Prologue: plan for today

• General overview of CCP4 (~40-50 minutes)

Then (for those who are interested):

• More technical overview/demo of using the software (open-ended):
  • tutorial presentation
  • demo of the software
An Overview of the CCP4 Project

More than just a software suite?

Peter Briggs
CCP4, CCLRC Daresbury Laboratory
About myself

• Based at Daresbury Laboratory

• CCP4 programmer since 1998
  • Maintain & develop core suite
  • Coordinate developments with CCP4 graphical user interface
  • Coordinate CCP4 involvement in BIOXHIT project
  • Manage three staff members
  • Organise workshops & exhibition booths
  • Various other odd-jobs

• In the past I’ve also
  • Edited CCP4 newsletter
  • Liaised with industrial customers
  • Managed several software releases (4.1, 4.2 and 5.0)
Outline of the presentation

• Description of the CCP4 Project
  • Structure/organisation, role of Daresbury group

• Core Activities
  • Website, resources, conferences, associated projects

• Overview of the CCP4 software suite
  • Philosophy, components, activities

• Software Releases
  • Next release 6.0

• Issues and Challenges
What is a CCP?

CCP’s = Collaborative Computational Projects

- First established in late 1970s/early 1980s by UK research councils
- Aim to provide infrastructure for “large-scale” computational projects in different areas of science
- 15 established – currently 12 are active

- Typical remit:
  - Implement flagship code development projects
  - Maintain and distribute code libraries
  - Organise training in use of codes
  - Hold meetings and workshops
  - Invite overseas researchers for lecture tours and collaborative visits
  - Issue regular newsletters
Collaborative Computational Project 4

- Established in 1979 by SERC

- Software for protein X-ray crystallography

- Aims:
  - “Support collaboration between researchers working on PX software in the UK”
  - “Assemble comprehensive collection of software to satisfy the computational requirements of the relevant UK groups”

- Gave rise to the CCP4 Program Suite

- Focus on collating, standardising, maintaining and distributing existing code
  - Some code developed “in-house”
  - Majority contributed by benevolent programmers
Current Status

- Core funding from BBSRC (5 year peer-reviewed grant)
  - Funds a small number of positions

- Software provided free of charge to academic/non-profit users

- Charge is made for commercial use
  - Provides funds for majority of staff/projects
Some observations ...

• Organisation is very fragmented
  • Diverse geographical locations, multiple projects involving contributions from different groups & different aims
  • Very difficult to manage
  • but it seems to work … most of the time

• Daresbury CCP4 group acts as a central focus for the project
  • make new programs available to general community
  • facilitate the development of new software
  • facilitate the general running of the project

• Much of software development takes place outside “core” group
  • Project couldn't exist without the (often unfunded) contributions from many different developers

• No single contribution can be said to be more important than any other
The Daresbury CCP4 Group

- Based at CCLRC Daresbury Laboratory in UK
  - currently home of the UK synchrotron source
  - also HPC centre
  - neutral location

- Group consists of:
  - four core programmers (Martyn Winn, Charles Ballard, Francois Remacle, Peter Briggs)
  - two programmers working part-time on core (Norman Stein, Daniel Rolfe)
  - executive assistant (Maeri Howard)

- Core activities split between these staff
Core CCP4 Activities

- **Maintenance and development of the core suite**
  - development of software libraries and infrastructure
  - graphical interface CCP4i
  - integrate in new software
  - bug fixing & technical support

- **Software releases**

- **Project administration**
  - commercial liaison
  - newsletters
  - maintain website and electronic mailing lists e.g. ccp4bb

- **Coordinate annual CCP4 Study Weekend**
  - also workshops and conference presence

- **Contributions to external projects**

19th July 2005

RUPX Meeting
CCP4 Website and Resources

http://www.ccp4.ac.uk

Links to useful information:

- News
- Downloads
- Documentation
- Courses
- Problems Pages
- Developer’s Resources
- Bulletin Boards
- Newsletters
- ...

19th July 2005

RUPX Meeting
E-mail distribution lists

CCP4 Bulletin Board ccp4bb

- Members from all over the world
- Fast expert help from colleagues
- Lively forum for discussing all aspects of PX
- “Sorry about the non-CCP4 posting …”
- Moderated list
- Online form for postings:
  http://www.ccp4.ac.uk/reportaprob.php

Other lists:

- Developers List ccp4-dev
- Also project-specific lists
CCP4 Newsletter on Protein Crystallography

- News items
- Reports on new software and algorithms
  - CCP4 and non-CCP4
- Web-based publication since 2001
  - Back issues available on-line
  - Next issue due this summer

- [http://www.ccp4.ac.uk/newsletters.php](http://www.ccp4.ac.uk/newsletters.php)
  - or go via the CCP4 website
CCP4 Study Weekend

- CCP4’s flagship event
- Held first weekend of each year
- Focus on particular aspect of PX
  - Next year: “Complex Crystallography and Crystallography that is Complex”
- Mixture of teaching material with new methods
- Proceedings published in Acta Cryst D
  - Older proceedings (pre 1998) available from Daresbury Library
Other Outreach Activities

• Exhibition booths at major conferences
  • ACA, IUCr, ECM, AsCA
  • Software demos and Q&A

• One-day satellite workshops
  • Focus on using the software
  • Last year: joint with CCP4-PDB (again this year)

• Longer one-off workshops
  • less frequent (< once a year)
  • e.g. one-week workshop in Bangalore, March 2005

• Travel Scholarships
  • 2-3 students from India & China
  • Visit UK labs for 2 weeks and attend Study Weekend
e-HTPX, DNA and Automation

**e-HTPX – E-science Resource for High Throughput Crystallography**
- BBSCR e-science initiative
- aim to provide a single interface to PX structure determination
- allow user to access experimental & computational resources from desktop
- pilot project started 2002

**DNA – automating PX data collection experiments**
- provides software to collect & process diffraction data
- includes rapid crystal screening
- unfunded collaboration, currently on release 1.0.0

**CCP4 Automation Project**
- aim to automate structure determination starting from processed data using CCP4 programs
- two programmers at Daresbury

19th July 2005 RUPX Meeting
BIOXHIT and PIMS

BIOXHIT - Integrated Technology Platform for High Throughput Protein Crystallography

- European project involving 20+ partner institutions
- CCP4 contribution involves development of CCP4i database for project tracking
- One full-time staff member at Daresbury

PIMS – “Protein Information Management System”

- Provide LIMS system suitable for high throughput crystallisation experiments in academic environment
- Funding from several institutions
The CCP4 Software Suite

Current release is version 5.0.2

Key components:
- Core software libraries
- Programs
- Documentation
- Graphical user interface

Languages:
- Programs: predominantly Fortran 77 (historical), increasingly also C/C++
- Libraries: predominantly C/C++ with Fortran 77
- Also Tcl/Tk (user interface), Java and shell

Platforms:
- Unix/Linux, Windows and Mac OS-X
- source code and binaries available
Philosophy of the CCP4 software suite

• Modular:
  • Large number of programs, each covers a small range of functionality
  • Data passed between programs via data files in standard formats
  • Keywords control program function and provide additional data
  • User decides on the sequence of programs to use for a particular task, e.g.
    
    *E.g. data reduction starting in CCP4:*
    Mosflm → Scala → Truncate
    
    *Or alternatively starting outside CCP4:*
    HKL2000/Scalepack → Combat → Scala → Truncate

  • Can easily be used in other software packages (*ARP/wARP, autoSHARP*)

• Inclusive & redundant:
  • Includes a number of different programs to do the same job
  • Allows user to choose from different approaches
Core CCP4 Libraries

Provide infrastructure for standardisation of programs
  • Data file formats: MTZ, PDB/mmCIF, maps ...
  • Keyword parsing
  • Handling symmetry and unit cell data
  • Features such as data harvesting

Freely available for use/ modification/ redistribution:
  • Mixture of LGPL & LGPL-compatible licensing
  • Enables SOLVE, SHARP, ARP/wARP to use MTZ files

Advantages of libraries
  • Less code to maintain
  • Easier/quicker to write new programs/ upgrade existing ones

Last release (5.0) focused on upgrading to new C/C++ based libraries
  • Move towards more modern code base e.g. scripting interfaces
The CCP4 Suite

- ~180 programs covering all aspects of PX structure determination
- Relatively small subset are significant for most users, e.g.:
  - Data processing and reduction \((\text{mosflm, scala, truncate} \ldots)\)
  - Experimental phasing \((\text{mlphare, acorn, rantan, oasis, bp3} \ldots)\)
  - Molecular replacement \((\text{amore, molrep, beast, phaser})\)
  - Density modification \((\text{dm, solomon, pirate})\)
  - Refinement \((\text{refmac5, sketcher})\)
  - Graphics and building \((\text{ccp4mg, coot})\)
  - Validation and analysis \((\text{sfcheck, procheck} \ldots)\)
  - Utility and conversion programs \((\text{e.g. fft, cad, revise, pdbset, \ldots})\)
Navigating the suite

**Key questions:**
- Which program(s) do I need?
- How do I run it/them?

**Documentation** ([http://www.ccp4.ac.uk/docs.html](http://www.ccp4.ac.uk/docs.html)):
- Roadmaps
- Tutorials (html)
- Individual program documentation (html)
  - Function index
  - General background on twinning, reindexing, ...
- Postscript manual
  - Slightly dated but still useful
  - Content distinct from program documentation

**Runnable example scripts**
**Graphical user interface**
Graphical User Interface CCP4i

- Sits on top of programs
  - Hides details of individual programs
  - User not locked in
  - Allows mix-and-match approach

- Philosophy:
  - “Task-driven” rather than “program-driven”
  - Interfaces to tasks use “natural language” approach
    - e.g. ![Input MAD data as F+ F- and prepare data for SHELXD](image)

- Key features:
  - Easy-to-use interfaces to major programs and utilities
  - Integrated tools for file viewing and basic project management
  - Customisable
  - Integrated help system
  - Some non-CCP4 programs also interfaced (e.g. ARP/wARP, SHELX)
Helpdesk – Technical Support

• Software is provided with “no warranty” clause  
  • in practice we aim to provide support

• Queries sent to ccp4@ccp4.ac.uk  
  • Problems with installation/usage  
  • Bugs in software and documentation  
  • Suggestions for improvements

• Central address acts as a clearing house for emails  
  • may be forwarded to authors of actively developed software

• DL programmers operate a rota (6 people/1-2 weeks each)

• Known bugs posted on Problems Pages  
  • http://www.ccp4.ac.uk/problems.php
Making Releases of CCP4

- Increasingly daunting task
### Growth of the CCP4 suite since 1998

<table>
<thead>
<tr>
<th>March 1998</th>
<th>August 2004</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCP4 version</td>
<td>v3.5</td>
</tr>
<tr>
<td>No. of programs</td>
<td>~150</td>
</tr>
<tr>
<td>Size</td>
<td>43.5 Mb</td>
</tr>
<tr>
<td>No. of files</td>
<td>1392</td>
</tr>
<tr>
<td>Lines of code</td>
<td>~1.3 million</td>
</tr>
<tr>
<td>Languages</td>
<td>Fortran, C, shell</td>
</tr>
<tr>
<td>Platforms</td>
<td>Unix &amp; VMS</td>
</tr>
<tr>
<td>Binaries provided</td>
<td>No</td>
</tr>
<tr>
<td>Subscribers to CCP4bb</td>
<td>~1000</td>
</tr>
</tbody>
</table>
Making Releases of CCP4

- Increasingly daunting task

- Software tools can help
  - CVS for source version control
  - Bugzilla track known issues prior to/after release
  - Automated building on various platforms to test

- Code freeze several weeks before test releases
  - can be difficult to enforce

- Beta releases for testing
  - Next version: test version via [http://www.ccp4.ac.uk/dev/releases.html](http://www.ccp4.ac.uk/dev/releases.html)

- Patch releases to fix known bugs after release

- Long process (typically now 3-4 months)
Improvements to installation/update

• Originally source code only
  • Provision of binaries (2001)
  • Windows version (2002)

• Automated installation and updates
  • Binary downloader scripts (2004)
  • Prototype updater (2005) – fetch & apply source patches
  • InstallShield for Windows (2005)
Next release of CCP4: version 6.0

New packages:
- **PHASER**: maximum-likelihood molecular replacement
- **Pirate**: statistical phase improvement
- **CCP4MG**: CCP4 Molecular Graphics package
- **Coot**: graphical model building tools
- **Superpose**: secondary structure alignment
- **BP3**: heavy atom phasing refinement
- **CHOOCH**: anomalous scattering factors from raw fluorescence spectra

Updates to **REFMAC5, MOLREP, SFCHECK, SCALA, PDBEXTRACT** and others

**CCP4i:**
- **CRANK**: automated structure solution via SAD, SIR, SIRAS
- **SHELXC/D/E** interface
- Database search and sort utility

Plus many bug fixes and minor improvements
Some major new programs in CCP4 v6.0

**Phaser**
- maximum likelihood phasing
- currently has methods for molecular replacement
- uses ensembles of many models (with estimate of similarity to target)
- experimental phasing methods under development

**Pirate**
- statistical phase improvement (replacement for dm)
- uses phases from known reference structure with features similar to unknown structure
- doesn’t require knowledge of solvent content
CCP4 Molecular Graphics

**CCP4MG**

- Focused on presentation graphics
- Rendering surfaces, superposition, structure analysis, movies …

**Coot**

- Platform for semi-automated model building tools
- Real space refinement, rigid-body fitting, ligand search, solvation, rotamers, Ramachandran plots …
Status of CCP4 version 6.0

Test version 5.99.2 available:
  • see http://www.ccp4.ac.uk/dev/releases.html

Downloads divided into a number of packages:
  • Basic CCP4 (about the same as v5.0)
  • Phaser
  • cctbx (libraries)
  • CCP4mg
  • Coot
  • CHOOCH
  • plus dependencies (Tcl/Tk/BLT, Python …)

New download pages:
  • allow user to select required packages and dependencies
  • download a single file for installation
  • source code and/or binaries
Future Developments

Software
- Increasing automation
  - programs
  - automated pipelines
- More integration of “external” software packages
- Increasing use of scripting to develop applications
  - already seen in other projects e.g. PHENIX
- Significant update/revision of user interface
- Associated projects start to deliver software for distribution

Delivery of suite
- More modular package
- Installation/update procedure will change
  - easier for users (to compensate for increasing complexity)
Challenges and Issues

• Legitious environments (software patents and licensing issues)

• Changing user expectations (user base)

• Changing code base (size/diversity)

• Changing ways that software is used (webservices/GRID)

• Use of complementary experimental techniques (NMR, EM, XAFS, CD …)
Acknowledgements

CCP4 working group members
  • Jim Naismith

CCP4 staff and contributors past and present
  • Everyone who has ever contributed to the project
  • Far too many to mention here

Martin Noble and Johan Turkenburg

Helen Berman & people at the RCSB-PDB

Cathy Lawson

You all for listening
  • Thank you!