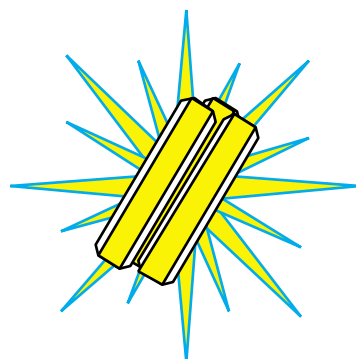




Recent Developments in **Software Tools for CIF and mmCIF** by

Herbert J. Bernstein and Frances C. Bernstein
Bernstein + Sons, 5 Brewster Lane, Bellport, NY 11713-2803, USA

**Work supported in part by IUCr (for HJB) and by US NSF, PHS, NIH, NCRR,
NIGMS, NLM and DOE (for FCB prior to 1998)**

What is CIF?

Why Tools?

Where to Get CIF Tools

CIF Software Tools

What's New in Our Tools

What is CIF?

Simple **Tag**-Value Language (STAR)

S. R. Hall, "The STAR File: A New Format for Electronic Data Transfer and Archiving", *J. Chem. Inform. Comp. Sci.*, Vol. 31, 1991, pp. 326-333

Plus Extensible Dictionaries (core CIF, mmCIF, pdCIF)

<code>_cell_length_a</code>	13.128(6)
<code>_cell_length_b</code>	22.438(15)
<code>_cell_length_c</code>	23.068(9)
<code>_cell_angle_alpha</code>	89.57(7)
<code>_cell_angle_beta</code>	91.08(6)
<code>_cell_angle_gamma</code>	119.53(7)

<code>loop_</code>	
<code>_struct_sheet.id</code>	
<code>_struct_sheet.number_strands</code>	
A	3
B	11

S. R. Hall, F. H. Allen and I. D. Brown, "The Crystallographic Information File (CIF): A New Standard Archive File for Crystallography", *Acta Cryst.*, Vol. A47, 1991, pp. 655-685.

Why Tools?

Check tags against dictionaries

Check syntax

tag for each value

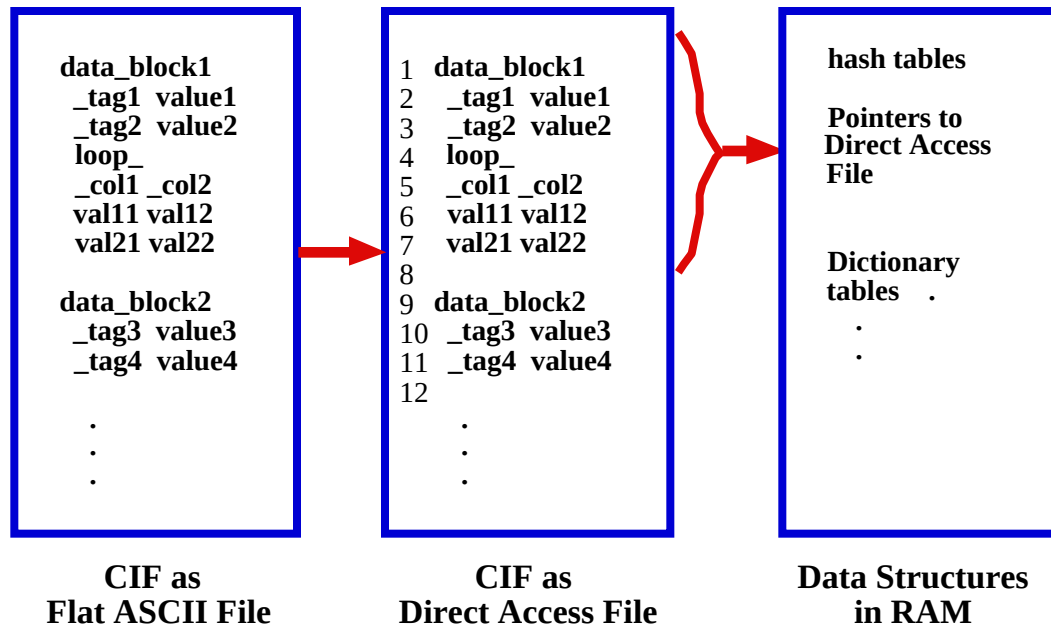
check types (numbers, character strings)

Build CIF-aware applications

Handle the CIF characteristics

free fields

arbitrary ordering



Where to Get CIF Tools

International Union of Crystallography (IUCr)

access to software, dictionaries, policy statements and documentation relating to CIF and mmCIF :

- **IUCr, Chester, England** (www.iucr.ac.uk/iucr-top/cif/)
- **University of Geneva, Geneva, Switzerland**
(<http://www.ch.iucr.org/iucr-top/cif/>)
- **Université P. et M. Curie, Paris, France** (<http://www.se.iucr.org/iucr-top/cif/>)
- **University of Stockholm, Stockholm, Sweden**
(<http://www.fr.iucr.org/iucr-top/cif/>)
- **SDSC, San Diego, California, USA** (<http://www.us.iucr.org/iucr-top/cif/>)
- **University of the Witwatersrand, South Africa**
(<http://www.za.iucr.org/iucr-top/cif/>)

Information and Software for STAR and CIF :

- **University of Western Australia, Nedlands, Australia**
(<http://www.crystal.uwa.edu.au>)

Nucleic Acid Database Project

software and documentation, mmCIF page, dictionary, mmCIF software tools:

- **Rutgers University, New Jersey, USA**
(<http://ndbserver.rutgers.edu/NDB/mmcif>)
- **EBI, Hinxton, England**, (<http://www.ebi.ac.uk/NDB/mmcif/>)
- **NIBH, Japan** (<http://www.nibh.go.jp/NDB/mmcif/>)
- **SDSC, San Diego, California, USA** (<http://ndb.sdsc.edu/NDB/mmcif/>)

CIF Software Tools

Available via IUCr Web Page[†]

(These tools are by many different groups and authors)

CIF Validation Tools

- vcif, a CIF syntax checker
- QUASAR, a CIF syntax checking and file manipulation tool
- Hiccup Validates CIF syntax and dictionary conformance. Multi-platform, but requires Python and Tcl/Tk
- **CYCLOPS2, checks apparent data names in text files against their occurrence in one or more CIF dictionaries (requires CIFtbx2)**

Software Libraries for Manipulating and Validating CIFs

- Converting Fortran Applications for Use with CIF
- **CIFtbx2 Version 2.6.2 (including CYCLOPS2), a Fortran toolbox and dictionary checker (DDL1 and DDL2 compliant) Replaces:**
 - CYCLOPS, a CIF dataname checker (Core CIF dictionary only)
 - CIFtbx, a toolbox of Fortran routines for handling CIF data
- The mmCIF/DDL2 libraries produced by the Nucleic Acid Database Project at Rutgers University, New Jersey. Comprises:
 - CIFLIB - a C Language application program interface
 - CIFPARSE - a library of access tools for mmCIF
 - CIFOBJ - a class library of mmCIF dictionary access tools

[†] This foil was derived from a portion of the IUCr CIF web page

CIF Software Tools (Cont.)

Available via IUCr Web Page[†]

(These tools are by many different groups and authors)

Visualization Tools

- Ciffilter, a utility for viewing online CIFs. Requires Xtal_GX and Tcl/Tk

Conversion Utilities

- cif2sx, a CIF to SHELXL93 converter
- DIFRAC, package to translate diffractometer output from many systems to CIF.

- | |
|--|
| <ul style="list-style-type: none">• pdb2cif Version 2.3.6, a filter to translate PDB format files to mmCIF• cif2pdb Version 1.0.1, a filter for translating mmCIF to PDB format |
|--|

Miscellaneous

- checkcif and printcif, e-mail interfaces to the IUCr CIF checking and typesetting facilities for prospective authors to Acta Crystallographica
- CIFtex, the IUCr typesetting tool [this requires QUASAR and TeX]

[†] This foil was derived from a portion of the IUCr CIF web page

What's New in Our Tools

Context

mmCIF adopted and accepted
EBI Data Harvest Proposal
CBF/imgCIF effort
NSF mandates mmCIF in MSD competition announcement

Past Efforts

CIFtbx2, Cyclops2, cif2cif, pdb2cif, cif2pdb

Recent Changes

CIFtbx2	-- speed improvements, better esd and category hooks
Cyclops2	-- stable
cif2cif	-- provides QUASAR capabilities for DDL2 data sets
pdb2cif	-- stable
cif2pdb	-- will now process fractional coordinates, including small molecule core CIFs (e.g. for RasMol) -- more complete embedded CIF (REMARK 960)

What's New in Our Tools (cont.)

Work in Progress Now

imgCIF -- hooks to MIME-encode binary files in a CIF
CIFtbx -- compression to save space/time

Work in Progress, Long Range

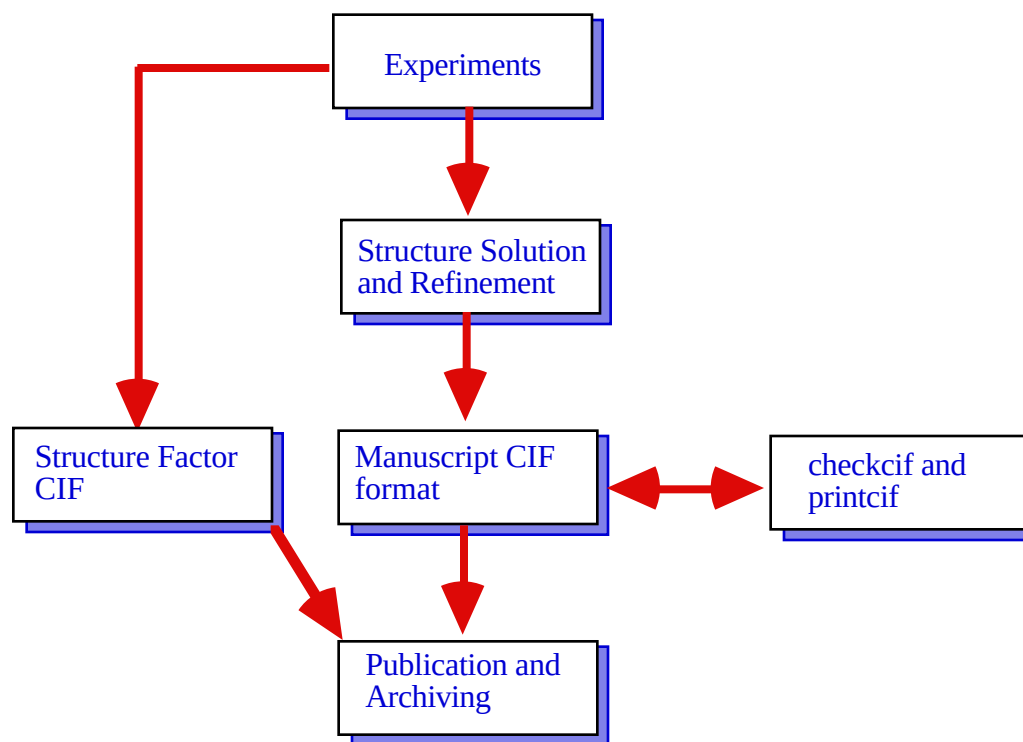
cifdiff -- WYSIWYG CIF differences (see ZINC)
cifrcs -- CIF revision control system
cifed -- CIF-aware editor (see Westbrook's CIF Input Tool)
cif2pdb/pdb2cif -- close the loop

...

Objective

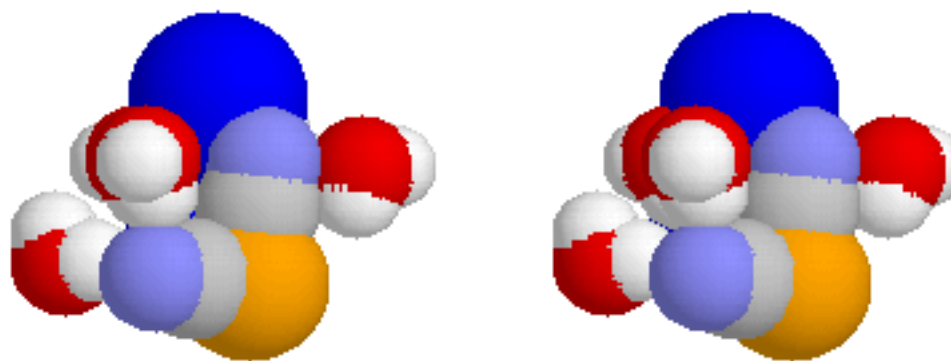
Smooth integration of DDL1 and DDL2 -based software
Publication tools for macromolecules and other domains

Simplified Block Diagram of Small Molecule Publication (Author's View)



Example of cif2pdb conversion of a Small Molecule

Tetrasodium Hexacyanoferrate(II) Decahydrate by J. J. Longridge, J. M. Rawson and J. E. Davies in Acta Crystallographica Section C, Volume 54, Part 7 (15 July 1998), converted by cif2pdb and rendered by RasMol.



Proposal to Embed Binary Information in a CIF

Create binary data type

Variant on semi-colon delimited text field

Starts with MIME boundary marker and header

Binary carried, as in email messages, as ASCII text

Work on CBF format (Hammersley, Westbrook, Ellis, etc.) provides parallel true binary format

```
TEXT bdry -->> ;
MIME bdry -->> --CIF-BINARY-FORMAT-SECTION--
MIME hdr  -->> Content-Type: application/octet-stream
              Content-Transfer-Encoding: BASE64
              X-Binary-Size: 3825041

Binary Data>> 0A12FF39872398479A2983FF487A10BAF98234147A10BAF98234147A103...
              9872398479A2983487A1098234140A1239872398479A2983487A1098002...

MIME bdry -->> ...
TEXT bdry -->> --CIF-BINARY-FORMAT-SECTION----
              ;
```

Proposed CIF Revision Control System

cifed, cifdiff, cifgrep, cifrcs

Companion editor to John Westbrook's CIF Input Tool.
Integrated with a revision-control system based on RCS [Tichy 85]
and replacements for Zinc-based cifdiff and cifgrep [Stampf 94]
Useful companion to EBI Data Harvest for end users
Useful for Editors and Archivists
Useful in Managing Dictionaries

[Stampf 94] D. R. Stampf, "ZINC -- Galvanizing CIF to work with UNIX," Protein Data Bank, Upton, NY 1994.

[Tichy 85] W. F. Tichy, "RCS -- A System for Version Control", Software Practice & Experience Vol. 15, 637-654, 1985.

References

- [BBB98] Bernstein, H.J., Bernstein, F.C., Bourne, P.E. "pdb2cif: Translating PDB Entries into mmCIF Format," , J. Appl. Cryst., 31, pp. 282-295, 1998, software available from <http://www.iucr.org/iucr-top/CIF> and <http://ndbserver.rutgers.edu>.
- [BB96] Bernstein, H. J. and Bernstein, F. C., "cif2pdb -- program to filter a CIF data set and produce a PDB entry," in preparation.
- [BH97] Bernstein, H.J, Hall, S.R., "CIFtbx2 Primer Plus," in preparation, 1997. See [CIFtbx_Primer.pdf](#).
- [HB96] Hall, S.R., Bernstein, H.J., "CIFtbx2: Extended Tool Box for Manipulating CIFs," J. Appl. Cryst., 29, pp 598-603, 1996.
- [HS93] Hall, S. R. & Sievers, R., "CIF Appl. I. QUASAR: for extracting CIF data", J. Appl. Cryst., 26, 469-473, 1993.
- [[H91] S. R. Hall, "The STAR File: A New Format for Electronic Data Transfer and Archiving", J Chem Inform Comp Sci., Vol. 31, 1991, pp. 326-333.
- [Hall, Allen, Brown 91] S. R. Hall, F. H. Allen and I. D. Brown, "The Crystallographic Information File (CIF): A New Standard Archive File for Crystallography", Acta Cryst., Vol. A47, 1991, pp. 655-685.
- [FBB96] Fitzgerald, P. M. D., Berman, H. M., Bourne, P. E., McMahon, B., Watenpaugh, K., Westbrook, J. "The MMCIF Dictionary: Community Review and Final Approval," 17th IUCR Congress and General Assembly, Seattle, Washington, USA, 8-17 August 1996, Abstract E1226. Version 0.8.02 available from <http://ndbserver.rutgers.edu>.
- [HS94] Hall, S.R., Spadaccini, N., "The STAR File: Detailed Specifications," J. Chem. Info. Comp. Sci., 34, 505-508, 1994. See http://www.crystal.uwa.edu.au/cc_star.html