Project Prospect
（プロジェクト・プロスペクト）

理化学出版に意味論を導入する

リチャード・キッド

RSC | Advancing the Chemical Sciences

プロジェクト・プロスペクトの狙い

- 著者の科学的活動を保存する
- 科学をXMLとして捕捉する
- 意味を見出すために、オントロジーを応用する
- 出版プロセスに組み込む
- 読者に新しい視点を提供する
- 機械可読な化学

RSC | Advancing the Chemical Sciences

意味のある検索

なぜ化学では難しいのか

- 検索
  - 無料テキストまたは所有されたテキスト
- 関連文献
  - 基準の欠如

RSC | Advancing the Chemical Sciences

Introduction

The generation of bespoke pseudopeptide sequences that exhibit highly ordered secondary and tertiary structures in both solution and solid phase has developed into a highly competitive field of research in recent years, with the ability to predict the conformation of a given peptide sequence from knowledge of its primary structure an elusive goal. While much effort has been directed towards understanding the factors that control the secondary structure of α-peptides, the utility of peptides incorporating the β-amino acid structural motif has recently been investigated widely, most notably by Seebach[1] and Gellman.[2] For instance, Gellman *et al.* have shown that β-peptides derived from *trans*-2-aminocyclopentanecarboxylic acid (transpentacin) **1** adopt a helical stucture in both the solid state and in solution,[3] while Fülop *et al.* have shown that homo-oligomers of *cis*-2-aminocyclopentanecarboxylic acid (cispentacin) **2** form a sheetlike secondary structure in solution (Fig. 1).[4] The ability of mixed α,β-peptides containing both α-amino and β-amino acid derivatives to adopt a preferred conformation in solution has also been reported recently.[5]

We have shown extensively that the conjugate addition of lithiumamides derived from α-methylbenzylamine to α,β-unsaturated acceptors may be used for the asymmetric synthesis of β-amino acid derivatives.[6] This methodology has recently been utilised for the synthesis of (1*R*,2*S*,3*R*)-3-methylcispentacin **5** in >98% de and 98 ± 1% ee and (1*S*,2*S*,3*R*)-3-methyltranspentacin **7** in >98% de and 97 ± 1% ee by the kinetic resolution of *tert*-butyl (*RS*)-3-methylcyclopentene-1-carboxylate **3** with lithium (*S*)-*N*-benzyl-*N*-α-methylbenzylamide (Scheme 1).[7]

The protocol that we use to understand fully the stereoselectivity observed in these kinetic resolution reactions

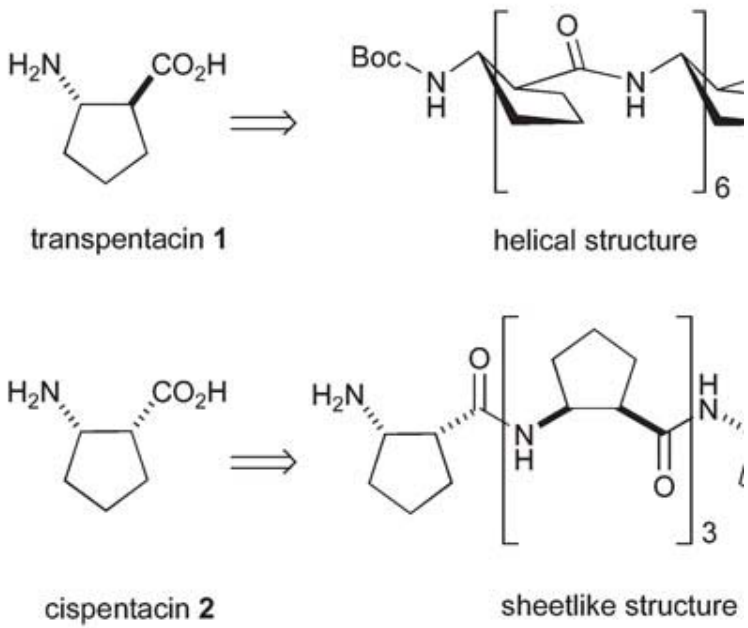


Fig. 1 Secondary structure of poly homo-pe

requires an initial evaluation of the level of offered by the chiral α,β-unsaturated ester un jugate addition, which is achieved through the a achiral lithium amide to the ester. If the α,β-uns shows high facial selectivity upon conjugate add of enantiorecognition between the chiral α,β-uns and a chiral lithium amide is evaluated throug kinetic resolution [addition of (*RS*)-ester to ar (*RS*)-lithium amide]. In this approach,[8] the ef action are eliminated, allowing the maximum st factor (*E*) for the reaction to be calculated inde reaction conversion, as it is identical to the diast observed in the reaction.[9] If high enantiorecog between the reacting partners in a mutual kine then efficient kinetic resolution may be expec

Introduction

The generation of bespoke pseudopeptide sequences that exhibit highly ordered secondary and tertiary structures in both solution and solid phase has developed into a highly competitive field of research in recent years, with the ability to predict the conformation of a given peptide sequence from knowledge

体系名

been investigated widely, most notably by
man.[2] For instance, Gellman *et al.* have shown that β-peptides derived from *trans*-2-aminocyclopentanecarboxylic acid (*trans*-pentacin) **1** adopt a helical structure in both the solid state and in solution,[3] while Fülop *et al.* have shown that homo-oligomers of *cis*-2-aminocyclopentanecarboxylic acid (cispentacin) **2** form a sheetlike secondary structure in solution (Fig. 1).[4] The ability of mixed α,β-peptides containing both α-amino and β-amino acid derivatives to adopt a preferred conformation in solution has also been reported recently.[5]

We have shown extensively that the conjugate addition of lithiumamides derived from α-methylbenzylamine to α,β-unsaturated acceptors may be used for the asymmetric synthesis of β-amino acid derivatives.[6] This methodology has recently been utilised for the synthesis of (1*R*,2*S*,3*R*)-3-methylcispentacin **5** in >98% de and 98 ± 1% ee and (1*S*,2*S*,3*R*)-3-methyltranspentacin **7** in >98% de and 97 ± 1% ee by the kinetic resolution of *tert*-butyl (*RS*)-3-methylcyclopentene-1-carboxylate **3** with lithium (*S*)-*N*-benzyl-*N*-α-methylbenzylamide (Scheme 1).[7]

The protocol that we use to understand fully the stereoselectivity observed in these kinetic resolution reactions

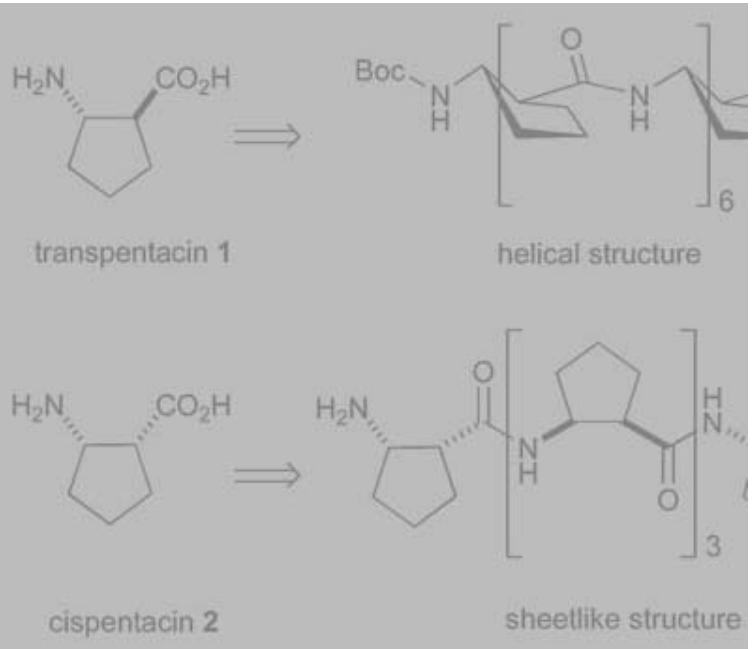


Fig. 1 Secondary structure of poly homo-pe

requires an initial evaluation of the level of offered by the chiral α,β-unsaturated ester un jugate addition, which is achieved through the achiral lithium amide to the ester. If the α,β-uns shows high facial selectivity upon conjugate add of enantiorecognition between the chiral α,β-uns and a chiral lithium amide is evaluated through kinetic resolution [addition of (*RS*)-ester to a (*RS*)-lithium amide]. In this approach,[8] the e action are eliminated, allowing the maximum st factor (*E*) for the reaction to be calculated indep reaction conversion, as it is identical to the diast observed in the reaction.[9] If high enantiorecog between the reacting partners in a mutual kine then efficient kinetic resolution may be expec

Introduction

The generation of bespoke pseudopeptide sequences that exhibit highly ordered secondary and tertiary structures in both solution and solid phase has developed into a highly competitive field of research in recent years, with the ability to predict the conformation of a given peptide sequence from knowledge of its primary structure an elusive goal. While much effort has been directed towards understanding the factors that control the secondary structure of α-peptides, the utility of peptides incorporating the β-amino acid structural motif has recently been investigated widely, most notably by Seebach[1] and Gellman.[2] For instance, Gellman *et al.* have shown that β-peptides derived from *trans*-2-aminocyclopentanecarboxylic acid (transpentacin) **1** adopt a helical stucture in both the solid state and in solution,[3] while Fülop *et al.* have shown that homo-oligomers of *cis*-2-aminocyclopentanecarboxylic acid (cispentacin) **2** form a sheetlike secondary structure in solution (Fig. 1).[4] The ability of mixed α,β-peptides containing both α-amino and β-amino acid derivatives to adopt a preferred conformation has also been reported recently.[5]

We have shown exte... lithiumamides deriv... unsaturated acce... of β-am... utilised ... in >98% de and 98 ... pentacin **7** in >98% de and 97 ± 1% ee by the kinetic resolution of *tert*-butyl (*RS*)-3-methylcyclopentene-1-carboxylate **3** with lithium (*S*)-*N*-benzyl-*N*-α-methylbenzylamide (Scheme 1).[7]

The protocol that we use to understand fully the stereoselectivity observed in these kinetic resolution reactions

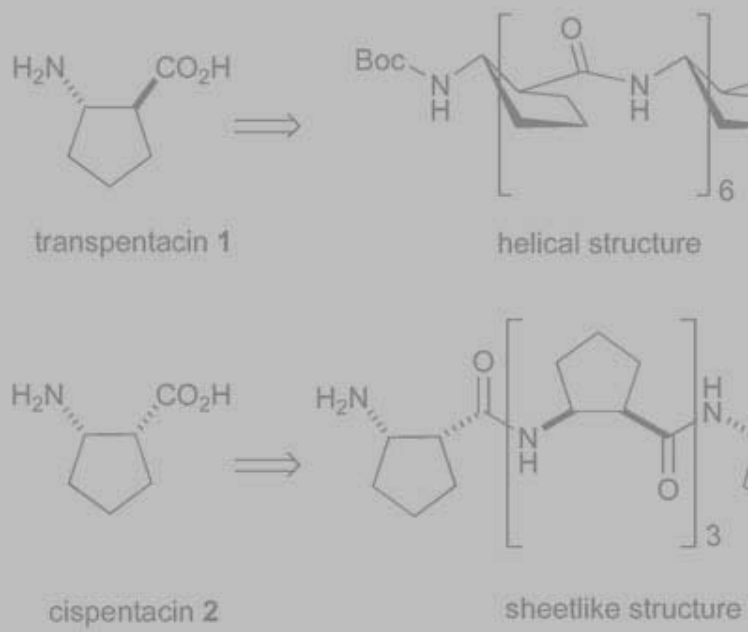


Fig. 1 Secondary structure of poly homo-pe

requires an initial evaluation of the level of ... offered by the chiral α,β-unsaturated ester un... gate addition, which is achieved through the a... hium amide to the ester. If the α,β-uns... igh facial selectivity upon conjugate add... nition between the chiral α,β-uns... ural lithium amide is evaluated through ... resolution [addition of (*RS*)-ester to ar... (*RS*)-lithium amide]. In this approach,[8] the ef... action are eliminated, allowing the maximum st... factor (*E*) for the reaction to be calculated indep... reaction conversion, as it is identical to the diast... observed in the reaction.[9] If high enantiorecog... between the reacting partners in a mutual kine... then efficient kinetic resolution may be expec...

慣用名

Introduction

The generation of bespoke pseudopeptide sequences that exhibit highly ordered secondary and tertiary structures in both solution and solid phase has developed into a highly competitive field of research in recent years, with the ability to predict the conformation of a given peptide sequence from knowledge of its primary structusive goal. While much effort has been directed towlerstanding the factors that control the secondary strf α-p incorporating the been investigated man.[2] For instancan e derived from *trans*-2-aminocyclopentanecarboxylic acid (transpentacin) **1** adopt a helical stucture in both the solid state and in solution,[3] while Fülop *et al.* have shown that homo-oligomers of *cis*-2-aminocyclopentanecarboxylic acid (cispentacin) **2** form a sheetlike secondary structure in solution (Fig. 1).[4] The ability of mixed α,β-peptides containing both α-amino and β-amino acid derivatives to adopt a preferred conformation in solution has also been reported recently.[5]

We have shown extensively that the conjugate addition of lithiumamides derived from α-methylbenzylamine to α,β-unsaturated acceptors may be used for the asymmetric synthesis of β-amino acid derivatives.[6] This methodology has recently been utilised for the synthesis of (1*R*,2*S*,3*R*)-3-methylcispentacin **5** in >98% de and 98 ± 1% ee and (1*S*,2*S*,3*R*)-3-methyltranspentacin **7** in >98% de and 97 ± 1% ee by the kinetic resolution of *tert*-butyl (*RS*)-3-methylcyclopentene-1-carboxylate **3** with lithium (*S*)-*N*-benzyl-*N*-α-methylbenzylamide (Scheme 1).[7]

The protocol that we use to understand fully the stereoselectivity observed in these kinetic resolution reactions

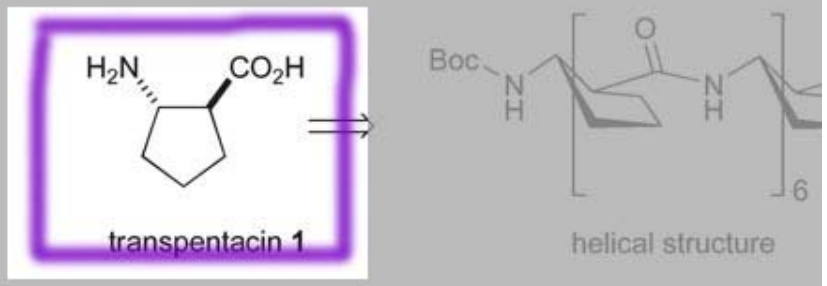


Fig. 1 Secondary structure of poly homo-pe

requires an initial evaluation of the level of offered by the chiral α,β-unsaturated ester un jugate addition, which is achieved through the achiral lithium amide to the ester. If the α,β-uns shows high facial selectivity upon conjugate add of enantiorecognition between the chiral α,β-uns and a chiral lithium amide is evaluated through kinetic resolution [addition of (*RS*)-ester to a (*RS*)-lithium amide]. In this approach,[8] the e action are eliminated, allowing the maximum st factor (*E*) for the reaction to be calculated indep reaction conversion, as it is identical to the diast observed in the reaction.[9] If high enantiorecog between the reacting partners in a mutual kine then efficient kinetic resolution may be expec

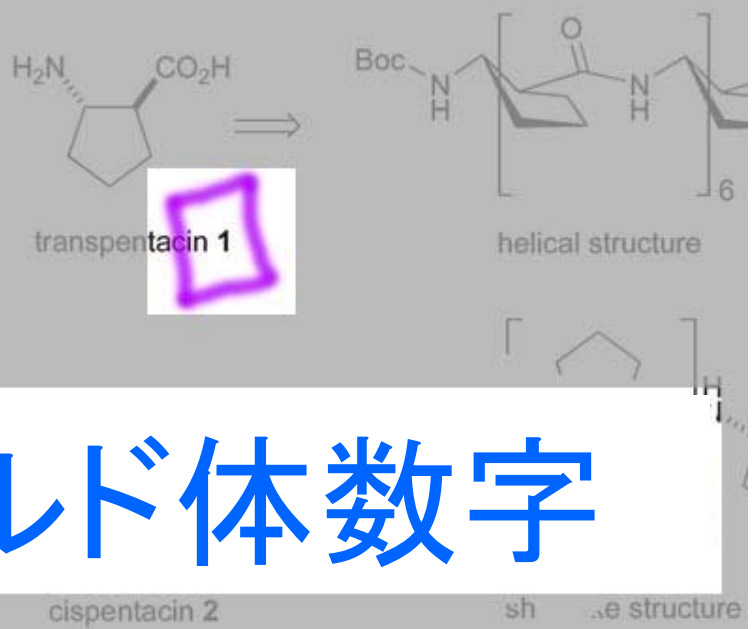
2次元構造図

Introduction

The generation of bespoke pseudopeptide sequences that exhibit highly ordered secondary and tertiary structures in both solution and solid phase has developed into a highly competitive field of research in recent years, with the ability to predict the conformation of a given peptide sequence from knowledge of its primary structure an elusive goal. While much effort has been directed towards understanding the factors that control the secondary structure of α-peptides, th
incorporating the β-amino acid
been investigated widely, most notabl
man.[2] For instance, Gellman *et al.* have
derived from *trans*-2-aminocyclope
pentacin) **1** adopt a helical stuctu
solution,[3] while Fülop *et al.* have shown that homo-oligomers of *cis*-2-aminocyclopentanecarboxylic acid (cispentacin) **2** form a sheetlike secondary structure in solution (Fig. 1).[4] The ability of mixed α,β-peptides containing both α-amino and β-amino acid derivatives to adopt a preferred conformation in solution has also been reported recently.[5]

We have shown extensively that the conjugate addition of lithiumamides derived from α-methylbenzylamine to α,β-unsaturated acceptors may be used for the asymmetric synthesis of β-amino acid derivatives.[6] This methodology has recently been utilised for the synthesis of (1*R*,2*S*,3*R*)-3-methylcispentacin **5** in >98% de and 98 ± 1% ee and (1*S*,2*S*,3*R*)-3-methyltranspentacin **7** in >98% de and 97 ± 1% ee by the kinetic resolution of *tert*-butyl (*RS*)-3-methylcyclopentene-1-carboxylate **3** with lithium (*S*)-*N*-benzyl-*N*-α-methylbenzylamide (Scheme 1).[7]

The protocol that we use to understand fully the stereoselectivity observed in these kinetic resolution reactions



ボールド体数字

Fig. 1 Secondary structure of poly homo-pe

requires an initial evaluation of the level of
offered by the chiral α,β-unsaturated ester un
jugate addition, which is achieved through the
achiral lithium amide to the ester. If the α,β-uns
shows high facial selectivity upon conjugate add
of enantiorecognition between the chiral α,β-uns
and a chiral lithium amide is evaluated through
kinetic resolution [addition of (*RS*)-ester to a
(*RS*)-lithium amide]. In this approach,[8] the e
action are eliminated, allowing the maximum st
factor (*E*) for the reaction to be calculated indep
reaction conversion, as it is identical to the diast
observed in the reaction.[9] If high enantiorecog
between the reacting partners in a mutual kine
then efficient kinetic resolution may be expec

プロスペクト･モデル

- 既存の出版物を強化する
- 既存のワークフローに直接組み込む
- 個々の製品を導入する際の問題を軽減する product

- 低分子ライブラリをマークした…
  …lysyl oxidase in zebrafish notochord morphogenesis

RSC | Advancing the Chemical Sciences

方法

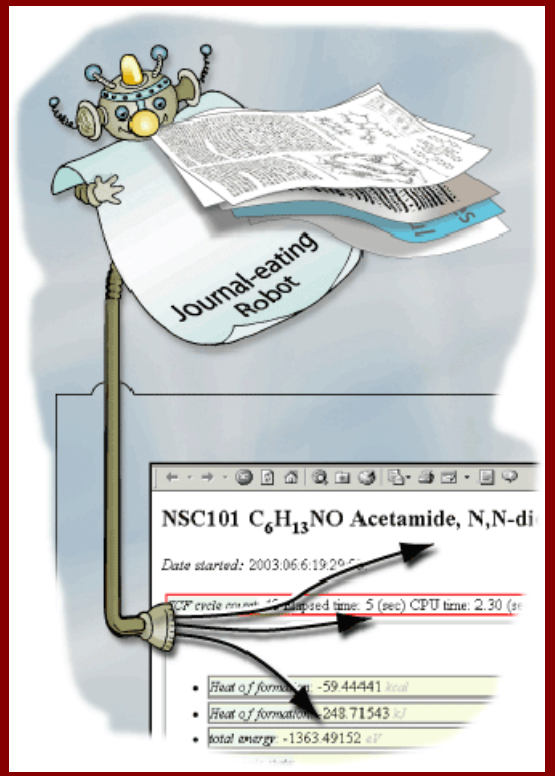
- 共同事業
  - ユニリーバセンター／コンピューターラボ
  - その他
- テクノロジー
  - OSCAR テキストマイニング
  - 基準
  - ワークフロー
  - RSS フィード

RSC | Advancing the Chemical Sciences

OSCAR テキストマイニング

OSCAR: Open Source Chemical Analysis Routines

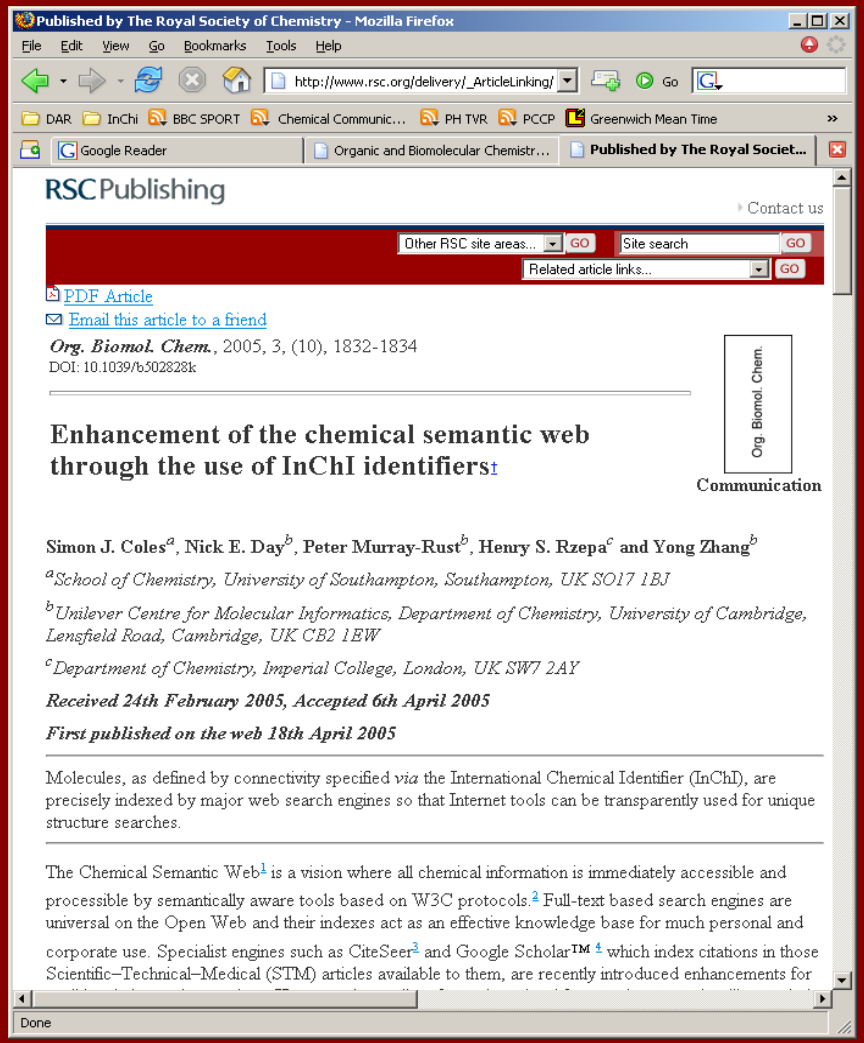
- ケンブリッジ大学:
  化学(ユニリーバ・センター)
  –コンピューターラボとの共同作業
- SciBorg
- 化学の単語や短い句の構造を
  分析して構造を決定する
- テキストに構造を付与する
- 各種テーマにも応用



RSC | Advancing the Chemical Sciences

InChI

- IUPAC International Chemical Identifier
- InChI=1/C8H10N4O2/c1-10-4-9-6-5(10)7(13)12(3)8(14)11(6)2/h4H,1-3H3
- 機械可読
- InChIkey RYYVLZVUVIJVGH-UHFFFAOYAW



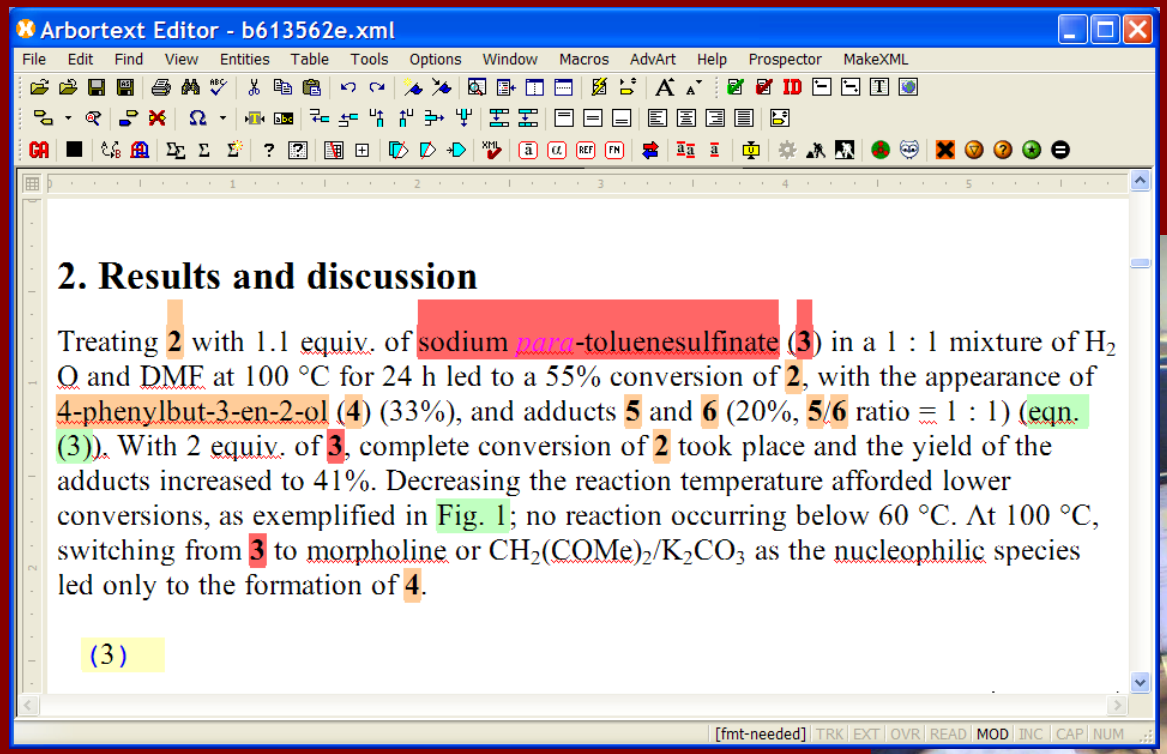
RSC | Advancing the Chemical Sciences

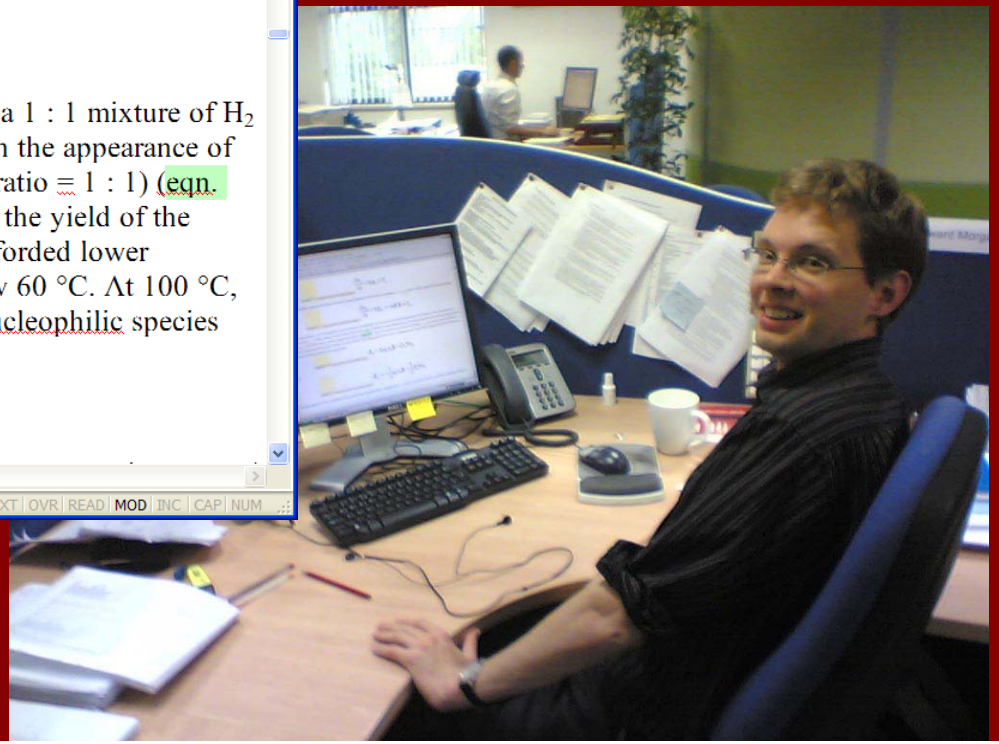
領域別の用語

- Open Biomedical Ontologies (OBO)
  - ジーンオントロジー (GO)
  - シークエンスオントロジー (SO)
  - セルオントロジー (CL)
- IUPAC ゴールド・ブック
  - 化学用語の辞書

RSC 出版プロセス

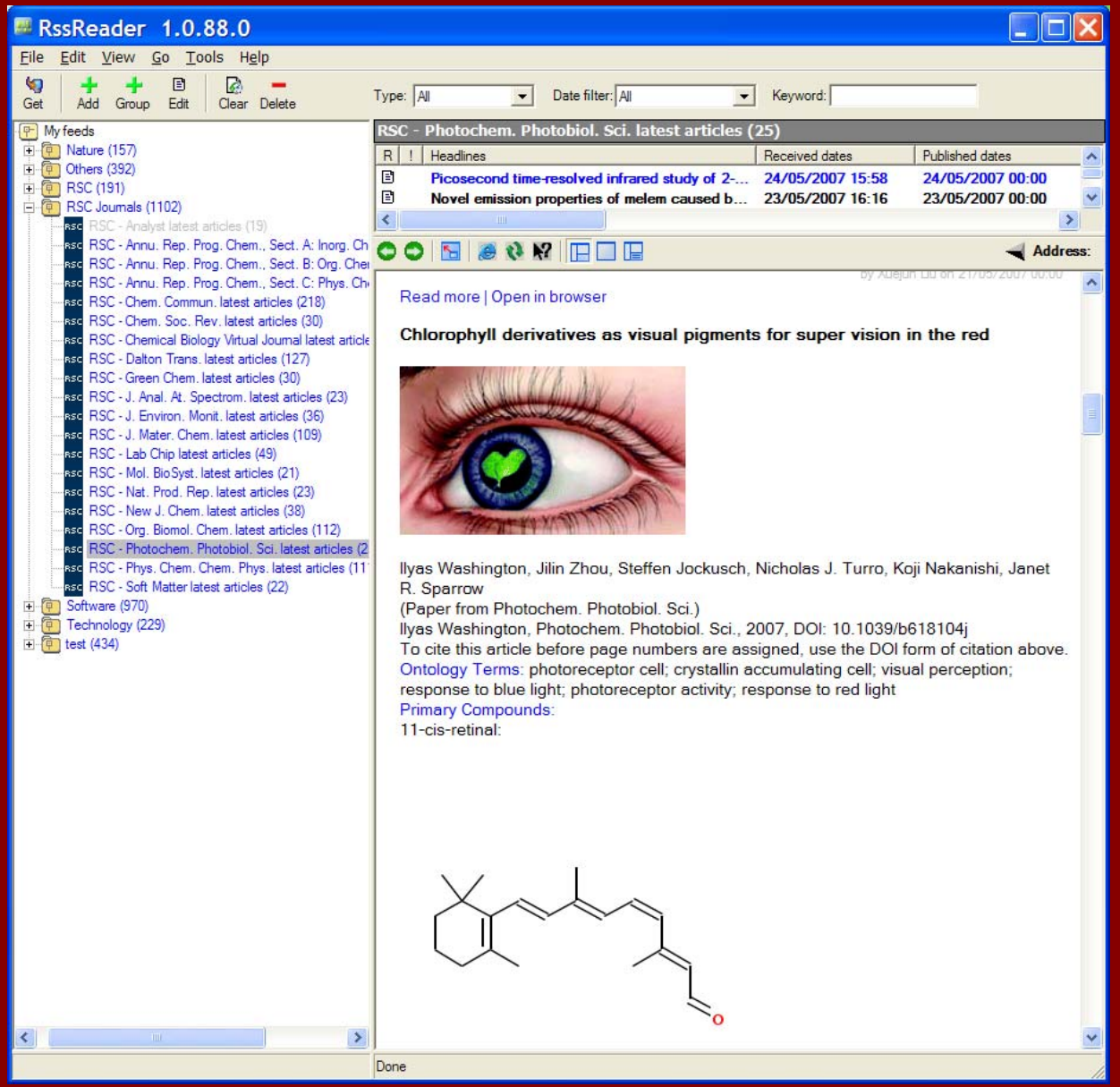
- 編集
- マイニング
- 更新と整理





RSC | Advancing the Chemical Sciences

強化された RSS フィード



RSC | Advancing the Chemical Sciences

問題点

- 意義のある、マニュアルの質疑応答の必要性
  - 古いデータはさらに困難
- 多くの分野におけるオントロジー開発の必要性
- 成果を最大化するための、広範囲での導入と相互リンクの必要性

データの新しい基準

- 実験的なデータ･チェッカー
  - 妥当性検証
  - 可視化
- 学会出版者のための役割

2-Methyl-2-hydroxymethyl-2,5-dihydrofuran **9** (50 mg, 0.51 mmol) was oxidised using OsO_4, TMEDA to yield the crude products [6 : 1 (*syn* : *anti*) by ^{1}H NMR]. The resulting oil was purified by flash chromatography (petrol-EtOAc, 4 : 1) to afford the pure title compound as a colourless oil (116 mg, 83%); ν_{max} (film)/cm^{-1} 2977, 2929, 1746, 1374, 1230, 1046; δ_H (300 MHz; $CDCl_3$) 5.43 (1H, q, *J* 5), 5.17 (1H, d, *J* 5), 4.26-4.09 (3H, m), 3.89 (1H, dd, *J* 10 and 5), 2.10 (6H, s), 2.08 (3H, s), 1.36 (3H, s); δ_C (75 MHz; $CDCl_3$) 170.6, 169.8, 169.5, 80.8, 76.2, 71.2, 68.9, 65.3, 22.3, 20.6, 20.4, 18.3; Found (CI) 292.1392, $C_{12}H_{22}NO_7 + NH_4$ requires 292.1396.

***anti*-2-Methyl-2-(acetoxymethyl)tetrahydrofuran *anti*-15**

2-Methyl-2-hydroxymethyl-2,5-dihydrofuran **9** (50 mg, 0.51 mmol) was oxidised using UpJohn conditions to yield the crude products [2 : 1 (*anti* : *syn*) by ^{1}H NMR]. The resulting oil was purified by flash chromatography (petrol-EtOAc, 4 : 1) to afford the product mixture as a colourless oil (110 mg, 79%) as an inseparable mixture of isomers (*anti* major compound); ν_{max} (film)/cm^{-1} 2989, 2944, 1746, 1379, 1232, 1046; δ_H (300 MHz; $CDCl_3$) 5.46-5.39 (1H, m), 5.26 (1H, d, *J* 6), 4.26-3.85 (4H, m), 2.11 (3H, s), 2.10 (6H, s), 1.26 (3H, s); δ_C (75 MHz; $CDCl_3$) 170.4, 169.8, 169.6, 81.2, 72.7, 71.9, 69.3, 67.9, 20.7, 20.5, 20.4, 18.3; Found (CI) 292.1400, $C_{12}H_{22}NO_7$ requires 292.1396,

***syn*-1,2,4-Triacetoxycyclohexane *syn*-16**

Cyclohex-3-en-1-ol **11** (50 mg, 0.51 mmol) was oxidised using OsO_4, TMEDA to yield the crude mixture of products [3 : 1 (*syn* : *anti*) by ^{1}H NMR]. The resulting oil was purified by flash chromatography (EtOAc) to afford a colourless oil (121 mg, 92%) as an inseparable mixture of isomers; mixture: ν_{max} (film)/cm^{-1} 2953, 1745, 1370, 1235, 1028; *syn*-**16** δ_H (300 MHz; $CDCl_3$) 5.17-5.11 (1H, m), 4.84-4.72 (2H, m), 2.04 (3H, s), 1.98 (3H, s), 1.86 (3H, s), 2.05-1.45 (6H, m); δ_C (75 MHz; $CDCl_3$) 170.1, 170.0, 169.9, 69.5, 69.3, 67.9, 31.4, 25.3, 24.7, 21.1, 20.9, 20.8; *anti*-**16**; δ_H (300 MHz; $CDCl_3$) 5.08-5.01 (1H, m), 4.81-4.67 (2H, m), 1.99 (9H, s), 2.05-1.45 (6H, m); δ_C (75 MHz; $CDCl_3$) 170.3, 170.2, 170.1, 69.7, 68.8, 68.6, 31.8, 25.9, 25.3, 23.6, 21.1, 20.9; mixture Found (CI) 276.1437, $C_{12}H_{22}NO_6$ requires 276.1447.

***anti*-1,2,4-Triacetoxycyclohexane *anti*-16**

Cyclohex-3-en-1-ol **11** (50 mg, 0.51 mmol) was oxidised using UpJohn conditions to yield the crude mixture of products [1 : 1 (*syn* : *anti*) by ^{1}H NMR]. The resulting oil was purified by flash chromatography (EtOAc) to afford the product mixture as a colourless oil (115 mg, 88%) as an inseparable mixture of isomers; All data are consistent with *syn*- and *anti*-**16** prepared previously.

***syn*-1,2,4,5-Tetraacetoxycyclohexane *syn*-17**

RSC | Advancing the Chemical Sciences

“使い始めてすぐにわかったのは（プロジェクト・プロスペクトの）賢さだ。化合物を見ることができたり、機械可読のSMILESやInChIsがあることが素晴らしい"

“この新しいシステムには非常に感銘を受けた。多人数のコミュニティにとって非常に便利となることは確かだ”

“これは素晴らしく、またエキサイティングだ!!数分試しただけで、大きな可能性に気付いて、すぐに学生にも紹介したよ！“

“これはこのコミュニティにとって素晴らしい資源であり、またGOやSOを有効活用している。すばらしい出来だ。"

”非常に直観的かつ使い方が簡単である。論文が読者によってより魅力的なものとなるはずだ。"

“It is fantastic. I’ve just seen the future of the journal”

Ed Pentz

Executive Director, CrossRef

RSC | Advancing the Chemical Sciences

何が違うのか

- テクノロジー
  - セマンティック・エンリッチメント
    - 化合物
    - オントロジー用語
  - RSSフィードの強化
- 既存のポートフォリオへの実装
  - ワークフローへ組み込み
  - 構造化された科学の可能性を示す

RSC | Advancing the Chemical Sciences

RSCにおけるプロジェクト・プロスペクトの展望

- 実験データ
- 言語学
- 広範囲なリンキング
- その他の領域
- eブック、データベース、バックファイル
- ピアレビュー
- プロスペクトによるサービス

RSC | Advancing the Chemical Sciences

RSC | Advancing the Chemical Sciences

Project Prospect
（プロジェクト・プロスペクト）

理化学出版に意味論を導入する

リチャード・キッド

RSC | Advancing the Chemical Sciences