

Product specification for



Version 3.0

By Paul Mostert
Appendices by Jeroen Hogendorp, adapted by Paul Mostert

Date June 2001

Contents

1	Introduction	1
1.1	Background	1
1.2	Overview of changes between EES versions 1.0 - 1.2, SDOS 2.0, 2.1 and 3.0	1
2	Dataset components	3
2.1	“dataset.toc” file	3
2.2	Acrobat PDF files	3
2.2.1	“True/Distilled” versus “Wrapped”	3
2.3	“Raw ASCII” files	4
2.4	SGML files	4
2.5	Artwork files	5
2.6	“stripin.toc” files	5
2.7	“checkmd5.fil” files	5
3	Dataset contents	6
3.1	Dataset directory structure	6
3.2	Available files	7
4	The “dataset.toc” file	8
4.1	The “Dataset”-level (_t0)	8
4.2	The “Journal Title”-level (_t1)	8
4.3	The “Journal Issue”-level (_t2)	10
4.4	The “Publication Item”-level (_t3)	12
4.5	Example of an article	15
4.6	Example of the associated “dataset.toc” file	17
4.7	Example of the associated SGML (<i>main.sgm</i>) file	18
5	Distribution	33
5.1	Fulfilment frequency	33
5.2	Media options	33
5.3	Printout	33
Appendix 1	SGML structure	34
1.1	Output specifications for SGML files	34
1.1.1	Regular SGML files	34
1.1.2	Linked textboxes	34
1.2	Conversion failures	34
1.2.1	Regular files	34
1.2.2	Linked textboxes	34
Appendix 2	Stripin components	35
2.1	Introduction	35
2.2	Affected SGML elements	35
2.2.1	Accent constructions	35
2.2.2	Arrays	37
2.2.3	Boxes and lines	37
2.2.4	Inline and displayed equations	37
2.2.5	Fence, fraction, radical and limit constructions	37
2.2.6	Tables	37
2.2.7	Certain fonts	37
2.2.8	Entities	37
2.3	stripin.toc file	38
Appendix 3	Output specifications for artwork	39

1 Introduction

In *ScienceDirect OnSite (SDOS)* datasets are structured and formatted according to the *EFFEFFECT - Exchange Format For Electronic Components and Texts* standard. The *EFFEFFECT* standard (detailed in *EFFEFFECT Technical Specifications Version 4.1*, September 1996, available via the ScienceDirect Support Website at support.sciencedirect.com ⇨ Technical ⇨ Effect) is a generic standard. A number of products apply to this standard, *SDOS* being one of those. *This* document describes specifically how to interpret *EFFEFFECT 4.1* for use in *SDOS*, and highlights particular details. A general understanding of *EFFEFFECT* is assumed.

1.1 Background

Elsevier Science converted its traditional production methods to generic computer-aided facilities, particularly aiming to be able to deal with the demands of electronic online publishing, imposed by the World Wide Web and associated facilities.

Journals were addressed one after the other, and the files contained in *SDOS* datasets reflect this distinction between several “generations” of publication production. Now we are moving one step forward again, by adding full text SGML and associated artwork files to *SDOS* datasets.

1.2 Overview of changes between EES versions 1.0 - 1.2, SDOS 2.0, 2.1 and 3.0

ScienceDirect OnSite and its predecessor *Elsevier Electronic Subscriptions* were and are available in the following versions:

- *EES version 1.0* datasets (now obsolete) contained:
 - ◆ TIFF files, each containing the scanned image of a single page;
 - ◆ Associated raw ASCII text files, one for each page;
 - ◆ SGML citation files, each containing the bibliographic data for an publication items;
 - ◆ **dataset.toc** file pertaining to *EFFEFFECT 4.0* specifications.
- *EES version 1.1* datasets (now obsolete) consisted of:
 - ◆ PDF files, each containing an publication item in “wrapped” format;
 - ◆ Raw ASCII text files, one for each page;
 - ◆ SGML citation files, each containing the bibliographic data for an publication item;
 - ◆ **dataset.toc** file pertaining to *EFFEFFECT 4.0* specifications.
- *EES version 1.2* datasets (now obsolete) carried:
 - ◆ PDF files, each containing an publication item in either “wrapped” or “true” format;
 - ◆ Raw ASCII text files, one for each page;
 - ◆ SGML citation files, containing bibliographic data for publication items;
 - ◆ **dataset.toc** file pertaining to *EFFEFFECT 4.0* specifications.

- *SDOS version 2.0* datasets carry:
 - ◆ PDF files, each containing an publication item in either “wrapped” or “true” format;
 - ◆ Raw ASCII text files, each containing an publication item in either “wrapped” or “true” format;
 - ◆ SGML citation files, containing bibliographic data for publication items;
 - ◆ **dataset.toc** file pertaining to *EFFEFFECT 4.0* specifications.
- *SDOS version 2.1* datasets carry:
 - ◆ PDF files, each containing an publication item in either “wrapped” or “true” format;
 - ◆ Raw ASCII text files, each containing an publication item in either “wrapped” or “true” format;
 - ◆ SGML citation files, containing bibliographic data for publication items, and also all article references in a structured format;
 - ◆ **dataset.toc** file pertaining to *EFFEFFECT 4.0* specifications.
- | ○ *SDOS version 3.0* datasets carry:
 - ◆ PDF files, each containing an publication item in either “wrapped” or “true” format;
 - ◆ Raw ASCII text files, each containing an publication item in either “wrapped” or “true” format;
 - ◆ Full article SGML files for publication items, and also all associated artwork files in web-enabled graphical formats;
 - ◆ **dataset.toc** file pertaining to *EFFEFFECT 4.1* specifications.

| The most significant technical format difference between *SDOS* version 3.0 and previous versions is the inclusion of full articles in the SGML files and together with all associated artwork. To accommodate this, the Effect envelope has been upgraded from version 4.0/issue-based to version 4.1/item-based.

| The structural changes of *SDOS* version 3.0 with regard to the previous version 2.1 are highlighted by a vertical bar in the left margin, as is the case with this paragraph. Minor textual changes or corrections are not highlighted.

2 Dataset components

This chapter details the different file components, which are available in *SDOS 3.0* datasets. The next chapters explain how these files are related.

2.1 “dataset.toc” file

Each *SDOS* dataset has one master index file, or “dataset.toc” file, with complete bibliographic information as well as all relevant cross reference data, e.g., which PDF and SGML files are related to which articles and/or journal issues. The **dataset.toc** file will be detailed further in Chapter 4 at page 8.

2.2 Acrobat PDF files

Portable Document Format (PDF) is a defacto standard set by *Adobe Corporation* to enable electronic distribution of documents without limitations on graphic design compared to a paper document. PDF is based on the very popular PostScript language, but has enhancements geared to network distribution.

Along with the creation of the PDF standard, *Adobe Corp.* developed a collection of programs, the *Acrobat* software suite, to create (“distill”), change, view and print PDF files. One of the components of the suite, the *Acrobat Reader* program, is also separately available for download free of charge from *Adobe Corp’s* Website **www.adobe.com**

PDF files in *SDOS 3.0* datasets are only item-based, whereas previous versions of *SDOS* also carried page-based PDF files. Item-based PDF files embody the articles in journals on a one-by-one basis, i.e. each PDF file represents one entire article. Typically, these files contain several pages, of which the first one starts with the article title, authors, abstract, etc.

A growing number of PDF files embedded in *SDOS 3.0* datasets are created with version 4.0 of *Acrobat*. This has the advantage that the size of PDF files can be substantially decreased because of better compression algorithms, yielding smaller file sizes and shorter network transmission times. Since older versions of *Acrobat Reader* are not capable of handling the new format, you must ensure that your users have version 4.0 or later of *Acrobat Reader*.

A number of frequently asked questions and their answers on *Adobe Acrobat* and PDF is available via the *EFFECT* support page in ScienceDirect’s Support Website at **support.sciencedirect.com**
⇒ Technical ⇒ PDF

2.2.1 “True/Distilled” versus “Wrapped”

In *SDOS* two different methods exist to create PDF files:

- In the new computer-aided production workflow, PDF files are produced by “distilling” those files from the original typesetter PostScript files. Further in this document these files are referred to as “True” or “Distilled” PDF files, because these files have been produced without any manual intervention and paper scanning steps. The files are in the same quality as the final paper journal issue, including high-fidelity texts and symbols, photographs in crisp grays and colors, and formulas, tables and further artwork in their original typeset quality.
- For journal issues not yet covered by the computer-aided workflow, PDF files are produced by performing image scanning of the printed paper journal issue, and “wrapping” the resulting (TIFF Fax Group IV) page images in PDF code, therefore these files are referred to in this document as “Wrapped PDF”. As a result of the scan-based production method, these files do not have the same quality and crispness as “True/Distilled” PDF files.

Regardless of their source, both “flavors” of PDF can be displayed and printed by the *Acrobat Reader* program.

2.3 “Raw ASCII” files

Raw ASCII files contain the full text of pages or articles as plain ASCII text, identified by the file extension **.raw**. As opposed to previous *EES* datasets, in which **.raw** files were delivered on a page-by-page basis, those files are now available on an item-by-item basis, i.e., each item will be represented by a single **.raw** file. The **_mf** information contained in the **dataset.toc** file (see Chapter 4 at page 8) supplies ample cross reference data to link raw ASCII text files with their respective publication items.

Raw ASCII files are generated by two different mechanisms:

- In the computer-aided production workflow, raw ASCII files are produced by converting those from the SGML files.
- For journal issues not yet covered by the computer-aided workflow, raw ASCII files are produced by performing Optical Character Recognition (OCR) technology on scanned page images. Text files from several pages are assembled together in one single **.raw** file. As a result of the scan-based production method, these files do not have the same quality as SGML-generated ASCII text files.

2.4 SGML files

The SGML files that are delivered in *SDOS 3.0* datasets contain the full article, including bibliographic data (article title, abstract, author names, keywords, etc.), and also the references such as these appear at the end of each article. Those files can be recognized by the file extension **.sgml**. An example of such a file can be found in chapter 4.7 at page 18.

The DTD's to which the SGML files pertain are available separately via the EFFECT support page in ScienceDirect's Support Website at **support.sciencedirect.com** ⇨ Technical ⇨ SGML/DTD or upon request.

Important note: Usually, the vast majority of all articles pertains to ‘regular’ article types such as full length articles, short communications, review articles, etc. that all are structured with an article title, author names, abstract, chapters, figures, acknowledgments, references, etc. There are however a number of articles that cannot be captured in the structure of the so-called ‘Full Length Article Document Type Definition’ (FLA DTD) of Elsevier Science. Examples of those article types are news items, event calendars, product reviews, etc. As there are no formal DTD's (yet) for those ‘odd-structured’ articles, these are not produced in the SGML workflow but by ‘traditional’ procedures.

The implication of this approach is that for those articles no SGML file would be available, making them unusable for electronic publishing purposes. To overcome this inadvertent effect, mechanisms have been put in place at Elsevier Science to capture at least the bibliographic details of such articles as SGML files. The same mechanism is also put in place as a fallback for resolving technical problems in the SGML workflow that could otherwise hamper publication deadlines, and for retrospective capturing of bibliographic details of articles that were produced in earlier years for which no electronic source material is available (any more).

The electronic production operation will strive to capture as much detail as possible. The underlying decision tree is:

- ◆ If available take the entire article, including bibliographic details (commonly known as the ‘head’), the body of the article, and the references (commonly called the ‘tail’);
- ◆ If this is not possible, capture the bibliographic details (‘head’) and the references (‘tail’);
- ◆ If this is not possible, capture the bibliographic details (‘head’-only);
- ◆ Otherwise, capture the article title (the entry that appears in the Table of Contents).

2.5 Artwork files

Many publication items not only consist of text, but several figures, plates, photographs, schemes, etc. are present to illustrate the meaning of the author. These artwork elements are considered an integral part of the publication item.

Artwork files are delivered in Compuserve GIF, Joint Photographers Expert Group (JPEG) formats, or in other accepted graphical Web formats.

2.6 “stripin.toc” files

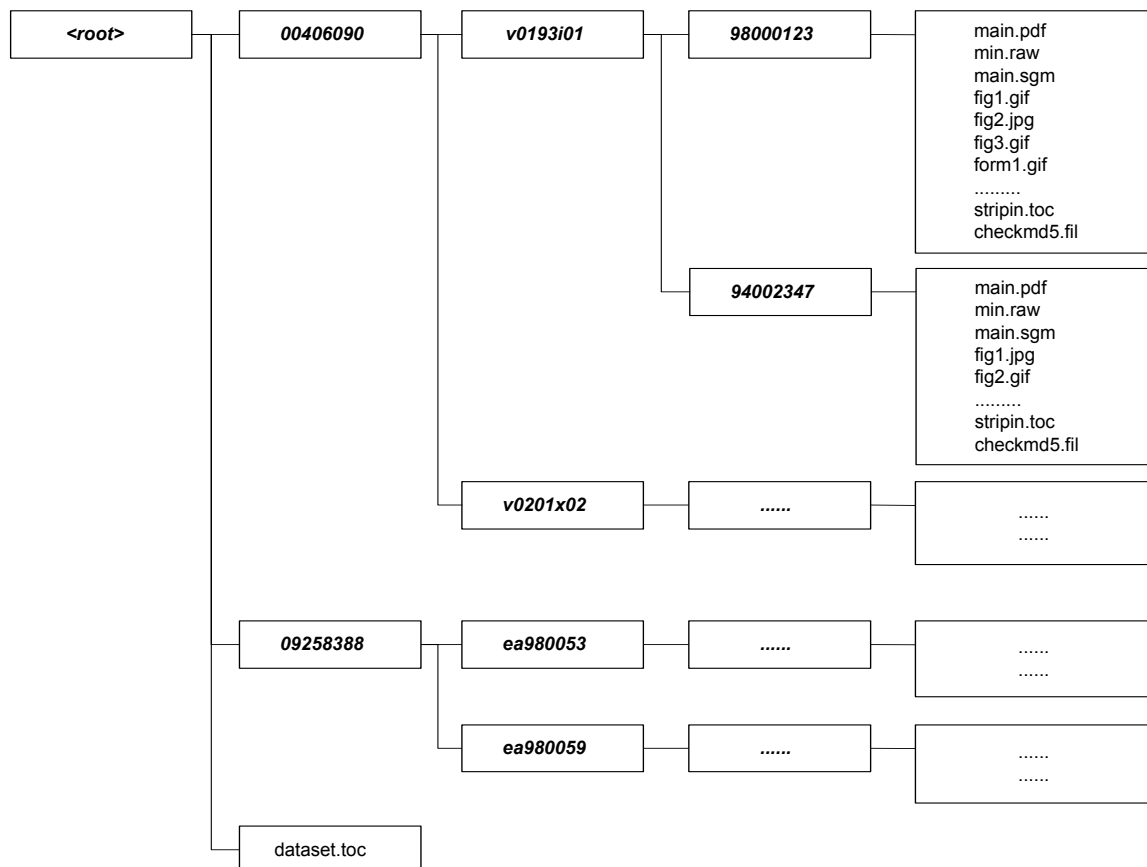
“Difficult” textual constructs that are impossible to represent in HTML, such as mathematical formulae or certain types of tables, are present as **.gif** files. A file called **stripin.toc** contains important associative information about how the “difficult” text parts are linked to their corresponding **.gif** representation. This is further worked out in Appendix 2 at page 35.

2.7 “checkmd5.fil” files

The **checkmd5.fil** files will follow the structure and format set forth in the *EFFECT Technical Specification 4.1* document.

3 Dataset contents

3.1 Dataset directory structure



An example of an *SDOS 3.0* dataset:

All material from a particular journal title is collected in a directory at the journal title level. The directory name is identified with an eight character code, formed by taking the ISSN of the journal omitting the dividing dash. In the above example, the directory /**00406090** holds material of the journal *Thin Solid Films* (ISSN is 0040-6090).

Every journal issue is available in a subdirectory within the journal title directory. The directory name is identified by a unique journal issue identifier, e.g., in the above example /**v0193i01** and /**ea940053**.

Every publication item is represented at a subordinate directory level within a journal issue directory. The directory name is identified by a unique publication item identifier. In the above example, the issue named /**v0193i01** holds two publication items, in the subdirectories /**98000123** and /**94002347**.

Please note that this publication item level is incompatible with earlier versions of EES/SDOS. In those versions without full SGML representations, there was a strict one-to-one relationship between the components. I.e., each article was represented by one single .sgc, one single .pdf and one single .raw file. This is not possible any more for full text SGML versions. Each publication item can have one or more SGML files, and often contains many artwork files. Therefore this additional directory level was introduced. All files of one publication item are kept separate from the files of other publication items. This eases the perception of which files belongs to what article, and it avoids possible naming conflicts.

3.2 Available files

Each directory at the third, publication item level holds a number of files:

- There is typically one PDF file for each publication item holding all the pages. This file is available as the file **main.pdf** in each item-directory.
- There is generally one raw ASCII file for each publication item holding the full text in a “bare” text format. This file is available as the file **main.raw** in each item-directory. Please note that not all publication items have an associated **main.raw** file.
- Files ending in **.sgm** are the SGML files holding the full text of an publication item, e.g. **main.sgm** in the above example. Generally, there is only one SGML file, typically with the name **main.sgm**. However, it can occur that larger articles are subdivided into several interrelated SGML files. This is for instance the case with magazine-like articles with “floating” text boxes, large articles that are divided into chapters, or a section of news items that are treated as “sub articles”. In these cases, the **main.sgm** file defines the overall structure of such a subdivided article and functions as the “starting point” for the navigation between the subordinate SGML files.
- Files ending in **.gif**, **.jpg**, or other accepted graphical formats indicate artwork files (e.g., figures, plates, photographs, schemes, etc.) that are part of the publication item.
- “Difficult” textual constructs that are impossible to represent in HTML, such as mathematical formulae or certain types of tables, are present as **.gif** files. A file called **stripin.toc** possibly present in an publication item directory, contains important associative information about how the “difficult” text parts are linked to their corresponding **.gif** representation. This is further worked out in Appendix 2 at page 35.
- Each directory, except the root directory, in principle holds the file **checkmd5.fil** which provides a “digital signature” for each file in that directory. If there is no further file in a directory, the file **checkmd5.fil** is also not available.
- The relationships between files and their location in the overall dataset structure of journal titles, journal issues and publication items is clearly defined with the **_mf** tags in the **dataset.toc** file (see page 13).

4 The “dataset.toc” file

The **dataset.toc** file, available in the <root> directory of an *SDOS 3.0* dataset, provides all relationships between journal titles, journal issues, the publication items and all their associated files. The structure and content of the **dataset.toc** file is described in detail in the *EFFECT Technical Specifications 4.1* document. This chapter lists the specific tags available in *SDOS 3.0* datasets. If a tag has a special meaning which is different from the *EFFECT 4.1* specifications or has other noteworthy properties, this is listed here. Otherwise, there is only the reference “*see EFFECT 4.1*”.

Note

Although all tags are affected because their description should be derived from the *EFFECT 4.1* specifications instead of *EFFECT 4.0*, only those tags that have actually changed are highlighted in this document. For the other ones, there is no difference in description between *EFFECT* versions 4.0 and 4.1.

4.1 The “Dataset”-level (_t0)

_t0 *{Mandatory}* Start of the “Dataset”-level. The 8-character **_t0** value has the following format:

- ◆ The first three characters form the customer code, e.g. **CUS**
- ◆ The following four characters is a sequence number, e.g. **0123**
- ◆ An extra character to ease splitting of large datasets across smaller several delivery media, and still get unique dataset identifiers. For instance, a large dataset can be spread across more than one CD-Rom or magnetic tape. The first CD-Rom or tape gets as the last character **A** or **0**, the second one gets **B** or **1**, etc. Each “sub”-dataset on a CD-Rom or tape is “complete” in that it contains its own **dataset.toc** file.

_vn *{Mandatory}* *EFFECT* version number of dataset. For *SDOS 3.0* this is fixed **4.1**; *see EFFECT 4.1*

_pd *{Mandatory}* Production date of dataset; *see EFFECT 4.1*

_st *{Mandatory}* Dataset Structure Type. For *SDOS 3.0* datasets this is **ITM**. Note that this tag is very relevant for distinguishing *SDOS 2.1* or lower datasets from *SDOS 3.0* ones; *see further EFFECT 4.1*

4.2 The “Journal Title”-level (_t1)

_t1 *{Mandatory}* Start of journal title; *see EFFECT 4.1*

_cd *{if available}* The CODEN code; *see EFFECT 4.1*

_jn *{Mandatory}* The full name of the journal; *see EFFECT 4.1*

_ja *{if available}* Abbreviated journal title; *see EFFECT 4.1*

_jo *{if available}* The former journal name

_io *{if available}* The former journal ISSN without the dividing dash

Notes

1. In the case of a “merger” of two or more journals, the **_jo** and **_io** fields will be repeated to denote the separate original titles. This is a widening of the *EFFECT 4.1* specifications. An example:

```
_t1 ABC0001A 13871811  
_jn Microporous and Mesoporous Materials  
_jo Zeolites  
_io 01442449  
_jo Microporous Materials  
_io 09276513
```

indicates that the journal *Microporous and Mesoporous Materials* (ISSN: 1387-1811) is a merger of the former journal titles *Zeolites* (ISSN 0144-2449) and *Microporous Materials* (ISSN 0927-6513).

2. The **_jo** and **_io** fields are not necessarily associated. If only the title **_jn** has changed and the ISSN remained the same, then only **_jo** is given. If only the ISSN has changed and the journal title was not changed, only the **_io** is given
3. Please note that if you consider applying the **_io** field to link to the former journal title, journal data of the former title is actually available in your library system, since the title change could have taken place before you started to subscribe to this journal.

_jf *{if available}* The full set journal name; *see EFFECT 4.1*
_if *{if available}* The full set journal ISSN without the dividing dash; *see EFFECT 4.1*
_pu *{if available}* The publisher; *see EFFECT 4.1*
_ci *{if available}* The publisher’s city and country; *see EFFECT 4.1*
_im *{if available}* The journal imprint; *see EFFECT 4.1*
_et *{if available}* The Editorial Board titles; *see EFFECT 4.1*
_em *{if available}* Editorial Board Member; *see EFFECT 4.1*
_ia *{if available}* The Instructions to Authors specification; *see EFFECT 4.1*
_cr *{if available}* The copyright notice of a journal; *see EFFECT 4.1*

4.3 The “Journal Issue”-level (**_t2**)

_t2 *{Mandatory}* Start of journal issue; *see EFFECT 4.1*

_ps *{Internal Elsevier Science code, used for internal workflow and debugging purposes; can usually be ignored by receiving SDOS customers}* Production Stage of the journal issue. Indicates the most early production stage of all transported issue items following at **_t3** level. The following production stages are currently defined:

_ps [S100]	publication item at first proof stage
_ps [S200]	publication item at page proof stage
_ps [S300]	publication item at final production stage
_ps [S350]	publication item at final production stage; journal issue is based on scanned page images

The **_ps** code used at **_t2** level is the lowest value of all **_ps** codes at the **_t3** level. See also page 12 for a description of **_ps** per publication item.

_v1 *{Mandatory}* Volume number(s); *see EFFECT 4.1*

_is *{if available}* Issue number(s); *see EFFECT 4.1*

The **_v1** and **_is** fields are always numeric and uniquely identify an issue. All alphabetic characters are removed from volume and issue as they appear on the cover. For instance *Volume A523* will be represented by **_v1 523**. The result is that those fields can differ from corresponding cover volume/issue numbers.

_pr *{if available}* The starting and ending page numbers; *see EFFECT 4.1*

_cf *{if available}* Certain journal issues are dedicated to a specific subject or are related to an event such as a conference, a symposium or a workshop. Generally these journal issues are referred to as “special issues”. The title of the subject and/or event is clearly presented on the front cover of special issue. The **_cf** field contains the title of the journal issue in the case of a special issue, i.e., the particular physical issue is devoted to the proceedings of a conference, symposium or another event, or the journal issue covers a certain subject.

The existing definition of the **_cf** field has been widened to cover a wider range of special issue types and titles. The original **_cf** field was restricted to event-based special issues (conferences, workshops, symposia, etc.). The definition has now been broadened to encompass also special issues about certain topics not related to a specific event. As a consequence, the **_cf [date]** subfield is no longer mandatory. This redefinition was necessary because the definition of the **_xt** field has been restricted for the indication of only Supplements, Indexes and Parts.

The **_cf** tag is followed by a code, enclosed in square brackets, which characterizes the individual details. Valid codes are:

[name]	<i>{Mandatory if there is a _cf group for this journal issue}</i> The full title of the special issue or the name of the conference.
[abbrev]	The official abbreviation of the conference name (if any).
[number]	The conference number if the conference is part of a repeating series of conferences.
[place]	The conference location (city, region, etc.).
[date]	The date of the conference/symposium/etc. in the format <i>YYYYMMDD</i> , if the special issue is dedicated to an event, in which <i>YYYY</i> denotes the year, <i>MM</i> the

month and *DD* the day in the month. This field follows the same rules for date ranges as the `_dt` field.

[editor] *{Repeating}* The editor(s) of the special issue. This field is repeated as often as there are editors for this issue, or the editors are listed together in this field.

Examples

`_cf [name] Proceedings of the Third International Conference
on Osteoporosis held by the Health Council on Osteoporosis`

`_cf [abbrev] Osteoporosis III`

`_cf [place] Paris, France`

`_cf [date] 19991020/24`

`_cf [editor] J. Jansma`

`_cf [editor] P. Petersen`

`_cf [name] Special Issue: Papers presented at the Second
European Seminar on Infrared Spectroscopy`

`_cf [abbrev] ESIS '98`

`_cf [place] Lyon, France`

`_cf [date] 19980630/0703`

`_cf [name] Special Issue: Flow Injection. Analysis
State of the Art Applied to Atomic Spectroscopy`

`_cf [name] Special Issue: Celebrating The Centenary of
Runge-Kutta Methods`

`_cf [name] Special Issue on Fuzzy Optimization`

`_cf [name] A Collection of Invited Papers by Japanese
Researchers`

`_cf [name] Special Issue: The Backus Fest: A Collection of
Papers To Honor George Backus`

`_cf [name] Accelerators, Spectrometers, Detectors and
Associated Equipment`

`_cf [name] Special Issue: Proceedings of the Workshop on
III-V Nitrides`

`_xt` *{if available}* Extra information about the particular journal issue; *see EFFECT 4.1*

`_dt` *{Mandatory}* The issue date of the journal issue; *see EFFECT 4.1*

`_np` *{Mandatory}* The total number of physical pages in the journal issue; *see EFFECT 4.1*

`_pn` *{Mandatory}* The actual printed page numbers of the journal issue; *see EFFECT 4.1*

`_ct` *{if available}* The pages in the journal issue on which the table of contents is printed; *see EFFECT 4.1*

`_mf` *{if available}* Manifestation at journal issue level. Not used in *SDOS 3.0*; *see EFFECT 4.1*

4.4 The “Publication Item”-level (**_t3**)

_t3 *{Mandatory}* Start of publication item; *see EFFECT 4.1*

Important note

In previous versions of *SDOS* the fourth field of the **_t3** field was a placeholder code for the publication item. Originally this was the same code as the name part of the SGML, raw ASCII and PDF files presented in the **_mf** field (see also page 13). In this new item-based *SDOS 3.0* version, with the additional directory level, there is a clear separation, *and the last field of the _t3 tag now indicates the directory name in which all material of the particular publication item is located.*

For instance: in dataset **tst0123a**, the files for the publication item starting with

```
_t3 TST0123A 00406090 V0193I01 98000123
```

can be found in directory **/00406090/v0193i01/98000123**

_ps *{Internal Elsevier Science code, used for internal workflow and debugging purposes; can usually be ignored by receiving SDOS customers}* Production Stage of the item. The following production stages are currently defined:

_ps [S100] publication item at author proof stage

_ps [S200] publication item at page proof stage

_ps [S300] publication item at final production stage

_ps [S350] publication item at final production stage; the publication item is based on scanned page images

See also page 10 for a description of **_ps** per journal issue.

_ii *{Mandatory}* Item Identifier (PII); *see EFFECT 4.1*

Note

In previous versions of *SDOS*, the **_ii** was typically identical to fields 2 + 4 from the **_t3** field with extra “() -” delimiters. This is not true any more. The information in the **_t3** tag may only be interpreted as directory names within one dataset. The information provided in the **_ii** tag identifies the official item identifier. In essence, the observation that **_t3** and **_ii** are identical should be regarded as a coincidence.

The reasoning behind this is increasing the flexibility in electronic journal production processes, and to make certain republication facilities possible, e.g., an article which is later presented in a conference and is republished in the conference proceedings.

Conceivably, the following snippet of a **dataset.toc** is possible:

```
_t3 ABC0001A 00406090 V0123I02 01234567  
_ii S0192-8388(98)12345-X
```

_ty *{Mandatory}* The type of the publication item; *see EFFECT 4.1* Please note that the list of item types has grown between EFFECT 4.0 and 4.1. Also note, that other item types can be added in the course of time.

_li *{if available}* The language of the item; *see EFFECT 4.1*

_ti *{if available}* The full title of the item; *see EFFECT 4.1*

_tf *{if available}* The foreign language title of the item; *see EFFECT 4.1*

_au *{if available}* Author name; *see EFFECT 4.1*

- _ca** *{if available}* The full address for correspondence with the author(s); *see EFFECT 4.1*
- _ab** *{if available}* The full English abstract of the item; *see EFFECT 4.1*
- _la** *{if available}* Language of the abstract; *see EFFECT 4.1*
- _kw** *{if available}* Keyword(s) which apply to the item; *see EFFECT 4.1*
- _su** *{if available}* Subject area of the publication item; *see EFFECT 4.1*
- _pg** *{if available}* Printed page numbers on which the item appeared in the journal issue; *see EFFECT 4.0*
- _br** *{if available}* Backward reference to another publication item; *see EFFECT 4.1*
- _mf** *{Mandatory}* Manifestation format.

_mf [**SGML** *{dtdname}* *{version}*] Indicates the SGML file(s) that contains the entire text of the publication item, including the bibliographic details (e.g., article title, abstract, keywords, author names, etc), the “body” text, and all article references such as appearing at the end of each article (when available in the article). The name of these files are associated with a file with extension **.sgm**. The *{dtdname}* and *{version}* indicate the particular Document Type Definition (DTD) and version to which this particular (set of) SGML files conforms.

_mf [**SGML ART 4.2.0**] This indicates that the included **.sgm** files adhere to the Elsevier Full Length Article (*ART*) DTD version *4.2.0*.

As has been noted in chapter 2.4 at page 4, not all articles are entirely available in SGML.

Additional to the version number information will be provided to enhance SGML processing and/or debugging. This information indicates the ‘weight’ of an SGML file. This is related to the decision tree described in chapter 2.4 at page 5:

- [**SGML ART** *{version}*] with no additional data identifies an entire article;
- [**SGML ART** *{version}* **HEAD-AND-TAIL**] indicates an SGML file with only bibliographic details (the title, abstract authors, etc; the so-called ‘head’) and references (the so-called ‘tail’);
- [**SGML ART** *{version}* **HEAD-ONLY**] indicates an SGML file that contains the bibliographic details;
- [**SGML ART** *{version}* **CONTENTS-ENTRY-ONLY**] indicates an SGML file that only contains the article title, and the author names, if any.

_mf [**STRIPIN** *{version}*] Indicates the **stripin.toc** file which contains important associative information about how the “difficult” text parts are linked to their corresponding **.gif** representation. The current *{version}* number for **Stripin.toc** is **1.0**.

If there is no **stripin.toc** file, then this **_mf** [**stripin 1.0**] tag is omitted.

For further details see Appendix 2 at page 35.

_mf [**Raw ASCII**] Full publication item text in an unedited “bare” text format. It is possible that a **.raw** files are not available for some particular publication items. See also Chapter 3.2 at page 7.

_mf [PDF {PDFversion} {WRAPPED | DISTILLED}] Indicates the Portable Document Format (PDF) file that holds the entire text and artwork presentation of one publication item in one single file. *{PDFversion}* indicates the version of PDF to which this PDF file adheres. *{WRAPPED}* or *{DISTILLED}* indicate the origin of the PDF file.

_mf [PDF 1.1 WRAPPED] This indicates that the included **.pdf** file has been produced via page image scanning procedures and it adheres to (the older) PDF version 1.1.

_mf [PDF 1.2 DISTILLED OPTIMIZED] This indicates that a **.pdf** file has been “distilled” directly from the PostScript file in the typesetting process and “optimized” to enhance downloading. It conforms to PDF version 1.2.

_mf [PDF 1.3 DISTILLED OPTIMIZED BOOKMARKED] As the previous one, but the PDF file conforms to version 1.3, which is related with Adobe Acrobat 4.0. This PDF file is not compatible with earlier versions of Acrobat. This PDF file is also “bookmarked”, i.e., the opening page displays the section headings instead of page “thumbnails”.

Note

It is possible that some **.pdf** files from the same journal issue are “wrapped” while others are “true”. This could be caused by the fact that there was a technical problem with some “true” **.pdf** files for which then the “wrapped” version has been included.

4.5 Example of an article

First page with the bibliographic details

<p>FEBS 23531</p> <p style="text-align: center;">Site-specific recombination in mammalian cells catalyzed by $\gamma\delta$ resolvase mutants: implications for the topology of episomal DNA</p> <p style="text-align: center;">Micha Schwikardi, Peter Dröge*</p> <p style="text-align: center;"><i>Institute of Genetics, University of Cologne, Weyertal 121, D-50931 Cologne, Germany</i></p> <p style="text-align: center;">Received 28 January 2000; received in revised form 8 March 2000</p> <p style="text-align: center;">Edited by Gianni Cesareni</p> <hr/> <p>Abstract We have transferred the prokaryotic $\gamma\delta$ resolvase system to mammalian cells and present a comparative analysis of recombination by wild-type and two mutant resolvases (E124Q and E102Y/E124Q). Transient co-transfection assays using β-galactosidase as reporter for recombination reveal that episomal DNA does not contain a significant level of unconstrained negative supercoiling, since only mutant resolvases are recombination-proficient. We also show that the efficiency of recombination by the resolvase double mutant is comparable to that observed with Cre, which indicates that resolvase can be used as a new tool for controlled manipulations of episomal DNAs.</p> <p>© 2000 Federation of European Biochemical Societies.</p> <hr/> <p>Key words: Site-specific recombination; $\gamma\delta$ resolvase; Mutant recombinase; Eukaryote; Episomal DNA; DNA topology</p> <hr/> <p>1. Introduction</p> <p>The bacterial $\gamma\delta$ (Tn1000) transposon-encoded resolvase belongs to the resolvase/invertase sub-family of conservative site-specific DNA recombinases. Its natural function is to resolve, through recombination, the co-integrate structure that is generated during transposition [1]. Resolvase recognizes a 114 bp DNA segment called res, which is composed of three binding sites (I–III) for resolvase dimers. Two res must be present as direct repeats on the same negatively (–) supercoiled DNA molecule. Only this orientation of sites allows for the formation of a functional synaptic complex, the so-called synaptosome, which entraps three (–) supercoils [2]. Intramolecular recombination catalyzed within the synaptosome thus results in excision of the DNA segment between res sites.</p> <p>An important prerequisite for efficient recombination by wild-type resolvase in <i>Escherichia coli</i> and in vitro is the presence of unconstrained (–) supercoiling in the substrate. Supercoiling is required for the assembly of a functional synaptosome and, in addition, energetically drives the DNA strand exchange reaction once the DNA has been cleaved at sub-binding sites I of the paired res [3]. We have transferred the resolvase system to mammalian cells and employ in this study wild-type and two $\gamma\delta$ resolvase mutants, E124Q and double mutant E102Y/E124Q. Both mutants are homologs of two previously characterized Tn3 resolvase mutants and</p>	<p>FEBS Letters 471 (2000) 147–150</p> <p>exhibit the same phenotype, i.e. they efficiently perform recombination in vitro on topologically relaxed DNA substrates containing two copies of res [4]. The double mutant, in addition, is proficient at catalyzing recombination on substrates containing two isolated I sites of res as direct or inverted repeats. However, the efficiency of the latter reaction is still significantly enhanced by (–) supercoiling of the substrate [4]. A comparative study of recombination by wild-type and mutant resolvase inside mammalian cells should therefore help to further elucidate the topological state of episomal and, eventually, genomic DNA.</p> <p>2. Materials and methods</p> <p>2.1. Expression and substrate vectors</p> <p>Expression vectors for wild-type and mutant resolvases are derivatives of pPGKCre (K. Fellenberg, University of Cologne). The resolvase genes were cloned by PCR using primers (P-γA) 5'-GATACCTG-CAGCATGCGACTTTTGGTTACGACCGGGTATCA-3', (P-γB) 5'-GATATCTAGATTAGTTGCTTTCATTTTACTTTATA-3', (P-δ124Q) 5'-CGACAGAGAATACTACAGCGTACCAATGAA-3', (P-δ124Qanti) 5'-TTCATTGGTACGCTGTAGTATTCTGT-CG-3', (P-δ102Y) 5'-AGTACCGATGGGTATATGGGTAATAAT-GGTT-3', (P-δ102Yanti) 5'-AACCATTTTACCCATATACCCAT-CGGTACT-3', and (P-2972) 5'-CATATCTAGACTATTAAC-CTTCTCTTCTTCTTAGGGTTGCTTTCATTTTACTTTATA-3'. Wild-type resolvase was generated with primers P-γA and P-γB, and DNA isolated from <i>E. coli</i> strain DH5α, which contains the $\gamma\delta$ transposon on the P' plasmid, served as template. The nuclear localization signal (NLS) variant (pPGK$\gamma\delta$NLS) was generated using P-2972 instead of P-γB. The resulting PCR fragments replaced the Cre gene in pPGKCre, using <i>Xba</i>I and <i>Pst</i>I. pPGKδ124 was generated by assembly PCR using P-γA and P-δ124Qanti, as well as P-γB and P-δ124Q as primer pairs. pPGK$\gamma\delta$ served as template. pPGKδ102 was also generated by assembly PCR using P-γA and P-δ102Y, as well as P-γB and P-δ102Yanti as primer pairs. pPGKδ124 served as template. The corresponding NLS-carrying variants were generated as described for pPGK$\gamma\delta$NLS.</p> <p>Substrate vectors are derivatives of pCH110 (Pharmacia). First, the prokaryotic promoter driving the <i>lacZ</i> gene in pCH110 was deleted by PCR and the <i>lacZ</i> gene was re-inserted. The neomycin gene was isolated from pSV2Neo [5] using <i>Sma</i>I and <i>Bgl</i>II, and subsequently flanked by two res sites which were generated by PCR employing pZWX1 [6] as template. After ligation, the res-neo-res cassette was cleaved with <i>Hind</i>III and inserted between the SV40 promoter and the <i>lacZ</i> gene of the modified pCH110 vector. The promoter proximal to res was modified at two residues in order to eliminate two ATGs, thus generating pCH-RNRZ, pCH-SNSZ and pCH-RLNRLZ are derivatives of pCH-RNRZ, and were generated by PCR. The corresponding recombinated product vectors of pCH-RNRZ and pCH-RLNRLZ were generated through transformation into DH5α or 294-Cre [7], respectively. pCH-SZ was generated by PCR.</p> <p>Plasmid DNAs were isolated from <i>E. coli</i> strains JC5547, BMH8117, DH5α or 294-Cre using affinity chromatography (Qiagen, Germany). Expression and substrate vectors were sequenced using the fluorescence-based 373A system (Applied Biosystems). PCRs were usually performed with 30 cycles using the 'Master MixQ Kit' (Qiagen, Germany). The reaction temperatures were calculated based on</p>
--	--

*Corresponding author. Fax: (49)-221-470 5170.
E-mail: p.droege@uni-koeln.de

Abbreviations: res, $\gamma\delta$ recombination site; lox, Cre recombination site; β -gal, β -galactosidase activity

0014-5793/00/\$20.00 © 2000 Federation of European Biochemical Societies. All rights reserved.
PII: S0014-5793(00)01394-6

Last page with the bibliographic references

150

The results show that the double mutant is at least as efficient as Cre at performing recombination on episomal substrates (Fig. 3C).

4. Discussion

We have demonstrated that two mutant $\gamma\delta$ resolvases can catalyze intramolecular recombination on episomal res inside mammalian cells. Recombination by wild-type resolvase, however, is not detectable. Based on the fact that only mutant resolvases recombine topologically relaxed substrates [4], a major conclusion is that episomal DNA does not contain a significant level of unconstrained (–) supercoiling over at least the time course of the experiment. This interpretation is supported by the result that only the double mutant recombines a substrate containing isolated I sites of res, albeit with a significantly reduced efficiency compared to recombination on a standard substrate.

Since we used (–) supercoiled plasmids for co-transfection, a second conclusion is that plasmid supercoiling must be relaxed once the DNA is introduced to the cells. This could be due to the action of topoisomerases [8] and/or the formation of nucleosomes on the internalized plasmids, a process which would consume unconstrained (–) supercoils. However, based on a topological analysis of plasmid DNA isolated from CHO cells 72 h post transfection, we found no evidence for nucleosome formation, i.e. the plasmid DNA was topologically relaxed (data not shown). Future comparative studies employing wild-type and mutant resolvase in reporter cell lines should also help to elucidate the topological state of genomic target sites and how it might be affected by, for example, transcription and the influence of topoisomerase or histone acetylase inhibitors [8,9].

A third conclusion from our study is that the resolvase

M. Schwikardi, P. Dröge/FEBS Letters 471 (2000) 147–150

double mutant performs recombination on episomal substrates with an efficiency comparable to that of Cre. The latter recombinase is used widely in studies elucidating the relevance and function of particular genes of interest in animal model systems [10]. If recombination by the resolvase double mutant is also efficient on genomic target sites, the available possibilities for controlled in vivo and ex vivo manipulations of eukaryotic genomes will be expanded.

Acknowledgements: Very special thanks go to N. Grindley for polyclonal antibodies raised against wild-type $\gamma\delta$ resolvase and to M. Boocock for communicating unpublished results. Thanks are due to K. Rajewsky and A. Tarakhovskiy who provided access to cell culture facilities, and to F. Edenhofer for comments on the manuscript. These studies were financed through SFB 274 Grant B10 to P.D., who is a DFG Heisenberg fellow (Dr187/8-1/8-2).

References

- [1] Stark, W.M., Boocock, M.R. and Sherratt, D.J. (1992) *Trends Genet.* 8, 432–439.
- [2] Dröge, P. and Cozzarelli, N.R. (1989) *Proc. Natl. Acad. Sci. USA* 86, 6062–6066.
- [3] Dröge, P., Hatfull, G.F., Grindley, N.D.F. and Cozzarelli, N.R. (1990) *Proc. Natl. Acad. Sci. USA* 87, 5336–5340.
- [4] Arnold, P.H., Blake, D.G., Grindley, N.D.F., Boocock, M.R. and Stark, W.M. (1999) *EMBO J.* 18, 1407–1414.
- [5] Southern, P.J. and Berg, P. (1982) *J. Mol. Appl. Genet.* 1, 327–341.
- [6] Wang, Z. and Dröge, P. (1997) *J. Mol. Biol.* 271, 499–510.
- [7] Buchholz, F., Angrand, P.-O. and Stewart, A.F. (1996) *Nucleic Acids Res.* 24, 3118–3119.
- [8] Wang, J.C. (1996) *Harvey Lect.* 81, 93–110.
- [9] Wang, Z. and Dröge, P. (1996) *EMBO J.* 15, 581–589.
- [10] Müller, U. (1999) *Mech. Dev.* 82, 3–21.
- [11] Lorbach, E., Christ, N., Schwikardi, M. and Dröge, P. (2000) *J. Mol. Biol.* 296, 1175–1181.

4.6 Example of the associated "dataset.toc" file

```
_t0 TST0123A
_vn 4.1
_pd 20001207
_st ITM

_t1 TST0123A 00145793
_jn FEBS Letters
_cr Copyright (c) 2000 Federation of European Biochemical Societies

_t2 TST0123A 00145793 V0471I02
_vl 471
_is 2-3
_cf [name] Special Issue: Topology of DNA
_pr 125-266
_dt 20000714
_np 152
_pn nil nil nil nil 125 126 127 128 129 130 131 132 133 134 135 136
    137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152
    153 154 .....etc..... 255 256 257 258 259 260 261 262 263 264
    265 266 nil nil nil nil nil nil
_ct 3 4

_t3 TST0123A 00145793 V0471I02 00013946
_ii S0014-5793(00)01394-6
_ty FLA
_li EN
_ti Site-specific recombination in mammalian cells catalyzed by
    @c@d resolvase mutants: implications for the topology of
    episomal DNA
_au Schwikardi, M
_au Droge, P
_ca P. Droge, Institute of Genetics, University of Cologne, Weyertal 121,
    D-50931 Cologne, Germany
_ab We have transferred the prokaryotic @c@d resolvase
    system to mammalian cells and present a comparative analysis of
    recombination by wild-type and two mutant resolvases (E124Q and
    E102Y/E124Q). Transient co-transfection assays using
    @b-galactosidase as reporter for recombination reveal that
    episomal DNA does not contain a significant level of unconstrained
    negative supercoiling, since only mutant resolvases are
    recombination-proficient. We also show that the efficiency of
    recombination by the resolvase double mutant is comparable to that
    observed with Cre, which indicates that resolvase can be used as a
    new tool for controlled manipulations of episomal DNAs.
_kw Site-specific recombination
_kw @c@d resolvase
_kw Mutant recombinase
_kw Eukaryote
_kw Episomal DNA
_kw DNA topology
_kw res, @c@d recombination site
_kw lox, Cre recombination site
_kw @b-gal, @b-galactosidase activity
_pg 147-150
_mf [SGML ART 4.2.0] main
_mf [PDF 1.2 DISTILLED] main
_mf [Raw ASCII] main
_mf [STRIPIN 1.0] stripin
```

4.7 Example of the associated SGML (*main.sgm*) file

(Line breaks and indentation are added for clarity)

```

<ART VERSION="4.1.0" JID="FEBS" AID="23531" PII="S0014579300013946" DOCSUBTY="FLA">
  <COPYRIGHT TYPE="SOCIETY" YR="2000">
    Federation of European Biochemical Societies
  </COPYRIGHT>
  <FM>
    <ATL>
      Site-specific recombination in mammalian cells catalyzed by
      &gamma;&delta; resolvase mutants: implications for the topology of
      episomal DNA
    </ATL>
    <DED>Edited by Gianni Cesareni</DED>
    <AUG>
      <AU>
        <FNM>Micha</FNM>
        <SNM>Schwikardi</SNM>
      </AU>
      <AU>
        <FNM>Peter</FNM>
        <SNM>Dr<A><AC>o</AC><AC>&uml;</AC></A>ge</SNM>
      </AU>
      <COR>Corresponding author. Fax: (49)-221-470 5170</COR>
      <EAD>p.droege@uni-koeln.de</EAD>
      <AFF>
        Institute of Genetics, University of Cologne, Weyertal 121,
        D-50931 <CTY>Cologne</CTY>, <CNY CNY-CODE="DE">Germany</CNY>
      </AFF>
    </AUG>
    <RE DAY="28" MO="1" YR="2000"> <RV DAY="8" MO="3" YR="2000">
    <ABS>
      <P>We have transferred the prokaryotic &gamma;&delta; resolvase
      system to mammalian cells and present a comparative analysis of
      recombination by wild-type and two mutant resolvases (E124Q and
      E102Y/E124Q). Transient co-transfection assays using
      &beta;-galactosidase as reporter for recombination reveal that
      episomal DNA does not contain a significant level of unconstrained
      negative supercoiling, since only mutant resolvases are
      recombination-proficient. We also show that the efficiency of
      recombination by the resolvase double mutant is comparable to that
      observed with Cre, which indicates that resolvase can be used as a
      new tool for controlled manipulations of episomal DNAs.</P>
    </ABS>
    <KWDG CLASS="KWD">
      <KWD>Site-specific recombination</KWD>
      <KWD>&gamma;&delta; resolvase</KWD>
      <KWD>Mutant recombinase</KWD>
      <KWD>Eukaryote</KWD>
      <KWD>Episomal DNA</KWD>
      <KWD>DNA topology</KWD>
    </KWDG>
    <KWDG CLASS="ABR">
      <KWD>res, &gamma;&delta; recombination site</KWD>
      <KWD>lox, Cre recombination site</KWD>
      <KWD>&beta;-gal, &beta;-galactosidase activity</KWD>
    </KWDG>
  </FM>

```

```

<BDY>
  <SEC>
    <NO>1</NO>
    <ST>Introduction</ST>
    <P>The bacterial  $\gamma$ ; $\delta$ ; (Tn<IT>1000</IT>) transposon-encoded resolvase belongs to the resolvase/invertase sub-family of conservative site-specific DNA recombinases. Its natural function is to resolve, through recombination, the co-integrate structure that is generated during transposition <CROSS-REF REFID="BIB1">[1]</CROSS-REF>. Resolvase recognizes a 114 bp DNA segment called res, which is composed of three binding sites (I&ndash;III) for resolvase dimers. Two res must be present as direct repeats on the same negatively ( $\ominus$ ; ) supercoiled DNA molecule. Only this orientation of sites allows for the formation of a functional synaptic complex, the so-called synaptosome, which entraps three ( $\ominus$ ; ) supercoils <CROSS-REF REFID="BIB2">[2]</CROSS-REF>. Intramolecular recombination catalyzed within the synaptosome thus results in excision of the DNA segment between res sites.</P><P>An important prerequisite for efficient recombination by wild-type resolvase in <IT>Escherichia coli</IT> and in vitro is the presence of unconstrained ( $\ominus$ ; ) supercoiling in the substrate. Supercoiling is required for the assembly of a functional synaptosome and, in addition, energetically drives the DNA strand exchange reaction once the DNA has been cleaved at sub-binding sites I of the paired res <CROSS-REF REFID="BIB3">[3]</CROSS-REF>. We have transferred the resolvase system to mammalian cells and employ in this study wild-type and two  $\gamma$ ; $\delta$ ; resolvase mutants, E124Q and double mutant E102Y/E124Q. Both mutants are homologs of two previously characterized Tn<IT>3</IT> resolvase mutants and exhibit the same phenotype, i.e. they efficiently perform recombination in vitro on topologically relaxed DNA substrates containing two copies of res <CROSS-REF REFID="BIB4">[4]</CROSS-REF>. The double mutant, in addition, is proficient at catalyzing recombination on substrates containing two isolated I sites of res as direct or inverted repeats. However, the efficiency of the latter reaction is still significantly enhanced by ( $\ominus$ ; ) supercoiling of the substrate <CROSS-REF REFID="BIB4">[4]</CROSS-REF>. A comparative study of recombination by wild-type and mutant resolvase inside mammalian cells should therefore help to further elucidate the topological state of episomal and, eventually, genomic DNA.</P>
  </SEC>
  <SEC>
    <NO>2</NO>
    <ST>Materials and methods</ST>
    <SEC>
      <NO>2.1</NO>
      <ST>Expression and substrate vectors</ST>
      <P>Expression vectors for wild-type and mutant resolvases are derivatives of pPGKCre (K. Fellenberg, University of Cologne). The resolvase genes were cloned by PCR using primers
      (P- $\gamma$ ; $\delta$ ;A)
      5&prime;;-GATACTGCAGCATGCGACTTTTTGGTTACGCACGGGTATCA-3&prime;;,
      (P- $\gamma$ ; $\delta$ ;B)
      5&prime;;-GATATCTAGATTAGTTGCTTTTCATTATTACTTTATA-3&prime;;,
      (P- $\gamma$ ; $\delta$ ;124Q)
      5&prime;;-CGACAGAGAATACTACAGCGTACCAATGAA-3&prime;;,
      (P- $\gamma$ ; $\delta$ ;124Qanti)
      5&prime;;-TTCATTGGTACGCTGTAGTATTCTCTGTCG-3&prime;;,

```

(P- γ ; δ ;102Y)
 5'prime;-AGTACCGATGGGTATATGGGTAAAATGGTT-3'prime;;
 (P- γ ; δ ;102Yanti)
 5'prime;-AACCATTTACCCATATACCCATCGGTACT-3'prime;; and (P-2972)
 AGACTATTAACCTTCTTCTTTCTTAGGGTTGCTTTCATTATTACTTTATA-3'prime;;.
 Wild-type resolvase was generated with primers P- γ ; δ ;A
 and P- γ ; δ ;B, and DNA isolated from *E. coli*
 strain DH5 α ;, which contains the γ ; δ ; transposon
 on the F' plasmid, served as template. The nuclear
 localization signal (NLS) variant (pPGK γ ; δ ;NLS) was
 generated using P-2972 instead of P- γ ; δ ;B. The
 resulting PCR fragments replaced the Cre gene in pPGKCre, using
 XbaI and PstI. pPGK γ ; δ ;124 was
 generated by assembly PCR using P- γ ; δ ;A and
 P- γ ; δ ;124Qanti, as well as P- γ ; δ ;B and
 P- γ ; δ ;124Q as primer pairs. pPGK γ ; δ ; served
 as template. pPGK γ ; δ ;102 was also generated by
 assembly PCR using P- γ ; δ ;A and P- γ ; δ ;102Y,
 as well as P- γ ; δ ;B and P- γ ; δ ;102Yanti as
 primer pairs. pPGK γ ; δ ;124 served as template. The
 corresponding NLS-carrying variants were generated as described
 for pPGK γ ; δ ;NLS. Substrate vectors are
 derivatives of pCH110 (Pharmacia). First, the prokaryotic
 promoter driving the lacZ gene in pCH110 was deleted by
 PCR and the lacZ gene was re-inserted. The neomycin
 gene was isolated from pSV2Neo
 using SmaI and BglII, and subsequently
 flanked by two res sites which were generated by PCR employing
 pZWX1 as template. After ligation, the res-neo-res cassette was
 cleaved with HinIII and inserted between the SV40
 promoter and the lacZ gene of the modified pCH110
 vector. The promoter proximal to res was modified at two
 residues in order to eliminate two ATGs, thus generating
 pCH-RNRZ. pCH-SNSZ and pCH-RLNRLZ are derivatives of pCH-RNRZ,
 and were generated by PCR. The corresponding recombined product
 vectors of pCH-RNRZ and pCH-RLNRLZ were generated through
 transformation into DH5 α ; or 294-Cre
 respectively. pCH-SZ was generated by PCR. Plasmid DNAs
 were isolated from *E. coli* strains JC5547, BMH8117,
 DH5 α ; or 294-Cre using affinity chromatography (Qiagen,
 Germany). Expression and substrate vectors were sequenced using
 the fluorescence-based 373A system (Applied Biosystems). PCRs
 were usually performed with 30 cycles using the 'Master
 MixQ Kit' (Qiagen, Germany). The reaction temperatures
 were calculated based on the respective primer sequences and the
 length of the expected product.

</SEC>

<SEC>

<NO>2.2</NO>

<ST>Cell culture and in vivo recombination assays</ST>

<P>Transient resolvase expression and recombination assays
 employed CHO cells cultured in Dulbecco's modified
 Eagle's medium containing 10% fetal calf serum, 2 mM
 L</SCP>

-glutamine, streptomycin (0.1 mg/ml), and penicillin (100 U/ml).
 Transient co-transfection assays were performed with supercoiled
 expression and substrate vectors at a molar ratio of about 3:1

and lipofection (FuGene[®]; Boehringer Mannheim, Germany). About 2 × 10⁵ cells were transfected with 2 μg of mixed plasmid DNA. Transfection efficiencies were usually in the range of 20–30%. Cells were analyzed 72 h post transfection by either cell staining or determining the β-galactosidase activity (β-gal).

</SEC>

<SEC>

<NO>2.3</NO>

<ST>β-Gal assays</ST>

Cells were stained after fixation with 0.5% (w/v) glutaraldehyde through incubation with a solution made up of 5 mM K₃Fe(CN)₆ · 3H₂O, 2 mM MgCl₂, and 4 mg/ml X-Gal in phosphate-buffered saline. The number of stained cells was determined microscopically after 7–8 h of incubation at 37°C. β-Gal activity was determined in cell lysates using the Galacto Light Kit (Tropix, Perkin Elmer), as recommended by the manufacturer. Relative light units (RLU) were determined in a Lumat LB9501 luminometer (Berthold), and normalized to the amount of protein (μg) in each lysate determined by the Bio-Rad protein assay.

</SEC>

<SEC>

<NO>2.4</NO>

<ST>Immunoblotting</ST>

Cell lysates from transiently transfected cells were prepared by boiling the harvested cells in sample buffer for 5 min. Proteins were separated through 12.5% SDS polyacrylamide gels and transferred onto a nitrocellulose membrane (Immobilon P, Millipore). The membrane was blocked with 1% blocking solution and incubated with mouse polyclonal antibodies raised against wild-type resolvase at a 1:3000 dilution. Peroxidase-conjugated secondary antibodies were used for detection (Boehringer Mannheim, Germany). Purified resolvase was used as a control

<CROSS-REF REFID="BIB6">[6]</CROSS-REF>.

</SEC>

</SEC>

<SEC>

<NO>3</NO>

<ST>Results</ST>

<SEC>

<NO>3.1</NO>

<ST>Expression of wild-type and mutant resolvases in CHO cells</ST>

We constructed eukaryotic expression vectors which contain the coding region for wild-type or mutant resolvase under the control of the phosphoglycerate kinase promoter (PGK). Two sets of vectors were generated: one containing an NLS and a corresponding set (Fig. 1A) lacking one fused to the C-terminal end of resolvase genes. At 72 h post transfection into CHO cells, we could detect resolvase by Western blotting using mouse polyclonal antibodies raised against wild-type resolvase (Fig. 2), lanes 3–8). The parental expression vector (pPGKCre) containing the Cre gene (lane 1) and purified resolvase (lane 2) served as controls. The results also show that the presence of the NLS has a profound effect on the amount of resolvase detectable in cell lysates (lanes 4, 6, and 8). However, the

amount of each resolvase variant generated from the same set of expression vectors seems to be comparable. Identical results were obtained after transfection into HeLa and NIH3T3 cells (data not shown). We conclude that presumably full-length resolvase protein is produced by different cell lines.</P>

</SEC>

<SEC>

<NO>3.2</NO>

<ST>Only mutant resolvases are recombination-proficient</ST>

<P>In order to test whether wild-type and/or mutant resolvase can perform recombination in CHO cells, we constructed a substrate plasmid termed pCH-RNRZ (

<CROSS-REF REFID="FIG1">Fig. 1B</CROSS-REF>).

This vector contains two res as direct repeats flanking the neomycin resistance gene (<IT>neo</IT>). This recombination cassette, in turn, is flanked by the SV40 promoter (SV40) and the <IT>lacZ</IT> gene. Hence, recombination between res should lead to deletion of the <IT>neo</IT> gene and the subsequent expression of β-gal.</P><P>We introduced pCH-RNRZ together with each of the six expression vectors for resolvase into CHO cells. Cell extracts were prepared 72 h after transfection and recombination was monitored by determining the normalized β-gal activity. As a control to determine the highest achievable level, set at 100%, we co-transfected the recombined product vector pCH-RZ with pPGKCre. The results (

<CROSS-REF REFID="FIG3">Fig. 3A</CROSS-REF>

) show that cells expressing mutant resolvase, either with or without NLS, exhibit about 60% β-gal activity. This level is even higher (about 80%) when compared to a second control, i.e. pCH-RZ co-transfected with the expression vector for the double mutant, pPGKγδ102. In contrast, cells co-transfected with pPGKγδ or pPGKγδNLS and pCH-RNRZ exhibit only negligible β-gal activity (5%). This low level is comparable to two other controls: the substrate pCH-RNRZ co-transfected with pPGKCre, and pPGKγδ102 transfected alone.</P><P>We also determined the fraction of X-gal-stained cells 72 h after co-transfection of substrate and expression vectors. The result of a representative experiment (

<CROSS-REF REFID="TBL1">Table 1</CROSS-REF>

) reveals that both mutants generate a significant fraction of stained cells if compared again to the control employing the recombined substrate (pCH-RZ), set at 100%. However, this fraction is rather low and comparable to a negative control (pPGKCre) when either pPGKγδ or pPGKγδNLS was used. We conclude that within the limits of detection by these methods both mutant resolvases, but not the wild-type, catalyze recombination on episomal substrates in mammalian cells.</P>

</SEC>

<SEC>

<NO>3.3</NO>

<ST>Recombination on isolated I sites of res</ST>

<P>We were interested to test whether mutant resolvases can recombine two copies of isolated I sites of res. We therefore constructed pCH-SNSZ, which is identical to pCH-RNRZ except that sub-sites II and III of each res are deleted (compare

<CROSS-REF REFID="FIG1">Fig. 1B</CROSS-REF>

). Recombination assays employing the corresponding product vector pCH-SZ and pPGKCre as control (100%) were performed as

described above. The results show ([Fig. 3B](#)) that only the resolvase double mutant recombines this substrate in CHO cells. However, the efficiency of this reaction is significantly reduced compared to recombination on pCH-RNRZ.

3.4

Cre and the resolvase double mutant exhibit comparable activities

In order to compare the recombination efficiency of the resolvase double mutant with that of Cre, we generated pCH-RLNRLZ. This substrate is identical to pCH-RNRZ, but contains in addition two recombination sites (lox) for Cre as direct repeats also flanking the [neo](#) gene ([Fig. 1C](#)). Hence, recombination by either resolvase or Cre deletes the [neo](#) gene and results in the subsequent expression of the [lacZ](#) gene. Co-transfections were performed with pPGK γ ; δ ;102NLS and pPGKCre as described above. The corresponding recombined product vector was used in each case as control (100%). The results show that the double mutant is at least as efficient as Cre at performing recombination on episomal substrates ([Fig. 3C](#)).

Discussion

We have demonstrated that two mutant γ ; δ resolvases can catalyze intramolecular recombination on episomal res inside mammalian cells. Recombination by wild-type resolvase, however, is not detectable. Based on the fact that only mutant resolvases recombine topologically relaxed substrates [\[4\]](#), a major conclusion is that episomal DNA does not contain a significant level of unconstrained (γ) supercoiling over at least the time course of the experiment. This interpretation is supported by the result that only the double mutant recombines a substrate containing isolated I sites of res, albeit with a significantly reduced efficiency compared to recombination on a standard substrate.

Since we used (γ) supercoiled plasmids for co-transfection, a second conclusion is that plasmid supercoiling must be relaxed once the DNA is introduced to the cells. This could be due to the action of topoisomerases [\[8\]](#) and/or the formation of nucleosomes on the internalized plasmids, a process which would consume unconstrained (γ) supercoils. However, based on a topological analysis of plasmid DNA isolated from CHO cells 72 h post transfection, we found no evidence for nucleosome formation, i.e. the plasmid DNA was topologically relaxed (data not shown). Future comparative studies employing wild-type and mutant resolvase in reporter cell lines should also help to elucidate the topological state of genomic target sites and how it might be affected by, for example, transcription and the influence of topoisomerase or histone acetylase inhibitors [\[8,9\]](#).

<P>A third conclusion from our study is that the resolvase double mutant performs recombination on episomal substrates with an efficiency comparable to that of Cre. The latter recombinase is used widely in studies elucidating the relevance and function of particular genes of interest in animal model systems <CROSS-REF REFID="BIB10">[10]</CROSS-REF>.

If recombination by the resolvase double mutant is also efficient on genomic target sites, the available possibilities for controlled in vivo and ex vivo manipulations of eukaryotic genomes will be expanded.

<FIG ID="FIG1">

<NO>Fig. 1</NO>

<CAPTION>

<P>Resolvase expression and substrate vectors. A: Relevant genetic elements present on expression vectors for Cre, wild-type or two mutant resolvases are schematized. A corresponding second set of vectors for resolvase, labelled with the postscript NLS, carries a SV40 large tumor antigen NLS fused to the C-terminal end of resolvase (not shown). B: Standard substrate for resolvase. C: Standard substrate for both resolvase and Cre. The orientation of recombination sites is indicated by arrows. The positions of resolvase sub-binding sites within res are demarcated by black rectangles and labelled I–III. See text for details.</P>

</CAPTION>

<LINK LOCATOR="gr1">

</FIG>

<FIG ID="FIG2">

<NO>Fig. 2</NO>

<CAPTION>

<P>Western analysis. Wild-type and mutant resolvases expressed in CHO cells. The higher and lower arrows point to the position of resolvase with and without an NLS, respectively.</P>

</CAPTION>

<LINK LOCATOR="gr2">

</FIG>

<FIG ID="FIG3">

<NO>Fig. 3</NO>

<CAPTION>

<P>Analysis of recombination by β-gal assays. A: Recombination on pCH-RNRZ containing two complete res. B: Recombination on pCH-SNSZ containing two isolated copies of I sites of res. C: Recombination on pCH-RLNRLZ containing res and lox sites. As an additional control, in C we used an expression vector for a phage λ integrase mutant, Int-h, co-transfected with pCH-RLNRLZ <CROSS-REF REFID="BIB11">[11]</CROSS-REF>.

In each case, data were collected from four to six separate transfections assays each employing between two and four wells containing about 2×10⁵ CHO cells. The graphs show the mean values with standard deviations indicated by vertical lines. See text for details.</P>

</CAPTION>

<LINK LOCATOR="gr3">

</FIG>

<TBL ID="TBL1">

<NO>Table 1</NO>

```

<CAPTION>
  <P>Recombination analyzed by cell staining</P>
</CAPTION>
<TBLBDY>
  <R>
    <C><TOP-BORDER STYLE="S">
      <BOTTOM-BORDER STYLE="S">
        Co-transfected plasmids
      </C>
    <C><TOP-BORDER STYLE="S">
      <BOTTOM-BORDER STYLE="S">
        Stained cells (%)
      </C>
    </R>
    <R>
      <C>None</C>
      <C CA="D">0</C>
    </R>
    <R>
      <C>pCH-RZ+pPGK&gamma;&delta;102NLS</C>
      <C CA="D">100</C>
    </R>
    <R>
      <C>pCH-RNRZ+pPGKCre</C>
      <C CA="D">9</C>
    </R>
    <R>
      <C>pCH-RNRZ+pPGK&gamma;&delta;</C>
      <C CA="D">4</C>
    </R>
    <R>
      <C>pCH-RNRZ+pPGK&gamma;&delta;NLS</C>
      <C CA="D">11</C>
    </R>
    <R>
      <C>pCH-RNRZ+pPGK&gamma;&delta;124NLS</C>
      <C CA="D">44</C>
    </R>
    <R>
      <C>
        <BOTTOM-BORDER STYLE="S">
          pCH-RNRZ+pPGK&gamma;&delta;102NLS
        </C>
      <C CA="D">
        <BOTTOM-BORDER STYLE="S">
          96
        </C>
    </R>
  </TBLBDY>
</TBL>
</P>
</SEC>
</BDY>

```

```

<BM>
  <ACK>
    <P>Very special thanks go to N. Grindley for polyclonal antibodies
    raised against wild-type  $\gamma$ ;  $\delta$ ; resolvase and to M. Boocock
    for communicating unpublished results. Thanks are due to K. Rajewsky
    and A. Tarakhovsky who provided access to cell culture facilities,
    and to F. Edenhofer for comments on the manuscript. These studies
    were financed through SFB 274 Grant B10 to P.D., who is a DFG
    Heisenberg fellow (Dr187/8-1/-8-2).</P>
  </ACK>

  <BIBL>
    <BIB ID="BIB1">
      <NO>[1]</NO>
      <BB>
        <CONTRIBUTION>
          <AUTHORS>
            <AUTHOR>
              <SNM>Stark</SNM>
              <FNM>W.M.</FNM>
            </AUTHOR>
            <AUTHOR>
              <SNM>Boocock</SNM>
              <FNM>M.R.</FNM>
            </AUTHOR>
            <AUTHOR>
              <SNM>Sherratt</SNM>
              <FNM>D.J.</FNM>
            </AUTHOR>
          </AUTHORS>
        </CONTRIBUTION>
        <HOST>
          <ISSUE>
            <SERIES>
              <TITLE>Trends Genet.</TITLE>
              <VOLUME-NR>8</VOLUME-NR>
            </SERIES>
            <DATE>1992</DATE>
            <PAGES>
              <FIRST-PAGE>432</FIRST-PAGE>
              <LAST-PAGE>439</LAST-PAGE>
            </PAGES>
          </ISSUE>
        </HOST>
      </BB>
    </BIB>

    <BIB ID="BIB2">
      <NO>[2]</NO>
      <BB>
        <CONTRIBUTION>
          <AUTHORS>
            <AUTHOR>
              <SNM>Dr<A><AC>o</AC><AC>&uml ;</AC></A>ge</SNM>
              <FNM>P.</FNM>
            </AUTHOR>
            <AUTHOR>
              <SNM>Cozzarelli</SNM>
              <FNM>N.R.</FNM>
            </AUTHOR>
          </AUTHORS>
        </CONTRIBUTION>
      </BB>
    </BIB>
  </BIBL>

```

```

    </AUTHORS>
  </CONTRIBUTION>
  <HOST>
    <ISSUE>
      <SERIES>
        <TITLE>Proc. Natl. Acad. Sci. USA</TITLE>
        <VOLUME-NR>86</VOLUME-NR>
      </SERIES>
      <DATE>1989</DATE>
      <PAGES>
        <FIRST-PAGE>6062</FIRST-PAGE>
        <LAST-PAGE>6066</LAST-PAGE>
      </PAGES>
    </ISSUE>
  </HOST>
</BB>
</BIB>

<BIB ID="BIB3">
  <NO>[ 3 ]</NO>
  <BB>
    <CONTRIBUTION>
      <AUTHORS>
        <AUTHOR>
          <SNM>Dr<A><AC>o</AC><AC>&uml;</AC></A>ge</SNM>
          <FNM>P.</FNM>
        </AUTHOR>
        <AUTHOR>
          <SNM>Hatfull</SNM>
          <FNM>G.F.</FNM>
        </AUTHOR>
        <AUTHOR>
          <SNM>Grindley</SNM>
          <FNM>N.D.F.</FNM>
        </AUTHOR>
        <AUTHOR>
          <SNM>Cozzarelli</SNM>
          <FNM>N.R.</FNM>
        </AUTHOR>
      </AUTHORS>
    </CONTRIBUTION>
  <HOST>
    <ISSUE>
      <SERIES>
        <TITLE>Proc. Natl. Acad. Sci. USA</TITLE>
        <VOLUME-NR>87</VOLUME-NR>
      </SERIES>
      <DATE>1990</DATE>
      <PAGES>
        <FIRST-PAGE>5336</FIRST-PAGE>
        <LAST-PAGE>5340</LAST-PAGE>
      </PAGES>
    </ISSUE>
  </HOST>
</BB>
</BIB>

<BIB ID="BIB4">
  <NO>[ 4 ]</NO>
  <BB>

```

```
<CONTRIBUTION>
  <AUTHORS>
    <AUTHOR>
      <SNM>Arnold</SNM>
      <FNM>P.H.</FNM>
    </AUTHOR>
    <AUTHOR>
      <SNM>Blake</SNM>
      <FNM>D.G.</FNM>
    </AUTHOR>
    <AUTHOR>
      <SNM>Grindley</SNM>
      <FNM>N.D.F.</FNM>
    </AUTHOR>
    <AUTHOR>
      <SNM>Boocock</SNM>
      <FNM>M.R.</FNM>
    </AUTHOR>
    <AUTHOR>
      <SNM>Stark</SNM>
      <FNM>W.M.</FNM>
    </AUTHOR>
  </AUTHORS>
</CONTRIBUTION>
<HOST>
  <ISSUE>
    <SERIES>
      <TITLE>EMBO J.</TITLE>
      <VOLUME-NR>18</VOLUME-NR>
    </SERIES>
    <DATE>1999</DATE>
    <PAGES>
      <FIRST-PAGE>1407</FIRST-PAGE>
      <LAST-PAGE>1414</LAST-PAGE>
    </PAGES>
  </ISSUE>
</HOST>
</BB>
</BIB>

<BIB ID="BIB5">
  <NO>[5]</NO>
  <BB>
    <CONTRIBUTION>
      <AUTHORS>
        <AUTHOR>
          <SNM>Southern</SNM>
          <FNM>P.J.</FNM>
        </AUTHOR>
        <AUTHOR>
          <SNM>Berg</SNM>
          <FNM>P.</FNM>
        </AUTHOR>
      </AUTHORS>
    </CONTRIBUTION>
    <HOST>
      <ISSUE>
        <SERIES>
          <TITLE>J. Mol. Appl. Genet.</TITLE>
          <VOLUME-NR>1</VOLUME-NR>
```

```

        </SERIES>
        <DATE>1982</DATE>
        <PAGES>
            <FIRST-PAGE>327</FIRST-PAGE>
            <LAST-PAGE>341</LAST-PAGE>
        </PAGES>
    </ISSUE>
</HOST>
</BB>
</BIB>

<BIB ID="BIB6">
    <NO>[ 6 ]</NO>
    <BB>
        <CONTRIBUTION>
            <AUTHORS>
                <AUTHOR>
                    <SNM>Wang</SNM>
                    <FNM>Z.</FNM>
                </AUTHOR>
                <AUTHOR>
                    <SNM>Dr<A><AC>o</AC><AC>&uml ;</AC></A>ge</SNM>
                    <FNM>P.</FNM>
                </AUTHOR>
            </AUTHORS>
        </CONTRIBUTION>
        <HOST>
            <ISSUE>
                <SERIES>
                    <TITLE>J. Mol. Biol.</TITLE>
                    <VOLUME-NR>271</VOLUME-NR>
                </SERIES>
                <DATE>1997</DATE>
                <PAGES>
                    <FIRST-PAGE>499</FIRST-PAGE>
                    <LAST-PAGE>510</LAST-PAGE>
                </PAGES>
            </ISSUE>
        </HOST>
    </BB>
</BIB>

<BIB ID="BIB7">
    <NO>[ 7 ]</NO>
    <BB>
        <CONTRIBUTION>
            <AUTHORS>
                <AUTHOR>
                    <SNM>Buchholz</SNM>
                    <FNM>F.</FNM>
                </AUTHOR>
                <AUTHOR>
                    <SNM>Angrand</SNM>
                    <FNM>P.-O.</FNM>
                </AUTHOR>
                <AUTHOR>
                    <SNM>Stewart</SNM>
                    <FNM>A.F.</FNM>
                </AUTHOR>
            </AUTHORS>
        </CONTRIBUTION>
    </BB>
</BIB>

```

```

        </CONTRIBUTION>
        <HOST>
          <ISSUE>
            <SERIES>
              <TITLE>Nucleic Acids Res.</TITLE>
              <VOLUME-NR>24</VOLUME-NR>
            </SERIES>
            <DATE>1996</DATE>
            <PAGES>
              <FIRST-PAGE>3118</FIRST-PAGE>
              <LAST-PAGE>3119</LAST-PAGE>
            </PAGES>
          </ISSUE>
        </HOST>
      </BB>
    </BIB>

    <BIB ID="BIB8">
      <NO>[ 8]</NO>
      <BB>
        <CONTRIBUTION>
          <AUTHORS>
            <AUTHOR>
              <SNM>Wang</SNM>
              <FNM>J.C.</FNM>
            </AUTHOR>
          </AUTHORS>
        </CONTRIBUTION>
        <HOST>
          <ISSUE>
            <SERIES>
              <TITLE>Harvey Lect.</TITLE>
              <VOLUME-NR>81</VOLUME-NR>
            </SERIES>
            <DATE>1996</DATE>
            <PAGES>
              <FIRST-PAGE>93</FIRST-PAGE>
              <LAST-PAGE>110</LAST-PAGE>
            </PAGES>
          </ISSUE>
        </HOST>
      </BB>
    </BIB>

    <BIB ID="BIB9">
      <NO>[ 9]</NO>
      <BB>
        <CONTRIBUTION>
          <AUTHORS>
            <AUTHOR>
              <SNM>Wang</SNM>
              <FNM>Z.</FNM>
            </AUTHOR>
            <AUTHOR>
              <SNM>Dr<A><AC>o</AC><AC>&uml ;</AC></A>ge</SNM>
              <FNM>P.</FNM>
            </AUTHOR>
          </AUTHORS>
        </CONTRIBUTION>
        <HOST>

```

```
<ISSUE>
  <SERIES>
    <TITLE>EMBO J.</TITLE>
    <VOLUME-NR>15</VOLUME-NR>
  </SERIES>
  <DATE>1996</DATE>
  <PAGES>
    <FIRST-PAGE>581</FIRST-PAGE>
    <LAST-PAGE>589</LAST-PAGE>
  </PAGES>
</ISSUE>
</HOST>
</BB>
</BIB>

<BIB ID="BIB10">
  <NO>[10]</NO>
  <BB>
    <CONTRIBUTION>
      <AUTHORS>
        <AUTHOR>
          <SNM>M<A><AC>u</AC><AC>&uml</AC></A>ller</SNM>
          <FNM>U.</FNM>
        </AUTHOR>
      </AUTHORS>
    </CONTRIBUTION>
  </HOST>
  <ISSUE>
    <SERIES>
      <TITLE>Mech. Dev.</TITLE>
      <VOLUME-NR>82</VOLUME-NR>
    </SERIES>
    <DATE>1999</DATE>
    <PAGES>
      <FIRST-PAGE>3</FIRST-PAGE>
      <LAST-PAGE>21</LAST-PAGE>
    </PAGES>
  </ISSUE>
</HOST>
</BB>
</BIB>

<BIB ID="BIB11">
  <NO>[11]</NO>
  <BB>
    <CONTRIBUTION>
      <AUTHORS>
        <AUTHOR>
          <SNM>Lorbach</SNM>
          <FNM>E.</FNM>
        </AUTHOR>
        <AUTHOR>
          <SNM>Christ</SNM>
          <FNM>N.</FNM>
        </AUTHOR>
        <AUTHOR>
          <SNM>Schwikardi</SNM>
          <FNM>M.</FNM>
        </AUTHOR>
      </AUTHORS>
    </CONTRIBUTION>
  </HOST>
</BB>
</BIB>
```

```

                <SNM>Dr<A><AC>o</AC><AC>&uml ;</AC></A>ge</SNM>
                <FNM>P.</FNM>
            </AUTHOR>
        </AUTHORS>
    </CONTRIBUTION>
<HOST>
    <ISSUE>
        <SERIES>
            <TITLE>J. Mol. Biol.</TITLE>
            <VOLUME-NR>296</VOLUME-NR>
        </SERIES>
        <DATE>2000</DATE>
        <PAGES>
            <FIRST-PAGE>1175</FIRST-PAGE>
            <LAST-PAGE>1181</LAST-PAGE>
        </PAGES>
    </ISSUE>
</HOST>
    </BB>
</BIB>
</BIBL>
</BM>
</ART>
```

5 Distribution

5.1 Fulfilment frequency

There is the option to receive electronic journals on a weekly, bi-weekly or monthly basis. CD-Rom's will normally be shipped within 24-48 hours of the date given to the customer. All CD-Rom's will be distributed via a service that can provide on-line tracking of package location and provide detailed information on final delivery address and signature.

5.2 Media options

CD-ROM, conforming to the ISO 9660 Mode I standard. CD-Rom's based on this standard are usable in MS-DOS, Apple Macintosh and UNIX operating systems.

5.3 Printout

SDOS 3.0 datasets will be accompanied by a printout containing dataset contents information.

An example of such a printout:

Dataset	: ESX0051A					
Production Date:	16-Aug-2001					
Report Date	: 16-Aug-2001					
01677799	Trends in Biotechnology					
	SZ982665	Vol 16	Iss 1	Date 199811	NP	56 NI 10
00032670	Analytica Chimica Acta					
	SZ972592	Vol 353	Iss 2-3	Date 19981030	NP	260 NI 30
03781119	Gene					
	V0206I01	Vol 206	Iss 1	Date 19980105	NP	164 NI 20
	SZ972760	Vol 205	Iss 1-2	Date 19971231	NP	364 NI 36
01686445	FEMS Microbiology Reviews					
	V0021I03	Vol 21	Iss 3	Date 199711	NP	76 NI 3
00052728	Biochimica et Biophysica Acta (BBA)/Bioenergetics					
	SZ972744	Vol 1322	Iss 2-3	Date 19971215	NP	188 NI 16
00052736	Biochimica et Biophysica Acta (BBA)/Biomembranes					
	V1368I01	Vol 1368	Iss 1	Date 19980105	NP	180 NI 16
00052760	Biochimica et Biophysica Acta (BBA)/Lipids and Lipid Metabolism					
	SZ982746	Vol 1389	Iss 2	Date 19980115	NP	112 NI 9

Note The codes **NP** and **NI** stand for "number of pages" and "number of items" respectively.

Appendix 1 SGML structure

1.1 Output specifications for SGML files

1.1.1 Regular SGML files

SGML files are normalized according to the SGML standard (ISO 8879:1986):

- Explicit insertion of optional beginning- or end-tags as implied by the context through application of the appropriate DTD,
- Explicit insertion of default attributes according to the appropriate DTD,
- No omission of literal delimiters (quotes) around attribute values,
- All entity references are written out in full (no missing “;”), and
- No tag abbreviations.

In addition to this, a special normalization rule applies:

- All tags and attribute names must be in upper case;
- From version 3.0 onward of the Elsevier Science FLA DTD, it is possible to have SGML elements within attributes. The above normalization rules also apply to these embedded elements.

1.1.2 Linked textboxes

In DTD 4.1 there is a facility known as linked textboxes **<TEXTBOX>**. This facility may hold a complete (short) article using the **<LINK>** element. The linked article is a local external object (SGML file). Each linked file may have incorporated artwork and this artwork may be cross-referenced from other linked documents.

The textbox should be handled as if it were a regular article.

The conversion handles each textbox as a separate SGML article, and this is reflected in the EFFECT structure. The file of each textbox is located within an individual subdirectory.

The subdirectory is given the name of the textbox file. The stripin.toc file is created as if it were a regular article. Therefore, the numbering within the stripin.toc file starts at 1.

1.2 Conversion failures

Elsevier verifies and converts SGML files according to the above-described specification. The conversion program attempts to deal with all eventualities, but a fallback mechanism is deployed in case of a conversion error.

1.2.1 Regular files

In case of a problem with a full SGML article, the article is transformed to contain only a head. To be precise, the body and possibly backmatter is removed, but also the nomenclature (if it exists) is taken out of the frontmatter. Any cross-references pointing outside the frontmatter are removed, as are entity declarations for entities which are not referred to in the frontmatter.

1.2.2 Linked textboxes

As with regular files, with linked textboxes there may be SGML conversion problems. Therefore, in case there is a problem within a linked textbox, the main article will be processed as a head-only file. Consequently, all linked textboxes are omitted from the article.

Appendix 2 Stripin components

2.1 Introduction

The Elsevier Science Full Length Article DTD versions 3.0 and higher have a number of SGML constructions which cannot be easily expressed in HTML. Elsevier delivers GIF images of this so-called *HTML-incompatible markup (HIM)*. These GIF images are further called “stripin” elements or GIFs as they are similar to the typesetter practice of stripping in small pictures in film or print plates for “difficult” constructions.

It should be noted that when a stripin element is nested inside another stripin element, only one GIF is supplied, namely of the outer (parent) construction. An example is an accent within an equation. Stripin elements within a table body constitute a special case, in that they are delivered both as separate GIFs and as part of the table body GIF.

The GIF images of stripin elements satisfy the following specifications:

- Character size of 12 pt,
- Resolution of 100 dpi,
- Border of 1 pt (0.351 mm),
- Transparent with a white background,
- Anti-aliased.

2.2 Affected SGML elements

2.2.1 Accent constructions

SGML accent constructions `<A>` are converted to a GIF image, except for a number of accents for which an equivalent HTML variant exists. These HTML equivalents are not converted into GIF images, because these can simply be translated into the HTML variant.

The following table lists SGML accent constructions and the equivalent HTML variants. The *Base character* and *Accent* fields are to be read (ignoring the valign attribute) as:

```
<A><AC>Base character</AC><AC>Accent</AC></A>
```

For instance, the SGML for í is

```
<A><AC>&inodot;</AC><AC>&acute;</AC></A>
```

The HTML equivalent for this character is simply `í`

Character	SGML Accent Constructions		
	Base character	Accent	HTML entity
í	ı	´	í
î	ı	ˆ	î
ì	ı	`	ì
ï	ı	¨	ï
Á	A	´	Á
Â	A	ˆ	Â
À	A	`	À
Å	A	˚	Å
Ã	A	˜	Ã
Ä	A	¨	Ä

SGML Accent Constructions				
Character	Base character	Accent	HTML entity	
Ç	C	¸	Ç	
É	E	´	É	
Ê	E	ˆ	Ê	
È	E	`	È	
Ë	E	¨	Ë	
Í	I	´	Í	
Î	I	ˆ	Î	
Ì	I	`	Ì	
Ï	I	¨	Ï	
Ñ	N	˜	Ñ	
Ó	O	´	Ó	
Ô	O	ˆ	Ô	
Ò	O	`	Ò	
Ø	O	&slash;	Ø	
Õ	O	˜	Õ	
Ö	O	¨	Ö	
Ú	U	´	Ú	
Û	U	ˆ	Û	
Ù	U	`	Ù	
Ü	U	¨	Ü	
Ý	Y	´	Ý	
ÿ	Y	¨	Ÿ	
á	a	´	á	
â	a	ˆ	â	
à	a	`	à	
å	a	˚	å	
ã	a	˜	ã	
ä	a	¨	ä	
ç	c	¸	ç	
é	e	´	é	
ê	e	ˆ	ê	
è	e	`	è	
ë	e	¨	ë	
ñ	n	˜	ñ	
ó	o	´	ó	
ô	o	ˆ	ô	
ò	o	`	ò	
ø	o	&slash;	ø	
õ	o	˜	õ	
ö	o	¨	ö	
ú	u	´	ú	
û	u	ˆ	û	
ù	u	`	ù	
ü	u	¨	ü	
ý	y	´	ý	
ÿ	y	¨	ÿ	

2.2.2 Arrays

When an array text construction **<AR>** is not part of an equation, it will be converted into a separate GIF image. Otherwise it is seen as an integral part of the equation.

2.2.3 Boxes and lines

Within the DTD, there are the box tag **<BOX>**, underline tags **<UNL>** and **<UN>**, and overline tags **<OVL>** and **<OV>**. All these items will be converted to a GIF image when not already part of an equation or array.

2.2.4 Inline and displayed equations

Both inline formulas **<F>**, and displayed formulas **<FD>**, are converted to a GIF image. All stripin elements that are present within these constructions will be part of the parent GIF image. This includes elements listed in this subsection. Formulas wider than A4 page width should be wrapped, not truncated.

Note: It is at your discretion to render inline and displayed formulas directly into HTML where possible, and therewith ignore the replacement GIF images.

2.2.5 Fence, fraction, radical and limit constructions

When a fence **<FEN>**, fraction **<FR>**, radical **<RAD>**, or limit **<LIM>**, construction is not already part of an array or a formula, it will be converted into a separate image. These constructions may also be nested, in which case one GIF is created for the outer construction (containing the inner one).

2.2.6 Tables

Tables require two forms of output: Not only should GIFs are created for any stripin element occurring within the table, but also the complete table body (i.e. not the caption, legend or footnotes) is supplied as a GIF image. This allows you to create an HTML table with the option to resort to the body GIF in case of problems.

As to complete body GIFs, if A4 page width is insufficient to contain the table, then the page size is increased to A0 and the resulting image is cropped to remove superfluous space. If A0 page width is still not sufficient, the table is spread over several pages but the resulting page images are joined together into one GIF image.

2.2.7 Certain fonts

Characters in the following fonts are to be converted into a GIF image:

- German font **<GE>**
- Script font **<SC>**
- Sans serif font **<SSF>**
- Open face **<OF>**
- Small caps **<SCP>**

2.2.8 Entities

SGML entities are only converted to GIF when they are within one of the constructions as listed in this subsection.

2.3 stripin.toc file

The stripin.toc file lists all constructions within the SGML file that were converted into a GIF image. Any element that is a child (sub-element) of an element that has a stripin.toc entry, does not have an entry for itself.

In case there are two or more identical SGML constructions within the article, then only one GIF image will be generated. The stripin.toc file could therefore have multiple references to one GIF image. The majority of multiple references to a GIF image relate to accent constructions.

Not all accent constructions have an entry within the stripin.toc file. The “regular” HTML equivalent variants listed in the table above are not included in the stripin.toc.

Each line of the stripin.toc file contains three fields, separated by one or more spaces. Figure 1 shows an example stripin.toc file.

```
fd 1 fd1.gif
fd 2 fd2.gif
f 1 f1.gif
fd 3 fd3.gif
fd 4 fd4.gif
f 2 f2.gif
f 3 f3.gif
a 1 a1.gif
a 2 a2.gif
a 3 a1.gif
f 4 f2.gif
fd 5 fd5.gif
a 4 a2.gif
fd 6 fd6.gif
fd 7 fd7.gif
fd 8 fd8.gif
```

The first field gives the general identifier of the SGML construct (fd, f, a, ...), the second field gives the ordinal of that type of construct in the SGML file (e.g. 3 for the third **<FD>** counting from the beginning of the file), and the last field gives the name of the image file for this construct (e.g. fd3.gif).

Appendix 3 Output specifications for artwork

The artwork of an article is produced internally in electronic journal production in one of three formats, TIFF Fax Group IV, JPEG (gray scale, RGB colour or CMYK colour), or Encapsulated PostScript. Before artwork is included in ScienceDirect OnSite datasets, the artwork files are converted into GIF 89a or JPEG format (dependent of the nature of the figure).

Besides format conversion, the artwork is also downsampled in two sizes, a poststamp image and a large image in GIF format, with extensions **.sml** and **.lrg**. The conversion rules in these cases are as follows:

Conversion rule for large image. The following formula is used to determine the LargeImageWidth:

$$\text{LargeImageWidth} = \text{ImageXDimension} * (\text{ThumbDefaultResolution} / \text{ImageXResolution})$$

where ThumbDefaultResolution = 100.

Conversion rule for poststamp image. The formula for the poststamp image is based on whether the image is in landscape or in portrait mode. The formula for the portrait mode is:

$$\text{PostStampImageWidth} = \text{ScreenAspect} * \text{WantedStampWidth} * (\text{ImageXDimension} / \text{ImageYDimension})$$

where ScreenAspect = 0.75 (because screens have a 4:3 ratio) and WantedStampWidth = 125.

The formula for the landscape mode is:

$$\text{PostStampImageWidth} = \text{WantedStampWidth}$$

where WantedStampWidth = 125.

Example. The following is an example of the usage of the scaling formulas as used within the artwork conversion. We have the following variables:

- Image in portrait,
- 715 ImageXDimension,
- 957 ImageYDimension,
- 300 dpi ImageXResolution and ImageYResolution

The following are the actual scaling formulas applied:

- $\text{LargeImageWidth} = 715 * (100 / 300) = 238$
- $\text{PostStampImageWidth} = 0.75 * 125 * (715 / 957) = 70$

So the conversion will scale the image to a 238 pixel wide Large image and a 70 pixel wide Poststamp image.