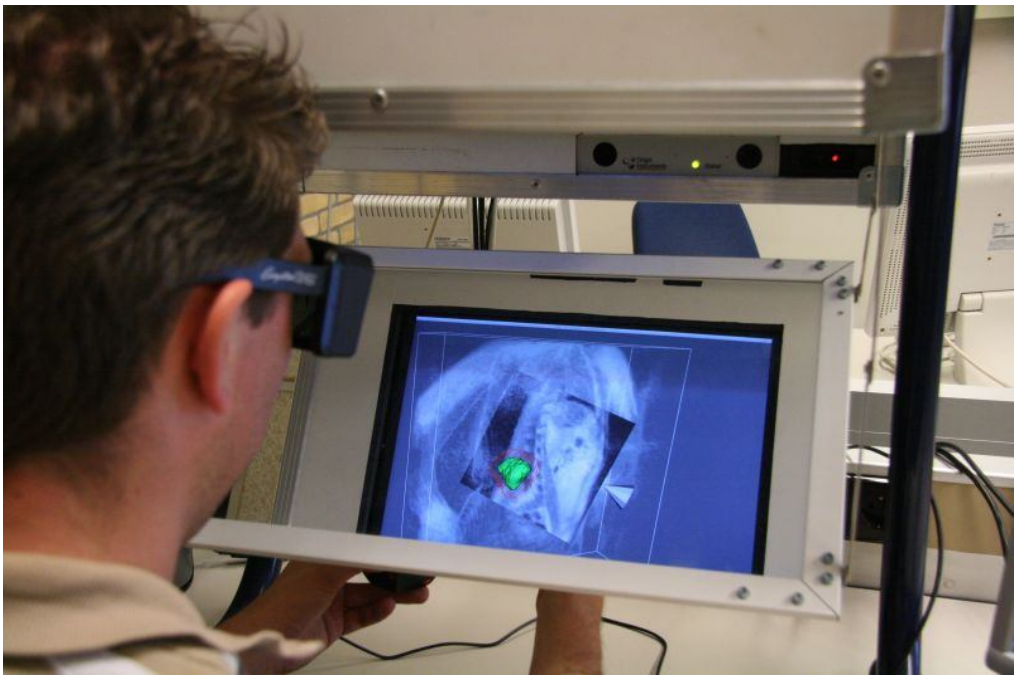


Figure V4. Interactive visualization of Oncosimulator data on a Virtual Reality device called the Personal Space Station (PSS). The system combines stereoscopic rendering with head and hand tracking to give the illusion that objects appear to float in front of the user.

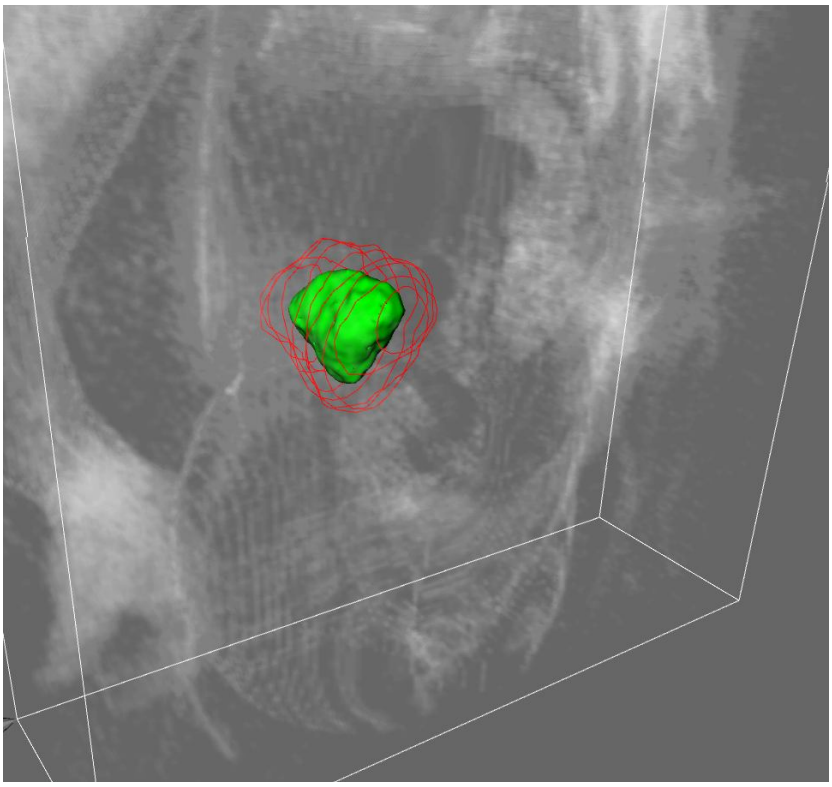


Figure V5. Simultaneous visualization of patient imaging data (white), pre-treatment tumour contours (red) and simulated post-treatment tumour (green).

9.6 Conclusions

This chapter describes the architecture developed for the exploratory visualization of Oncosimulator results through lightweight interactive visualization services. The “visualization as a service” paradigm is a powerful and flexible way of allowing visualizations to be used by different end-user applications. Our architecture allows interactive 3D visualizations from third party applications such as web portals, desktop applications and Virtual Reality devices.

The performance of the visualization services depends largely on the distance between the end-user and the visualization architecture. We have had good results with the use of our services in local and regional settings. Applications that are widely distributed have shown significant decreases in responsiveness of the visualization services. Table V1 shows timing measurements for five different geographical locations. We used four datasets in these measurements, with each dataset being available through a separate service instance. The instances all ran on the same physical server located in Amsterdam, which is a single-CPU 3.06 GHz machine using an NVIDIA Quadro FX 4000 graphics card. In each measurement the resolution of the images produced was 300x250 pixels, compressed in PNG format. This might seem small compared to the average user’s screen resolution, but in a setup with the RecipeSheet as service user we display four images at the same time together with GUI elements, so the image size is constrained by the available screen space.

Table V1. Timings for changing view parameters of the visualization service, followed by retrieving the corresponding updated image. The view update and image retrieval times are in milliseconds, averaged over 1,000 updates, with standard deviation between parentheses. “Total time” (in seconds) measures the total time taken to complete the updates and retrievals, “FPS” is the corresponding average number of frames per second.

Set	View upd.	Image retr.	Total time (s)	FPS
	LAN			
1	43.3 (0.9)	20.8 (1.2)	64.0	15.6
2	44.6 (1.6)	18.6 (2.6)	63.3	15.8
3	42.0 (1.0)	18.3 (2.5)	60.3	16.6
4	44.8 (1.6)	18.4 (1.6)	63.2	15.8
	Amsterdam			
1	64.1 (1.7)	47.1 (1.7)	111.2	9.0
2	63.8 (1.7)	44.1 (8.0)	107.9	9.3
3	63.9 (1.7)	43.3 (2.6)	108.8	9.2
4	64.3 (1.6)	43.7 (1.7)	108.0	9.3
	Eindhoven			
1	43.3 (1.7)	27.3 (1.8)	70.6	14.2
2	44.8 (0.8)	25.2 (2.0)	70.0	14.3
3	45.2 (0.6)	24.8 (0.9)	70.0	14.3
4	45.1 (2.5)	30.5 (28.3)	75.5	13.2
	Poznan			
1	96.5 (1.7)	75.2 (3.2)	171.6	5.8
2	95.1 (1.1)	73.1 (1.1)	168.2	5.9
3	95.5 (1.5)	72.6 (2.9)	169.8	5.9
4	95.4 (1.3)	73.0 (3.1)	168.4	5.9
	Hokkaido			
1	758.7 (4.5)	582.4 (3.5)	1344.3	0.7
2	760.3 (4.2)	582.6 (9.5)	1346.1	0.7
3	759.3 (3.5)	579.2 (2.3)	1341.2	0.7
4	759.8 (4.1)	582.5 (2.5)	1346.6	0.7

The current implementation of our architecture uses the hardware resources from a limited number of workstations. To provide a scalable solution that can be used by multiple users from different locations, a collection of resources would need to be created. The University of Amsterdam is currently in the process of acquiring new hardware resources to satisfy this requirement.

9.7 References

[BDG_04] BRODLIE K., DUCE D., GALLOP J., WALTON J., WOOD J.: Distributed and collaborative visualization. Computer Graphics Forum 23 (2004), 223–251.

[FGM_99] FIELDING R., GETTYS J., MOGUL J., FRYSTYK H., MASINTER L., LEACH P., BERNERS-LEE T.: Hypertext transfer protocol – http/1.1. http://www.faqs.org/rfcs/rfc2616.html.

[Gel85] GELERNTER D.: Generative communication in linda. ACM Transactions on Programming Languages and Systems 7 (January 1985), 80–112.

[GNU] Gnuplot. http://www.gnuplot.info.

[HHN_02] HUMPHREYS G., HOUSTON M., NG Y., FRANK R., AHERN S., KIRCHNER P., KLOSOWSKI J.: Chromium: A stream processing framework for interactive graphics on clusters. In ACM SIGGRAPH (2002).

[HK03] HEINZLREITER P., KRANZLMÜLLER D.: Visualization services on the grid: The Grid Visualization Kernel. Parallel Processing Letters 13 (June 2003), 135–148.

[Jav] Sun Developer Network, Java 3D API. http://java.sun.com/javase/technologies/desktop/java3d.

[JOG] Java.net, The JOGL API Project. http://jogl.dev.java.net/.

[LH06] LUNZER A., HORNBÆK K.: RecipeSheet: Creating, Combining and Controlling Information Processors. In Proceedings of the 19th annual ACM symposium on user interface software and technology (UIST 2006) October 2006), pp. 145–153.

[RK04] ROSMANITH H., KRANZLMULLER D.: glogin - a multifunctional, interactive tunnel into the grid. In Fifth IEEE/ACM International Workshop on Grid Computing (November 2004), pp. 260–265.

[SDG_07] STAMATAKOS G., DIONYSIOU D., GRAF N., SOFRA N., DESMEDT C., HOPPE A., UZUNOGLU N., TSIKNAKIS M.: The "oncosimulator": a multilevel, clinically oriented simulation system of tumour growth and organism response to therapeutic schemes. towards the clinical evaluation of in silico oncology. Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE (August 2007), 6628 – 6631.

[SML06] SCHROEDER W., MARTIN K., LORENSEN B.: The Visualization Toolkit An Object-Oriented Approach To 3D Graphics, 4 ed. Kitware, Inc., December 2006.

[TKP_06] TSIKNAKIS M., KAFETZOPOULOS D., POTAMIAS G., MARIAS C., ANALYTI A., MANGANAS A.: Building a European Biomedical Grid on Cancer: The ACGT Integrated Project. In Proceedings of HealthGRID 06 Conference (2006), pp. 247–258.

[Viz] SGI VizServer. http://www.sgi.com/software/vizserver/.

[VNC] RealVNC Limited. http://www.realvnc.com/.

[WSS_07] WEGENER D., SENGSTAG T., SFAKIANAKIS S., RUPING S., ASSI A.: GridR: An R-Based Grid-Enabled Tool for Data Analysis in ACGT Clinico-Genomics Trials. In IEEE International Conference on e-Science and Grid Computing (December 2007), pp. 228–235.

10. A front-end interface for the ACGT Oncosimulator validation

(Code letter: F)

10.1 Introduction

In this chapter we describe the front-end facilities that have been created to support the developers of the Oncosimulator in their work on validating the simulator's behaviour. The facilities are designed to provide secure access, for multiple users across multiple partner sites, to ACGT's internal Grid, and thus to let users benefit from high-performance computation, a shared repository for completed simulations, and sophisticated on-demand visualisations.

It is perhaps important to clarify the sense of "validation" being used here. In the context of this chapter, validation of the Oncosimulator is narrowly taken as meaning the work of confirming that the simulator behaves appropriately as a simulation of cancer. That is, that it delivers credible results, in terms of both the cell counts and tumour shape, when applied to tumours of a range of sizes and shapes, and that adjustments to each of its parameters, singly or in combination, have the expected impact on the results – an impact appropriate to the parameters' roles within the simulator's underlying model of tumour growth and treatment.

It should be stressed that such validation requires human-guided visualisation rather than automated optimisation. Even if in its eventual deployment the simulator might be called upon to produce just a single value – the predicted volume of the tumour after treatment – at this stage its developers must confirm that each simulation follows a legitimate and credible course to its final value. This calls for inspection of the time plots of all significant elements of the simulator's internal state to confirm that there are no suspicious discontinuities, and likewise of the modelled 3D tumour shape to confirm that it is not deformed in unlikely ways.

As explained in deliverable D8.3, the decision was made to base the validation interface on the RecipeSheet [12], an end-user programming environment that explicitly includes facilities to help users request, compare, and interact with multiple alternative results simultaneously. These facilities were felt likely to be valuable for the validation work, due to the need for rigorous examination and comparison of the differing outcomes obtained when parameters are varied. What we believe is special about the facilities offered by the RecipeSheet, deriving from its built-in subjunctive-interface features [8,13], is how they support the user in rapidly configuring and reconfiguring the set of alternatives that are to be viewed. Given a collection of parameters on which choices are to be made, the user is free to create up to 12 "scenarios", each comprising a selection on every parameter, and these scenarios' results are then displayed side by side. As a substrate for building a simple calculation flow and comparing its visual outputs, the RecipeSheet is thus more flexible than previously proposed environments such as the more strictly spreadsheet-like systems proposed by Chi et al. [1] or by Jankun-Kelly and collaborators [5,6].

The initial research question driving this deployment of the RecipeSheet therefore related to the usability of the subjunctive-interface mechanisms: would the Oncosimulator developers find effective ways of using these mechanisms to obtain informative visualisations of sets of simulation results, and thereby uncover

previously unknown properties of the simulator? The main challenge in pursuing this question was that the RecipeSheet had not previously been designed to support a calculation with such a large number of parameters, and hence such a large potential result space, as the Oncosimulator.

However, the above question assumes that the simulation results to be explored already exist, and leaves unanswered the issue of how users would be supported in requesting new results. In the interests of presenting the Oncosimulator developers with a single, unified interface through which they could both browse existing results and request new ones, it was decided to extend the functionality of the RecipeSheet application into, in effect, a simple form of problem-solving environment [20] or even online laboratory notebook [19], that lets its users benefit from the capabilities of ACGT's secure Grid mechanisms. For this goal the main challenges turned out to be performance and error tolerance: providing services that responded quickly enough for interactive use, and protecting the users from the many possible failure modes (transient or long-lived) of a heterogeneous, multi-site Grid that was itself still a work in progress. Our initial findings have been reported in [3,10,9].

The remainder of this chapter is organised as follows: First we give an overview of all the work carried out in this module of the workpackage over the course of the ACGT project. Next is a section going into greater detail on the particular developments since the previous report (in deliverable D8.3), and the work planned for the final months of the project. Finally we summarise our findings to date regarding the validation work, and make some suggestions regarding future testing, development and potential deployment.

10.2 Review of work since the start of the project

As stated in the Introduction, the scope of this aspect of Oncosimulator work has expanded greatly since the early days of the project. In this section we give an overview of the sequence of development and evaluation tasks carried out since UHok became involved with the workpackage.

10.2.1 Early proof-of-concept

The first proof-of-concept demonstration of how the RecipeSheet could help users to explore simulator results was produced at the end of 2007, using results from a simulation of the growth and radiotherapeutic treatment of glioblastoma multiforme. A screenshot is shown in Fig. F1. For this example we fetched results from a repository, held on the user's local disk, of the results of pre-running the simulation on all combinations of the parameter values of interest. All the simulations started from the same tumour shape, built from a scan of a previously treated patient.

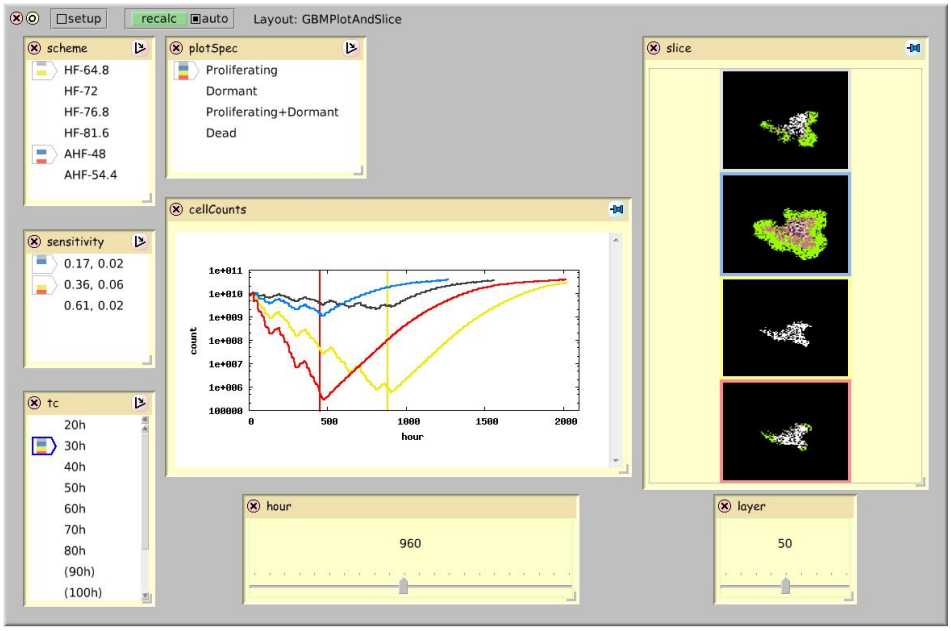


Fig. F1 A demonstration of browsing simulator results using multiple user-created scenarios on an early version of the RecipeSheet. This figure shows the results of four simulations of radiotherapy treatment of a glioblastoma multiforme tumour. The simulations are for two treatment modes (accelerated or not), each combined with two levels of tumour sensitivity. In the centre is a plot of the calculated number of proliferating cells over time, and on the right are slices – one for each scenario – through the tumour.

10.2.2 First use of remote services

At the ACGT annual review in May 2008 we demonstrated using the RecipeSheet to request visualisations from a server (at UvA) on which the simulation results were stored. By this stage we were using the simulator that had been built to model chemotherapeutic treatment of nephroblastoma, but the results still related to a single anonymised patient, and were again created by pre-running exhaustive combinations of values on each of a handful of simulation parameters.

A snapshot of the demonstrated system is seen in Fig. F2. As reported in D8.3, by this stage the views delivered from the visualisation server were not only static images. In particular, in collaboration with UvA we had created a mechanism supporting synchronised dynamic update of the 3D views in response to manipulation by the user. A result cell on a RecipeSheet can have the full behaviour of an Internet Explorer view, and this was exploited by embedding Javascript code in the individual views. This mechanism is described more fully in the visualisation chapter of this deliverable. The manipulations supported in our demonstration were simple – horizontal or vertical rotation of the view, by dragging with the mouse – but the mechanism showed the feasibility of supporting potentially complex synchronised view manipulation, with all the code for that manipulation provided by the server. The views' substrate, in this case the RecipeSheet, only needs to detect when the views have been updated, and to ensure that they are appropriately redrawn.

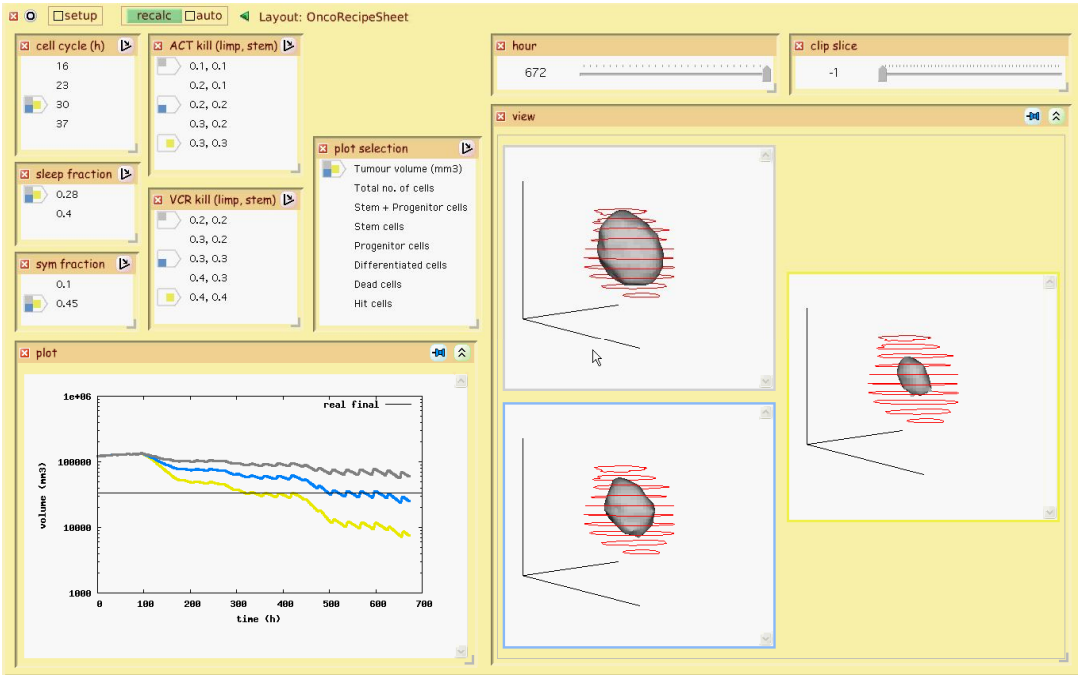


Fig. F2 Early demonstration of interaction with results from simulation of nephroblastoma. Both the line plot and 3D visualisations were requested and delivered on the fly from a server in Amsterdam. The 3D view cell included mechanisms for synchronised manipulation of all scenarios' views: dragging the mouse to rotate any tumour image would result in the same transformation on all tumour views.

10.2.3 Full Gridification

The next major step, which began in earnest at the start of 2009 and involved intensive collaboration among all the workpackage's technical partners, was to design the Grid-based services and communications that would give a RecipeSheet user control over the full life-cycle of simulation results: seeing which parameter-value combinations had already been processed, choosing new combinations, submitting them for execution, seeing when the results became available, and requesting visualisations based on new and previously obtained results.

The requirements for this Grid setup, which became known as the Integrated Oncosimulator, included the following:

- **security:** Only authenticated users should be able to run simulations, or to access original patient data (DICOM datasets, etc).
- **centralised storage of large data:** Users should not need to download simulation results to their local PCs in order to view them.
- **low-latency access to visualisations:** Users should have fast (sub-second) access to at least simple visualisations; likewise to information showing the parameter combinations for which results are already available.
- **scalability:** System performance should not degrade unduly when the repository reaches a level of tens of thousands of simulation results (our initial estimate was that validation could easily involve up to 50,000).
- **service access without advanced code or settings:** The RecipeSheet is written in Smalltalk – a high-level, dynamic language – so the services intended for

use from the RecipeSheet needed to be accessible without relying on low-level system libraries, and preferably without complex protocols (these are commonly accepted requirements in provision of services in the field of bioinformatics, for example, where service users typically write their code in dynamic languages such as Perl; see [17]). Similarly, user sites should not be required to set up non-standard configurations for network access, such as unblocking particular ports in a firewall to allow their PCs to be signalled directly by a remote server (this being one communication mechanism offered by PSNC's GRMS).

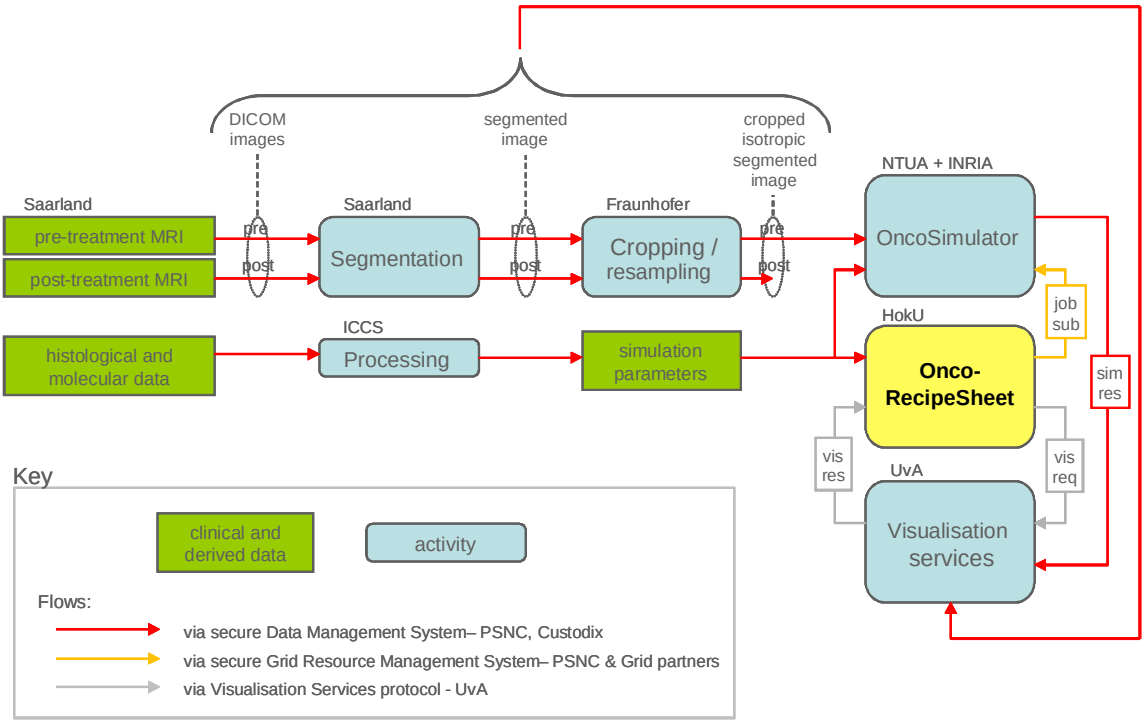


Fig. 3: The overall structure of the multi-site Integrated Oncosimulator setup for the validation phase

Fig. 3 shows the overall flows of data and commands through the Integrated Oncosimulator, as set up for simulator validation based on the use of data from previously treated patients. It includes the role for the RecipeSheet – which is hereafter referred to as the OncoRecipeSheet (ORS), meaning the RecipeSheet environment as extended to support the operations needed for the Oncosimulator. An explanation of this figure, adapted from detailed notes supplied by Robert Belleman in a WP8 internal document, is given below.

First, the main elements in the diagram have the following meanings:

- green boxes represent data sets (i.e. files);
- blue rounded boxes are tasks, annotated with the responsible institute;
- red arrows reflect the use or transfer of data that is stored and maintained in a secure central data repository called the Data Management System (DMS);
- yellow arrows represent secure commands sent to the Grid Resource Management System (GRMS);
- grey arrows represent the exploration loop involving the ORS and the Visualisation services.

The activities represented are:

1. **Importing reference data** Partners at Saarland (USAAR) provide data from ~20 previously treated patients. For each patient the following data is collected:

a. Pre-treatment MRI (DICOM)

b. Post-treatment MRI (DICOM)

c. Clinical and genetic information (Excel sheet)

2. **Segmentation** Saarland use the Fraunhofer SegmentationTool to create a segmentation of the pre-treatment and post-treatment MRI. This results in the following datasets:

a. Pre-treatment segmented image (MHD/RAW)

b. Post-treatment segmented image (MHD/RAW)

Remarks: After additional processing in step 3, the pre-treatment segmented image is used as the “initial condition” for the Oncosimulator: it represents the simulated tumour at $t=0$, the start of the simulation. The post-treatment segmented image will be used as a validation image: in an ideal situation, the simulated tumour will converge to the exact same morphology of the post-treatment tumour.

3. **Cropping/resampling** Fraunhofer (FhG) take the pre- and post-treatment segmented images and process them (cropping and resampling to isotropic voxels). This results in the following datasets:

a. Pre-treatment cropped isotropic segmented image

b. Post-treatment cropped isotropic segmented image

Remarks: The reason for this is that the Oncosimulator requires an input image that (1) is isotropic, i.e. each voxel must be cubic, and (2) the image must be cropped to fit only the tumour with a small amount of "empty" space around it to allow the simulated tumour to expand. The pre-treatment image is ready for use by the Oncosimulator. The post-treatment image is used in visualisation for a visual comparison of the simulated tumour with the post-treatment tumour.

4. **Processing non-image data** ICCS process the clinical and genetic information into simulation parameters that will be used by the Oncosimulator. This results in simulation data and parameters. UHok use this simulation data to derive "free" parameters and their value ranges. These are the parameters that span a solution space that is to be explored using the ORS. In the ORS, the free parameters are represented by graphical widgets. Facilities are available for advanced ORS users to edit the parameter ranges.

5. **Running simulations** ORS users request new simulations, by selecting desired combinations of parameter values, including the patient whose tumour is to be simulated. This leads to submission of new Oncosimulator jobs ("job sub"). Each job is assigned a unique ID at the time of submission. Simulation results ("sim res") generated by the Oncosimulator are of two kinds: (1) the tumour morphology over time as multiple 3D volumes, and (2) text data that represents the internal state of the OncoSimulator over the progress of its execution (such as cell counts).

Remarks: The Grid can only be accessed by users who have the appropriate security credentials, as administered within ACGT by the partner Custodix. All individuals who want to work with the ORS must first register with Custodix, and must then prepare a certificate, stored on their local PC, that is used by the GRMS and DMS access clients to persuade the Grid servers that the user is entitled to the access. These certificates expire, and must be renewed periodically by having the user input a dedicated password. The ORS includes an automatic mechanism for checking the validity of its user's certificate before attempting to access Grid resources, and if necessary guiding the user through the renewal.

6. **Preparing result visualisations** UvA receive the results ("sim res") from each completed simulation, and store them locally to the Visualisation server to enable rapid generation of visualisations when later requested by users. Some pre-processing is also carried out at this stage to improve performance of the visualisation service. The Visualisation server also maintains copies of other data from which visualisations can be requested, such as the segmented images derived from the patients' DICOM studies.

7. **Viewing simulation results** ORS users select desired combinations of parameter values for results that they wish to see, leading to visualisation requests ("vis req") being sent to the Visualisation services. The requests identify a simulation result using the ID that was assigned to the simulation job. Similarly, visualisations based on a patient's DICOM studies can only be requested in association with a known simulation result for that patient; thus visualisation requests never need to contain a patient ID. The Visualisation server returns to the ORS a visualisation result ("vis res") – either a graphical object, or a handle to an online resource – to be displayed.

Details of how simulation jobs are specified to the GRMS, and the directory structure used on the DMS for storing results, are to be found in the chapter on Grid services. Likewise, details of the protocols supported by the visualisation

server are given in the visualisation chapter. The designs of these protocols are not especially innovative or noteworthy in themselves, but constitute aspects of an overall design by which we have sought to make efficient use of the available resources, in particular by balancing the division of labour between the services and the client software (in this case, the ORS) using them. Some other significant aspects of this design are as follows:

- **Metadata is handled by clients.** The DMS includes facilities for attaching metadata to files, and for searching for files by their metadata. However, given our stringent performance requirements for the interactive browsing of simulation results, it was clear that on-the-fly querying would not be fast enough. Therefore we designed a simple mechanism, using textual index files held on the DMS, for keeping a record of all simulations and their parameter values. Each user's copy of the ORS consults these index files to build an in-memory map of the simulations, that can be queried rapidly to find simulations that use particular parameter values: with the current repository of 3000 simulations, a query selecting on the values for 10 parameters takes less than 100ms (on a 3GHz Windows PC). When a user submits new simulation runs, the ORS updates the relevant index file with the data for the new runs.

- **Temporary DMS files are used for signalling.** Simulations submitted to the Grid are handled asynchronously: the ORS that submits them does not suspend its other work while waiting for them to finish. This implies that the ORS needs a way to discover when a simulation has finished, at which point it will check the outcome and, if the simulation was successful, trigger the next step of transferring the results to the visualisation server. In addition there are activities that update data that is shared between many users' ORS clients, and therefore require a form of lock to prevent multiple ORSs trying to make changes at the same time. This includes the submission of runs, where a lock is needed to ensure that a given combination of parameters is submitted only once, and that each run is given a unique ID. Both these situations – checking status, and the use of locks – require a form of signalling that can be read by ORS instances; our chosen approach was to set up a DMS directory where short-lived files are written to act as the signals.

- **Grid jobs are tuned to reduce overheads.** It is a fact of life that submitting processing to a computational Grid incurs overheads: the use of the Grid becomes worthwhile only if the benefits obtained – access to many high-powered processors, massive shared disk space, etc – outweigh the cost of these overheads. In this respect the Oncosimulator work poses something of a challenge, in that it involves neither massive volumes of data nor massive amounts of processing. Concretely, simulating the complete treatment of a tumour on a desktop-sized PC can take as little as 30 seconds, rising to an hour or more only in the most extreme cases of large tumours and fine-grained analysis. A typical simulation carried out within the Oncosimulator validation might take 15 minutes on a local PC, reducing to 10 minutes on a high-performance Grid node. But the overheads – of job submission, GRMS task scheduling, data transfer between the DMS and the execution node, and so on – can easily turn the speed advantage into a deficit.

It is important not to forget that running on the Grid brings ORS users other advantages, including having a visualisation server available to produce sophisticated result visualisations, and the fact that these visualisations and the raw results are available to multiple collaborators across the world. However,

purely in terms of job turnaround we needed to consider how to make the most of the resources available.

One policy has been to batch several runs together into a single job; this reduces the number of jobs submitted, thus saving some of the overhead incurred by each job submission. Since we encode the runs within a job to be executed in series rather than in parallel, this policy also reduces the risk of a Grid node “thrashing” through being assigned too many tasks to execute at once. Another policy is that we make an estimate of the total memory size required for each simulation, and declare this so that the GRMS can assign even the smallest available Grid nodes to run appropriately small simulations.

10.2.4 Early validation results

Based on the above architecture and policies, and various extensions to the ORS itself as described below in the section “OncoRecipeSheet developments since D8.3”, in the past year we have been able to start the validation work proper.

Fig. F4 shows examples of using multiple scenarios on the ORS to gather sets of simulation results so that comparisons can be made between them. The examples illustrate the various forms of combined result display supported by the RecipeSheet: the coloured line graphs overlaid in the time plot at the top right of each sheet; the spatially arranged numeric readouts in the cell at bottom left, with their coloured frames; the 3D tumour displays, again arranged spatially and framed; and the overlaid tumour slices in the sheet at the bottom of the figure, each scenario's slice rendered in a translucent shade of the scenario colour.

The sheet at top right in Fig. F4 shows a result comparison that led us to one particular insight about the simulator. Its results show, in the plots of tumour volume against time, a dramatic discontinuity in tumour response for G0 probability values above 0.34. Although these values (in combination with the default values used for the undisplayed parameters) are known to represent biologically impossible situations, it is valuable to observe how the simulator behaves when driven to these extremes. This case also illustrates the importance of viewing the whole time-sequence of a simulation, rather than just a figure for overall volume change; the final volumes for the two out-of-range cases appear plausible, but the volume plot as a whole shows why they cannot be believed.

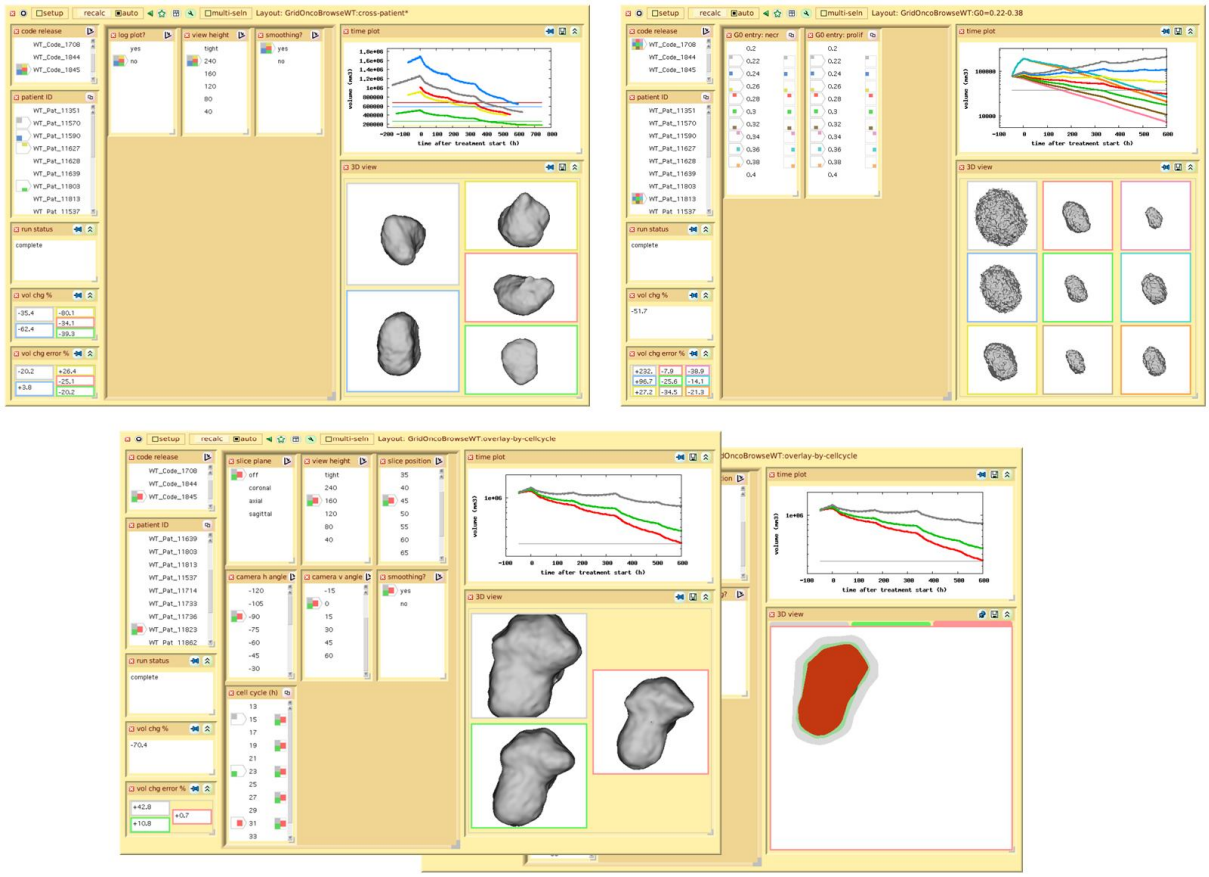


Fig. F4 Three multi-scenario setups of the OncoRecipeSheet. At top left, the user has requested to view the equivalent simulation results for five patients; at top right, the results for a single patient using nine alternative values for probability of cells entering the G0 state. At the bottom is a sheet that has been set up with three results (based on alternative cell-cycle durations), shown in two states: one showing for each case an external 3D view of the tumour after treatment, the other showing colour-coded overlaid slices through the centre of the tumour mass.

Note that setting up and switching between sheet states such as those in Fig. F4 requires no recoding of the sheet: scenarios are set up by choosing the desired parameters from the parameter menus, followed by selection of the desired values within the parameter cells; visualisation choices such as the overlay of the 3D slices are made using switches on the result cells. It is also a crucial property of subjunctive interfaces, and hence of RecipeSheet setups, that all scenarios can be manipulated in parallel. In the ORS this means, for example, that a complex arrangement of scenarios set up for one patient can be applied to another simply by clicking on the desired entry in the patient-selection cell. Similarly, one could add to the sheet any parameter that is currently not displayed, and is therefore being supplied as a default value, and select a different value instead. This facility greatly accelerates the rate at which one can sweep through a region of the result space, checking for equivalent features in a wide range of cases.

10.3 OncoRecipeSheet developments since deliverable D8.3

Here we describe some of the most important additions to the OncoRecipeSheet that have been made in the final stages of the ACGT project, in which it has been extended from an interface for browsing pre-calculated simulation results to a tool for managing the simulator validation process.

10.3.1 A tool for submitting simulation runs

The first requirement for this phase of the project was that users should be able to submit new simulation runs. As mentioned earlier, we wanted this facility to be integrated with the browsing, so that switching from browsing to submission would be as intuitive and painless as possible. To this end, we introduced the rule that if the user selects a combination of parameter values for which no result is yet available, that parameter-value combination is automatically picked up as a candidate for submission to the Grid. It would not make sense to take the next step – of actually submitting the run – without user intervention, since this could result in unintentional submission of large numbers of runs just through experimenting with parameter selections. Therefore the user is given the chance to review the candidates, and to choose which of them to submit. Rather than trying to incorporate this two-phase process (and other submission-related functionality) into the recipe sheet itself, we implemented a new tool, called a Run Manager, shown in Fig. F5.

Fig. F5 Submission and harvesting of simulation runs using the Run Manager. The same window is shown in two states: in the state shown at the top, the manager is part-way through submitting the 13 runs listed in the Submissions view; in the state at the bottom, all runs have been submitted and 7 have already completed. The Candidates view on the left of the manager contains a further 18 runs that the user has specified but not yet chosen to submit. This example illustrates that runs do not necessarily complete in the order in which they were submitted to the Grid; this reflects the impact of batching the runs into jobs, and the general fact that simulations are executed in parallel.

The operation of the Run Manager in conjunction with the result-browsing recipe sheet proceeds as follows:

- **Run specification.** Using the same parameter-setting operations that are used to select results for browsing – including the ability to create multiple scenarios with independent selections – the user sets up parameter-value combinations that are of interest. If there is a Run Manager open on screen (if not, the user can open one at any time), its Candidates view dynamically reflects the latest selection(s) made on a browsing sheet, listing one candidate run for each combination of parameter values on which the simulator has not yet been run.

- **Candidate review.** In the Candidates list, each specified run is listed as a concatenation of just the non-default parameter settings in its specification. For example, if the user has set up a cell for adjusting the number of LIMP stages, but in some scenario the parameter settings include the default LIMP-stages setting of 3, then the run description corresponding to that scenario does not mention LIMP stages. Given that most runs will use the default value for most of the parameters,

this allows for a relatively compact representation. The patient ID and the code version, neither of which has a default value, both appear in every run description.

Each run description also includes an estimate of the amount of memory that will be required for its execution. Since this depends primarily on the number of geometrical cells that will be allocated in the 3D simulation space, and calculated in each time step, the memory requirement is also a useful indicator of the likely overall execution time. In our validation work to date, these memory estimates have ranged from a few tens of MB up to over 10GB.

• **Candidate selection.** The user makes selections in the Candidates list and presses the "submit selected" button to pass them to the next stage (or just presses "submit all" to accept all candidates). The chosen runs are then added to the end of the Submissions list, to await submission.

• **Run submission and harvesting.** There are "Stop" and "Go" buttons to enable final control over when the runs in the Submissions list are processed. If the user has pressed "Go", the Run Manager submits runs in the order in which they appear in the list. Once submitted, a run's status is checked periodically (about once a minute) by the Run Manager, and status updates are shown in the list. The possible status conditions are as follows:

➢ **(no status)** The run is just waiting to be processed

➢ **submitting** Submission is in progress.

➢ **submitted as [run ID]** The run has been given the ID shown, and submitted to the Grid. For each submitted run an empty file is set up on the DMS; the simulation job writes to this file to indicate that execution has finished – which, as described below, does not guarantee successful completion. The Run Manager watches for the change of file status from "empty" to "permanent" (the DMS ensures that the file cannot be read until its contents are complete), then downloads the file and analyses its contents to detect the success or otherwise of the simulation. This procedure is referred to as harvesting the run, and leads to one of the states below.

➢ **completed as [run ID]** Simulation finished, and the results appear to be complete. A few minutes after this status appears, the run's results will be available for visualisation.

➢ **partial as [run ID]** Simulation was started, and produced results corresponding to at least two days' elapsed time, but then stopped for some reason – for example, because the tumour overflowed the bounds of the allocated simulation volume.

➢ **failed as [run ID]** For some reason, the simulation attempt failed before producing even two days' output. This can indicate a premature exit from the simulator, or it can be due to a problem with some aspect of the Grid services – such as failing to authenticate the user (for example, because the security service is offline), or failure of the assigned Grid node to access the files on the DMS.

At the time of writing, we are working with ICCS to expand this status information to indicate the reasons for failures (and partial completions) that are due to premature exit from the simulator. Each of the simulator's various exit points has been associated with an exit code, which will in future be shown as part of the run status.

Diagnosing failure caused by a problem with a Grid service is more complex. At this point we have not attempted to provide automatic diagnosis. If repeated attempts to re-submit the run (see below) are also unsuccessful, the user should contact the administrators of the Grid.

- **Re-submission.** When a run fails, the user has the option of re-submitting it. If the failure was due to a temporary fault with the Grid services, re-submission might lead to successful completion.

Note: The alert reader will have noticed something surprising about the candidate list shown in Fig. F5: it holds 18 runs, even though the ORS supposedly supports a maximum of 12 scenarios. The limit on scenarios is indeed 12, but the RecipeSheet includes a switch that enables each parameter cell to support multiple selections within a single scenario. The recipe implemented for interpreting Oncosimulator parameter selections responds to such cases of multi-selection by generating multiple simulation-run specifications, using every selected value in combination with every other. For example, a user could make two selections in the *patient ID* cell and three selections in *cell cycle time*; this would create six run specifications. A multitude of specifications has no meaning in terms of a request to browse results (so the result-display cells go blank), but is meaningful as a request to generate new candidate runs. Thus a user can specify tens or hundreds of runs with a single scenario – and is also free to use multiple scenarios, whose candidate-run sets are then merged (with duplicates being removed automatically) by the Run Manager.

10.3.2 Support for working with large numbers of parameters

The initial demonstrations of using recipe sheets to browse simulation results used fixed setups involving cells for adjusting just a small number of parameters. The versions of the simulator code that are now being validated have between 35 and 40 parameters, and validation requires confirming that every parameter is having the intended effect. Given such a large number of parameters, from a simple consideration of screen display space it is not practical for a user to view and adjust all parameters at the same time. We therefore developed a means for the user to choose dynamically which parameters should be shown.

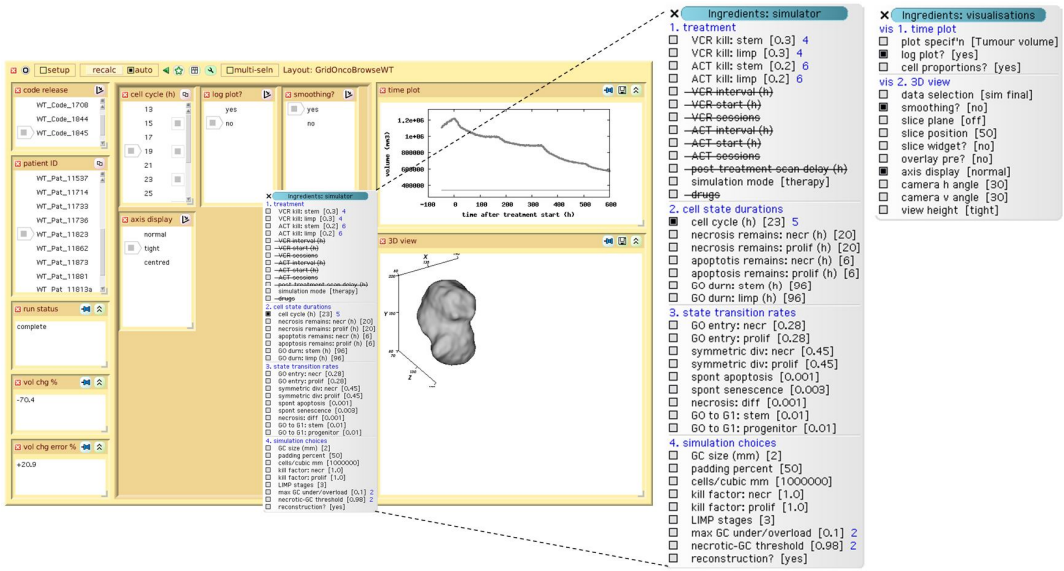


Fig. 6: A sheet being used to browse a single simulation result, and showing the pop-up (but pinnable) menus that let the user choose which parameters to display on the sheet.

The dark-brown central region of the sheets seen in Fig. F4 and Fig. 6 is a dynamic-cell zone, for holding the particular set of parameter cells that the user has chosen to work with. The user chooses the desired parameters by using the pop-up menus shown in Fig. 6: one menu is for parameters to the simulation itself, the other for parameters determining how results are visualised. Each parameter that is not on display takes its default value.

Unlike the manual placement of cells on the normal substrate of a RecipeSheet, the dynamic-cell zone attempts to lay out cells automatically in a space-efficient arrangement. This layout is updated on the fly as the user adds, deletes or resizes cells within the zone. A layout zone of the size shown in these figures can hold 8 to 12 parameters; this is plenty for the early validation stages, in which the simulator developers need first to confirm that changing one or two parameters at a time while keeping all others fixed has the expected impact on the simulation results. On a larger screen the layout zone could be expanded by the user, thus enabling more parameter cells to be on view at the same time.

Users are also able to customise for themselves, using XML-based configuration files in the RecipeSheet's working directory, the contents of the menus and the ranges of values offered for each simulation parameter. For example, in the standard configuration that we have distributed up to now the range of cell-cycle durations is set to use increments of 2 hours; a user wanting to explore the effects of 1-hour differences could edit the XML to set up the desired range. These configuration files are specific to each code release, since a new release can potentially involve the addition or removal of simulation parameters. When a new release is made, the users create a new configuration file, typically by copying and editing the file for the previous release.

A second issue when working with a large number of parameters is that of managing the overall exploration of the parameter space. The Oncosimulator validation task involves up to 40 parameters, many having at least 10 settings that are potentially worth trying; without some form of assistance, it would be extremely laborious for a user to try to keep track of which of the vast number of possible parameter-value combinations have been evaluated so far. We therefore implemented a form of guidance, using annotation of the parameter menu and of the parameter cells.

In the parameter menu, each entry shows in square brackets the default value that the parameter will take if not displayed as a cell. In the case of parameters representing facts about an individual patient (such as the duration of actual treatment received), the item is crossed out; this reminds the user that the appropriate default value will be used for whichever patient is selected. Each menu item also shows a count of how many distinct values of that parameter have been used in simulations that are compatible with the parameter values currently selected on the sheet. For example, in Fig. 6 the default value for “VCR kill: stem” is 0.3, and among the runs performed with the currently selected code release, patient ID and cell-cycle time, four values have been used for this parameter (if the count is 1, it is not displayed).

When a parameter is on view as a cell, its individual values include markup that is updated dynamically to show which values are compatible with the other parameter selections the user has made. A sequence of changes to this markup is shown in the series of pictures in

Fig. F7. When no selection has been made in either cell, the cells show all values that have been used with the patient and code version that are currently selected. When the user selects 0.3 for VCR, the markup for ACT changes to show only those values that are still compatible; in this case, evidently there are no results that combine an ACT setting of 0.15 with the VCR of 0.3. Likewise, when the user selects 0.2 for ACT, the change in markup in the VCR cell reveals which of the VCR values have been run with that value for ACT.

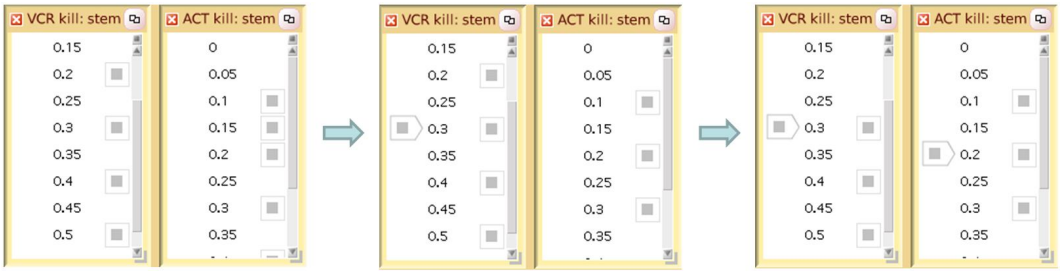


Fig. F7 A sequence of views of two parameter cells, including the markup that shows where simulation results are available, as the user progressively makes selections.

However, in early tests it was found that this markup mechanism was less helpful than expected. In particular, the disappearance of the markup as soon as the user makes a few selections made it hard to keep track of the available values. Fig. F8

shows an updated design in which, in addition to the previous markup, each cell shows “hints” of other values that are compatible with the chosen patient and code release, but not with the selections made on other parameters. In this figure the user has created two scenarios (for different patients) and has selected a value for just the parameter controlling VCR kill ratio for stem cells. The cell for VCR kill: limp shows only a single compatible value with that selection, and the cell for ACT shows three values available in the grey scenario and one available for the blue. But the additional hint markers – smaller than the main markers, and somewhat faded – show that there are not just three but six ACT values that could be chosen for the grey-scenario patient, and three values for the blue scenario. This updated design is now being evaluated.

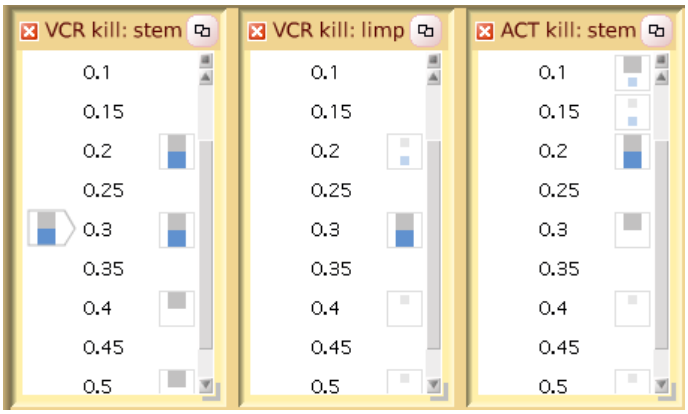


Fig. F8 Updated design of the parameter-value markup, which now includes “hints” (the markers with faded, smaller colour regions) on values that are compatible with the non-defaultable parameter values – patient ID and code version – even if not with the other parameter choices.

It should be noted that we are by no means the first developers to encounter the challenge of assisting users in making non-empty queries in a sparse result space. One early demonstration of marking up an input widget to show which values would correspond to available results is Eick's data visualisation sliders [2]; a more recent design, dealing explicitly with multiple discrete-valued inputs, is Backward Highlighting [18]. The mechanism shown here is similar in presentation to the latter, though the RecipeSheet has to deal with the additional complexity of the user potentially setting up multiple mutually independent scenarios on the same sheet.

10.3.3 Support for downloading data files to local PC

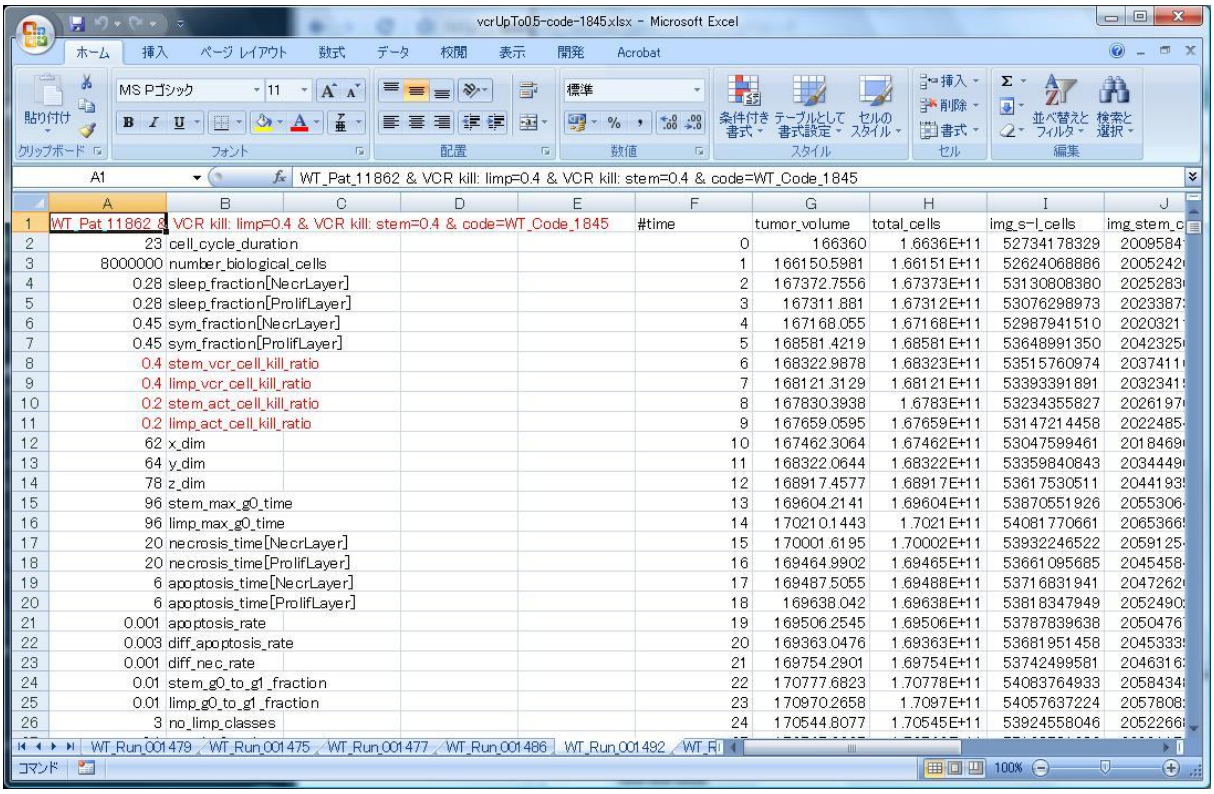
ICCS researchers commented that although the time plots in the form displayed on the ORS provide a reasonable overview of how a simulation proceeded, there are times when they would like to examine the numerical data in detail, or to produce charts (e.g., for publication) that have different styles, resolution etc. They asked whether the ORS could be extended to support this work, for example by making the charts editable as they are in Microsoft Excel. They also specified that they sometimes need to work with other simulation-result files, including the raw 3D data from which tumour images can be produced.

Extending the ORS to support Excel-style editable charts would have been a major undertaking, beyond the scope of the ACGT project. However, providing a facility to download data to the local PC, where it could then be manipulated using Excel or other tools, was clearly more feasible. There was already a simple Web service at UvA for requesting some elements of simulation-result data, so we designed an extension to that service to allow the downloading of various result files, requested by name. The ORS uses this service as the basis for two new facilities. One is to let the user download to the local PC a set of result files (currently hard-coded) for each of the simulation runs currently on view; the other is to create a Microsoft Excel workbook containing just the tumour-evolution data for those runs.

The basic file-download facility is accessed through a button on the ORS time-plot cell. At present the files downloaded include the log of supplied parameters, the tumour-evolution data file (which holds the counts of all cell types at each hourly time step), the initial and final tumour shape, and the standard-output and standard-error logs from the simulator. The files are saved in a dedicated “oncosim data” subdirectory of the ORS’s working directory, with names that are prefixed by the IDs of the runs to which they correspond: for example, WT_Run_001234_oncosim-parameters.log.

The facility to build an Excel workbook based on the tumour-evolution data uses the same file-download service to obtain for each run just the parameter log and the tumour-evolution data. It writes a single sheet for each run (up to 12 sheets, the maximum number of scenarios that can be viewed simultaneously). An example is shown in

Fig. F9. As seen in the figure, each sheet contains the parameter values and the tumour evolution data; the non-default parameter values used to define the run on that sheet are shown in cell A1, and entries in the parameter list that differ between the sheets in this workbook are highlighted in red.



	A	B	C	D	E	F	G	H	I	J
1	WT_Pat_11862 &	VCR kill: limp=0.4 & VCR kill: stem=0.4 & code=WT_Code_1845				#time	tumor_volume	total_cells	img_s-l_cells	img_stem_c
2	23	cell_cycle_duration				0	166360	1.6636E+11	52734178329	2009584
3	8000000	number_biological_cells				1	166150.5981	1.66151E+11	52624068886	2005242
4	0.28	sleep_fraction[NecrLayer]				2	167372.7556	1.67373E+11	53130808380	2025283
5	0.28	sleep_fraction[ProlifLayer]				3	167311.881	1.67312E+11	53076298973	2023387
6	0.45	sym_fraction[NecrLayer]				4	167168.055	1.67168E+11	52987941510	2020321
7	0.45	sym_fraction[ProlifLayer]				5	168581.4219	1.68581E+11	53648991350	2042325
8	0.4	stem_vcr_cell_kill_ratio				6	168322.9878	1.68323E+11	53515760974	2037411
9	0.4	limp_vcr_cell_kill_ratio				7	168121.3129	1.68121E+11	53393391891	2032341
10	0.2	stem_act_cell_kill_ratio				8	167830.3938	1.6783E+11	53234355827	2026197
11	0.2	limp_act_cell_kill_ratio				9	167659.0595	1.67659E+11	53147214458	2022485
12	62	x_dim				10	167462.3064	1.67462E+11	53047599461	2018469
13	64	y_dim				11	168322.0644	1.68322E+11	53359840843	2034449
14	78	z_dim				12	168917.4577	1.68917E+11	53617530511	2044193
15	96	stem_max_g0_time				13	169604.2141	1.69604E+11	53870551926	2055306
16	96	limp_max_g0_time				14	170210.1443	1.7021E+11	54081770661	2065366
17	20	necrosis_time[NecrLayer]				15	170001.6195	1.70002E+11	53932246522	2059125
18	20	necrosis_time[ProlifLayer]				16	169464.9902	1.69465E+11	53661095685	2045458
19	6	apoptosis_time[NecrLayer]				17	169487.5055	1.69488E+11	53716831941	2047262
20	6	apoptosis_time[ProlifLayer]				18	169638.042	1.69638E+11	53818347949	2052490
21	0.001	apoptosis_rate				19	169506.2545	1.69506E+11	53787839638	2050476
22	0.003	diff_apoptosis_rate				20	169363.0476	1.69363E+11	53681951458	2045333
23	0.001	diff_nec_rate				21	169754.2901	1.69754E+11	53742499581	2046316
24	0.01	stem_g0_to_g1_fraction				22	170777.6823	1.70778E+11	54083764933	2058434
25	0.01	limp_g0_to_g1_fraction				23	170970.2658	1.7097E+11	54057637224	2057808
26	3	no_limp_classes				24	170544.8077	1.70545E+11	53924558046	2052266

Fig. F9 A Microsoft Excel workbook built by the ORS from multiple scenarios being viewed by the user. Each scenario is written onto one sheet, with parameter values listed in the first two columns and the tumour-evolution data starting from column six. Parameter values that differ between the sheets are highlighted in red. In the example seen here, the saved runs have various values for the kill ratios of both VCR and ACT, but the definition of the run being viewed includes only the VCR kill values since it uses the default values for ACT.

10.4 Findings and future work

We installed a preliminary version of the ORS at UvA and ICCS in the summer of 2009, and have distributed new releases every few months as the facilities of the ORS and of the services on which it depends have been upgraded. Between these sites and UHok we have so far gathered roughly 3,500 simulation results for the two cancer types being examined. Although in general the simulator has proven to be robust and predictable across a broad range of parameter settings, the ability to run and to inspect large numbers of simulations has already helped us to uncover some unexpected behaviours. One problem in the BRCA code was found when we noticed that simulations run on the Grid behaved differently from those run on a local PC; it turned out that parameter settings for the PC runs always followed a pattern that was hiding a bug in the code. In the WT simulator we found that results differed depending on which Grid node happened to be used for the execution; this led to the discovery of another previously undetected but decidedly harmful bug.

These can be considered early successes in the validation work; there is still much to be done in testing the influence of all parameters, and in confirming that the simulator responds appropriately whether the tumours and treatment plans it is fed are realistic or stretch the limits of its applicability. We plan to continue to work with

the team at ICCS to ensure that they can see this crucial phase in Oncosimulator development through to a satisfactory conclusion.

There are three areas in which we feel that further work is needed in supporting the validation efforts: faster simulation processing, richer result visualisations, and improved exploration support. We now briefly address each of these areas in turn.

10.4.1 Faster simulation processing

The generation of simulation results over the past year has constituted a useful stress-test of the ACGT Grid infrastructure, bringing to light many issues that we were then able to resolve, thus improving the reliability and performance of the overall system. However, because of the issue that simulation runs are relatively short compared to the overheads imposed by use of the Grid, the benefits gained from using the Grid are not as impressive as we would like. We are working to reduce those overheads.

One form of run-time overhead that we would like to reduce is that of creating the DMS files for holding the results from each simulation. At present this preparatory work, carried out before submitting a GRMS job, takes tens of seconds for each run (in fact, running from Japan it typically takes a full minute). Rather than creating the files on demand, it should be possible to create a pool of DMS file nodes in advance (including setting the appropriate permissions), and just move and rename them to reflect how they end up being used. If the moving and renaming can be carried out in parallel with the execution, i.e., after the job has been submitted, then the submission itself will become almost instantaneous.

A second potential way to reduce overhead is to batch multiple simulations into a single execution. Rather than running just a single simulation within a GRMS task, it should be possible to run a script that processes a batch of independent simulations in rapid sequence. In general the use of scripts is counter to Grid philosophy, since the GRMS services have no way of telling what resources a script will end up using, and therefore of balancing this against other resource demands. However, taking the point of view of a user who considers a set of, say, 20 or 30 simulations as a single heavyweight processing task, this kind of batch processing could nonetheless be argued as a valid use of Grid resources.

At present a batch of 100 runs takes roughly four hours to process. Our goal is that at least for small simulations we should be able to process such a batch, from submission to viewable results, within 30 minutes.

10.4.2 Richer result visualisations

In terms of the visualisations that can be requested, especially when merging multiple results, there is plenty of scope for more advanced techniques than those demonstrated on the ORS to date. For example, to help the Oncosimulator developers find the significant differences between 3D tumour shapes produced under different simulation conditions it may be profitable to apply techniques for highlighting differences between such structures, such as those surveyed by Pang et al. [15].

The ORS now includes an experimental mechanism for creating merged views of tumour slices in the 3D view. This mechanism requires some improvements to make it more usable, and also to find an appropriate way of overlaying the various scenarios' contributions so that their individual outlines and the differences

between them are easily discernable. One technique for such overlay that we are interested in trying was presented recently by Luboschik et al. [7].

The ICCS researchers have suggested that we implement a facility for helping them to create “videos” – in other words, animated sequences – of tumour views. For example, they might want to show the day-to-day change in a tumour over the course of a treatment, or rotate a 3D view through 360° to clarify the overall tumour shape. This is not the same as being able to step quickly through a range of visualisation-parameter settings on the ORS: first, the overheads in recalculating the RecipeSheet make it difficult to support smooth stepping; second, the researchers want to be able to “take away” the animations, for example to show them when giving talks.

UHok and UvA have produced an initial design for such a facility, taking into account trade-offs related to, for example, where the animation should be created. The basic choice is between creating the animation in its final form at the UvA server, or having the ORS create the animation on the local computer by obtaining individual images from the server and then stitching them together. The former is likely to be more time-efficient (both through doing the work on a high-performance graphics engine, and in needing to download only an efficiently compressed video file), whereas the latter would give more flexibility in terms of making comparisons between simulation scenarios. In particular, as reported above we are working to support composite views created by overlaying tumour-slice outlines from multiple scenarios; if responsibility for creating animations rests with the ORS, it could also create animations of these composite views. By contrast, it would be unrealistic to expect a general visualisation server to support the creation of such application-specific compositions. The current design allows for both approaches.

UvA have implemented a prototype service that allows a client to request an animation by specifying the data set (run ID), the parameters to hold fixed, and one parameter to vary – for example, fixing a timepoint and view direction, then varying the angle of rotation. The client can also select whether to receive the results as a compressed archive containing PNG images, or as a playable Windows Media Video. We are now testing these facilities.

10.4.3 Improving exploration support

There are many potentially valuable directions for extending the RecipeSheet to provide more advanced support for exploring simulation results. One would be to migrate from a fully manual, user-driven interface towards one in which the system provides some level of assistance in the choice of parameter combinations. Even without abdicating investigation responsibility to the computer, we can at least have it reduce the burden experienced by a user in making a suitably thorough coverage of a result space. A first step would be to introduce parameter-setting operations at a higher level of abstraction than the current requirement to click each specific parameter value; one illustration of how this might be done is given by the *cogito* [4], in which the user can specify the next set of cases to examine by clicking directly on interesting results within the current set. A further step could be to apply some heuristics to select sets of available results that are in some way important for the user to check by eye. One way to tackle this would be using a mechanism similar to that underlying Design Galleries [14], which works on the basis of measures for determining how similar results are to each other. However, in contrast to the Design Galleries goal of finding results that, according to the supplied measures, cover as large a range of divergence as possible, in the Oncosimulator studies it is perhaps more valuable to enlist help in finding results

that – despite being based on broadly different parameter settings – turn out to be effectively indistinguishable. This might reveal unexpected interactions between parameters, and could help to arrive at confidence levels for use in the future work of "tuning" the simulator results to match those of actual patient cases.

10.5 Conclusion

These are still very early days in a process towards development of an Oncosimulator that might one day be used in clinical practice. We have reported the beginnings of the work of validating the simulator code, by testing its response to parameter variation. We believe that being able to obtain and compare results from multiple alternative simulations, using facilities like those of the ORS, has been shown to be useful for the validation and could also play a useful role in future phases of Oncosimulator testing and investigation. For example, it is hoped that one day the simulator will reach a stage where it makes sense to try to "tune" it so as to match the results from clinical treatments. Given that the results from the existing simulator show that apparent increases or decreases in therapy effectiveness can be induced by variation in a number of the model's parameters, independently or in concert, the key to such tuning will be to back it up with realistic confidence levels. Being able to run multiple simulations will help in exploring the diverse parameter-value combinations that would all lead to similar predictions, and thus in avoiding premature over-confidence regarding the ability of the simulator to predict the course of treatments on previously unseen patients. Further in the future, an interface that shows a selection of the diverse treatment outcomes that would be predicted in the face of typical (and unavoidable) inaccuracies in patient-data measurements is likely to be more credible than a 'black box' that produces a single prediction result. Such a tool would also help in educating clinicians about what tools like the Oncosimulator can tell us... and what they can't.

10.6 References

1. E. H. Chi, J. Riedl, P. Barry, and J. Konstan. Principles for Information Visualization Spreadsheets. *IEEE Computer Graphics and Applications* **18**(4), Jul/Aug 1998, 30-38. IEEE Press, 1998.

2. S. Eick. Data Visualization Sliders. *Proceedings of ACM UIST '94*, 119--120. ACM Press, 1994.

3. N. Graf, A. Hoppe, E. Georgiadi, R. Belleman, C. Desmedt, D. Dionysiou, M. Erdt, J. Jacques, E. Kolokotroni, A. Lunzer, M. Tsiknakis, G. Stamatakos. In Silico Oncology for Clinical Decision Making in the Context of Nephroblastoma. *Klinische Pädiatrie* 221:141-149. 2009.

4. D. H. Hepting. Interactive evolution for systematic exploration of a parameter space. *Proceedings of ANNIE 2003*, 125-131. American Society of Mechanical Engineers, 2003.

5. T. J. Jankun-Kelly, O. Kreylos, K. Ma, B. Hamann, K. I. Joy, J. Shalf, and E. W. Bethel. Deploying Web-Based Visual Exploration Tools on the Grid. *IEEE Computer Graphics and Applications* **23**(2), Mar 2003, 40-50. IEEE Press, 2003.

6. T. J. Jankun-Kelly and Kwan-Liu Ma. Visualization Exploration and Encapsulation via a Spreadsheet-Like Interface. *IEEE Transactions on Visualization and Computer Graphics* **7**(3):275-287. IEEE Press, 2001.

7. M. Luboschik, A. Radloff, and H. Schumann. A new weaving technique for handling overlapping regions. *Proceedings of AVI 2010*, 25-32. ACM Press, 2010.

8. A. Lunzer. Choice and comparison where the user wants them: Subjunctive interfaces for computer-supported exploration. *Proceedings of the IFIP TC. 13 International Conference on Human-Computer Interaction (INTERACT '99)*, 474-482. IOS Press, 1999.

9. A. Lunzer, R. Belleman, P. Melis, J. Pukacki, P. Spychała, and G. Stamatakos. Validating the ACGT Oncosimulator with a Grid-Supported Visualisation Environment. To appear in *Proceedings of 4th International Advanced Research Workshop on In Silico Oncology and Cancer Investigation*, Athens, Greece, Sept 2010.

10. A. Lunzer, R. Belleman, P. Melis, and G. Stamatakos. Preparing, Exploring and Comparing Cancer Simulation Results Within a Large Parameter Space. To appear in *Proceedings of 3rd International Symposium on Information Visualization in Biomedical Informatics (IVBI)*, London, UK, Jul 2010.

11. A. Lunzer and J. Fujima. Building and Exploring with the RecipeSheet. *Proceedings of the 7th International Conference on Creating, Connecting and Collaborating through Computing (C5 2009)*, Kyoto, Jan 2009. 41-47. IEEE Press, 2009.

12. A. Lunzer and K. Hornbæk. RecipeSheet: Creating, Combining and Controlling Information Processors. *Proceedings of the 19th Annual ACM Symposium on User interface Software and Technology (UIST '06)*, Montreux, Switzerland, Oct 2006, 145-153. ACM Press, 2006.

13. A. Lunzer and K. Hornbæk. Subjunctive interfaces: Extending applications to support parallel setup, viewing and control of alternative scenarios. *ACM Transactions on Computer-Human Interaction* **14**(4), January 2008.

14. J. Marks, B. Andalman, P. A. Beardsley, W. Freeman, S. Gibson, J. Hodgins, T. Kang, B. Mirtich, H. Pfister, W. Ruml, K. Ryall, J. Seims, and S. Shieber. Design Galleries: A General Approach to Setting Parameters for Computer Graphics and Animation. *Proceedings of SIGGRAPH 97*, Los Angeles, CA, USA, 389-400. ACM Press, 1997.

15. A. T. Pang, C. M. Wittenbrink, and S. K. Lodha. Approaches to Uncertainty Visualization. *The Visual Computer* **13**, 370-390. 1996.

16. Georgios S Stamatakos, Vassilis P Antipas, and Nikolaos K Uzunoglu. A spatiotemporal, patient individualized simulation model of solid tumour response to chemotherapy in vivo: the paradigm of glioblastoma multiforme treated by temozolomide. *IEEE Trans Biomed Eng.* **53**(8):1467-77. August 2006

17. L. Stein. How Perl Saved the Human Genome Project. *The Perl Journal* **1**(2), Summer 1996. Available from http://bioperl.org/wiki/How_Perl_saved_human_genome

18. M. L. Wilson, P. Andre, and mc schraefel. Backward highlighting: enhancing faceted search. *Proceedings of the 21st Annual ACM Symposium on User interface Software and Technology (UIST '08),* Monterey, CA, USA, Oct 2008, 235-238. ACM Press, 2008.

19. A. J. Winfield. A Virtual Laboratory Notebook for Simulation Models. *Proceedings of Pacific Symposium on Biocomputing 1998*. 177-188. World Scientific, 1998.

20. H. Wright, K. Brodlie, and M. Brown. The dataflow visualization pipeline as a problem solving environment. *Proceedings of the Eurographics Workshop on Virtual Environments and Scientific Visualization '96,* Monte Carlo, Monaco. 267-276. Springer-Verlag, 1996.

11. The ACGT Oncosimulator as an ACGT-compliant grid service

(Code letter: A)

The Oncosimulator can be integrated into the ACGT environment as an ACGT-compliant Grid Service that can be integrated into a high-level ACGT workflow. This section gives a brief overview of this Grid Service, outlines its design rationale, and introduces some exemplary exercises that demonstrate the potential of integration the Oncosimulator approach to cancer research and treatment with the traditional bio-statistical approaches that prevail in the ACGT workflow environment.

As a standalone application, the Oncosimulator is a complex software suite that can be parameterized to a set of different tasks and application cases (i.e., cancer types), by a parameter file. This basic architecture bears similarity to the ACGT services of GridR and command-line analysis services (see Deliverable D6.6), which are also based on parameterizing some generic software (such as the R statistical toolkit) by a complex parameter file (e.g., a script written in the R language) to solve a specific data analysis tasks, thereby defining a specific Grid service with static inputs and outputs, that corresponds to the execution of the parameter file. Following the experience of setting up these services, the Oncosimulator software is easily integrated into the ACGT environment (Fig. A1) by the following steps

1. Selection of a specific Oncosimulator use case (e.g, by selecting the simulation target, the corresponding parameters, and the type of expected input and output)
2. Setup of a parameter script that executes the selected use case
3. Definition of an ACGT service from a template that

a. Has inputs and outputs in ACGT-compliant data types that correspond to the selected input and output types (usually references to files in the data management system DMS)
b. Interfaces with the DMS for data input and output to other ACGT services
c. Executes the generic Oncosimulator standalone application on basis of the specific parameter script

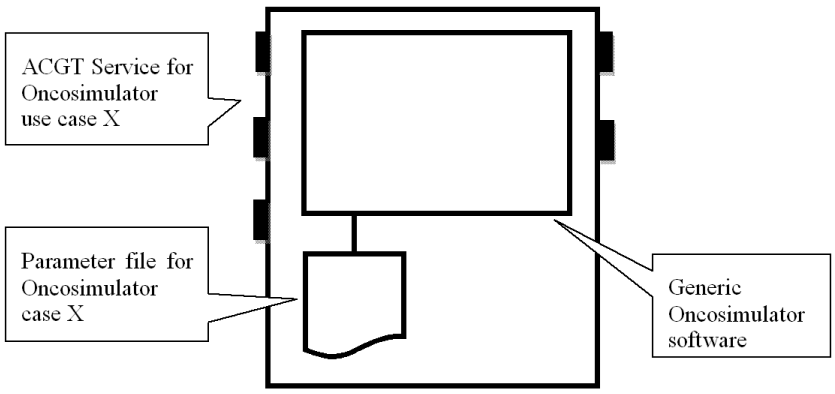


Fig. A1 Integration of the Oncosimulator software into the ACGT environment

11.1 Use Cases of an Integrated Workflow

In the following, we will briefly present some potential use cases that demonstrate the possibilities of integration of the Oncosimulator approach with the traditional bio-statistical approaches.

11.1.1 Black-box evaluation of the Oncosimulator

In order to prepare a future use of the Oncosimulator as a predictive tool in clinical studies, it must be subjected to a strict statistical evaluation. In addition to the white-box evaluation of the Oncosimulator during its development which strives to evaluate the scientific correctness of the underlying model, i.e. the inspection of the relation of the Oncosimulator model to the real biomedical processes that it represents, it is also necessary to evaluate the Oncosimulator in a black-box setting, i.e. checking only the correctness of its inputs and outputs on independent patient data. In medical statistics, many standard statistical tests and measures are defined for the proper testing of a new procedure, and most of these are readily available in the R toolkit, thus allowing for easy testing of the predictive performance of the Oncosimulator.

11.1.2 The Oncosimulator as a Data-Enriching Tool

The success of knowledge discovery approaches and decision support on biomedical data very much depends on the availability of meaningful, informative data. The Oncosimulator can be used as a tool to enrich available data with high-level knowledge in the form of simulations results. These information can be added to the data that is already available for a patient and can significantly contribute to

more complete predictions and recommendations based on statistical and data mining algorithms.

12. A scenario for the future immunological extension of the Oncosimulator

(Code letter: M)

12. 1. Introduction

The interplay between tumour and immune system (IS) which takes place because tumour cells are characterized by a vast number of genetic and epigenetic events leading to the appearance of specific antigens, called neoantigens, triggering antitumoural actions by the IS [1]. Indeed transformed cells that acquired such an antigenicity are recognized as nonself. These observations provided an experimental theoretical basis [2] to the old empirical hypothesis of immune surveillance, i.e. that the IS may act to eliminate tumours [3]. The story of cancer immunobiology is, in fact, very old, but only in recent years, thanks to new molecular techniques (and to large epidemiological studies) a sufficient amount of evidence has been accumulated in favor of this hypothesis [4].

The competitive interaction between tumour cells (TCs) and the IS, involves a considerable number of events and molecules, and as such is extremely complex and, as a consequence, the IS is not able to eliminate a neoplasm in all cases, which may escape from IS control. Of course, a dynamic equilibrium may also be established, such that the tumour may survive in a microscopic steady state (MISS), which is undetectable by diagnostic equipment [5]. However, consider a tumour which is constrained by the IS in a MISS. Over a long period of time (a significant fraction of the mean life span in men, according to [4]), the neoplasm may develop multiple strategies to circumvent the action of the IS [1,4,6], which, in the long term, may allow it to evade immune surveillance and to re-commence growing to its carrying capacity [5]. The tumour has adapted itself to survive in a hostile environment, in which antitumour immune response is activated. In other words the immunogenic phenotype of the tumour is "sculpted" [4] by the interaction with the host's IS. For this reason, the theory of IS-T interaction has been called immunoediting theory by Dunn et al. [4] .

An interesting therapeutic approach is immunotherapy [7,8], consisting in stimulating the IS in order to better fight, and hopefully eradicate, a cancer. The basic idea of immunotherapy is simple and promising, but the results obtained in medical investigations are globally controversial [9,10,11], even if in recent years there has been evident progress. The various forms of immunotherapy fall into three main categories: monoclonal antibodies, immune response modifiers, and vaccines. However, similarly to what is observed in chemotherapy, it is likely that the best strategy to combat cancer may require multiple immunotherapeutic strategies.

From a theoretical point of view, a large body of research has been devoted to mathematical models of cancer-immune system interactions and related therapies [12-26,28-29]. However, at the best of our knowledge, no effective multiscale and clinically validable model has been proposed.

12. 2 The underlying hybrid biophysical meta-model

In order to extend the oncosimulator, by following and improving recent literature on multi-scale biophysical models of tumour grwoth and antitumour therapies, we propose here a hybrid modelization approach.

Indeed, in recent literature, a number of proposed multiple-scale models are composed by two different but inter-communicating modules (by adopting rather informatics therminology). The first is represented by a detailed stochastic model, whose spatial scale arrives to the individual agent, for example tumour cells represented as a cellular automaton. This stochastic individual based model is then interlinked with a module coordinating a series of mean field models (ODE, PDE) to describe other important processes that are different from the cell proliferation.

Many examples of models implementing such approach were published in recent Systems Biomedicine literature.

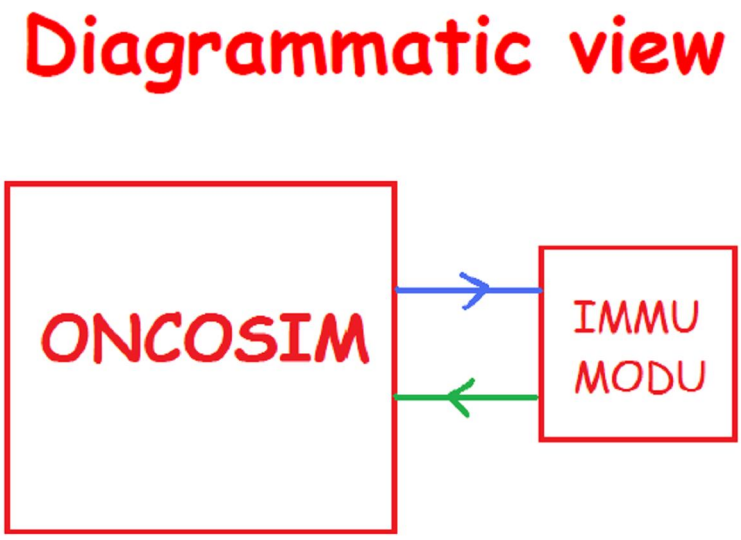


Fig. M1. Diagrammatic view of the communications between the Oncosimulator and its IMMUnologic module

In defining a module for the inclusion in the oncosimulator of the dynamical effects of immune systems in tumours, we shall follows this approach, which has a number of important advantages:

- It is no necessary to intensively change the core code of the Oncosimulator, since minimal modification are needed to be added, which essentially involve three process of bidirectional communication from the core of the oncosimulator to the IMMU (-nologic) module, and vice versa (Fig.M1).
- The IMMU code is relatively simple in comparison with he more detailed computational engine of the oncosimulator

As far as the modelization of the dynamics of the immune system, we decided to implement the module as a sort of meta-module where the user may define and test (in a well pre-established framework) the system he/she elaborated, or he/she may choose a model in an embedded thesaurus of models stored in the module.

We remark here that, any case, the resulting multi-scale model will be deeply different from the original single-scale considered model since here the dynamics

of the tumour is described by the oncosimulator, which is inherently multi-scale and extremely realistic.

12.3 Synthetic description of the module and of its interaction with the main core module of the Oncosimulator

The flux of information can be summarized as follows (Fig. M2):

- The Oncosimulator [30] has to transmit to the IMMU module a series of informations (biological nature and, in case, also of geometric nature) on thetumour state
- These information are processed in the IMMU module
- The IMMU module send to the Oncosimulator a changed set of parameters

In other words, the IMMU module is invisible for the Oncosimulator, since the role of the former is limited to dynamically update the values of a well-defined subset of parameters used by the oncosimulator.

In turn, the IMMU module will be composed of a series of sub-modules that will be interlinked by following a well defined skeleton, but which will allow the user to implement a specific model of the evolution of the immune system reaction to the tumour.

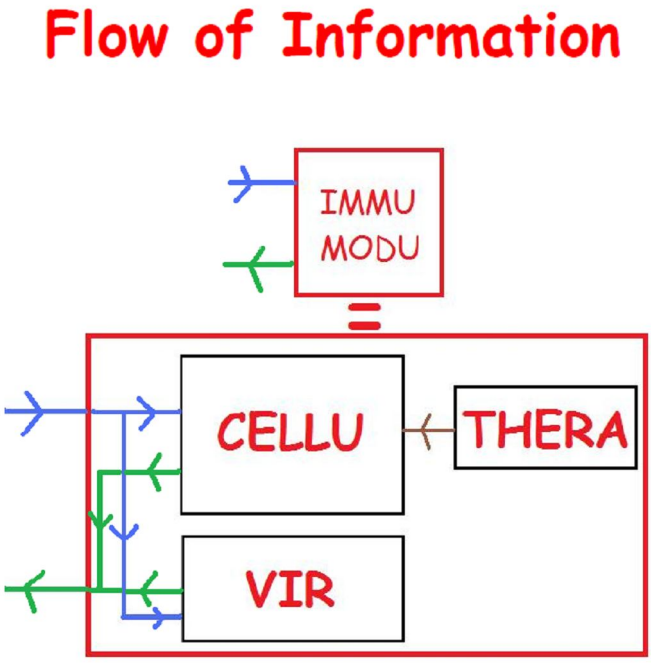


Fig.M2 The suggested immunologic module and its interaction with the main core module of the ACGT Oncosimulator

Namely, the names and functions of these internal modules will be the following:

- CELLU: This unit will be devoted to the management of the dynamics of the various kind of immune system cells (natural killers, antigen presenting cells) as well as of the proteins produced by them, such as interleukins etc.

- THERA: this component will manage the dynamics of the time-varying immunotherapies (see the remark after this list)

- VIR: this module processes the evolution of the viruses that may be used to enhance the visibility of tumour cells by the immune system

We remark here that we preferred to reserve the term Immunotherapy to classical immune-system based anti-tumour therapies. This separation of terminology has also been adopted since immunotherapetic acts only have an indirect effect (through immune system improvement) on the tumour, whereas viruses have a more or less intense onco-lytic activity.

The flow of communications that is internal to the IMMU module is structured as follows:

- The THERA unit sends information to the CELLU unit

- CELLU and VIR exchange information

- Both CELLU and VIR send information tot he "external world"

As far as the details of flow of information from the Oncosimulator main core to this module are concerned, it may be summarized as follows:

- At each of its updates OS_MAIN calculates the total Number X(t) of tumour cells and communicate it to IMMU. X may be, depending upon the request of the specific user-defined model of the Immune system, either a scalar quantity or a vector of the various kinds of cells present in the tumour in study, e.g. Stem cells, progenitors, mature cells, quiescent cells etc.

- Values of the apotptosis and necrosis rates (with teference to the code: apoptosis_rate, diff_apopt_rate and diff_nec_rate) are also communicated

- In the important case where a chemotherapy is delivered to the tumour, some other important parameters (e.g. Those related to the time-scheduling of the drug delivery) have to be defined by the user and communicated by the main core to IMMU, since those therapies influences the dynamics of the Immune system, e.g. Killing its proliferating cells and acting on the bone marrow

Once that the information have been received, IMMU uses them to modify the state of the immune system and of other variables (such as viral burden etc..) and accordingly it modifies the Oncosimulator parameters and send it back to the main module.

12.4 Characteristics of the models that can be embedded in the IMMU module

The basic biological skeleton of IMMU is as follows:

- The dynamics of the tumour cells depend on the dynamics of the IS effector(s) via a prey-predator-like interaction: the immune effectors increase the death rates of the Tumour Cells

- The presence of tumour cells stimulates the proliferation of effectors, with

the possibility, of course, of representing the dynamics of inactivated and of activated antigen-presenting cells

- These processes are mediated by interleukins (I(t))
- The influx of effectors depends on the Tumour Burden: large tumours are considered at least partially immunosuppressive
- There may be a presence of a Immunotherapy

As far as the mathematical structure of the admissible user-defined models is concerned, both deterministic and stochastic models can be defined. We shall allow the following mathematical structures:

- Ordinary differential equations
- Delay differential equations, if delays in the tumour-immune system inter-signalling are thought to be important for the tumour in study
- Stochastic differential equations, in order to model exogenous noises resulting in bounded stochastic fluctuations of the parameters of the specific model implemented by the user. Note that the user will not be allowed to use unbounded perturbations of the parameters, in order to preserve thir positivity and, in such a way, the inherent biological coherence of the models
- Stochastic birth and death processes with discrete state variables (for the cells) and continuous time. Even in case of definition of deterministic models, this modality will automatically be switched on if the total number of effector cells goes under an user-defined threshold. The stochastic process will be simulated by means of the Gillespie exact algorithm.

As far as the immunoherapies are concerned, their scheduling will be visually defined by the user via OBTIMA.

Spatial effects can be included, but in such a case the main module of the oncosimulator will have to send information that are related to each specific Geometric Cells, including not only biological informations but also geometrical informations. Note that since the oncosimulator is a meso-scale model that does not give spatial information on the distribution of tumour cells in the Geometric Cells, the same information has to be utilized in the spatial model. Thus indicating with

$$S=(S1,S2,S3)$$

the generic point in the tridimensional space, we shall assume that in the G-th Geometric Cell the density of tumour cells is constant and equal to the number of tumour cells divided by the volume of that GC. On the contrary, the immune cells as well as the immune-related proteins will be computed in all the tumour space, whereas to update the oncosimulator main module parameters their spatial average in each geometric cell will be utilized.

12. 5 The thesaurus

As far as the database of pre-set specific hybrid models, we shall implement the following well known and validated (in the single-scale) models:

- The de Pillis-Radunskaja model
- The Agur model

- The Kuznetsov model
- The de Boer model
- The Kirschener-Panetta model
- The Chaplain-Matsavinos-Kutsnetosv spatial model

Of those models we shall consider both the original ODE version and stochastic variants, such as the stochastic version of the Panetta-Kirschner model we recently proposed [27].

Moreover, since the specific tumour will need a targeted modelling, the user will also be able to test his/her own variants of the model by defining user defined nonlinear rates, which, however, will have to be coherent with the "original model".

12.6 Proposed pre-clinical and clinical validation

A cornerstone of the development of a multi-scale tools for simulating biomedical phenomena is its validation, and related adaptation. To this respect the IMMU model, being linked to the oncosimulator, and being intrinsically a meta-model, might be a good framework allowing the users to test the biological realism of the underlying specific models implemented by them.

As far as the models contained in the thesaurus, our aim is to directly validate at least some of them by using both (ina first phase of the implementation) literature data and, then, direct collaboration with experimental and clinical studies on immunotherapy.

In doing this, we shall follows the rule of validation adopted for the main module of the Oncosimulator, and established in the framework of the present ACGT project. For each model (and its variants), the pre-clinical and clinical validation procedure will involve comparison of the model predictions with pertinent pre-clinical and clinical data before, during and after the immunotherapy course, and in case of absence of therapy, during the experiments on the mice.

It should be stressed that no modifications to the standard clinical or laboratory practice are necessary for the clinical validation. The simulation model can just “follow” the laboratory/clinical practice and activate a self-optimization procedure. Hence no major ethical concerns arise during the clinical validation-adaptation procedure, neither during the phase involving experiments on animals.

12. 7 Possible applications in basic science

With the design of a new additional module for the Oncosimulator we had two main aims of some interest in basic science. The first and most important was of biomedical nature, the second of algorhythmic nature.

As far as the second is concerned, we introduced a new class of hybrid multi-scale models that is able to conjugate the detailed multi-scale modelling provided by the framework of the Oncosimulator with well-established or new mathematical models that were originally developed at macro-scale (PDE, and stochastic or delayed PDE) or at super-macroscale (ODE, DDE, SDE and birth-death sotchastic processes). This may provide an invaluable opportunity of testing those models by using real data and to assess whether dynamics predicted by them in large-scale may be sufficiently robust to realistically mimick the tumour in study. Accessorily, this may allow to better tune the numerical values of the parameters of these models. On this topic, it might be interesting to verify whether a model better

performs in this multi-scale framework than in its original macroscale. If so, this might be an evidence of its intrinsic biological soundness.

Coming now at the first an most relevant aim, the biomedical one, we were interested to extend the oncosimulator in order to get a sufficiently flexible and powerful tool to give a computational contribution in the complex process of assessing the influence of Immune System on unperturbed an perturbed (by human intervention) tumour dynamics.

Namely:

- The assessment of the action of IS on a tumour in the phase immediately following a chemotherapy based on cytotoxic or antiproliferating agent. As far as this issue is concerned, we are mainly interested in answering to the following question: to which extent is the IS, which is usually severely affected by such a therapy, able to suppress a residual colony of tumour cells that cannot be detected with the current available biomedical technologies ? May a durable, therapeutic antitumour immune response be achieved and maintained over the course of a patient's lifespan through a concerted effort of multiple immunotherapies [27] ?

- The role of IS in influencing the natural history a tumour in absence of therapies of any time. A first issue, where the stochastic nature of the involved processes is fundamental, is to better characterize the event of suppression of small tumours through immune surveillance [27]. Of course, in this case the investigation has to be focused on microscopic tumours, although it is of the outmost interest the investigation of the sporadic cases of spontaneous remission of macroscopic tumours. Moreover, independently from the initial dimensions of the neoplasia, we are also aimed to assess if and, in case, to which extent the dynamics of a macroscopic tumour may be influenced by the IS, for example with respect to the case of total lack of immune response.

- The simulation, in a realistic settings, of various kinds of anti-tumour therapies, including immuno-viro-therapies, based on the up-regulation of the immune system in order that it may better control the tumour. This both in the case of immunotherapies delivered as co-adjuvant of chemotherapies, and in the case of immunotherapy-based monotherapy.

- Assessing the relevance of geometrical shape of the tumour in the success or failure of the attack of the immune system.

12.8 Possible applications in pre-clinical and clinical studies

Once that a specific IMMU-compliant model defined by an user and implemented in IMMU has been validated first qualitatively and then quantitatively on real preclinical or clinical data regarding a specific tumour for which a specific and well established immunotherapy or viro-immunotherapy has been defined (alone or in combination with to her therapies), the may start having an active role in the preclinical or clinical studies regarding those specific therapies.

Indeed, similarly to the main module of the oncosimulator [28], the expected use of an IMMU-compliant model following thorough clinical adaptation, optimization and validation is to simulate either several candidate treatment schemes for a particular patient and support the selection of the optimal one or to simulate the expected extent of tumour shrinkage for a given time instant and decide on the adequacy or not of the simulated scheme.

However, another application that is strongly dependent of the peculiar nature of the tumour-iimmune system is possible. Indeed, since the immunotherapies (at least, those that do not involve lytic viruses) have and indirect effect on the tumour, the lysing effects of an immunotherapy may initially be quite delayed, as well as the tumour shrinking. In this complex scenario, an IMMU-compliant model may be useful in assessing whether the non-responsiveness of a specific patient is linked to kynetic effects or it depends on a non-adequate treatment scheme.

Moreover, given the important role of Immune System in tumour control and, sometime, suppression, a second scenario of preclinical and clinical applications may be envisaged for an IMMU-compliant model: the monitoring of the follow-up of a specific patient after a chemotherapy or a radiotherapy.

Finally, since an adequately validated IMMU-compliant model may predict the levels of immune reaction, it might also contribute to the predictions of immune-systems related side effects of an immunotherapy.

12.9 References

1. {PA03} D. Pardoll, Does the Immune System See Tumours as Foreign or Self?, Ann. Rev. of Immun. 21, 807-839, (2003).

2. {IS57} F.M. Burnet, Cancer - a biological approach, Brit. Med. J. 1, (1957) 841-847.

3. {IS09} P. Ehrlich, Ueber den jetzigen Stand der Karzinomforschung, Ned. Tijdschr. Geneeskd. 5, (1909) 273-290.

4. {Dunn2004} G.P. Dunn, L.J. Old and R. D. Schreiber, The three ES of Cancer Immunoediting, Ann. Rev. of Immun. 22, (2004) 329-360.

5. A. d'Onofrio "Tumour Evasion from Immune Control: strategies of a MISS to become a MASS", Chaos, Solitons and Fractals 31 261-268 (2007)

6. vi02 A. P. Vicari, G. Caux and G. Trinchieri, Tumour escape from immune surveillance through dendritic cell inactivation, Sem. Canc. Biol 12, (2002) 33-42.

7. {PPV} M. Pekham, H. M. Pinedo and U. Veronesi, Oxford Textbook of Oncology, (Oxford Medical Publications, 1995)

8. {VHR} V. T. de Vito Jr., J. Hellman and S. A. Rosenberg, Cancer: principles and practice of Oncology, (J. P. Lippincott,2005).

9. {Seminar} New Applications of Cancer Immunotherapy, S. A. Agarwala (Guest Editor), Seminars in Oncology -Special Issue n. 29-3 Suppl. 7 (2003).

10. {Euronc} I. Bleumer, E. Oosterwijk, P. de Mulder and P. F. Mulders, Immunotherapy for Renal Cell Carcinoma, European Urology 44, (2003) 65-75.

11. {Kaminski} J. M. Kaminski , J. B. Summers, M. B. Ward, M. R. Huber and B. Minev, Immunotherapy and prostate cancer, Canc. Treat. Rev. 29 (2004) 199-209.

12. DeBoer, R.J., Hogeweg, P., Hub, F., Dullens, J., DeWeger, R.A., DenOtter, W., 1985. Macrophage T Lymphocyte interactions in the anti-tumour immune response: a mathematical model. J. Immunol. 134, 2748–2758

13. De Lisi, C., Rescigno, A., 1977. Immune surveillance and neoplasia: a minimal mathematical model. Bull. Math. Biol. 39 (2), 201–221

14. Stepanova, N.V., 1980. Course of the immune reaction during the development of a malignant tumour. Biophysics 24, 917–923.

15. Kirschner, D.,Panetta,J.C.,1998.Modelingimmunotherapyofthe tumour—immune interaction.J.Math.Biol.37,235–252

16. Kuznetsov, V.A., Makalkin, I.A., Taylor, M.A., Perelson, A.S., 1994. Nonlinear dynamics of immunogenic tumours: parameter estimation and global bifurcation analysis. Bull. Math. Biol. 56, 295–321

17. Matzavinos, A., Chaplain, M.A.J., Kuznetsov, V.: Mathematical modelling of the spatio-temporal response of cytotoxic T-lymphocytes to a solid tumour. Math. Med. Biol., 21, 1–34 (2004)

18. De Pillis, L.G., Radunskaya, A.E., Wiseman, C.L., 2005. A validated mathematical model of cell-mediated immune response to tumour growth. Cancer Res. 65, 7950–795

19. Wodarz D and Komarova N, Computational biology of Cancer. World Scientific (2005)

20. A. d'Onofrio "A general framework for modeling tumour-immune system competition and immunotherapy: analysis and medical inferences", "Physica D" 208, 220-235, (2005).

21. A. d'Onofrio "The role of the proliferation rate of effectors in the tumour-immune system competition" Mathematical Models and Methods in Applied Sciences 16, 1375-1401 (2006)

22. A. d'Onofrio "Metamodeling tumour-immune system interaction, tumour evasion and immunotherapy" Mathematical and Computer Modelling 46 614-637 (2008)

23. A. d'Onofrio, F. Gatti, P. Cerrai and L. Freschi, "Delay-induced Oscillatory dynamics of Tumour-Immune System Interaction". In press on Mathematical and Computer Modelling

24. C. Cattani, A. Ciancio and A. d'Onofrio "Metamodeling the learning-hiding competition between tumours and immune system: a kinematic approach In press on Mathematical and Computer Modelling

25. A. d'Onofrio "Tumour evasion from immune system control as bounded-noise induced transition". Physical Review E 81 art.n. 021923

26. G. Caravagna, A. d'Onofrio, P. Milazzo and R. Barbuti. "Tumour suppression by Immune system through stochastic oscillations" in press on Journal of Theoretical Biology

27. Jamila K. Adam, Bharti Odhav, Kanti D. Bhool, Immune responses in cancer, Pharmacology & Therapeutics 99 (2003) 113– 132

28. Cappuccio, A., Elishmereni, M., Agur, Z., 2006. Immunotherapy by IL-21: potential treatment strategies evaluated in a mathematical model. Cancer Res. 66 (14), 7293–7300.

29. Natalie Kronik, Yuri Kogan, Vladimir Vainstein, Zvia Agur, Improving alloreactive CTL immunotherapy for malignant gliomas using a simulation model of their interactive dynamics, Cancer Immunol Immunother (2008) 57:425–439

30. G.S.Stamatakos, E.A.Kolokotroni, D.D.Dionysiou, E.Ch.Georgiadi and C. Desmedt,An advanced discrete state–discrete event multiscale simulation model of the response of a solid tumour to chemotherapy: Mimicking a clinical study, Journal of Theoretical Biology, (2010) doi:10.1016/j.jtbi.2010.05.019

13. The ACGT Oncosimulator and the Oncosimulator team in the 4th International Advanced Research Workshop on *In Silico* Oncology and Cancer Investigation

(*Code letter: O)*

The 4th International Advanced Research Workshop on In Silico Oncology and Cancer Investigation which was also the ContraCancrum workshop (www.4th-iarwisoci.iccs.ntua.gr) took place in Athens on September 8-9 2010. Five out of the eight Organizing Committee Members, including the General Chair, are participants in the *in silico* oncology action of the ACGT project (workpackage WP8). About 40 leading investigators from three continents participated in the event. The workshop was sponsored by the European Commission through the ContraCancrum project, co-sponsored by the IEEE-Engineering in Medicine and Biology Society (EMBS) and endorsed by the International Federation for Medical and Biological Engineering (IFMBE). It turned out to be a very successful and fruitful event attracting the attention of the American Association for Cancer Research (AACR) with which an important interaction was initiated.

The following papers and/or addresses have been prepared and presented and/or delivered by members of the WP8 team of the ACGT project.

1. Welcome address - *In Silico Oncology: A Platonic approach to medical science*

G.Stamatakos, ICCS-NTUA

2. Greeting by K. Marias, FORTH

3. Greeting by N. Graf, USAAR

4. *Scientific and Technological Background of In Silico Oncology*

N.Uzunoglu, ICCS-NTUA

5. *Tumour segmentation: The impact of standardized signal intensity*

histograms in glioblastoma

J.Zepp, N.Graf, E.Skounakis, R.Bohle, E.Meese, H.Stenzhorn, Y.-J. Kim, C.Farmaki, V.Sakkalis, W.Reith, G.Stamatakos, K.Marias

6. I*n Silico Oncology: A Hypermatrix–Operator Formulation of a Top-Down Multiscale Simulation Model of Tumour Response to Treatment. The Oncosimulator Concept.*

G.Stamatakos

7. *The ISOG, NTUA Tumour Response to Treatment Discrete Simulation Models: A Review of Basic Concepts and Algorithms*

D.Dionysiou

8. *Breast Cancer Modeling in the Clinical Context: Parametric Studies*

E.Kolokotroni, D.Dionysiou, E.Georgiadi, N.Uzunoglu, G.Stamatakos

9. *Discrete Event Based Modeling of Nephroblastoma. Sensitivity Considerations*

E.Georgiadi, D.Dionysiou, E.Kolokotroni, N.Uzunoglu, N.Graf, G.Stamatakos

10. *Molecular Personalization of Cancer Treatment via a Multiscale Simulation Model of Tumour Response to Therapy. The Paradigm of Glioblastoma Treated with Temozolomide.*

A .Folarin, G.Stamatakos

11. *Clinically Oriented Translational Cancer Multilevel Modelling*

K.Marias, V.Sakkalis, A.Roniotis, I.Karatzanis, G.Stamatakos, D.Dionysiou, S.Giatili, N.Uzunoglou, N.Graf, R.Bohle, E.Messe, H.Stenzhorn, Y.-J.Kim, P.Coveney, S.Zasada, S.Wan, A.Folarin, P.Büchler, T.Bardyn, S.Bauer, M.Reyes, G.Clapworthy, E.Liu, T.Bily, V.Bednar, M.Karasek, A.Franz, R.Grewer, and J.Sabczynski

12. *The ContraCancrum Oncosimulator: Integrating Biomechanisms Across Scales in the Clinical Context*

G.Stamatakos, D.Dionysiou, S.Giatili, E.Kolokotroni, E.Georgiadi, A.Roniotis, V.Sakkalis, P.Coveney, S.Wan, S.Manos, S.Zasada, A.Folarin, P.Büchler, T.Bardyn, S.Bauer, M.Reyes, T.Bily, V.Bednar, M.Karasek, N.Graf, R.M.Bohle, E.Meese, Y.-J.Kim, H.Stenzhorn, G.Clapworthy, E.Liu, J.Sabczynski, K.Marias

13. *Glioma diffusive modeling: Calculating diffusion coefficients from*

atlases with proportional tissue information

A.Roniotis, V.Sakkalis, G.Stamatakos, M.Zervakis, K.Marias

14. *An Explicit Boundary Condition Treatment of a Diffusion – Based Glioblastoma Tumour Growth Model*

S.Giatili, N.Uzunoglu , G.Stamatakos

15. *Using the GPU for Simulating Spatiotemporal Tumour Growth*

B.Liu, G.Clapworthy, E.Kolokotroni, G.Stamatakos

16. *A framework supporting sharing and reuse of data and tools in translational cancer research: Lessons learned for VPH research*

M.Tsiknakis, S.Sfakianakis, G.Zacharioudakis, L.Koumakis

17. *The ACGT Oncosimulator: from Conceptualization to Development via Multiscale Cancer Modeling*

G.Stamatakos, D.Dionysiou, E.Georgiadi, E.Kolokotroni, S.Giatili, A.Hoppe, C.Desmedt, A.Lunzer, M.Erdt, J.Jacques, J.Pukacki, R.Belleman, P.Melis, A.d'Onofrio, F.Buffa, B.Claerhout, S.Rueping, K.Marias, M.Tsiknakis, N.Graf

18. Validating the ACGT Oncosimulator with a Grid-Supported Visualisation Environment

A.Lunzer, R.Belleman, P.Melis, J.Pukacki, P.Spychała, G.Stamatakos

19. *A collaborative system for the in silico oncology domain: Requirements, solutions and guidelines*

I.Lykourentzou, D.Dionysiou, G.Stamatakos

14. Conclusion
(Code letter: N)

In the present document a brief recapitulation of all the basic features of the ACGT Oncosimulator (OS) has been provided along with the latest advances concerning its clinical positioning, scientific foundation, analysis, modelling and technological implementation and development. Particular emphasis has been put on the procedure of clinical adaptation and validation of the simulation models and the integration of the OS into the ACGT architecture.

It should be noted that the entire system constitutes the *first* large scale clinical trial driven and clinically adaptable and testable multiscale cancer model based simulator *on the global level*. It has also proved a highly successful intercontinental venture that combines in an unprecedented way basic science, technology and medicine.

The invaluable experience that has been accumulated during the OS development and adaptation has already been exploited by other oncosimulator centered Virtual Physiological Human (VPH) research and development actions concerning other tumour types, other clinical trials, other treatment modalities, other investigation contexts etc. The European Commission (EC) co-funded projects ContraCancrum and TUMOUR are good examples of the latter.

The importance of the joint effort has been supported enthusiastically by the international scientific community on many occasions and within the framework of several initiatives including the 1st Transatlantic Workshop on Multiscale Cancer Modeling (1st TWMSCM) that was co-funded by the EC and the National Institutes of Health (NIH) – National Cancer Institute (NCI) of the United States and took place in Brussels in 2008. The latter was a highlight of the ICT 2008 strategic event. Other occasions, facts and events include the 4th International Advanced Research Workshop on *In Silico* Oncology and Cancer Investigation that took place in Athens in Sept. 2010, the special interest expressed by the American Asscociation for Cancer Research (AACR) for both the Athens workshop and *in silico* oncology in general and numerous scientific and broad audience publications including a CRC multi-author book inspired by the presentations made in the 1st TWMSCM that is to be published in Dec. 2010.

It has to be admitted, however, that due to the extremely complex requirements of the endeavour in scientific, technological and clinical terms and despite the committed and systematic work done by of all involved partners, neither the clinical adaptation nor the clinical validation of the OS have reached a mature stage that would allow a direct translation of the system into clinical routine. Nevertheless, the invaluable foundational and preparatory work already done in conjunction with the provision of proofs for several principles lying at the heart of the clinical adaptation and validation procedures have certainly paved the way for the next clinical translation stage to be performed obviously within the framework of other related research projects.

In summary it may be claimed that the Oncosimulator development and testing component of the ACGT project has proved a foundational, successful and globally pioneering endeavour driving multiscale cancer modelling and *in silico* oncology into clinical reality and practice. Therefore, in the future a new biologically grounded and scientifically trustable weapon for personalized optimization of

cancer treatment will eventually be added to the anti-cancer arsenal of the clinical routine.

Appendix 1 - Abbreviations and acronyms *(Code letter: R)*

AACR	*American Asscociation for Cancer Research*
ACGT	*Advancing Clinico-genomic Trials on Cancer*
ACT	*Actinomycin*
AJAX	*Asynchronous Javascript*
CELLU	*Cellular (module)*
CKR	*Cell Kill Ratio*
COG	*Children's Oncology Group*
CRF	*Case report form*
CT	*Computerized Tomography*
DICOM	*Digital Imaging and Communications in Medicine*
DMS	*Data Management System (or Suite)*
EC	*European Commission*
FhG	*Fraunhofer Gesellschaft*
FDA	*Food and Drug Administration*
GAS	*Gridge Authorization Service*
GC	*Geometrical Cell*
GPOH	*Gesellschaft für Pädiatrische Onkologie und Hämatologie*
GRMS	*Gridge Resource Management System*
GVK	*Grid Visualization Kernel*
FORTH	*Foundation for Research and Technology Hellas*
HTTP	*Hypertext Transfer Protocol*
ICCS	*Institute of Communication and Computer Systems- National Technical University of Athens*
IJB	*Institut Jules Bordet*

IEO	*Istituto Europeo di Oncologia*
IMMU	*Immunologic (module)*
IS	*Immune System*
LIMP or Limp or limp	*Limited Mitotic Potential (cells)*
MODU	*Module*
MRI	*Magnetic Resonance Imaging*
MISS	*Microscopic Steady State*
NBC	*Number of Biological Cells*
NTUA	*National Technical University of Athens*
ODE	*Ordinary Differential Equation*
ORS	*OncoRecipeSheet*
OS	*Oncosimulator*
PC	*Personal Computer*
PDE	*Partial Differential Equation*
PET	*Positron emission tomography*
RECIST	*Response Evaluation Criteria in Solid Tumours*
PSNC	*Poznan Supercomputing and Networking Center*
PSS	*Personal Space Station*
REST	*Representational State Transfer*
SIOP	*Société Internationale d'Oncologie Pédiatrique*
SOA	*Service Oriented Architecture*
TC	*Tumour Cell*
THERA	*(Immuno-)therapeutic (module)*
TOP	*Trial of Principle*
URI	*Uniform Resource Identifier*

UHoK	*University of Hokkaido*
U/S or US	*Ultrasound*
UvA	*University of Amsterdam*
USAAR	*University Hospital of Saarland*
VNC	*Virtual Network Computing*
VCR	*Vincristine*
VIR	*Virus (evolution)*
VTK	*Visualization Toolkit*
VPH	*Virtual Physiological Human*