Static Analysis of a Model of the LDL Degradation Pathway

Henrik Pilegaard, Flemming Nielson, Hanne Riis Nielson

The Technical University of Denmark
{hepi,nielson,riis}@imm.dtu.dk

Abstract. BioAmbients is a derivative of mobile ambients that has shown promise of describing interesting features of the behaviour of biological systems. As for other ambient calculi static program analysis can be used to compute safe approximations of the behaviour of system models. We use these tools to model and analyse the production of cholesterol in living cells and show that we are able to pinpoint the difference in behaviour between models of healthy systems and models of mutated systems giving rise to known diseases.

1 Introduction

BioAmbients [18, 20] is a sibling to the ambient calculus [5] designed to model biological systems. The ambients are used to model chemically active sub-systems (compartments) bound by biological barriers (membranes). A set of capabilities allows the modelling of biological reactions that may happen between the sub-systems. These include movement capabilities that influence the hierarchical structure of the system as well as communication capabilities that allow the exchange of information between sub-systems.

When defining BioAmbients Regev gave a set of abstraction guidelines and numerous examples of their use [18]. Using these guidelines and static program analysis we develop and subsequently analyse a model of the uptake of *cholesterol* [1, 9]. In its initial configuration the model contains two ambients, one modelling a mammalian cell and another modelling Low Density Lipo-proteins; the processes enclosed in the two ambients describe how they react with one-another to create cholesterol. The disorder known as *familiar hypercholesterolemia* is due to a defect in the receptor proteins that prevents the uptake of cholesterol; the disorder exists in two variants and both are easily modelled in BioAmbients. Syntactically the difference between these models is nothing but a typo; biologically, however, the difference is severe as it increases the risk of *cardiovascular disease*.

This paper demonstrates how program analysis technology [11] may be used to quickly identify pathways affected by system perturbations. More specifically, we are able to reveal crucial differences between healthy systems and those that suffer from familiar hypercholesterolemia.

Dagstuhl Seminar Proceedings 06161
Simulation and Verification of Dynamic Systems
http://drops.dagstuhl.de/opus/volltexte/2006/723

Overview of paper. After having introduced the syntax and semantics of BioAmbients in Section 2 we outline the static program analysis in Section 3. The development and analysis of the cholesterol example is described in Section 4. Concluding remarks are given in Section 5.

2 BioAmbients

BioAmbients [20, 18, 3] inherit the notion of ambients as bounded mobile sites of activity from Mobile Ambients [5], thereby allowing the biological concept of compartment to be modelled in an intuitive manner. Contrary to mobile ambients, however, bioambients are cast as nameless entities - the roles of which *may* be indicated by annotations. Both communication and ambient interaction are facilitated by having capability/co-capability pairs react with each other as in [8, 15]. As a consequence all reactions are synchronous; it is demanded that the process exposing the capability and the process exposing the corresponding co-capability must agree on a reaction for it to happen. Such an agreement can be reached only if the two parties expose compatible capabilities and share the same name.

The set of control structures for processes is slightly larger than what is traditionally studied for Mobile Ambients. It includes non-deterministic (external) choice as well as a general recursion construct in the manner of CCS [10] in order to facilitate more faithful models of biological systems.

Before presenting the formal details we should like to stress that formalisms like BioAmbients constitute limited models. As a consequence modelling artifacts necessarily arise whenever the formalism cannot perfectly represent reality. Being aware of this allows us to know when and how our model is imprecise and, thus, how to interpret the formal behaviour in a biological setting.

2.1 Syntax

The full syntax of BioAmbients is defined in Table 1, where we write P for *processes*, $M(\in \mathbf{Cap})$ for *capabilities*, and $n, m, p(\in \mathbf{Name})$ for *names*. We associate each ambient with a *role* $\mu(\in \mathbf{Role})$ and annotate the ambient constructs accordingly. These roles, which may be thought of as identities, have no semantic significance but are useful as "pointers" both when modelling and analysing actual biosystems.

We shall write $\mathsf{fn}(P)$ for the *free names* of P; the notion is defined by straightforward structural induction over the syntax. We write $P[m/n]$ for the process that is as P except that all free occurrences of n are replaced by m (subject to α-renaming of bound names) and similarly $P[Q/X]$ for the substitution of a process Q for all free occurrences of the process identifier X in P (subject to α-renaming of bound names and process identifiers).

$P ::=$	$\mathsf{0}$	inactive process
	$(n)P$	binding box for the constant n
	$[P]^{\mu}$	ambient P with the role μ
	$M.P$	prefixing with capability M
	$P \mid P'$	parallel processes
	$P + P'$	non-deterministic external choice
	$\mathsf{rec}\, X.\, P$	recursive process (*i.e.* $X = P$)
	X	process identifier
$M ::=$	$\mathsf{enter}\ n \mid \mathsf{accept}\ n$	enter movement
	$\mathsf{exit}\ n \mid \mathsf{expel}\ n$	exit movement
	$\mathsf{merge}{-}\ n \mid \mathsf{merge}{+}\ n$	merge movement
	$n!\{m\} \mid n?\{p\}$	local output and input binding the variable p
	$n\_!\{m\} \mid n\hat{}?\{p\}$	parent to child output and input binding the variable p
	$n\hat{}!\{m\} \mid n\_?\{p\}$	child to parent output and input binding the variable p
	$n\#!\{m\} \mid n\#?\{p\}$	sibling output and input binding the variable p

Table 1. Syntax of BioAmbients.

Alpha-renaming of bound names and process identifiers:

$P \equiv Q$ if P may be α-renamed to Q

Reordering of parallel processes:

$$P \mid P' \equiv P' \mid P$$
$$(P \mid P') \mid P'' \equiv P \mid (P' \mid P'')$$
$$P \mid \mathsf{0} \equiv P$$

Scope rules for name bindings:

$$(n)\mathsf{0} \equiv \mathsf{0}$$
$$(n_1)(n_2)P \equiv (n_2)(n_1)P$$
$$(n)(P \mid P') \equiv ((n)P) \mid P' \quad \text{if } n \notin \mathsf{fn}(P')$$
$$(n)(P + P') \equiv ((n)P) + P' \quad \text{if } n \notin \mathsf{fn}(P')$$
$$(n)([P]^{\mu}) \equiv [(n)P]^{\mu}$$

Reordering of sum processes:

$$P + P' \equiv P' + P$$
$$(P + P') + P'' \equiv P + (P' + P'')$$
$$P + \mathsf{0} \equiv P$$

Recursion:

$$\mathsf{rec}\, X.\, P \equiv P[\mathsf{rec}\, X.\, P/X]$$

Table 2. Axioms for structural congruence $P \equiv Q$.

2.2 Semantics

The semantics is defined in the standard way as a reduction system based on a congruence relation, $\equiv$, defined for processes in general and a transition relation, $\rightarrow$, defined only for programs. We define $\equiv$ as the least congruence relation induced by the axioms of Table 2 and the transition relation is presented in Table 3.

Note that the semantics of the recursion construct is given as a congruence. Contrary to the original Mobile Ambients [5] we also allow constant introductions (n) to migrate in and out of non-deterministic external choice constructs in much the same way as is customary for parallel composition.

3 Control Flow Analysis of BioAmbients

We shall analyse the LDL pathway model of this paper using a *context sensitive control flow analysis*, which belongs to a broad class of techniques called *Static*

Movement of ambients:

$$[(\mathsf{enter}\ n.\, P + P') \mid P'']^{\mu_1} \mid [(\mathsf{accept}\ n.\, Q + Q') \mid Q'']^{\mu_2} \rightarrow [\,[P \mid P'']^{\mu_1} \mid Q \mid Q'']^{\mu_2}$$

$$[\,[(\mathsf{exit}\ n.\, P + P') \mid P'']^{\mu_1} \mid (\mathsf{expel}\ n.\, Q + Q') \mid Q'']^{\mu_2} \rightarrow [P \mid P'']^{\mu_1} \mid [Q \mid Q'']^{\mu_2}$$

$$[(\mathsf{merge}{-}\ n.\, P + P') \mid P'']^{\mu_1} \mid [(\mathsf{merge}{+}\ n.\, Q + Q') \mid Q'']^{\mu_2} \rightarrow [P \mid P'' \mid Q \mid Q'']^{\mu_2}$$

Communication between ambients:

$$(n!\{m\}.\, P + P') \mid (n?\{p\}.\, Q + Q') \rightarrow P \mid Q[m/p]$$

$$(n\_!\{m\}.\, P + P') \mid [(n\hat{}?\{p\}.\, Q + Q') \mid Q'']^{\mu} \rightarrow P \mid [Q[m/p] \mid Q'']^{\mu}$$

$$[(n\hat{}!\{m\}.\, P + P') \mid P'']^{\mu} \mid (n\_?\{p\}.\, Q + Q') \rightarrow [P \mid P'']^{\mu} \mid Q[m/p]$$

$$[(n\#!\{m\}.\, P + P') \mid P'']^{\mu_1} \mid [(n\#?\{p\}.\, Q + Q') \mid Q'']^{\mu_2} \rightarrow [P \mid P'']^{\mu_1} \mid [Q[m/p] \mid Q'']^{\mu_2}$$

Execution in context:

$$\frac{P \rightarrow Q}{(n)P \rightarrow (n)Q} \qquad \frac{P \rightarrow Q}{[P]^{\mu} \rightarrow [Q]^{\mu}} \qquad \frac{P \rightarrow Q}{P \mid R \rightarrow Q \mid R} \qquad \frac{P \equiv P' \quad P' \rightarrow Q' \quad Q' \equiv Q}{P \rightarrow Q}$$

Table 3. Transition relation: $P \rightarrow Q$.

analyses. Based on *formal languages* the aim of static analysis is to predict *safe and computable approximations* to the behaviours captured by any given (valid) expression of the input language [11].

Over-approximating analyses are *safe* if and only if all valid behaviours are actually captured by the analysis, i.e. negative answers to enquiries regarding behaviour are provably reliable. This is exactly the promise of the analysis techniques introduced in the following and, as we shall see in Section 4.3, this makes them useful for quickly identifying pathways that are negatively affected by system perturbations.

3.1 Context Insensitive Analysis

The principles underlying the static analysis of Ambient Calculi are well understood [12, 7, 6, 14]. The customary non-contextual Control Flow Analyses (0-CFAs) for BioAmbients have already been treated [13, 16]. For a program $P_\star$ these analyses approximate what ambients may turn up inside what other ambients. To collect this information they extract the following information:

- An approximation of the contents of ambients:

  $$\mathcal{I} : \mathbf{Role} \rightarrow \mathcal{P}(\mathbf{Role} \cup \mathbf{Cap})$$

  Here $\mu' \in \mathcal{I}(\mu)$ means that μ' may be a sub-ambient of the ambient μ and $M \in \mathcal{I}(\mu)$ means that the capability M may be within the ambient μ.
- An approximation of the relevant name bindings:

  $$\mathcal{R} : \mathbf{Name} \rightarrow \mathcal{P}(\mathbf{Name})$$

  Here $\nu' \in \mathcal{R}(\nu)$ means that the (constant) name ν' may be bound to the (variable) name ν.

The judgements of the analysis take the form

$$(\mathcal{I}, \mathcal{R}) \models^{\mu} P$$

and express that when the sub-process P (of $P_\star$) is *initially* enclosed within an ambient with the identity $\mu \in \mathbf{Role}$ then $\mathcal{I}$ and $\mathcal{R}$ correctly capture the behaviour of P — meaning that $\mathcal{I}$ will reflect the contents of the ambients as P evolves inside $P_\star$ and $\mathcal{R}$ will contain all the bindings of names that take place.

3.2 Context Sensitive Version

The analysis results presented in this paper have been obtained using an improvement of this analysis scheme. Essentially the same information is extracted but whenever the analysis makes an entry into $\mathcal{I}$ or $\mathcal{R}$, corresponding either to the initial configuration or to a semantic action that changes the system, it also records the roles of the two ambients immediately enclosing the site of change. Thus, this 2-CFA analysis extracts the following information:

- A localised approximation of the contents of ambients:

$$\mathcal{I} : \mathbf{Role^2} \times \mathbf{Role} \to \mathcal{P}(\mathbf{Role} \cup \mathbf{Cap})$$

  Here $\mu' \in \mathcal{I}(\mu_{gp}, \mu_p, \mu)$ means that μ' may be a sub-ambient of the ambient μ whenever μ is inside μ_p and μ_p is inside μ_{gp}, and $M \in \mathcal{I}(\mu_{gp}, \mu_p, \mu)$ similarly means that the capability M may be within the ambient μ whenever μ is inside μ_p and μ_p is inside μ_{gp}.
- A localised approximation of the relevant name bindings:

$$\mathcal{R} : \mathbf{Role^2} \times \mathbf{Name} \to \mathcal{P}(\mathbf{Name})$$

  Here $\nu' \in \mathcal{R}(\mu_{gp}, \mu_p, \nu)$ means that the (constant) name ν' may be bound to the (variable) name ν inside μ_p whenever μ_p is inside μ_{gp}.

The judgements of this analysis have the form

$$(\mathcal{I}, \mathcal{R}) \models^{\mu_{gp}\mu_p\mu} P$$

and express that when the sub-process P (of $P_\star$) is *initially* enclosed within a nesting of the ambient roles $(\mu_{gp}, \mu_p, \mu) \in \mathbf{Role^3}$ then $\mathcal{I}$ and $\mathcal{R}$ correctly capture the behaviour of P — meaning that $\mathcal{I}$ will reflect the contents of the ambients as P evolves inside $P_\star$ and $\mathcal{R}$ will contain all occurring bindings of names. The "outermost" ambients μ_{gp} and μ_g are said to represent the *context*.

The amount of context information was chosen with care. For the analysis to be precise and maximally informative there must be enough information to uniquely locate all entities tracked in the ambient hierarchy of the model. For computational reasons, however, the amount of information cannot be very large.

We intend the analysis for uni-cellular systems and want it to distinguish compartments only up to their roles. Thus, we can safely assume all cellular compartments to be at fixed positions within the nesting hierarchy of the modelled cell. Under these assumptions, the analysis need consider only a few levels of context information: the maximal nesting depth of foreign bodies that may enter the cell - minus one.

Some biological processes, though, can only be modelled in BioAmbients if non-compartment entities are also encoded as ambients - thus increasing the nesting depth. During the course of our work we have considered numerous biological examples. For some of these, as indeed for the cholesterol case study of this paper, one level of context information has been sufficient while others have required two levels. We have, however, never found the need to consider three or more levels of context (perhaps because we have not considered phagocytosis[1]). Hence we have focused our work on contexts of length two.

4 A Model of Cholesterol Uptake

In the following we shall model and investigate an *endocytic pathway* facilitating a biological process called *receptor mediated endocytosis*. This process is common in mammalian cells, where it is a general mechanism for subsuming particles from the blood stream.

The best known example of this process is the *LDL degradation pathway* shown in Fig. 1. By this mechanism cells acquire the *cholesterol* required for the membrane synthesis that occurs during cell growth [1, 9]. It is also a common source of medical conditions as even small errors in the active components greatly increase the risk of cardiovascular disease. As we shall see our analysis is able to illustrate the more immediate effects of such component defects.

4.1 The LDL Degradation Pathway

Cholesterol is mainly obtained from *Low-density Lipoproteins* (LDLs), which carry cholesterol in the form of tightly packed *cholesteryl esters*.

Specialised *transmembranal receptor proteins* that perform free lateral diffusion in the *plasma membrane* of the cell recruit the LDLs from the blood. When the *extra-cellular* domain of such an *LDL receptor* encounters the *ApoB* domain exposed by an LDL particle the two particles will bind to each other by *complexation*.

Meanwhile, and independent of this, *clathrin particles* continuously assemble on the *cytosolic* side of the plasma membrane - thereby forcing it to form *clathrin coated pits* that grow progressively deeper until released into the cytosol as separate *clathrin coated vesicles*.

The diffusing receptors tend to associate with clathrin coated pits because their *intra-cellular* domain binds to complementary *adaptin* domains (molecules really) exposed by the clathrin coat. Such associated receptors and the LDLs that they bind, if any, are internalised when the coated vesicle is formed.

Once internalised, coated vesicles shred their clathrin coat and become *early endosomes*. At this stage the LDL/receptor complex is still intact. This changes, however, when the early endosome merges with a *late endosome*. The acidic environment in this compartment makes the receptors separate from the LDLs.

[1] The ingestion of large foreign bodies by macrophages.

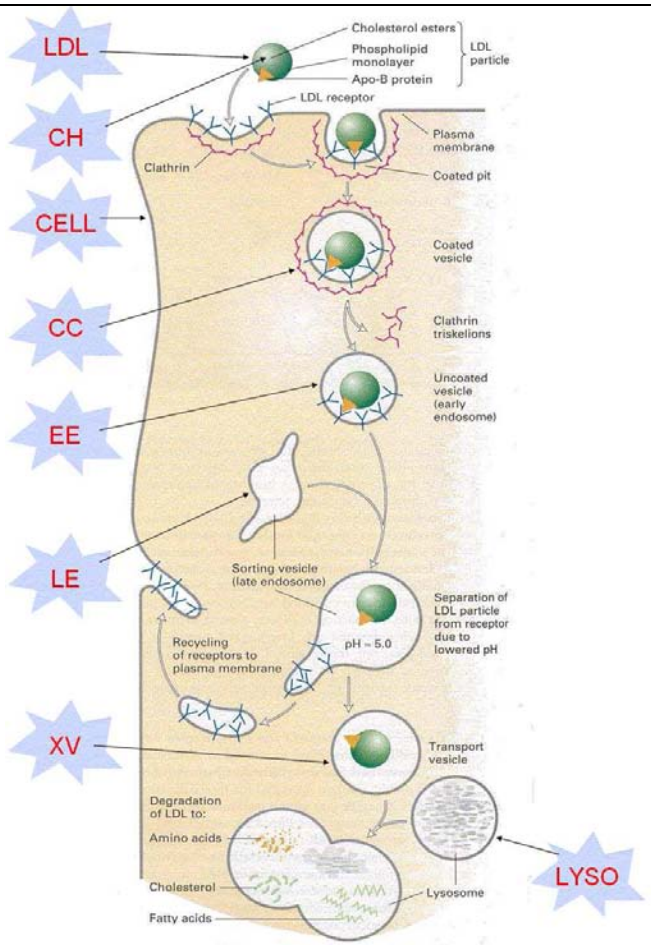


Fig. 1. The LDL cholesterol degradation process [9].

From the late endosomes the receptor proteins are recycled to the plasma membrane. The LDL molecules, however, are transferred by vesicles to *lysosomes* where they are finally hydrolysed to free the necessary cholesterol.

4.2 The BioAmbients Model

In order to subject this biological system to formal analysis we have modelled it in the BioAmbients language as shown in Table 4.

In accordance with Regev's examples and guidelines we take the approach that each kind of physical compartment as well as each kind of multiprotein complex should correspond to one ambient role. Thus, we have the following correspondences:

- The LDL role models LDL particles.
- The EE role models early endosomes (true compartments).
- The LE role models late endosomes (true compartments).

$\mathsf{LipoProtein} \triangleq$
$[LDLrcpt\#!\{ApoB\}.\mathsf{enter}\ ApoB.\mathsf{enter}\ ee.\mathsf{enter}\ xv.\ syncXV\hat{}!\{Le\}.\ proc\hat{}?\{Hydr\}.$
$(\mathsf{expel}\ Hydr \mid [\mathsf{exit}\ Hydr]^{\mathrm{CH}})]^{\mathrm{LDL}}$
$\mathsf{EarlyEndo} \triangleq$
$[\mathsf{accept}\ ee \mid \mathsf{enter}\ AP2.\ syncCCEE\hat{}?\{le\}.\mathsf{exit}\ AP2.\mathsf{merge}{-}\ le]^{\mathrm{EE}}$
$\mathsf{ClathrinCoat} \triangleq$
$[EErcpt\hat{}?\{ap2\}.\mathsf{accept}\ ap2.\ syncCCEE\_!\{Le\}.\mathsf{expel}\ ap2]^{\mathrm{CC}}$
$\mathsf{XferVesicle} \triangleq$
$[\mathsf{accept}\ xv.\ syncXV\_?\{le\}.\mathsf{exit}\ le.\mathsf{merge}{-}\ lyso]^{\mathrm{XV}}$
$\mathsf{LateEndo} \triangleq$
$[\mathsf{merge}{+}\ Le.\mathsf{expel}\ Le \mid \mathsf{XferVesicle}]^{\mathrm{LE}}$
$\mathsf{Lysosome} \triangleq$
$[\mathsf{merge}{+}\ lyso.\ proc\_!\{hydr\}]^{\mathrm{LYSO}}$
$\mathsf{Cell} \triangleq$
$[\ LDLrcpt\#?\{apob\}.\mathsf{accept}\ apob.\ EErcpt!\{AP2\}.\mathsf{EarlyEndo}$
$+EErcpt\_!\{AP2\}.\ LDLrcpt\#?\{apob\}.\mathsf{accept}\ apob.\mathsf{EarlyEndo}$
$+EErcpt\_!\{AP2\}.\mathsf{EarlyEndo}$
$\mid \mathsf{ClathrinCoat}$
$\mid \mathsf{LateEndo}$
$\mid \mathsf{Lysosome}]^{\mathrm{CELL}}$

$\mathsf{System} \triangleq$
$(LDLrcpt)(EErcpt)(ApoB)(AP2)(ee)(cc)(lyso)(xv)(Le)(syncCCEE)(syncXV)(proc)(hydr)$
$\mathsf{LipoProtein} \mid \mathsf{Cell}$

Table 4. The BioAmbients encoding the LDL degradation pathway.

- The CC role models clathrin coats (coating the EE in the coated vesicle).
- The LYSO role models lysosomes (true compartments).
- The CELL role models cells (true compartments).
- The XV role models transfer vesicles (true compartments).
- The CH role models cholesterol.

When we can do so without ambiguity, we will use the abbreviated ambient roles also when referring to the biological entities that they model. As will be evident from the explanation below, the model emphasises the receptor dynamics that facilitates the initial LDL binding but (for lack of space) ignores the details of receptor recycling. Nonetheless this allows the analysis to highlight certain medical issues. In some places the model has explicit synchronisation points. These are not important for the accuracy of the model, but alleviate some of the imprecision inherent in the over-approximations of the static analysis. In Nature each compartment and reaction would be present in the thousands. The analysis we perform here, however, is not able to track the number of occurrences and therefore it suffices for us to model a single representative for each biological entity.

In Table 4 the LDL (in LipoProtein) is initially located outside of the CELL (in Cell). Here it offers an *ApoB* signal via the channel *LDLrcpt* that corresponds to the extra-cellular binding site of the transmembranal LDL receptor of the CELL.

At this stage the early endosome has not been formed yet. We model, however, the membrane patch and the transmembranal LDL receptors, which are later going to fold into the early endosome, as a process capable of evolving into the EE ambient. As explained, the clathrin coated early endosome may be formed with or without bound LDL particles. We model this as a non-deterministic external choice such that one of the following three binding scenarios may occur before the EE ambient is released:

1. The extra-cellular part *LDLrcpt* of the LDL receptor binds the *ApoB* signal of the LDL thus forcing LDL to enter the CELL. Subsequently the intra-cellular part *EErcpt* of the receptor is bound by the *AP2* domain exposed by the CC bound adaptins.
2. The intra-cellular part *EErcpt* of the receptor is bound by the *AP2* domain exposed by the CC bound adaptin. Subsequently the extra-cellular part *LDLrcpt* of the LDL receptor binds the *ApoB* signal of the LDL thus forcing LDL to enter the CELL.
3. The intra-cellular part *EErcpt* of the receptor is bound by the *AP2* domain exposed by the CC and the extra-cellular part *LDLrcpt* is never bound.

If the LDL is in place inside the CELL after the binding scenario it may enter the EE otherwise not. Either way, the internalisation of the clathrin coated pit may be completed by the EE entering the CC.

In Nature this internalisation process is atomic since the $[[[[]^{LDL}]^{EE}]^{CC}]^{CELL}$ (or $[[[]^{EE}]^{CC}]^{CELL}$) configuration arises instantaneously when the coated vesicle is completed and internalised. By modelling this as a sequence of events we are introducing modelling artifacts. Most importantly, for the LDL to enter the EE we have to allow it into the CELL, which is biologically unsound. We must keep this in mind when interpreting the analysis results. Also the EE must pass the CC in order to enter the CELL. We enforce this by synchronising the CC and the EE via an exchange of the token *Le* on the channel *syncCCEE*. Once synchronised the EE can freely leave the CC. This corresponds to the internalised early endosome shredding its clathrin coat.

Knowing the token *Le* from the synchronisation the EE is now able to merge with the LE. This releases the LDL into the LE from where it may enter the XV. Once the LDL is inside the XV they are able to synchronise by reusing the token *Le* for an exchange on the channel *syncLDLXV*. This synchronisation gives XV the ability to leave the LE taking the LDL cargo with it.

Finally, the XV may merge with the LYSO, thus releasing the LDL cargo into its final destination where it may be hydrolysed into CH.

4.3 Analysing the LDL Degradation Pathway

When subjecting this model to the context dependent analysis we obtain a result that can be represented graphically as shown in Fig. 2. Except for the special $\top$ node with triple borders, which represents the super-environment, the nodes represent ambients and the edges represent the containment relation $\mathcal{I}$. The

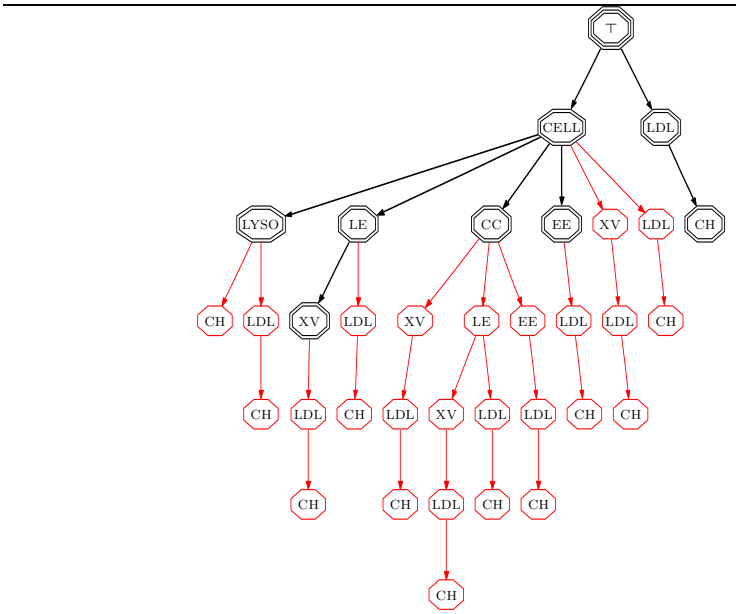


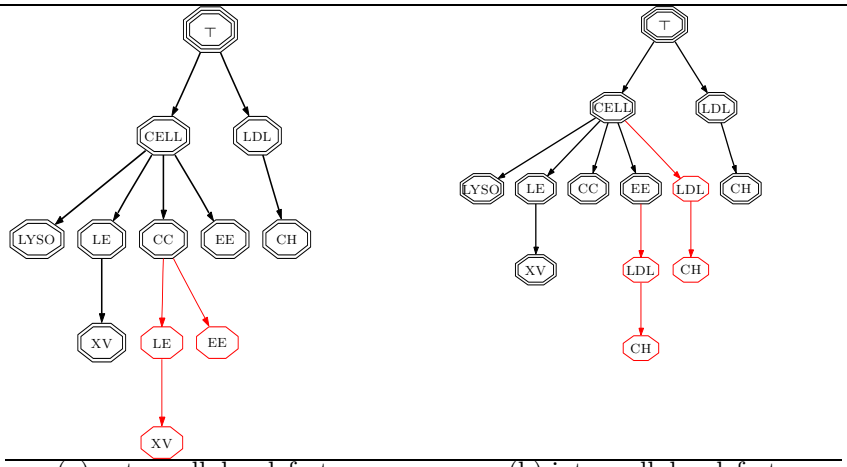
Fig. 2. Normal receptors.

nodes with double borders connected with bold edges represent the system in its initial state. The remaining nodes and edges account for the dynamic evolution of the system.

As the shown in Fig. 2 the analysis reveals a system that largely behaves as expected; in particular we notice that CH may be released inside LYSO. A few edges showing XV and LE occurring inside CC are analysis artifacts caused by the fact that the analysis is not able to take the sequencing of capabilities fully into account. Furthermore, in the biological system the LDL is never able to float freely in the cytosol (the top-level fluid of the CELL); as mentioned in Section 4.2 the occurrence in the figure is caused by a modelling artifact. Also note that the LDL *can* occur inside an uncoated EE after the CC has been shredded; in this case the double bordered EE represents both the initial configuration and a later stage of evolution.

Disorders. Some mammals suffer from the inherited disorder *familial hypercholesterolemia*, which dramatically increases the risk of the cardiovascular disease *atherosclerosis*. This disorder is caused by defects in the LDL receptor proteins that originate from inherited mutations.

When such a defect is located in the extra-cellular part of the receptor it is no longer able to bind an LDL particle. We model this phenomenon simply by introducing a spelling mistake in the sending end of the *LDLrcpt* channel, i.e. $(LDLrcpt\hat{}?\{apob\}, LDLrcp\_!\{ApoB\})$. As can be seen from figure 3(a) the analysis reveals that the cell can no longer internalise LDL particles. It still internalises early endosomes but only empty ones (EE may occur inside CC within CELL but EE carries no LDL cargo). Still, the LE occurring inside CC is an analysis artifact.



(a) extra-cellular defects (b) intra-cellular defects

Fig. 3. Analysis results for defect systems.

Finally, when the defect is located in the intra-cellular part of the receptor protein it can bind but not internalise an LDL particle. Again we model this by introducing a spelling mistake in the receiving end of the *EErcpt* channel, i.e. $(EErcp?\{ap2\}, EErcpt!\{AP2\})$. The analysis result shown in Fig. 3(b) shows the effect: LDL may occur inside EE but EE never enters CC. Again, the LDL floating freely in the cytosol is caused by a modelling artifact.

These results indicate that the system, as modelled here, cannot perform its normal function if the receptors are somehow defective. At this level of interpretation the the results coincide completely with biological reality. Since the analysis is very efficient this leads us to hypothesise that this approach can be used as a first means to identify pathways affected by this kind of system perturbations.

5 Conclusion

In this paper we have demonstrated the ability of static analysis to provide interesting information about the behaviour of biological systems as formalised in BioAmbients. In particular, the analysis has clearly pinpointed behavioural differences between biological systems that correctly perform the LDL degradation process for supplying cells with cholesterol, and mutated or otherwise defect systems that are likely to lead to diseases.

Other calculi with stochastic elements, such as stochastic π [19] and PEPA [2], have been successfully used for modelling and analysing regulatory feed-back and feed-forward mechanisms governing gene expression. As far as we are aware, however, our paper constitutes the first successful demonstration of the potential of programming language technology for investigating medical disorders that affect cellular transport systems.

Other modelling languages, more similar in aims and scope to BioAmbients, are also available [4, 17]. The Brane Calculus by Cardelli [4] is undeniably a

stronger abstraction of membrane based systems. It is, however, also more complex and it is not clear if a static analysis retaining both the high precision and low complexity of the one outlined here can be developed for it.

Once the modelling methodology and analysis algorithms were in place, the modelling and analysis tasks described were completed with relative ease. This suggests that BioAmbients-like calculi and static program analysis together provide a strong tool for the early investigation of transport-centred systems such as endocytic and biosynthetic-secretory pathways - in particular for the identification of pathways affected by system perturbations.

As mentioned, the static analysis used here mainly targets uni-cellular systems. This potentially rules out the analysis of immune responses to complex cell types. Consider, for example, a model comprising three eukaryotic cell types - two belonging to the immune system of a given host and the third being an invader. If one host cell, the defender, ingests the invader by phagocytosis three levels of context would be required to track all invader compartments during the digestion process. However, as even cells of different types have identical organelles (up to their roles) it would require another four to five levels of context to distinguish the defender host cell from the non-defender host cell when tracking the ingested compartments.

In terms of specification this is not a problem as our 2-CFA development generalises straightforwardly to k-CFA of higher magnitudes. Thus, in principle, the presented approach extends to the analysis of multi-cellular systems also. It is unclear, however, if the resulting algorithm would be computationally viable.

References

1. B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. *Molecular Biology of The Cell.* Garland Science, 4th edition, 2002.
2. M. Calder, J. Hillston, and S. Gilmore. Modelling the influence of rkip on the erk signalling pathway using the stochastic process algebra pepa. In *Proc. of Bio-CONCUR'04*, ENTCS. Elsevier, To appear.
3. L. Cardelli. Bioware languages. In *Computer Systems - Papers for Roger Needham.* Springer, 2003.
4. L. Cardelli. Brane calculi. In *Proc. of Bio-CONCUR'03*, ENTCS. Elsevier, To appear.
5. L. Cardelli and A. D. Gordon. Mobile ambients. *Theoretical Computer Science*, 240(1):177–213, 2000.
6. C.Bodei, P.Degano, C. Priami, and N. Zannone. An enhanced cfa for security policies. In *Proc. of the Workshop on Issues on the Theory of Security (WITS'03)*, 2003.
7. F. Levi and S. Maffeis. An abstract interpretation framework for analysing Mobile Ambients. In *Proc. of Static Analysis Symposium 2001*, volume 2126 of *LNCS*, pages 395–411. Springer, 2001.
8. F. Levi and D. Sangiorgi. Controlling interference in ambients. In *Proc. of Principles of Programmng Languages 2000*, pages 352–364. ACM Press, 2000.
9. H. Lodish, A Berk, P. Matsudaira, C. A. Kaiser, M. Krieger, M. P. Scott, S. L. Zipursky, and J. Darnell. *Molecular Cell Biology.* W.H. Freeman and Company, 4th edition, 1999. Fig. 6 reprinted with the permission of the publishers.

10. R. Milner. *Communicating and Mobile Systems: The π-Calculus*. Cambridge University Press, 1999.
11. F. Nielson, H. Riis Nielson, and C. Hankin. *Principles of Program Analysis*. Springer, 1999.
12. F. Nielson, H. Riis Nielson, and R. R. Hansen. Validating firewalls using Flow Logics. *Theoretical Computer Science*, 283(2):381–418, 2002.
13. F. Nielson, H. Riis Nielson, C. Priami, and D.S. da Rosa. Control flow analysis for BioAmbients. In *Proc. of Bio-CONCUR'03*, ENTCS. Elsevier, To appear.
14. H. Riis Nielson and F. Nielson. Shape analysis for mobile ambients. In *Proc. of Principles of Programmng Languages 2000*, pages 142–154. ACM Press, 2000.
15. H. Riis Nielson, F. Nielson, and M. Buchholtz. Security for mobility. In *FOSAD'01/'02 Tutorial Lecture Notes*, volume 2946 of *LNCS*. Springer, 2004.
16. H. Riis Nielson, F. Nielson, and H. Pilegaard. Spatial analysis of BioAmbients. In *Proc. of Static Analysis Symposium 2004*, volume 3148 of *LNCS*, pages 69–83, 2004.
17. C. Priami and P. Quaglia. Beta binders for biological interactions. In *Proc. of Computational Methods in Systems Biology 2004*, Lecture Notes in Bioinformatics. Springer, To appear.
18. A. Regev. *Computational Systems Biology: A Calculus for Biomolecular Knowledge*. PhD thesis, Tel Aviv University, 2003.
19. A. Regev, W. Silverman, and E. Shapiro. Representation and simulation of biochemical processes using the π-calculus process algebra. In *Proc. of Pacific Symposium on Biocomputing 2001*, pages 459–470, 2001.
20. Aviv Regev, Ekaterina M. Panina, William Silverman, Luca Cardelli, and Ehud Shapiro. Bioambients: An abstraction for biological compartments. *Theoretical Computer Science*, 325(1):141–167, September 2004.