

*На правах рукописи*

**Тарасов Денис Станиславович**

**Формальные языки для представления  
биохимических процессов**

**03.00.04 - биохимия**

**05.13.18 – математическое моделирование,  
численные методы и комплексы программ**

Автореферат  
диссертации на соискание ученой степени  
кандидата биологических наук

Казань, 2007

Работа выполнена на кафедре генетики ГОУ ВПО «Казанский государственный университет им В.И. Ульянова-Ленина», г. Казань.

Научный руководитель: *кандидат биологических наук  
Акберова Наталья Ивановна*

Официальные оппоненты: *доктор биологических наук  
Чернова Ольга Александровна*

*доктор технических наук  
Латыпов Рустам Хафизович*

Ведущая организация: *Институт биологии гена РАН,  
г. Москва*

Защита состоится 26 апреля 2007 г. в 13 ч 00 мин. на заседании диссертационного совета Д. 212.081.08 при Казанском государственном университете по адресу г. Казань Кремлевская 18

С диссертацией можно ознакомиться в Научной библиотеке им. Н.И. Лобачевского Казанского государственного университета

Автореферат разослан “ \_\_\_ ” марта 2007 года

Ученый секретарь  
диссертационного совета,  
доктор биологических наук

З.И. Абрамова

ая Пушкинская школа-конференция молодых ученых "Биология - наука XXI века".- Пушино.- 2005.- с. 223.

28. Тарасов Д.С. Интерпретатор языка CONCEPTSYSTEM//Программа для ЭВМ. Зарегистрирована в Реестре программ для ЭВМ 18.01.2007. Свидетельство о регистрации № 2007610350

### Актуальность темы

В настоящий момент существует тенденция к расширению использования *формальных языков* для представления биологических знаний разного рода (исходных экспериментальных данных, гипотез, теорий). Другими словами имеет место процесс кодификации знаний - перевода информации в форму, пригодную для компьютерной обработки. Эта тенденция обусловлена рядом причин.

1. *Проблема хранения и поиска информации.* Централизованные базы данных биологической информации, будут наращивать свои объемы, и их роль будет постоянно возрастать. Типичная на сегодня стратегия работы с этими базами "найти информацию -> загрузить к себе в компьютер -> проанализировать" становится с ростом объемов данных непрактичной. Поэтому в базы данных будут посылаться не запросы на определенные сведения, а программы, производящие анализ данных на месте и передающие только конечные результаты. Эти операции невозможны, если данные представлены только на естественном языке.

2. *Моделирование биохимических процессов.* Предсказание эффектов взаимодействия различных лекарственных препаратов *in vivo*, изучение регуляции каскадных и циклических процессов, выяснение механизмов развития наследственных заболеваний – это далеко не полный перечень точек приложения компьютерного моделирования в биохимии. Для того чтобы модель была понятна компьютеру, необходимо использовать формальный язык.

4. *Классификация и систематика.* Классификация организмов на основе последовательности ДНК является важнейшим методом в современной биологической систематике. Большинство методов подобного рода ориентируются на сравнение одного короткого фрагмента ДНК у различных организмов. Этот подход может давать некорректные результаты, поскольку короткий фрагмент ДНК не отражает всю совокупность признаков, присущих организму. Поэтому необходимо производить сравнение больших областей ДНК с точки зрения их значения для того или иного организма. Количественное применение такого метода требует наличия способов формального описания «смысла» фрагментов генома, их связи с фенотипом, что, в свою очередь, создает потребность в средствах формального описания фенотипических признаков.

5. *Обработка результатов эксперимента.* В молекулярной биологии происходит рост объемов экспериментальных данных. Для обработки данных необходимы программные средства. Часто для этой цели применяются готовые программные пакеты с закрытым кодом. Их

использование становится причиной "отчуждения" научного работника от научного процесса, когда собственно сам процесс обработки данных и формирования выводов оказывается скрытым от исследователя внутри "черного ящика" программного пакета. В результате снижается степень понимания исследователем методов обработки данных, а сами эти методы применяются неправильно, поскольку компьютерные программы не способны сами выбрать подходящие средства анализа информации. Как следствие, падает уровень научной работы. Такая ситуация создает необходимость в создании специального языка программирования, ориентированного на обработку биохимических данных, который исследователи могли бы использовать в повседневной работе.

6. *Понимание информационной природы живого.* Ряд авторов предлагает рассматривать живую клетку как вычислительное устройство или совокупность вычислительных устройств, выполняющих программу, записанную на клеточном языке. В этом случае для понимания функционирования живой клетки необходимо изучить архитектуру клетки как вычислительного устройства, а также выяснить принципы построения клеточного языка. Это требует создания некоторого формального языка описания программ для клеточного вычислительного устройства.

Вышеизложенное демонстрирует потребность в разработке новых формальных языков для молекулярной биологии, которые были бы пригодны одновременно для создания баз данных, моделирования биохимических реакций, биологической систематики и создания программ обработки данных. Очень желательно, чтобы все эти свойства сочетались в одном языке, или, существовало бы семейство языков, объединенных принципами построения и сходным синтаксисом. Необходимость унификации средств формального представления биохимических знаний для всех указанных выше целей, а также соображения теоретического характера, высказанные рядом авторов, приводят к заключению о том, что перспективным является представление биохимических знаний в форме вычислительных процессов, записанных в виде программ на особом языке.

#### **Цель работы**

Решение проблемы создания средств формального представления биохимических знаний, пригодных для разработки баз данных, моделирования биохимических реакций и создания программ обработки биохимических данных.

19. Тарасов Д.С. Применение новых достижений молекулярной биологии при проектировании современных устройств микроэлектроники/Д.С. Тарасов, Н.И. Акберова//Новые методологии проектирования устройств микроэлектроники.- Владимир - 2002.- С. 75-76
20. Тарасов Д.С. Технологии молекулярной биокibernетики: использование системы CDPL/CDS в конструировании ПЦР-амплификаторов/Д.С. Тарасов, Н.И. Акберова //III Научная конференция молодых ученых, аспирантов и студентов научно-образовательного центра КГУ "Материалы и технологии XXI века".- Казань.- 2003.- С. 84
21. Тарасов Д.С. Язык описания клеточных программ CDPL-1 и его применение/Д.С. Тарасов, Н.И. Акберова, А.Ю. Леонтьев//6-ая Пущинская школа-конференция молодых ученых "Биология - наука XXI века".- Пущино.- 2002.- С. 195-196
22. Tarasov D.S. The model of molecular biological computational device and its application to automatic genome annotation/D.S. Tarasov, N.I. Akberova, A.Y. Leontiev//International Moscow conference on computational molecular biology.- М.- 2003.- P. 225-226
23. Тарасов Д.С. Компьютерное моделирование структурно-функциональной организации ogi-сайтов бактерий/Д.С. Тарасов, Н.И. Акберова//12-я международная конференция "Ферменты микроорганизмов", Казань.- 2001.- с. 83-84
24. Тарасов Д.С. Регуляция и контроль инициации репликации: лингвистический подход/Д.С. Тарасов, Н.И. Акберова, А.Ю. Леонтьев//Материалы XI международной научной студенческой конференции «Студент и научно-технический прогресс»:Биология.- Новосибирск.- 2002.- С. 149-150
25. Леонтьев А.Ю. Алгоритм построения образа функциональных областей генома/А.Ю. Леонтьев, Д.С. Тарасов//Материалы международной научной конференции, посвященной 70-летию образования зооинженерного факультета. Казанская государственная академия ветеринарной медицины.- 2000.- С. 203-205
26. Тарасов Д.С. Язык представления описания морфологии грибов на примере Trichoderma/Д.С. Тарасов, Р.И. Тухбатова//9-ая Пущинская школа-конференция молодых ученых "Биология - наука XXI века".- Пущино.- 2005.- С. 331.
27. Шишкин А.В. Построение нетривиальной классификации грибов рода Trichoderma/А.В. Шишкин, Р.И. Тухбатова, Д.С. Тарасов// 9-

International conference of bioinformatics and genome regulation and structure.- Novosibirsk.- 2004.- P. 146-149

10. Tarasov D.S. Architecture of cell device/D.S. Tarasov, A.Y. Leotiev, N.I. Akberova//Third International conference of bioinformatics and genome regulation and structure.- Novosibirsk.- 2002.- P. 216-218
11. Акберова Н.И. Исследование симметричной структуры геномов вирусов HSV/Н.И. Акберова, Д.С. Тарасов //Первая всероссийская Internet-конференция "Компьютерное и математическое моделирование в естественных науках".- Тамбов.- 2001.- С. 23
12. Акберова Н.И. Компьютерный дизайн ПЦР-праймеров различной специфичности/Н.И. Акберова, Д.С. Тарасов //Четвертая всероссийская Internet-конференция "Компьютерное и математическое моделирование в естественных науках".- Тамбов.- 2002.- С. 32
13. Акберова Н.И. Метод симметричного моделирования структуры ДНК-текстов/Н.И. Акберова, А.Ю. Леонтьев, Д.С. Тарасов//Первая всероссийская Internet-конференция "Компьютерное и математическое моделирование в естественных науках".- Тамбов.- 2001.- С. 24
14. Тарасов Д.С. Архитектура клеточного устройства и гибридные биокбернетические системы/Д.С. Тарасов, Н.И. Акберова//7-ая Пушкинская школа-конференция молодых ученых.- Пушкино.- 2003.- С. 256
15. Тарасов Д.С. Компьютерный дизайн праймеров для ПЦР/Д.С. Тарасов, Н.И. Акберова//II научная конференция молодых ученых, аспирантов и студентов научно-образовательного центра КГУ.- Казань.- 2001.- С. 91
16. Тарасов Д.С. Молекулярно-биологическое вычислительное устройство и клеточное киберпространство/Д.С. Тарасов, Н.И. Акберова//8-ая Пушкинская школа-конференция молодых ученых "Биология - наука XXI века".- Пушкино.- 2003.- С. 24
17. Тарасов Д.С. Объекто-ориентированная система описания, классификации и моделирования биологических объектов и ее применение к грибам рода *Trichoderma*/Д.С. Тарасов//XII Международная конференция студентов, аспирантов и молодых ученых "Ломоносов".- М.- 2005.- С. 40-41
18. Тарасов Д.С. Организация базы знаний для молекулярно-биологических исследований/Д.С. Тарасов, Н.И. Акберова //IV Научно-практическая конференция молодых ученых и специалистов Республики Татарстан.- Казань.- ,2001 с. 110

В соответствии с поставленной целью определены задачи диссертационной работы:

1. *Изучить* основные существующие языки программирования общего назначения и формальные языки, применяемые для моделирования биохимических реакций, с целью выявить набор свойств, которыми должен обладать специализированный формальный язык для представления и обработки биохимических знаний.
2. *Изыскать* принципы построения формального языка для представления и обработки биохимических знаний.
3. *Разработать спецификацию* для формального языка, или семейства языков, предназначенных для представления и обработки биохимических знаний, моделирования биохимических реакций и представления описаний живых организмов с целью создания более точных биологических систематик.
4. *Реализовать интерпретатор* для специфицированных языков, который мог бы осуществлять моделирование биохимических процессов по их описанию, представленному на разработанном языке, а также мог бы выполнять программы обработки экспериментальных данных.
5. *Проверить* функционирование языка и интерпретатора путем моделирования нескольких биохимических процессов, применения языка для создания систематических описаний некоторой группы живых организмов и написания программы для обработки экспериментальных данных.

#### **Научная новизна**

Впервые предлагается семейство формальных языков, предназначенных для представления биохимических знаний, моделирования биохимических реакций и представления описаний живых организмов с целью создания биологических систематик, основанное на принципах организации клеточного языка, полученных в результате построения и изучения модели функционирования клеточного вычислительного устройства.

Впервые предложена модель архитектуры клеточного устройства и языка описания клеточных программ. Построен язык высокого уровня (CDPL), принципы организации которого повторяют принципы организации клеточного языка. Создан интерпретатор программ на языке CDPL, позволяющий моделировать биохимические системы, записанные с помощью CDPL.

Впервые рассматривается использование представления описаний микробиологических объектов по морфологическим и биохимическим признакам в форме программ на специализированном (разработанном в данной работе) объектно-ориентированном языке программирования.

Впервые предлагается использовать классификацию программ для систематики биологических объектов. Рассмотрены возможности применения принципов объектно-ориентированного программирования для систематики биологических объектов. Разработан специализированный язык *ConceptSystem* для составления программ-описаний микробиологических объектов, создан интерпретатор данного языка и создана интерактивная среда разработки таких программ.

#### **Практическая значимость работы**

1. Интерпретатор языка CDPL может использоваться:

1.2. Для моделирования как отдельных биохимических реакций, так и сложных взаимодействий.

1.3 В учебном процессе:

1.3.1. Для иллюстрации функционирования биохимических систем

1.3.2. Для иллюстрации принципов кинетики ферментативных реакций

2. Новый метод представления описаний биологических объектов и построения систематик может быть использован для уточнения существующих биологических классификаций.

3. Язык *ConceptSystem* можно использовать для создания программ обработки экспериментальных данных.

#### **Апробация работы**

Результаты работы докладывались на международных конференциях *Bioinformatics of Genome Regulations and Structure'2002* и *2004*, *Moscow Conference on Computational Molecular Biology*, 2003.

Кроме того, результаты работы докладывались на 6-ой (2002), 7-ой (2003), 8-ой (2004) и 9-ой (2005) Пущинских школах-конференциях молодых ученых «Биология - наука XXI века», секция «Математическая биология», и XII Международной конференция студентов, аспирантов и молодых ученых «Ломоносов 2005».

#### **Публикации**

По теме диссертации опубликовано 28 печатных работ, в том числе 17 тезисов конференций, 4 трудов международных конференций, 6 статей в научных журналах (в т. ч. 1 в издании, рекомендованном ВАК), 1 зарегистрированная программа для ЭВМ.

данных программ, описывающих изменение морфологических и биохимических признаков гриба рода *Trichoderma*, что позволило сформулировать новый подход к его систематике.

#### **ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ**

1. Тарасов Д.С., Формат представления биологических описаний гриба и его применение на примере рода *Trichoderma*/Д.С. Тарасов, Р.И. Тухбатова, Н.И. Акберова, Ф.К. Алимова//Вестник Татарстанского отделения российской экологической академии.- 2005.- Т. 24.- №2.- С. 44-49
2. Тарасов Д.С. Молекулярно-биологическое устройство – принципы организации//Д.С. Тарасов, Н.И. Акберова//Ученые записки КГУ, серия естественные науки.- 2005.- Т.147.- кн. 2.- С. 180-196
3. Тарасов Д.С. Увеличение интеллектуальных ресурсов научного исследования в биологических областях/Д.С. Тарасов, Н.И. Акберова//Ученые записки КГУ, серия естественные науки.- 2006.- Т.-148.- кн 1.- С. 138-155
4. Тарасов Д.С. Применение принципов объектно-ориентированного программирования к описанию и классификации биологических объектов на примере грибов рода *Trichoderma*/Д.С. Тарасов, Н.И. Акберова, Р.И. Тухбатова, Ф.К. Алимова//Ученые записки КГУ, серия естественные науки.- 2006.- Т. 148.- кн. 3.- С. 125-148
5. Тарасов Д.С. Виртуальные машины для исследования молекулярно-биологических процессов/Д.С. Тарасов, Н.И. Акберова//Георесурсы.- 2006.- Т.21.- №. 4.- С. 45-48
6. Tarasov D.S. Virtual machine technology for studying living cells/ D.S. Tarasov N.I. Akberova//Georesources.- 2006.- in print.
7. Тарасов Д.С.Молекулярно-биологическое вычислительное устройство: архитектура и язык управления/Д.С. Тарасов, Н.И. Акберова//Новая Геометрия Природы.- Казань.- 2003.- С. 341-250
8. Тарасов Д.С.Применение концепции молекулярно-биологического устройства для создания современных обучающих программ нового поколения/Д.С. Тарасов, Н.И. Акберова//Новая Геометрия Природы.- Казань.- 2003.- С. 332-334
9. Tarasov D.S.A Language for modeling generic regulation in prokaryotes/D.S. Tarasov, A.Y. Leontiev , N.I. Akberova //4th

## ВЫВОДЫ

1. Проведенное сравнение существующих формальных языков и их применения к биохимическим задачам показало, что формальный язык биохимии должен обладать свойством биологической конгруэнтности – понятия и конструкции языка должны соответствовать биохимическим сущностям (объектам и процессам). В своей основе этот язык должен быть объектно-ориентированным языком программирования с поддержкой баз данных на уровне языка, сильной типизацией, интуитивно понятным синтаксисом и содержать средства обработки больших массивов данных (подобно имеющимся в языках APL и J)
2. Рассмотрение живой клетки как вычислительного устройства позволяет изучать архитектуру такого вычислительного устройства и соответственно построить язык программирования, оптимально подходящий для программирования такого устройства.
3. В работе разработана и представлена спецификация семейства из двух языков CPDL и ConceptSystem. Язык CPDL построен исходя из модели архитектуры биохимического вычислительного устройства и позволяет компактно описывать различные биохимические процессы. CPDL обладает простым и понятным синтаксисом, так что смысл программы на нем может быть понят практикующим биохимиком без длительного обучения и тренировки. ConceptSystem является развитием идей CPDL и основывается на тех же принципах, однако в него введены практические модификации, позволяющие использовать его для создания баз данных и программ обработки экспериментальных данных.
4. Разработан эмулятор биохимического вычислительного устройства, способный выполнять CPDL программы. Разработан интерпретатор языка ConceptSystem вместе с графической оболочкой, поддерживающей интерактивный режим работы.
5. На языке CPDL были описаны процессы протекания простой ферментативной реакции, регуляции лактозного оперона и литического каскада фага лямбда. С помощью языка ConceptSystem были написаны программы анализа данных о генетической структуре популяции, а также создана база

## Объем и структура диссертации

Диссертация состоит из введения, трех глав, заключения и списка литературы, изложена на 117 страницах. Список литературы содержит 103 наименования. Работа содержит 55 рисунков и 2 таблицы.

## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

### Введение

Во введении обосновывается актуальность проблемы, сформулированы цели и задачи, дан краткий обзор содержания диссертации, перечислены полученные новые результаты исследования.

### Первая глава

Первая глава представляет собой обзор литературы по теме работы. Она состоит из трех разделов.

В первом разделе рассматриваются понятия теоретической информатики, лежащие в основе данной работы. Разъясняются такие понятия, как «формальный язык», «представление информации», «интерпретация», «абстрактная машина». Далее рассматриваются их приложения к молекулярной биологии: понятие о клеточном языке и моделирование биохимических реакций как дискретных процессов.

Во втором разделе дается обзор основных парадигм программирования на примере четырех различных языков. Рассматривается, как в этих языках решаются задачи обработки данных биохимического эксперимента, моделирования молекулярно-биологических систем и создания баз данных биологической информации.

Третий раздел посвящен обзору существующих формальных языков для представления биохимических процессов.

### **Вторая глава. Язык описания молекулярно-биологических систем и его интерпретатор для моделирования биохимических процессов.**

Во второй главе излагается разработанная модель архитектуры клеточного устройства и описывается созданный на ее основе язык описания клеточных программ CPDL.

Известно, что любая информационная система может быть рассмотрена с двух точек зрения. Первый подход имеет дело с **физической организацией** системы, а второй – с ее **логической организацией**. Живая клетка как информационная система тоже может быть рассмотрена с двух точек зрения.

Ряд авторов предлагает рассматривать живую клетку как вычислительное устройство или совокупность вычислительных устройств, выполняющих программу, записанную на клеточном языке. Живая клетка содержит программу в форме последовательности нуклеотидных звеньев ДНК. Практически использовать эту форму для представления биохимических знаний неудобно. Проводя аналогию с компьютерами, можно сказать, что ДНК содержит своеобразные «машинные коды», описывающие низкоуровневые операции клеточных механизмов. Однако можно предположить, что существуют клеточные языки «высокого уровня», в которых части ДНК-кода организованы в более крупные семантические единицы.

Изучив принципы логической организации клеточного языка, мы можем построить искусственный язык высокого уровня, выражения которого в равной степени пригодны для чтения человеком и компьютером и который по своей природе хорошо подходит для представления биохимических знаний.

Чтобы изыскать принципы построения такого языка, нужно сначала создать модель архитектуры клеточного вычислительного устройства, а затем сформулировать принципы языка, оптимально подходящего для программирования этого устройства.

#### Модель архитектуры клеточного устройства

Построенная модель клеточного устройства изображена на Рис 1. Ее организация определяется следующими принципами:

1. Нуклеиновые кислоты, белки и другие внутриклеточные компоненты образуют клеточное вычислительное устройство, состоящее из пулов данных, активных компонентов и пассивных компонентов.
2. Активные компоненты представлены процессами и метапроцессами. Клеточное вычислительное устройство может содержать несколько активных компонентов.
3. Процессы и метапроцессы являются источником активности в системе.
4. Активные компоненты содержат одновременно данные и программы, записанные на специальном клеточном языке в форме их химической структуры.
5. Процессы могут связывать, трансформировать и освобождать любые компоненты. Это происходит в соответствии с их встроенными программами
6. Метапроцессы активируются наличием или отсутствием определенных процессов. При активации метапроцессы могут

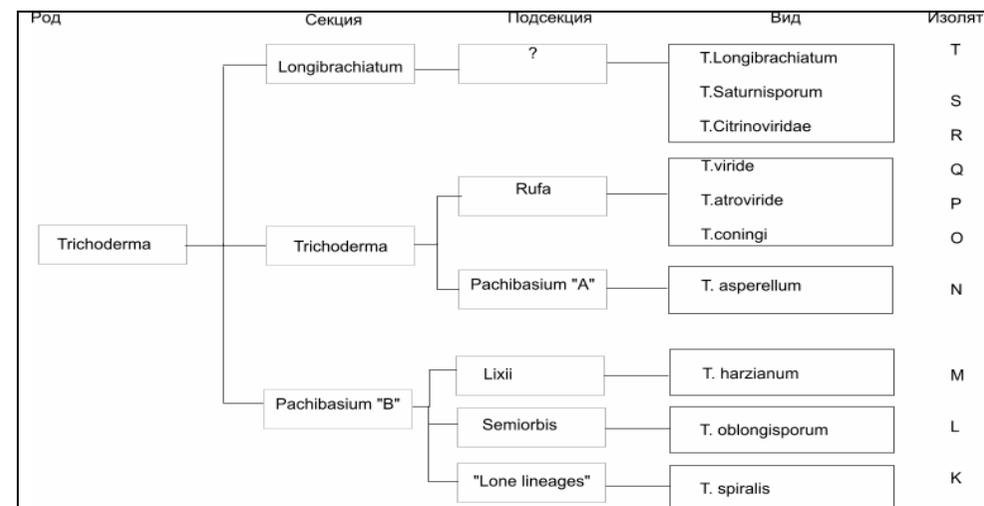


Рис 6. Фрагмент одной из «традиционных» классификаций рода Trichoderma. Справа приведено систематическое положение изолятов, которые использовались при построении ОО-систематики.

## ЗАКЛЮЧЕНИЕ

Разработан язык CDPL, обладающий способностью представления информации о метаболических путях в прокариотической клетке и отдельных метаболических путей в эукариотической клетке. Создан интерпретатор языка CDPL, пригодный для моделирования биохимических реакций и их регуляции.

Предложен новый способ описания биохимических и морфологических признаков живых организмов, основанный на принципах объектно-ориентированного программирования. Установлена предпочтительность описаний, полученных таким способом, по сравнению с принятыми в настоящее время описаниями на естественном языке.

Предложен новый метод построения биологической систематики, использующий принципы объектно-ориентированного программирования.

Разработан язык ConceptSystem для представления описания биологических объектов и создания программ обработки экспериментальных биохимических данных.

Создан интерпретатор, выполняющий программы на языке ConceptSystem

этом рисунок не содержит никакой информации о смысле различий между таксонами.

ОО-версия классификации (рис 6) содержит 11 абстракций на 10 экземпляров из которых имеется 9 видов 1 род и 1 промежуточный таксон. Между тем, как можно видеть из рисунка, ОО-классификация описывает взаимоотношения между рассматриваемыми организмами более подробно. В частности, в «стандартной» версии виды к которым относятся изоляты T, S и R поставлены в один ряд и дальнейшая информация об их взаимоотношениях отсутствует, в то время как в ОО-версии класс TExPig (R) является наследником класса TPigmented (S) и TExudated (T) из чего сразу следует что R имеет признаки как S так и T.

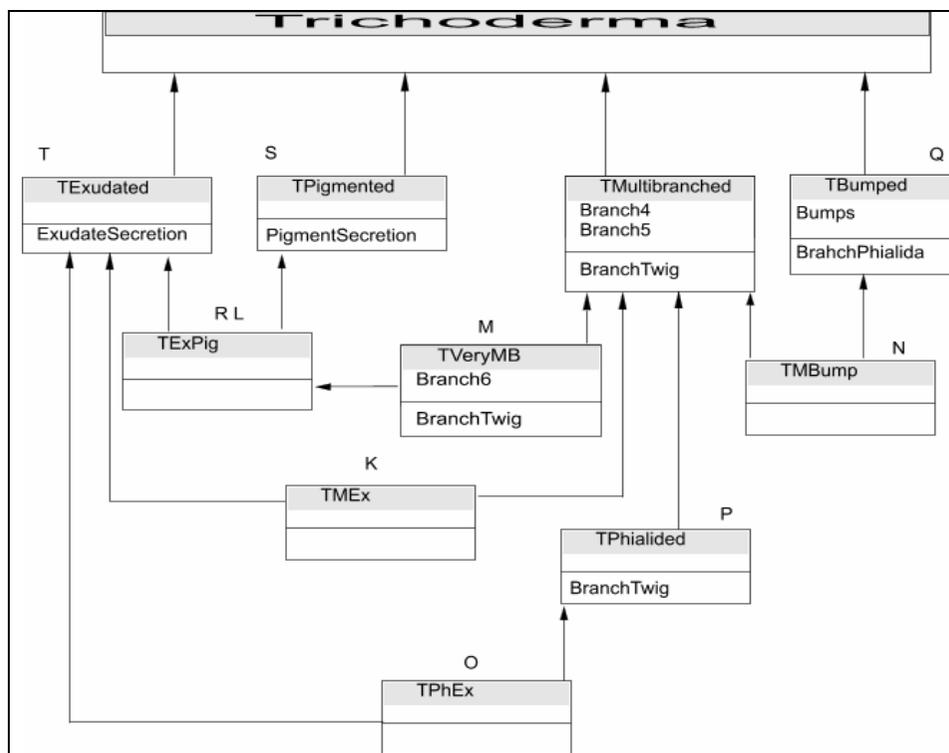


Рис 5. ОО-версия фрагмента классификации Trichoderma. Рядом с каждым классом обозначены буквами (К-Т) соответствующие экземпляры (изоляты) Trichoderma

запускать дополнительные процессы в соответствии со своими встроенными программами.

7. Все компоненты помещены в пулы данных. Компоненты из различных пулов не могут прямо взаимодействовать, но они могут перемещаться из одного пула в другой специальными процессами.



Рис 1. Модель архитектуры клеточного устройства.

### Язык описания клеточных программ CPDL

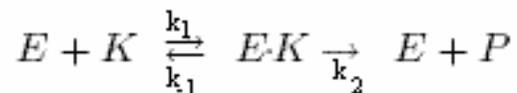
Принципы организации:

1. Функционирование клеточного вычислительного устройства описывается программой, записанной на языке описания клеточных программ CPDL.
2. Программы на CPDL содержат несколько частей. Эти части отвечают за описание начального состояния пула значений, типы данных, а также за встроенные программы для процессов и метапроцессов.
3. Существует прямое соответствие между сущностями языка и компонентами живой клетки.
4. Программы на CPDL обычно имеют декларативную и процедурную интерпретацию. Каждое выражение внутри встроенной программы активных компонентов имеет форму <действие>:- <условие>.

## Интерпретатор CPDL и моделирование биохимических процессов

Возможность использования интерпретатора CPDL лучше всего пояснить на простом примере.

Химическое уравнение типичной ферментативной реакции выглядит следующим образом:



Здесь K- субстрат, E- фермент, (EK) - комплекс фермента и субстрата, P – продукт,  $k_1, k_{-1}$  – константы скорости прямой и обратной реакций образования EK,  $k_2$  – константа скорости образования P.

Соответствующая CPDL программа выглядит так (после //комментарии):

```
Declare          //Объявления
tSubstrate = basic //Субстрат
tProduct  = basic  // Продукт
tEnzyme   = process // Фермент
Pool      //Содержание пула в начале
(1):tSubstrate(free) //Субстрат, концентрация 1
(0.1):tEnzyme(free) //Фермент, концентрация 0.1
tEnzyme process pE (tSubstrate,tProduct) //Описание функционирования фермента
bound(S,P,k1) :- not bound(P) //Связывает субстрат, со скоростью k1
transform(S,P,k2) :- bound(S). //Превращает связанный субстрат в продукт
release(_,P). // освобождает продукт
release(S,_,k_1). //освобождает субстрат со скоростью k_1
end.
```

Результаты работы программы изображены на Рис 2 и 3

Таксон высших порядков	Абстрактный класс
Организм	Экземпляр класса

Для представления описаний и классификации биологических объектов был разработан специальный объектно-ориентированный язык ConceptSystem, который был применен для описания и классификации микроскопических грибов рода *Trichoderma*.

### Применение разработанного языка для описания и классификации грибов рода *Trichoderma*

Существует множество групп живых организмов, систематика которых находится в процессе постоянного изменения. Одной из таких групп являются грибы рода *Trichoderma*. Грибы рода *Trichoderma* представляют ценность для применения в различных областях, в частности, для биологической защиты растений против грибных возбудителей болезней.

Предлагаемый в настоящей работе способ описания базируется на рассмотрении развития организмов рода *Trichoderma* как последовательности деления и дальнейшей специализации отдельных клеток

Описание представляет собой программу, которую необходимо задать исходной клетке (споре) чтобы получить ее развитие сначала в колонию *Trichoderma*, а в итоге снова в спору. Такое описание является естественным, поскольку отражает реальное положение вещей и позволяет разделить признаки на *свойственные единичным клеткам и появляющиеся в результате взаимодействия большого числа клеток*. Описание-программа следует логике развития самого организма, а не логике проведения наблюдений, особенностей экспериментов и исторических обстоятельств, нередко оказывающих влияние на обычные текстовые описания.

С использованием предложенных методов были построены программы описания и классификация для 10 изолятов грибов рода *Trichoderma*. На Рис.5. приведен пример фрагмента систематики, полученной с использованием *описаний-программ* и принципа множественного наследования. Буквами K,L,M,N,O,P,Q,R,S,T обозначены конкретные изоляты *Trichoderma*.

Для сравнения на рисунке 3.3. приведен один из существующих вариантов «стандартной» биологической классификации *Trichoderma*, который содержит 19 абстракций на 10 экземпляров из них 10 видов и 1 род и 8 промежуточных таксонов. Различные по смыслу таксоны имеют одинаковые названия (*Trichoderma* – одновременно род и секция). При

ООП может быть оболочкой, «оберткой» для любой существующей систематики. Конвертация любой систематики в ООП дает ей все преимущества использования формального языка без ущерба гибкости подхода. С другой стороны, возможно построение чистых ОО - систематик, основывающихся на принципах представленного подхода и использующих возможности ОО, для которых в биологических систематиках нет аналогов (например, множественное наследование).

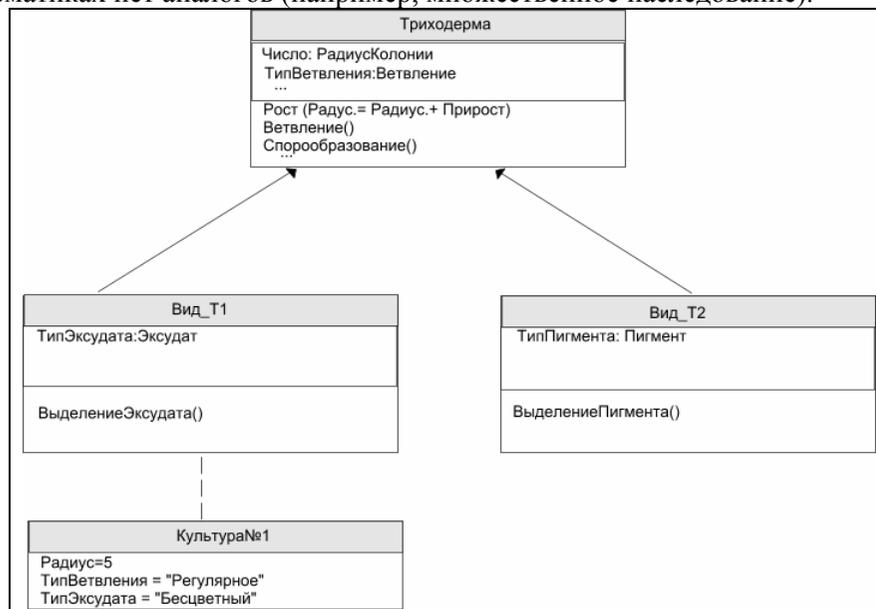


Рис 4. Наследование классов и биологическая систематика. «Триходерма» является абстрактным классом (не может иметь экземпляров), в котором определены атрибуты и методы, свойственные всем организмам рода *Trichoderma*. Классы Вид\_Т1 и Вид\_Т2 являются подклассами «Триходерма», определяющими дополнительные атрибуты. Классы Вид\_Т1 и Вид\_Т2 являются видами с точки зрения биологической систематики, поскольку из них могут образовываться конкретные экземпляры (Культура№1)

Таблица 1. Соответствие понятий биологической классификации и ООП

Биологическая классификация	ООП
Таксон	Класс
Вид	Неабстрактный класс, который может иметь экземпляры.

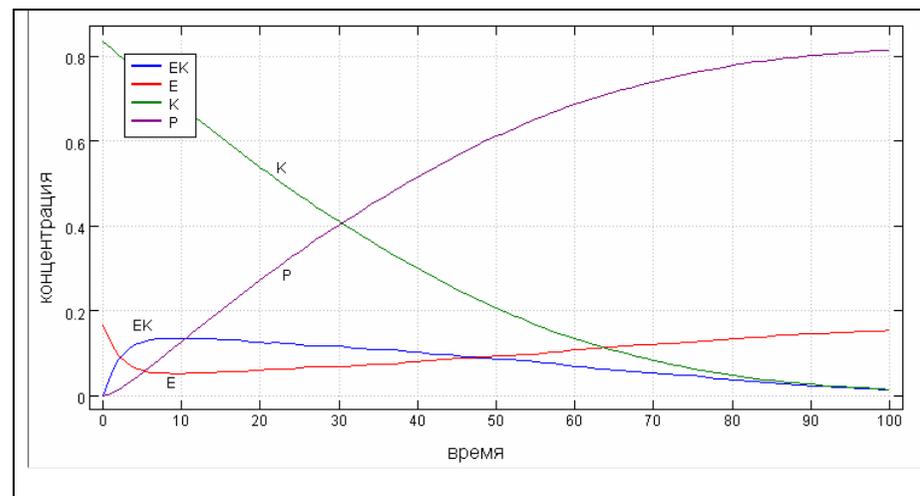


Рис. 2. Результаты выполнения программы на CDPL. Концентрация компонентов системы в зависимости от времени.

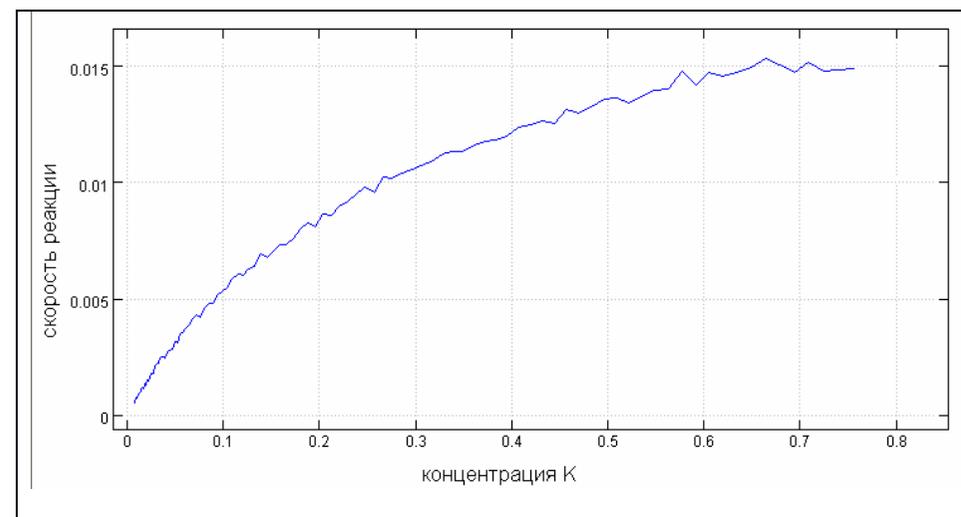


Рис 3. Результаты выполнения программы на CPDL. Зависимость концентрации продукта от скорости реакции

Помимо описания простой ферментативной реакции, пригодность разработанного языка была исследована путем описания с его помощью регуляции лактозного оперона, каскадной регуляции фага лямбда.

### Сравнение CPDL с другими языками описания биохимических моделей.

В последние годы активно разрабатываются языки описания биохимических моделей, такие как SBML и CellML. Эти языки являются производными популярного стандарта XML. Хотя теоретически информация представленная на этих языках доступна для прочтения человеком, с практической точки зрения, разбирать описания на этих языках затруднительно. Таким образом, данные языки являются в большей степени стандартами для автоматизированной обработки, они написаны для «выполнения машинами», а не для прочтения людьми. Языки типа SBML представляют интерес для специалистов по компьютерному моделированию и программистов, создающих программные пакеты, работающие с данными, представленными в SBML формате.

В отличие от этих языков, CPDL является концептуальным инструментом для людей, прежде всего для практических биохимиков. В отличие от громоздкого XML-описания CPDL программа компактна и легка для восприятия. Она имеет одновременно декларативную и процедурную семантику. Таким образом, CPDL подходит для формулировки идей о функционировании биохимических систем, передачи этих идей другим людям, т.е. является прежде всего средством мышления и общения. Одновременно CPDL является исполняемым языком, т.е. программы на нем могут быть выполнены на компьютере.

### Третья глава. Применение разработанных языковых средств для классификации организмов по биохимическим и морфологическим признакам.

#### Язык ConceptSystem

Представленный в предыдущей главе язык CPDL хорошо подходит для описания биохимических процессов, но плохо пригоден для построения баз данных и обработки экспериментальных биохимических данных. Поэтому на основе CPDL был разработан язык ConceptSystem, в котором предусмотрены средства для выполнения этих задач.

Язык ConceptSystem был применен для формализации описаний биохимических и морфологических признаков грибов рода *Trichoderma*, что дало возможность развить новый подход к систематике данной группы организмов.

### Формальное описание биохимических и морфологических признаков организма с помощью языка программирования.

Основным способом описания признаков организма или таксона до настоящего времени является описание на естественном языке, т.е. в форме последовательности предложений, записанных на (русском, английском или другом) языках общения людей. В то же время, известно, что любое описание характеристик вида должно включать в себя описания характеристик различных стадий жизненного цикла. Большинство существующих способов и форматов описания напоминает мгновенные фотографические снимки – срезы всех характеристик в определенные моменты времени. Такие описания не отражают достаточно полно специфику изучаемого организма.

Как говорилось ранее, живая клетка может быть рассмотрена как вычислительное устройство. В этом случае последовательную смену жизненных стадий можно описывать как развертывание вычислительного процесса, выполняемого под управлением клеточного компьютера. Для задания последовательности действий (последовательности вычислений) используются программы, записанные на специальных языках. Программы являются более компактным и естественным способом представления вычислительного процесса, чем последовательное описание всех состояний системы. Кроме того, содержание программы не зависит от конкретной стадии развития.

Основываясь на этом, мы можем построить язык программирования, предназначенный для представления описаний живых организмов. С помощью такого языка возможно не только представить динамику развития морфологических и физиологических характеристик отдельных организмов, но также и представить свойства абстрактных систематических единиц – видов и таксонов высших порядков.

#### Использование принципов объектно-ориентированного программирования для представления описаний и систематики биологических объектов

Если мы сопоставим объектно-ориентированное программирование (ООП) и биологическую систематику, то можем идентифицировать некоторые соответствия. Любая существующая биологическая система может быть представлена средствами ООП. Чтобы понять эти соответствия, обратимся к рисунку 4, на котором представлен фрагмент биологической классификации в форме диаграммы классов. Любому таксону может быть сопоставлен класс, определяющий общие характеристики данного таксона (табл 1).