
Exploring and Visualisation of Chemistry in Patent Documents

with **Marvin** and **Instant JChem**

Alexander Klenner
aklenner@scai.fraunhofer.de

Outline

Challenge & Motivation

Exploration Tools

NER, N2S, I2S

Workflow

UIMA (Unstructured Information Management)

UNICORE (Uniform Interface to Computing Resources)

Visualisation

PDF

Instant Jchem

Summary & Outlook

Challenge & Motivation

- 472.000 Chemistry Patents at EPO^[1]
- Patents contain chemical knowledge of 10+ decades
- Multimodal Information → Image and text
- Information is not easily human accessible,
 - Mostly unstructured
 - Hidden in noise
 - Different languages (e.g. Russian, Chinese...)
 - Not searchable (text & structure search)
- **No *on demand* large scale processing on the market!**

[1] <http://epo.org> (2011)

Exploration Task

- Pre-Processing

Separating Image and Text (chemoCR^[2])

Optical Character Recognition (Abbyy^[3])

- Processing

chemical NER (e.g. Linda^[4], OSCAR^[5], ProMiner^[6], Peregrine^[7])

Patent Zoning (e.g. Synthesis Detector)

- Post-Processing (planned)

Co-Reference solving

^[2] <http://www.scai.fraunhofer.de/en/business-research-areas/bioinformatics/products/chemocr.html>

^[3] <http://abbyy.com> ^[5] <http://sourceforge.net/projects/oscar3-chem>

^[6] <http://www.scai.fraunhofer.de/en/business-research-areas/bioinformatics/products/prominer.html>

^[7] <https://trac.nbic.nl/data-mining/>

Exploration: Pre-Processing

EP 1 464 708 A1

Description

FIELD OF THE INVENTION

5 [0001] The present invention relates to a process for the preparation of fludarabine phosphate (I), in particular to a process for the preparation of fludarabine phosphate from 2-fluoroadenine and 9-β-D-arabinofuranosyluracil using *Enterobacter aerogenes*.

10

15

20

(I)

TECHNOLOGICAL BACKGROUND

25 [0002] Fludarabine (9-β-D-arabinofuranosyl-2-fluoroadenine) (II) is a purine nucleoside antimetabolite resistant to adenosine deaminase, employed for the treatment of leukemia.

30

35

40

(II)

45 [0003] Fludarabine is usually administered as a pro-drug, fludarabine phosphate, which is also the natural metabolite. Fludarabine was firstly synthesised by Montgomery (US 4,186,378 and US 4,210,745) starting from 2-aminoadenine. The method comprised acetylation of 2-aminoadenine, reaction with a benzyl-protected chlorosugar, deacetylation of the amino groups, diazotization and fluorination of the 2-amino group followed by deprotection of the sugar residue.

50 [0004] Fludarabine phosphate can be obtained according to conventional phosphorylation methods, typically by treatment with trimethylphosphate and phosphoryl chloride. Recently, a method for preparing highly pure fludarabine, fludarabine phosphate and salts thereof has been disclosed by Tilstam et al. (US 6,46,322).

55 [0005] Enzymatic synthesis has been regarded as a valid alternative to conventional methods for the synthesis of nucleosides and nucleotides derivatives. EP 0 867 516 discloses a method for the preparation of sugar nucleosides from sugar 1-phosphates and nucleosides monophosphates by use of yeast cells having nucleoside diphosphate-sugar pyrophosphorylase activity. EP 0721 511 B1 discloses the synthesis of vidarabine phosphate and fludarabine phosphate by reacting an arabinonucleotide with an arylphosphate in the presence of a microorganism able to catalyse the phosphorylation of nucleosides. This method is particularly convenient in that it does not require purified enzymes, but it does not allow to synthesise vidarabine and fludarabine.

2

chemoCR



EP 1 464 708 A1

Description

FIELD OF THE INVENTION

5 [0001] The present invention relates to a process for the preparation of fludarabine phosphate (I), in particular to a process for the preparation of fludarabine phosphate from 2-fluoroadenine and 9-β-D-arabinofuranosyluracil using *Enterobacter aerogenes*.

10

15

20

(I)

TECHNOLOGICAL BACKGROUND

25 [0002] Fludarabine (9-β-D-arabinofuranosyl-2-fluoroadenine) (II) is a purine nucleoside antimetabolite resistant to adenosine deaminase, employed for the treatment of leukemia.

30

35

40

(II)

45 [0003] Fludarabine is usually administered as a pro-drug, fludarabine phosphate, which is also the natural metabolite. Fludarabine was firstly synthesised by Montgomery (US 4,186,378 and US 4,210,745) starting from 2-aminoadenine. The method comprised acetylation of 2-aminoadenine, reaction with a benzyl-protected chlorosugar, deacetylation of the amino groups, diazotization and fluorination of the 2-amino group followed by deprotection of the sugar residue.

50 [0004] Fludarabine phosphate can be obtained according to conventional phosphorylation methods, typically by treatment with trimethylphosphate and phosphoryl chloride. Recently, a method for preparing highly pure fludarabine, fludarabine phosphate and salts thereof has been disclosed by Tilstam et al. (US 6,46,322).

55 [0005] Enzymatic synthesis has been regarded as a valid alternative to conventional methods for the synthesis of nucleosides and nucleotides derivatives. EP 0 867 516 discloses a method for the preparation of sugar nucleosides from sugar 1-phosphates and nucleosides monophosphates by use of yeast cells having nucleoside diphosphate-sugar pyrophosphorylase activity. EP 0721 511 B1 discloses the synthesis of vidarabine phosphate and fludarabine phosphate by reacting an arabinonucleotide with an arylphosphate in the presence of a microorganism able to catalyse the phosphorylation of nucleosides. This method is particularly convenient in that it does not require purified enzymes, but it does not allow to synthesise vidarabine and fludarabine.

2

Exploration: Pre-Processing

EP 1 464 708 A1

Description

FIELD OF THE INVENTION

[0001] The present invention relates to a process for the preparation of fludarabine phosphate (I), in particular to a process for the preparation of fludarabine phosphate from 2-fluoroadenine and 9-β-D-arabinofuranosyluracil using *Enterobacter aerogenes*.

10

15

20

(I)

TECHNOLOGICAL BACKGROUND

25 [0002] Fludarabine (9-β-D-arabinofuranosyl-2-fluoroadenine) (II) is a purine nucleoside antimetabolite resistant to adenosine deaminase, employed for the treatment of leukemia.

30

35

40

(II)

45 [0003] Fludarabine is usually administered as a pro-drug, fludarabine phosphate, which is also the natural metabolite. Fludarabine was firstly synthesised by Montgomery (US 4,188,378 and US 4,210,745) starting from 2-aminoadenine. The method comprised acetylation of 2-aminoadenine, reaction with a benzyl-protected chlorosugar, deacetylation of the amino groups, diazotization and fluorination of the 2-amino group followed by deprotection of the sugar residue.

50 [0004] Fludarabine phosphate can be obtained according to conventional phosphorylation methods, typically by treatment with trimethylphosphate and phosphoryl chloride. Recently, a method for preparing highly pure fludarabine, fludarabine phosphate and salts thereof has been disclosed by Tilstam et al. (US 6,46,322).

55 [0005] Enzymatic synthesis has been regarded as a valid alternative to conventional methods for the synthesis of nucleosides and nucleotides derivatives. EP 0 867 516 discloses a method for the preparation of sugar nucleotides from sugar 1-phosphates and nucleosides monophosphates by use of yeast cells having nucleoside diphosphate-sugar pyrophosphorylase activity. EP 0721 511 B1 discloses the synthesis of vidarabine phosphate and fludarabine phosphate by reacting an arabinonucleotide with an arylphosphate in the presence of a microorganism able to catalyse the phosphorylation of nucleosides. This method is particularly convenient in that it does not require purified enzymes, but it does not allow to synthesise vidarabine and fludarabine.



[...]

[0003] Fludarabine is usually administered as a pro-drug, fludarabine phosphate, which is also the natural metabolite. Fludarabine was firstly synthesised by Montgomery (US 4,188,378 and US 4,210,745) starting from 2-aminoadenine.

The method comprised acetylation of 2-aminoadenine, reaction with a benzyl-protected chlorosugar, deacetylation of the amino groups, diazotization and fluorination of the 2-amino group followed by deprotection of the sugar residue.

[0004] Fludarabine phosphate can be obtained according to conventional phosphorylation methods, typically by treatment with trimethylphosphate and phosphoryl chloride. Recently, a method for preparing highly pure fludarabine, fludarabine phosphate and salts thereof has been disclosed by Tilstam et al. (US 6,46,322).

[0005] Enzymatic synthesis has been regarded as a valid alternative to conventional methods for the synthesis of nucleosides and nucleotides derivatives. EP 0 867 516 discloses a method for the preparation of sugar nucleotides from sugar 1-phosphates and nucleosides monophosphates by use of yeast cells having nucleoside diphosphate-sugar pyrophosphorylase activity. EP 0721 511 B1 discloses the synthesis of vidarabine phosphate and fludarabine phosphate by reacting an arabinonucleotide with an arylphosphate in the presence of a microorganism able to catalyse

the phosphorylation of nucleosides. This method is particularly convenient in that it does not require purified enzymes,

but it does not allow to synthesise vidarabine and fludarabine.

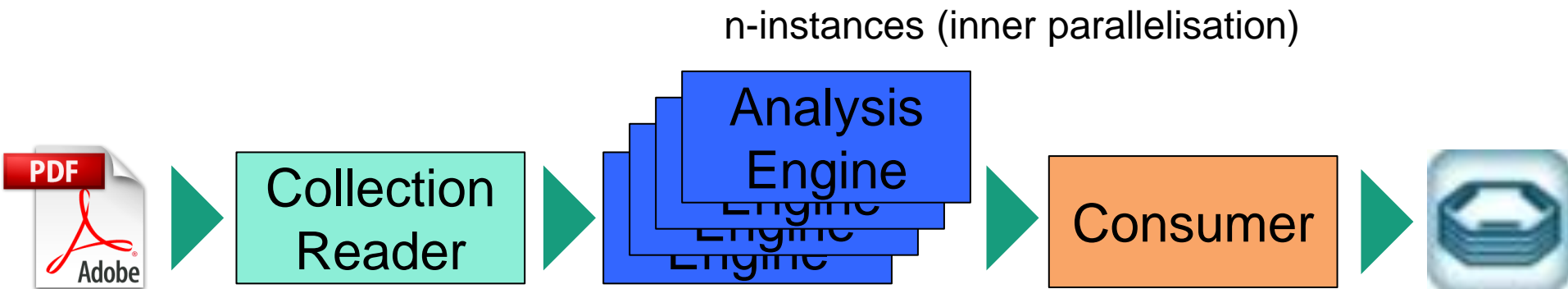
[...]

Exploration: Processing

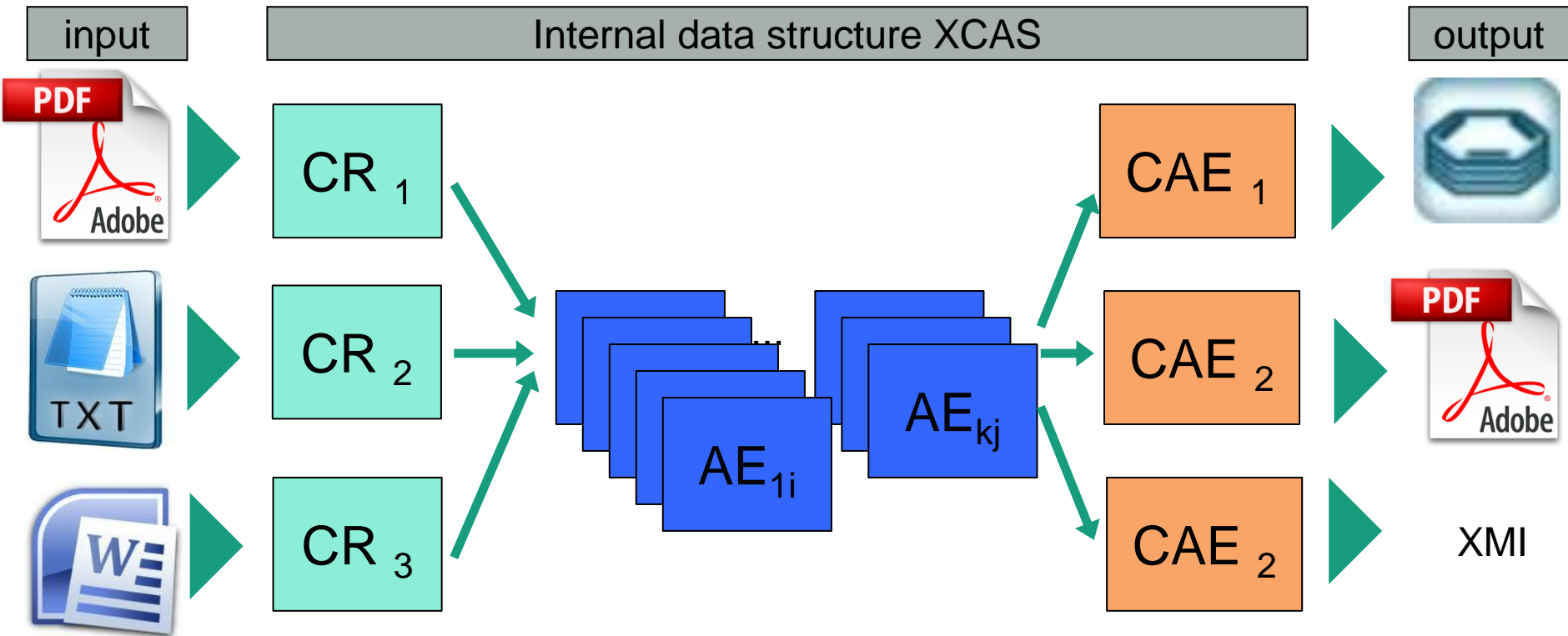
- Chemical NER
 - Linda CRF model trained for IUPAC terms
 - ProMiner Dictionary based chemistry tagger
 - OSCAR Open source NER Tagger
- Zoning
 - Synthesis Classifier
 - Claims / Description Tagger

Workflow – Unstructured Information Management (UIMA)

- Collection Reader (CR)
 - reads documents and converts into Framework CAS objects
- AnalysisEngines (AE)
 - adds annotations to the CAS objects
- ConsumerAE (CAE)
 - presents CAS (common analysis structure) object to user



Workflow – Unstructured Information Management (UIMA)



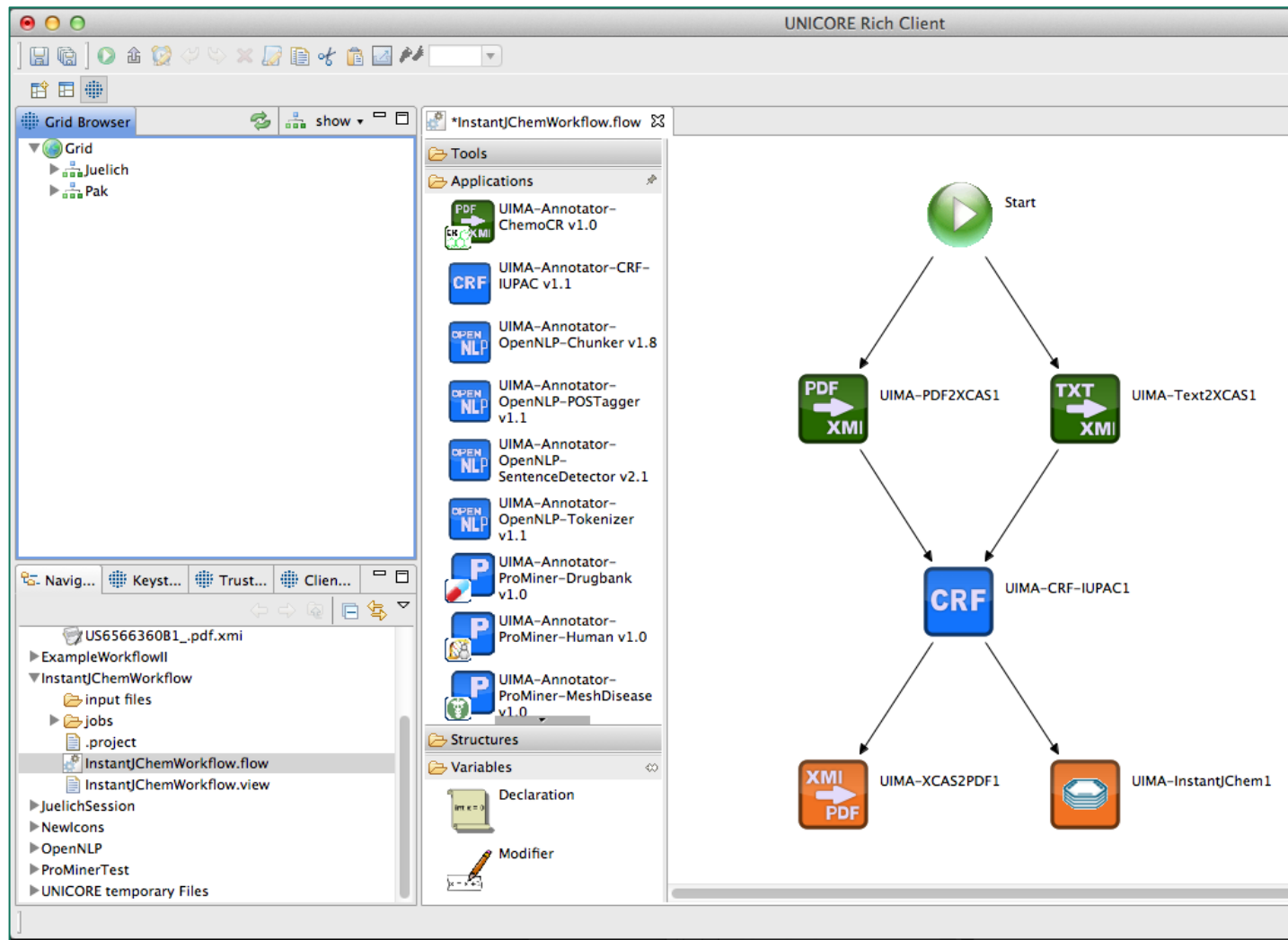
Workflow – UNICORE^[9]

- Access to grid resources
- Safe communications → certificates
- Outer parallelisation
- Scheduling of submitted jobs

- RichClient – Graphical User Interface
 - Drag&Drop functionality
 - Automatic Application Download
 - CAS-Editor
 - Web-Interface (in development)

[9] <http://unicore.eu>

Workflow – UNICORE Rich Client





30

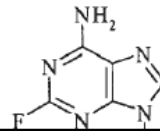
35

40

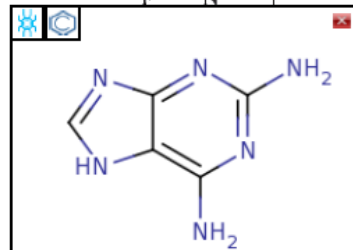
45

50

55



Created with Marvin API



2-aminoadenine

[0003] Fludarabine is usually administered as the monophosphate, which is also the natural metabolite. Fludarabine was firstly synthesised by [US 2,110,745] starting from 2-aminoadenine. The method comprised acetylation of the amino groups, protection of the sugar residue, deacetylation of the amino groups, diazotization and fluorination of the 2-amino group followed by deprotection of the sugar residue.

[0004] Fludarabine phosphate can be obtained according to conventional phosphorylation methods, typically by treatment with trimethylphosphate and phosphoryl chloride. Recently, a method for preparing highly pure fludarabine, fludarabine phosphate and salts thereof has been disclosed by Tilstam et al. (US 6,46,322).

[0005] Enzymatic synthesis has been regarded as a valid alternative to conventional methods for the synthesis of nucleosides and nucleotides derivatives. EP 0 867 516 discloses a method for the preparation of sugar nucleotides from sugar 1-phosphates and nucleosides monophosphates by use of yeast cells having nucleoside diphosphate-sugar pyrophosphorylase activity. EP 0721 511 B1 discloses the synthesis of vidarabine phosphate and fludarabine phosphate by reacting an arabinonucleotide with an arylphosphate in the presence of a microorganism able to catalyse the phosphorylation of nucleosides. This method is particularly convenient in that it does not require purified enzymes, but it does not allow to synthesise vidarabine and fludarabine.

Visualisation – PDF Writer



EP1464708A1_abby.pdf

2 / 11 58.2%

Comment Share

Bookmarks

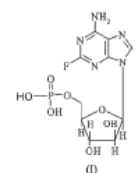
- ▼ Concept
 - ▶ InChI=1/C5H4FN5/c6-5-10-3(7)2-4(11-5)9-1-8-2/h1H(H37891011)/f/h8H7H2
 - ▶ na
 - ▼ InChI=1/C5H6N6/c6-3-2-4(9-1-8-2)11-5(7)10-3/h1H(H567891011)/f/h8H6-7H2
 - ▶ 2-aminoadenine p.2
 - ▶ 2-aminoadenine p.2

EP 1 464 708 A1

Description

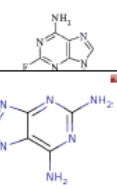
FIELD OF THE INVENTION

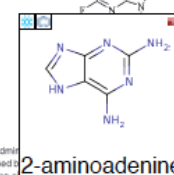
[0001] The present invention relates to a process for the preparation of fludarabine phosphate (I), in particular to a process for the preparation of fludarabine phosphate from 2-fluoro-9- β -D-arabinofuranosyladenine and 9- β -D-arabinofuranosyluracil using *Entrobacter aerogenes*.

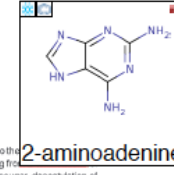


TECHNOLOGICAL BACKGROUND

[0002] Fludarabine (9- β -D-arabinofuranosyl-2-fluoro-9H-purine (II)) is a purine nucleoside antimetabolite resistant to adenosine deaminase, employed for the treatment of leukemia.







[0003] Fludarabine is usually administered as its phosphate salt, which is also the form in which it is used in the treatment of leukemia. Fludarabine was firstly synthesized by [10,745] starting from 2-aminoadenine. The method comprised acetylation of the amino groups, diazotization and fluorination of the 2-amino group followed by deprotection of the sugar residue.

[0004] Fludarabine phosphate can be obtained according to conventional phosphorylation methods, typically by treatment with trimethylphosphate and phosphoryl chloride. Recently, a method for preparing highly pure fludarabine, fludarabine phosphate and salts thereof has been disclosed by Tilsten et al. (US 6,46,322).

[0005] Enzymatic synthesis has been regarded as a valid alternative to conventional methods for the synthesis of nucleosides and nucleotides derivatives. EP 0 867 516 discloses a method for the preparation of sugar nucleotides from sugar 1-phosphates and nucleosides monophosphates by use of yeast cells having nucleoside diphosphate-sugar pyrophosphorylase activity. EP 0721 511 B1 discloses the synthesis of vidarabine phosphate and fludarabine phosphate by reacting an arabinonucleoside with an arylphosphate in the presence of a microorganism able to catalyse the phosphorylation of nucleosides. This method is particularly convenient in that it does not require purified enzymes, but it does not allow to synthesize vidarabine and fludarabine.

2

Visualisation – Instant JChem Writer

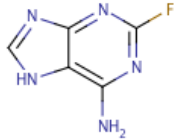
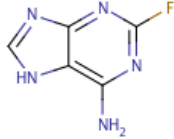
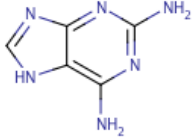
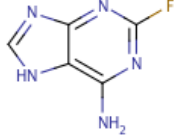
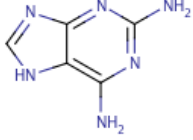


Instant JChem 5.9.2

1-19 / 19

test [as admin] | Grid view for instantJChemTable | Grid view for instanJChemTable

Query Browse

Cdid	Structure	Mol Weight	Formula	Field 5
1		153.12	C5H4FN5	EP1464708A1_.pdf
2		153.12	C5H4FN5	EP1464708A1_.pdf
3		150.14	C5H6N6	EP1464708A1_.pdf
4		153.12	C5H4FN5	EP1464708A1_.pdf
5		150.14	C5H6N6	EP1464708A1_.pdf

instantJChemTable: 19 out of 19 rows.

Summary

- UIMA framework compliance (active growing community)
- Implicit **parallelisation for large scale** applications
- Intuitive drag&drop workflow creation
- Various output formats (**InstantJChem**, MySQL, RDF-TripleStore)
- **Easy to integrate (your) 3rd party software in existing workflows**

Acknowledgments



Dr. Marc Zimmermann

Carina Haupt

Ludwig Wildner



Dr. Mathilde Romberg

Sandra Bergmann

Sebastian Demuth



Dr. Alexander Piechot

Dr. Christian Janssen



Guy Lonsdale

Thorsten Bartelth

Federal Ministry of Education and Research (01IH11012A)

Demo

<http://uima-hpc.org/demo/>