

Community Consultation on Research Data Standards

Comments are invited on this outline for starting to develop biomedical informatics standards for common use in antibody research.

Please send responses to Richard Begent (r.begent@ucl.ac.uk)

Developing Standards for Recording, Sharing and Integrating Data from Antibody Research

(version 1, 10th March 2008)

Two of the Antibody Society's aims are to:

- Work for development and acceptance of formats for the interoperability of data, databases, and computational resources underpinning this field
- Develop guidelines that help to ensure the safety of antibody-related therapeutics, during preclinical and clinical testing, and beyond

Work presented below begins this process for antibody therapy and seeks to provide common platform adaptable for addressing other aspects of antibody research.

Recording, sharing and integrating data

In antibody research, individual research groups generate large amounts of data of diverse types. Value can be added by collecting data together and analysing it further. This may involve sharing data of one type and integrating data of different types. However in many areas of research sharing and integrating data is invalid or error prone because of lack of data standards and infrastructure for

Establishing informatics standards for interoperability of data in this context is challenging because it covers so many research domains from genetics through to clinical trials and population studies. However, existing standards for instance in genomics, proteomics and randomised clinical trials are applicable to antibodies and are backed in some cases by databases. The NCRI Planning Matrix identifies these resources as data elements and controlled vocabularies and ontologies (cancerinformatics.org.uk/matrix/matrix.htm)]. However, in other areas, such as preclinical and clinical development of antibody-based drugs, agreed standards are lacking and need to be defined.

Complex Applications of Antibodies.

Antibody therapies are having a major impact in treating human disease but can also produce life-threatening side effects in certain circumstances. This was most tragically demonstrated in the TGN1412 trial in 2006 when healthy volunteers had very severe reactions to an antibody which was being developed for treatment of rheumatoid arthritis and other conditions. The subsequent investigation showed that a more thorough understanding of the way the antibody worked could have predicted the problem but this insight was only evident after analysis of complex interacting parameters.

It is crucial for this that information from different sources is compatible so that it is valid to integrate it, helping to predict and avoid adverse effects and to accelerate the development of effective new drugs.

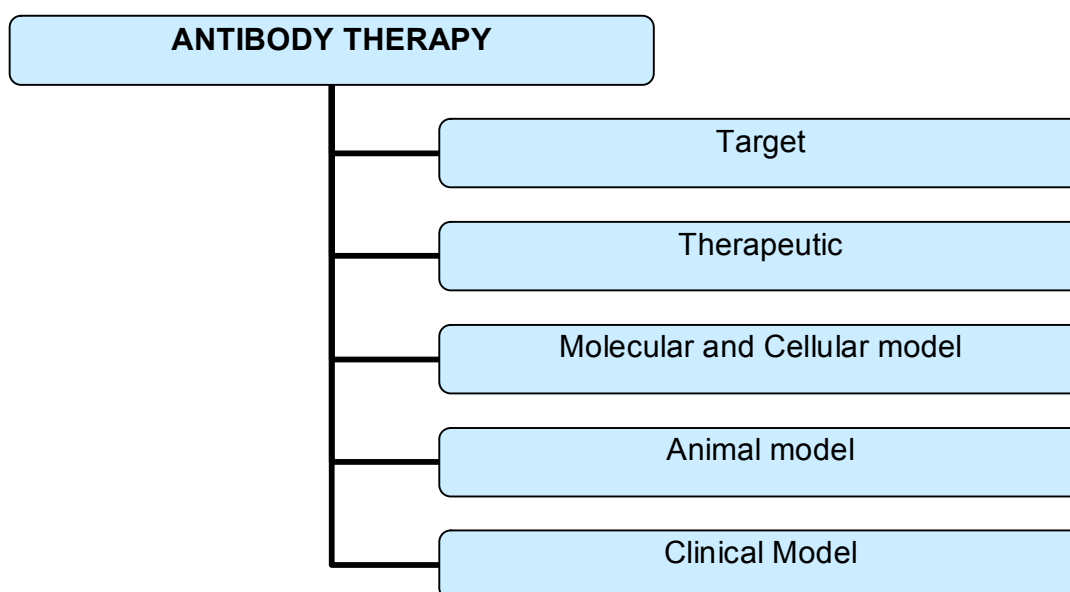
Common Data Elements

The first step to enable this is to use common data elements which are re-usable so that data from one source can validly be integrated or compared with that from another. These are preferably documented to ISO11179 standards and made available in a controlled vocabulary. The NCI Thesaurus provides such a controlled vocabulary which is in extensive use within the NCRI family as well as in the USA.

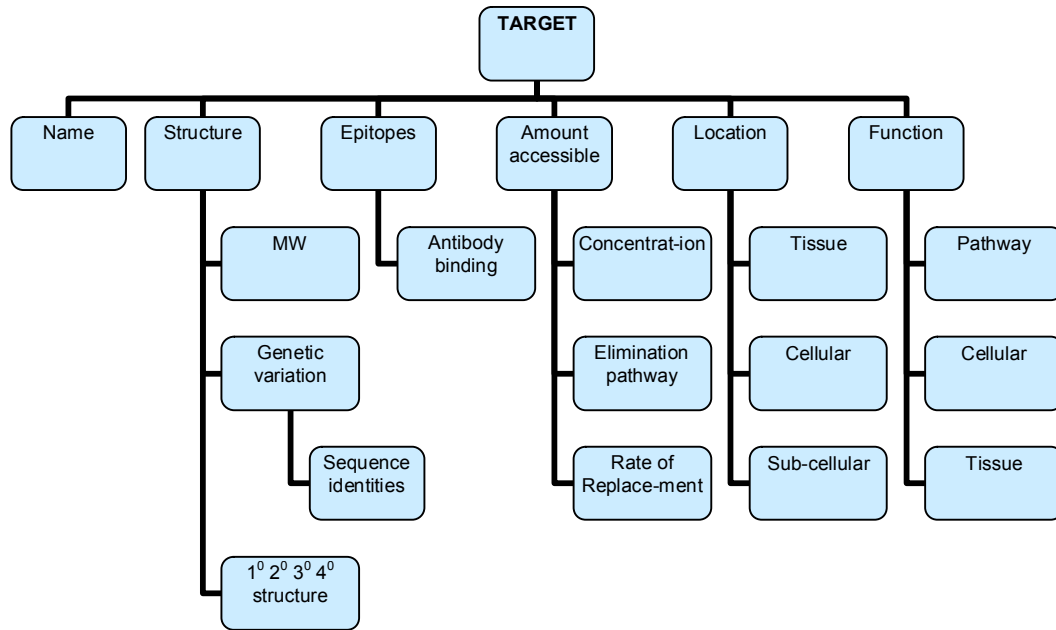
Defining Information which should be collected:

The relevant common data elements for antibody therapy must be identified, thus generating guidelines for information about antibody therapy experiments (GIAATE). These are comparable to the minimum information advocated for reporting of data from microarray experiments (MIAME) developed by the MGED Society. Other groups are addressing the same issue for other types of molecular and cellular data. The problem is generally more complex for antibody therapy because many diverse types of data are relevant from molecules to the whole person or population. For this reason guidelines are advocated rather than a minimum information requirement.

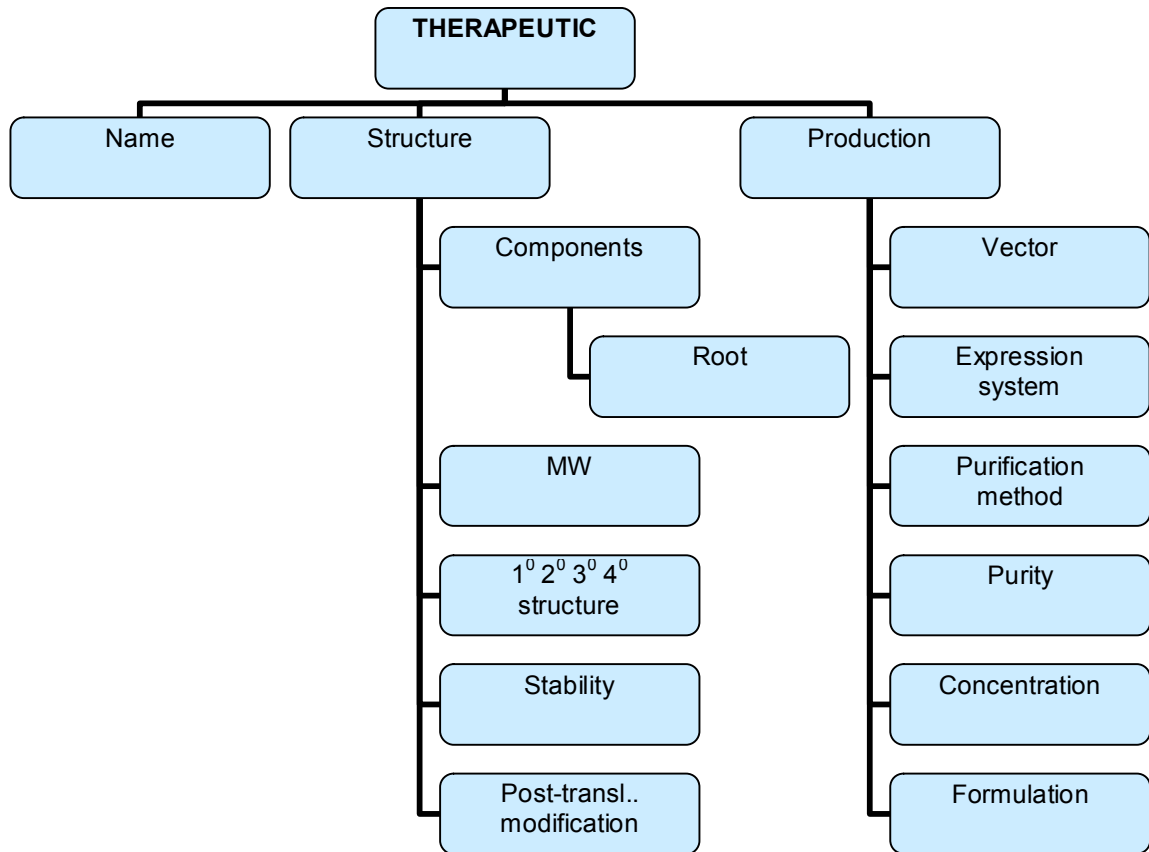
The Antibody Society through its diverse experts has started to develop these guidelines. The main groups have been identified as relating to the target for therapy, the antibody-based therapeutic, a molecular and cellular model of function, an animal model and a clinical model. An example of a prototype of information suggested for each group is given below. Each box contains a common data element which may already exist in the US National Cancer Institute Thesaurus or may need to be defined de novo. Further detail can be added in additional levels of the hierarchy. When data elements are defined and shared in the thesaurus they can be used in the protocols for experiments relating to antibody therapy and a body of data which can validly be compared will begin to build. If these data are shared and re-used (subject to protection of intellectual property and confidentiality), the development of effective and safe therapeutics will be accelerated and made more cost-effective.



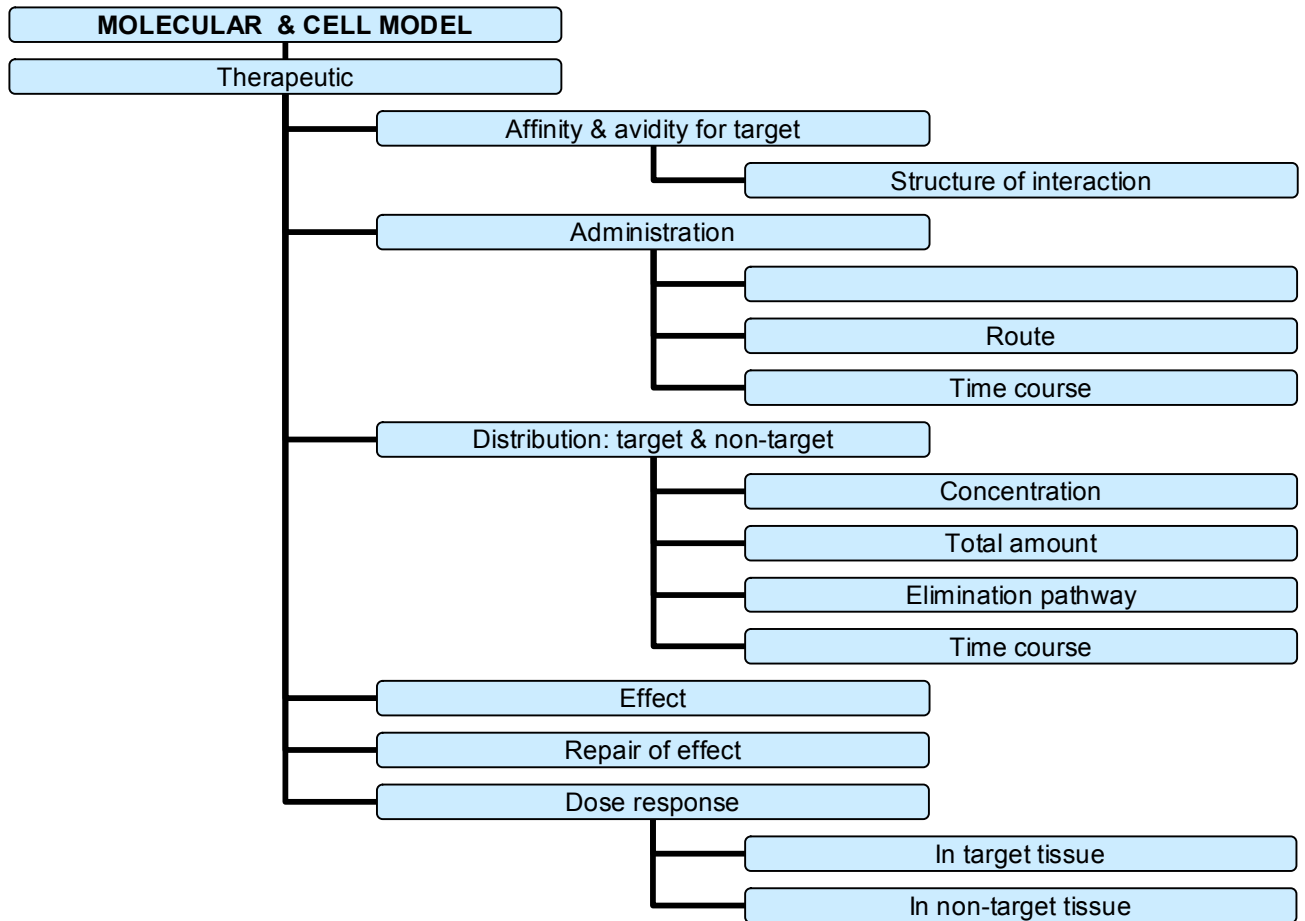
Target



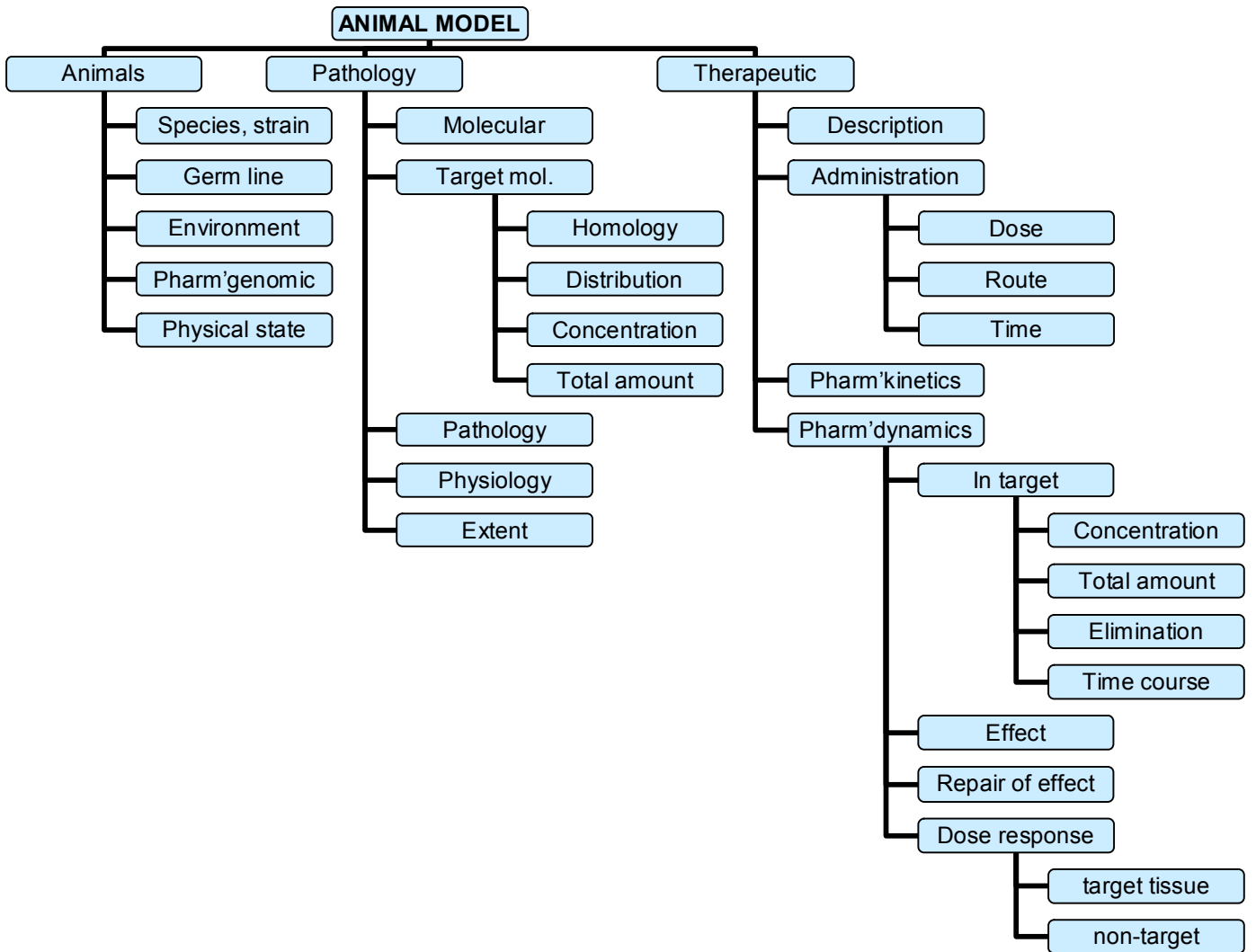
Therapeutic



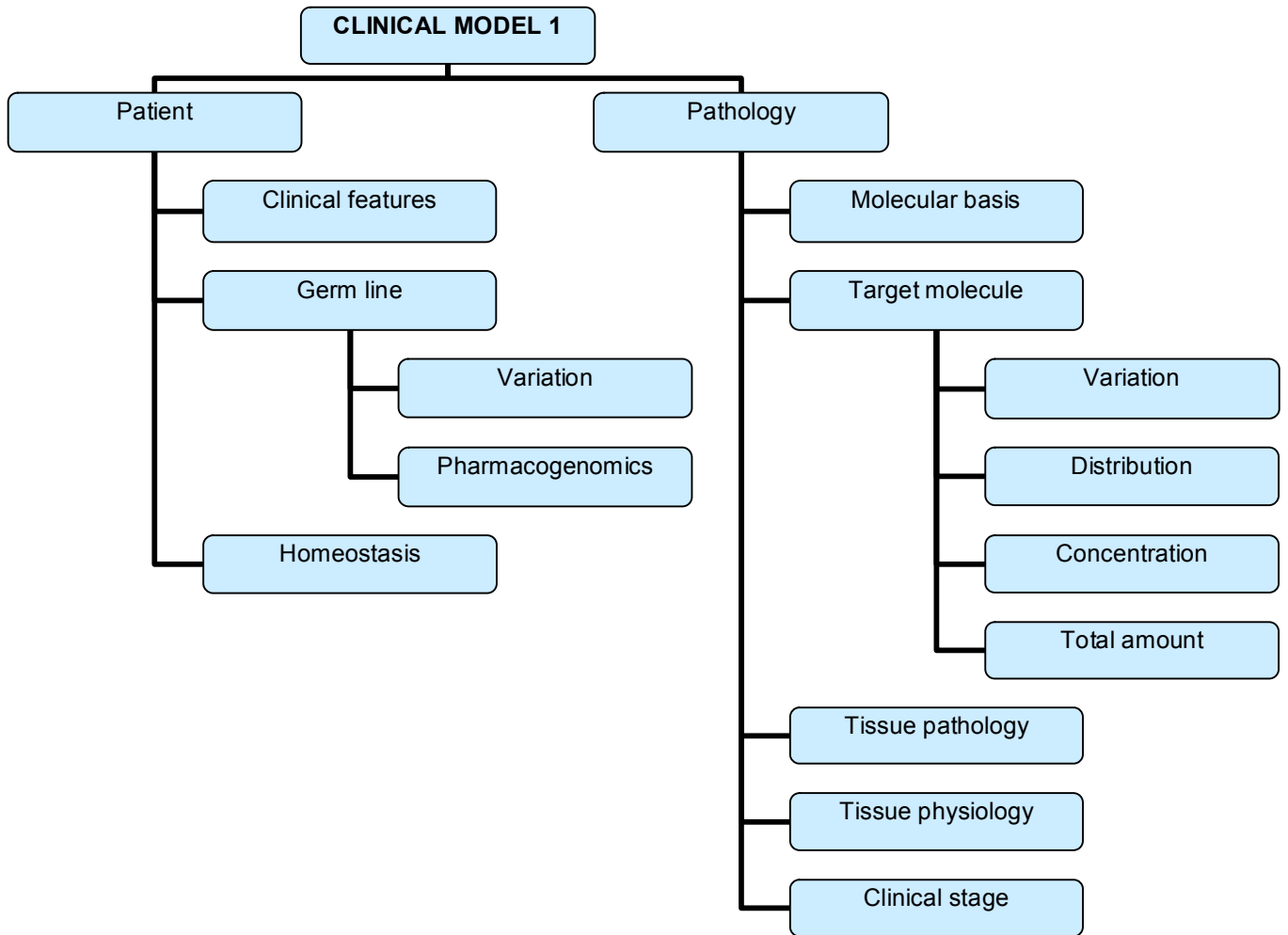
Molecular and Cell

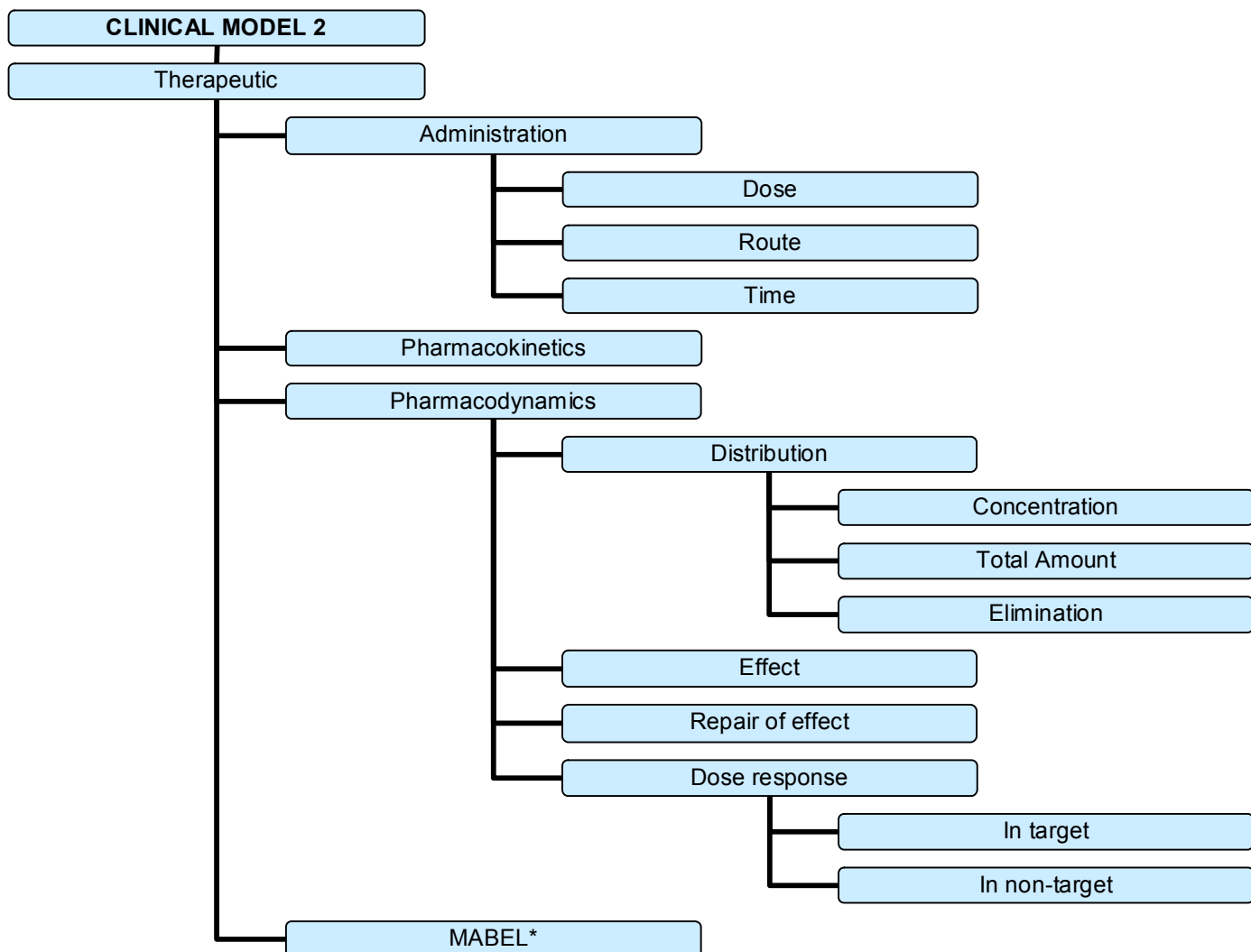


Animal Model



Clinical Model





*MABEL – minimum anticipated biological effect level see Expert Scientific Group on Phase One Clinical Trials Final Report 30th November 2006 ISBN-10 0 11 703722 2, ISBN-13 978 0 11 703722 9 www.tsoshop.co.uk