



SGC

www.thesgc.org

Scarab – data management in ICM

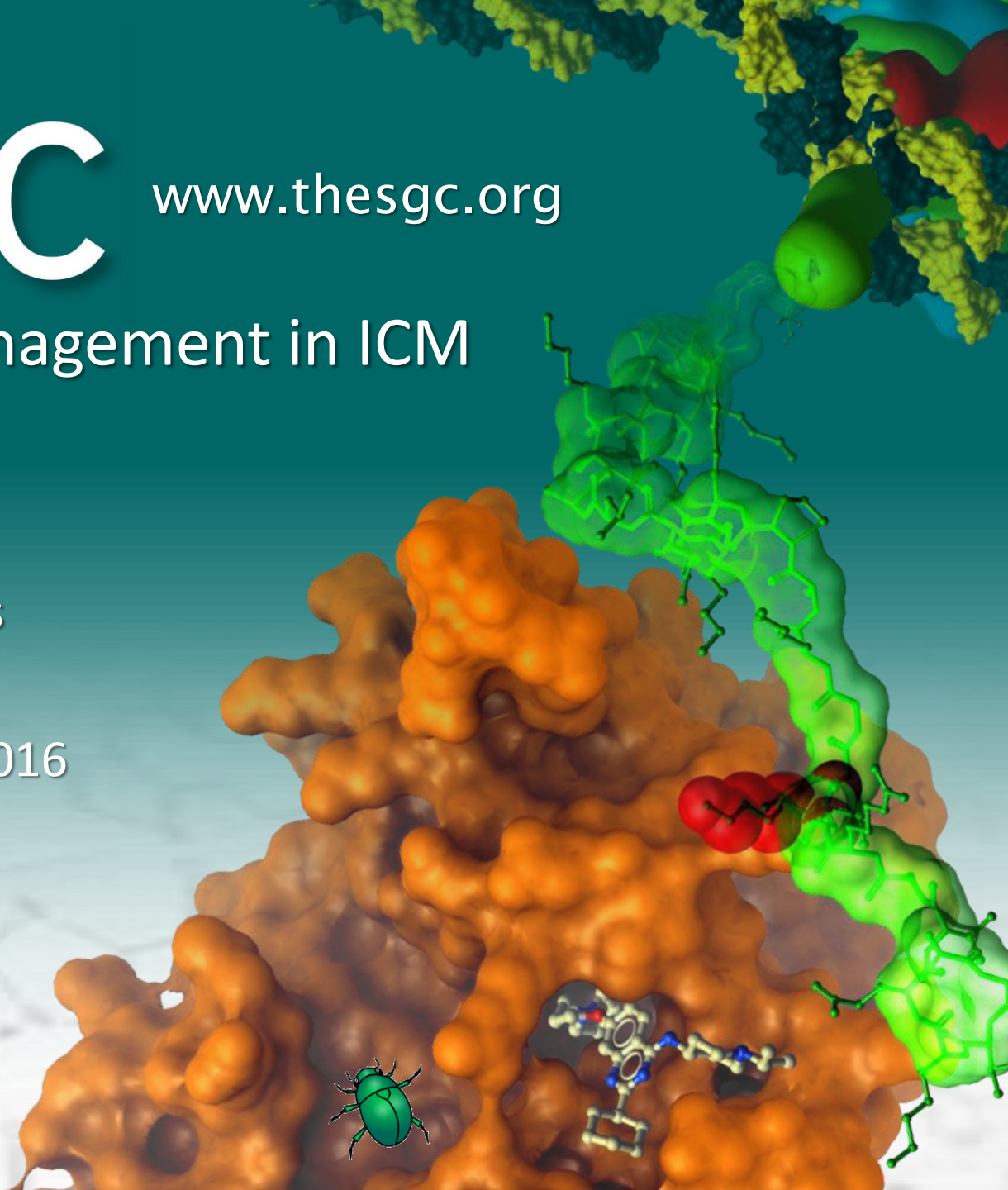
Brian Marsden

PI, Research Informatics

SGC & Kennedy Institute

University of Oxford

MolSoft UGM, March 2016



Our Core Outputs

High Throughput Structural Biology

Protein families of relevance to drug discovery

Human and human parasites

>1850 structures deposited since 2004

~15% Yearly Global Output

Chemical biology

Chemical probes to explore protein function and enable target validation

37 Epigenetic probes made available to date

Renewable Antibodies / Binders

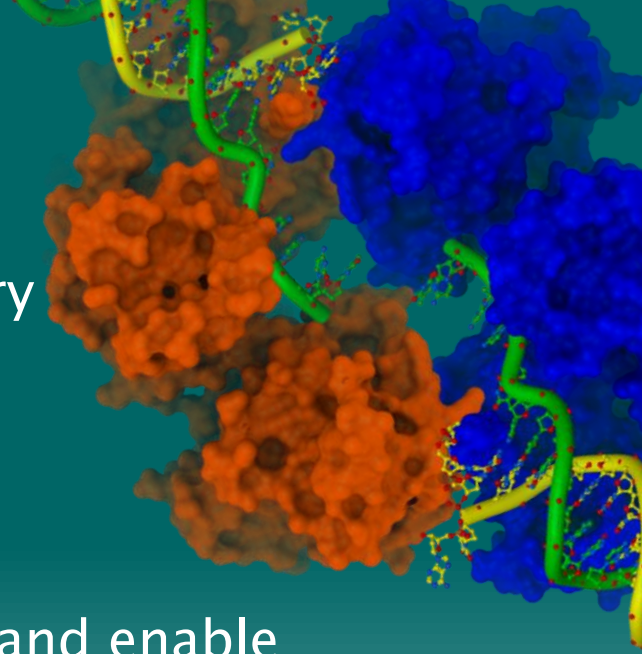
Purified proteins as antigens for recombinant antibodies using phage display technologies

Patient-Cell Derived Assays

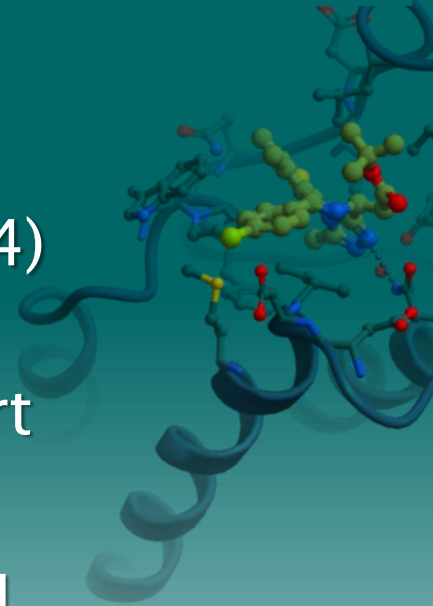
Tissue platforms

Develop assays and test systems based on primary cells directly from patients and healthy volunteers

CNS, inflammation and oncology



SGC at a glance



- Operations started in June 2004
- Government agencies, Wellcome Trust, charities & leading pharma companies
- Aggregate funding in excess of \$300M (2014)
- +260-strong team in Oxford, Toronto, Stockholm, Campinas, Chapel Hill & Frankfurt
- Open Access Policy:
 - Promptly placing results, reagents and know-how in the public domain
 - SGC scientists **never** file patents



GenomeCanada



BILL & MELINDA
GATES foundation

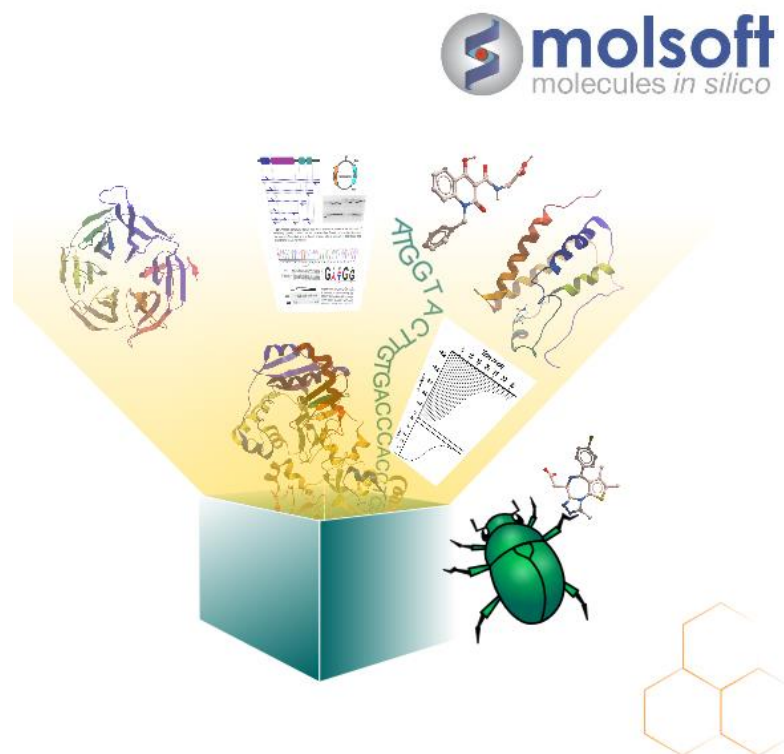


abbvie



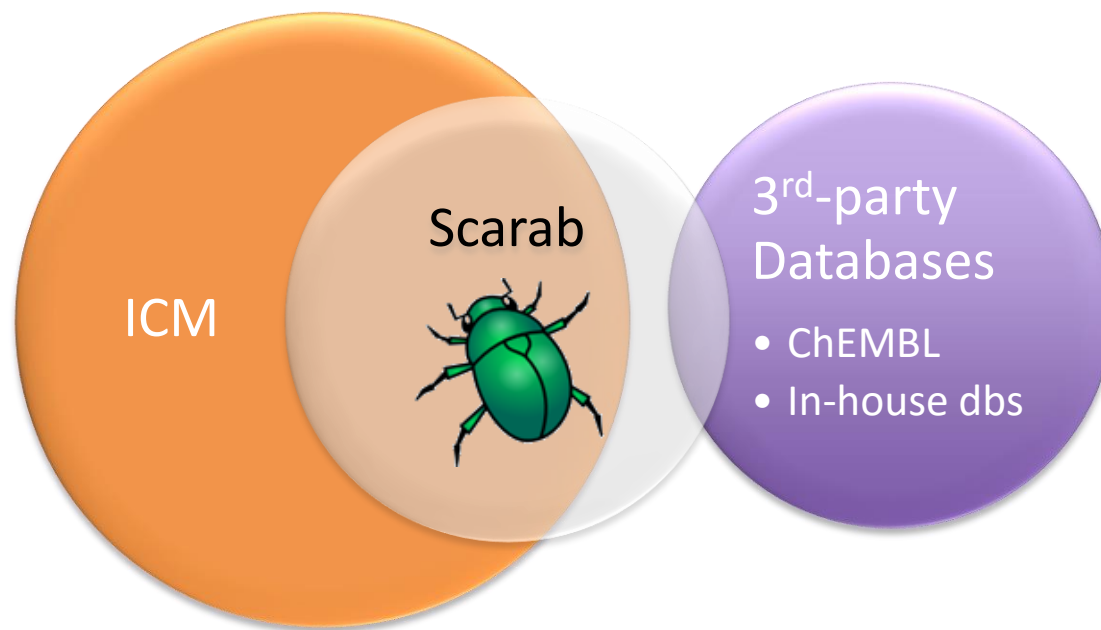
LOTS OF (HETEROGENEOUS) DATA

- SGC employs moderately high-throughput pipeline processes
- Capturing data & pipeline progress is key
- Making the data easily available and interpretable is difficult
- We use a LIMS
- Collaboration with MolSoft for over 12 years.





- Part of the ICM family
- Built-in Laboratory Notebook
- Data mining, data extraction and insertion
- Multi-user management in scientific labs and groups
- A unified way of managing all scientific data in one place
- Eliminating data clutter
- Integration with ICM product features





Pages: A section-based NoteBook

PAGE16-00002

PAGE16-00002

Exit Edit mode

Export to PDF Submit Authors

Export to Word Preferences Refresh

Print View Audit

Text Image Excel Attach Reference Resultset ICB Session

Insert

Experiment No.1 [\[edit title...\]](#)

PAGE16-00002

Author: **arman**

Created: 2016-Mar-07

Last updated: 2016-Mar-10 15:45:57

Experiment started: [\[edit date...\]](#)

Projects: Writer [\[edit projects...\]](#)

Tags: [\[edit tags...\]](#)

Related Pages: [\[edit related pages...\]](#)

Referenced by:

- ✓ Many types of data supported
- ✓ Including whole ICB sessions
- ✓ Exporting to popular formats

UNSTRUCTURED DATA – PAGES



SGC

ICM Pages
PAGE14-05114 PAGE14-05599 PAGE14-05570 PAGE14-05574

Export to PDF Submit Authors
Export to Word Preferences Refresh
Print View Audit

File Edit View

Text Image Excel Attach

PX-XE-Plate-2: Rosetta transformation to test express

PAGE14-05574

Author: claired
Created: 2014-Dec-18
Last updated: 2015-Jan-15 08:48:15
Experiment started:
Projects: Cloning
Tags:
Related Pages:
Referenced by: PAGE15-00172

Transforming Rosetta

5µl of the miniprep DNA were used to transform 40µl Rosetta cells.

Test expression protocol

Test expression and purification of soluble proteins (E. coli) - DDM method.docx

Plate details

Well	Construct ID	Clone ID	Expression ID	Vector	MW	Epitope	Description
A01	ATAD2A-c009	ATAD2A-c060	ATAD2A-e090	pNIC28-BsaI	17896.2	-	WT
B01	ATAD2A-c091	ATAD2A-c061	ATAD2A-e091	pNIC28-BsaI	17921.2	1	D1096E
C01	ATAD2A-c032	ATAD2A-c062	ATAD2A-e092	pNIC28-BsaI	17962.3	2	V1022R
D01	ATAD2A-c033	ATAD2A-c063	ATAD2A-e093	pNIC28-BsaI	17954.2	3	V1022R
E01	ATAD2A-c034	ATAD2A-c064	ATAD2A-e094	pNIC28-BsaI	17912.2	4	Q1099D
F01	ATAD2A-c035	ATAD2A-c065	ATAD2A-e095	pNIC28-BsaI	17992.3	5	G1070E
G01	ATAD2A-c036	ATAD2A-c066	ATAD2A-e096	pNIC28-BsaI	17952.3	6	S979Q
H01	ATAD2A-c037	ATAD2A-c067	ATAD2A-e097	pNIC28-BsaI	17927.2	7	I891E
A02	CERC2A-c008	CERC2A-c028	CERC2A-e064	pNIC28-BsaI	16175.3	-	WT
B02	CERC2A-c031	CERC2A-c029	CERC2A-e065	pNIC28-BsaI	16219.3	1	K487R
C02	CERC2A-c032	CERC2A-c030	CERC2A-e066	pNIC28-BsaI	16200.3	2	M536R
D02	CERC2A-c033	CERC2A-c031	CERC2A-e067	pNIC28-BsaI	16176.3	3	Q473E
E02	CERC2A-c034	CERC2A-c032	CERC2A-e068	pNIC28-BsaI	16217.3	4	S517E
F02	CERC2A-c035	CERC2A-c033	CERC2A-e069	pNIC28-BsaI	16275.4	5	K486R
G02	CERC2A-c036	CERC2A-c034	CERC2A-e070	pNIC28-BsaI	16203.3	6	S517D
H02	CERC2A-c037	CERC2A-c035	CERC2A-e071	pNIC28-BsaI	16149.3	7	N470S
A03	FAM83AA-c008	FAM83AA-c014	FAM83AA-e022	pNIC28-BsaI	23603.3	-	WT
B03	FAM83AA-c013	FAM83AA-c015	FAM83AA-e023	pNIC28-BsaI	23656.3	1	C229R
C03	FAM83AA-c014	FAM83AA-c016	FAM83AA-e024	pNIC28-BsaI	23626.3	2	N183H
D03	FAM83AA-c015	FAM83AA-c017	FAM83AA-e025	pNIC28-BsaI	23686.4	3	D206E
E03	FAM83AA-c016	FAM83AA-c018	FAM83AA-e026	pNIC28-BsaI	23639.3	4	T156H
F03	FAM83AA-c017	FAM83AA-c019	FAM83AA-e027	pNIC28-BsaI	23622.3	5	V159G
G03	FAM83AA-c018	FAM83AA-c020	FAM83AA-e028	pNIC28-BsaI	23659.3	6	Q202R
H03	FAM83AA-c019	FAM83AA-c021	FAM83AA-e029	pNIC28-BsaI	23590.2	7	Q277D
A04	JMJD18A-c066	JMJD18A-c225	JMJD18A-e348	pNIC28-BsaI	42928.6	-	WT
B04	JMJD18A-c076	JMJD18A-c226	JMJD18A-e361	pNIC28-BsaI	42967.5	1	V1710E
C04	JMJD18A-c077	JMJD18A-c227	JMJD18A-e362	pNIC28-BsaI	42984.6	2	Q1656R
D04	JMJD18A-c078	JMJD18A-c228	JMJD18A-e363	pNIC28-BsaI	43028.7	3	Q1601E
E04	JMJD18A-c079	JMJD18A-c229	JMJD18A-e364	pNIC28-BsaI	42999.7	4	E1530R
F04	JMJD18A-c080	JMJD18A-c230	JMJD18A-e365	pNIC28-BsaI	42915.5	5	E1624R
G04	JMJD18A-c081	JMJD18A-c231	JMJD18A-e366	pNIC28-BsaI	43028.7	6	I1577V
H04	JMJD18A-c082	JMJD18A-c232	JMJD18A-e367	pNIC28-BsaI	42983.7	7	Q1436R
A05	JMJD18A-c087	JMJD18A-c150	JMJD18A-e504	pNIC28-BsaI	42728.4	-	WT

ICM Pages
PAGE14-05114 PAGE14-05599 PAGE14-05570 PAGE14-05574

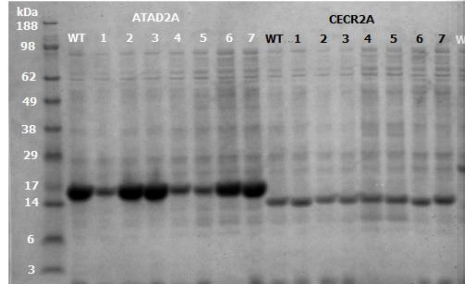
Export to PDF Submit Authors
Export to Word Preferences Refresh
Print View Audit

File Edit View

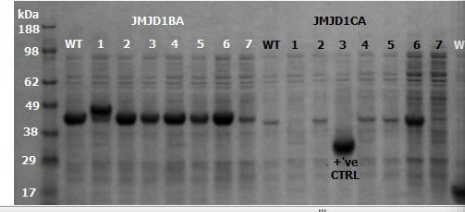
Text Image Excel Attach

Well	Construct ID	Clone ID	Expression ID	Vector	MW	Epitope	Description
B11	SPIN1A-c021	SPIN1A-c025	SPIN1A-e034	pNIC-CTHF	27462.8	1	T133R
C11	SPIN1A-c022	SPIN1A-c026	SPIN1A-e035	pNIC-CTHF	27489.8	2	E51D-N53R
D11	SPIN1A-c023	SPIN1A-c027	SPIN1A-e036	pNIC-CTHF	27414.8	3	S199N
E11	SPIN1A-c024	SPIN1A-c028	SPIN1A-e037	pNIC-CTHF	27376.6	4	R113H-S117T
F11	SPIN1A-c025	SPIN1A-c029	SPIN1A-e038	pNIC-CTHF	27385.7	5	K132R
G11	SPIN1A-c026	SPIN1A-c030	SPIN1A-e039	pNIC-CTHF	27409.7	6	G129S
H11	SPIN1A-c027	SPIN1A-c031	SPIN1A-e040	pNIC-CTHF	27432.8	7	G129S-H131E
A12	WSPRPAA-c002	WSPRPAA-c002	WSPRPAA-e002	pNIC-CTHF	33244.3	-	WT
B12	WSPRPAA-c003	WSPRPAA-c003	WSPRPAA-e003	pNIC-CTHF	32102.9	1	G306D-G308D
C12	WSPRPAA-c004	WSPRPAA-c004	WSPRPAA-e004	pNIC-CTHF	32038.9	2	L197H-Q200R
D12	WSPRPAA-c005	WSPRPAA-c005	WSPRPAA-e005	pNIC-CTHF	32084.3	3	T277E-S280R
E12	WSPRPAA-c006	WSPRPAA-c006	WSPRPAA-e006	pNIC-CTHF	32043.4	4	Q321R-Q325R
F12	WSPRPAA-c007	WSPRPAA-c007	WSPRPAA-e007	pNIC-CTHF	31984.8	5	E317D-M318D
G12	WSPRPAA-c008	WSPRPAA-c008	WSPRPAA-e008	pNIC-CTHF	32029.9	6	Q325R-N329E
H12	WSPRPAA-c009	WSPRPAA-c009	WSPRPAA-e009	pNIC-CTHF	32087.7	7	Q231R-G234E

Gel 1 - ATAD2A, CECR2A and FAM83AA



Gel 2 - JMJD18A, JMJD18A and JMJD2AA

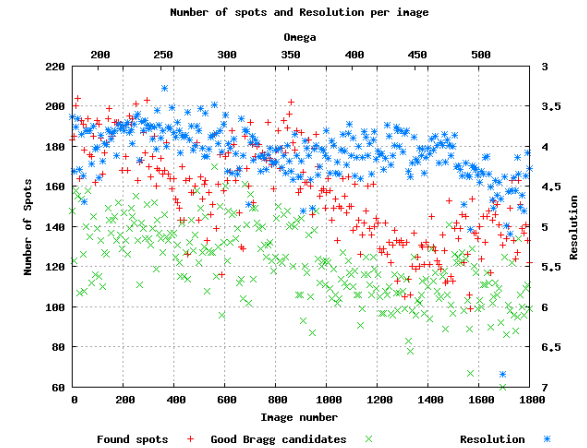


ICM Pages
PAGE15-03885 PAGE15-03403 PAGE15-03274 PAGE15-03275 PAGE15-03005 PAGE15-01219 PAGE15-00401 PAGE14-05114

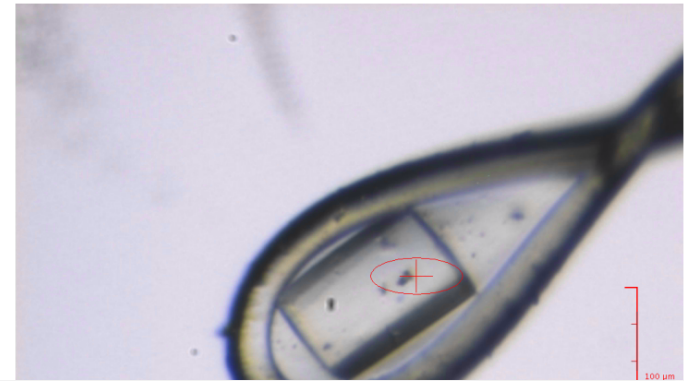
Export to PDF Submit Authors
Export to Word Preferences Refresh
Print View Audit

File Edit View

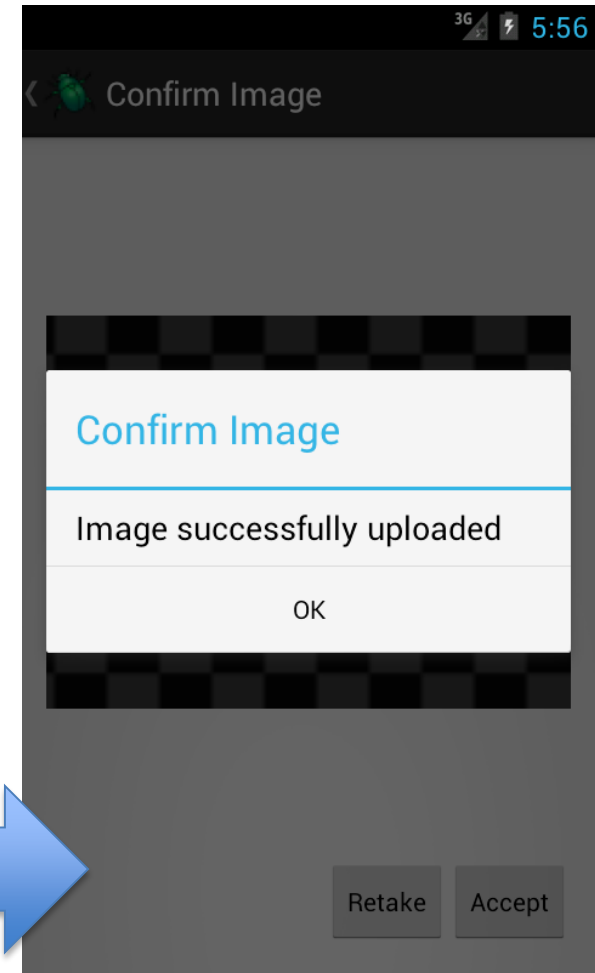
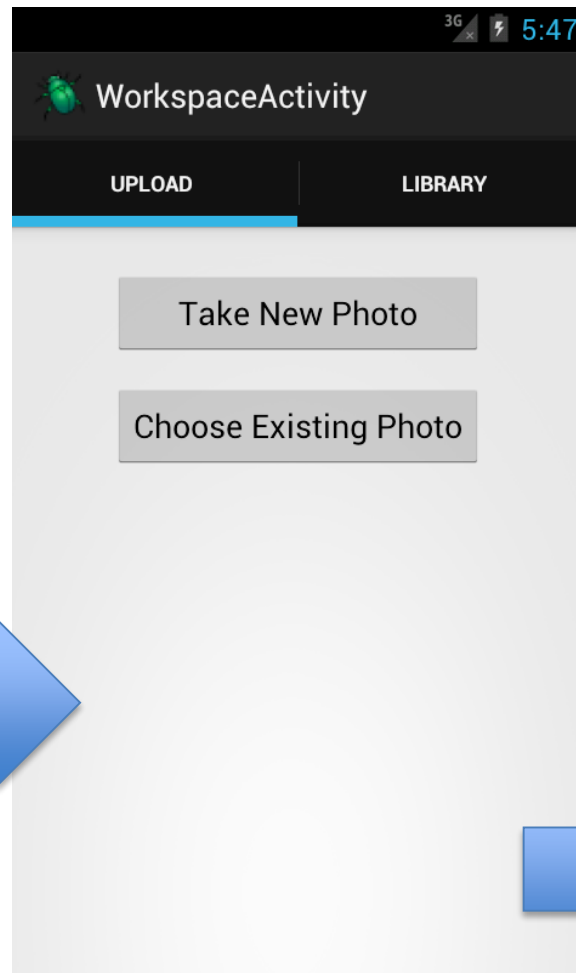
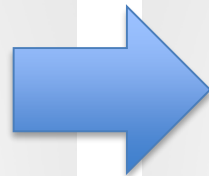
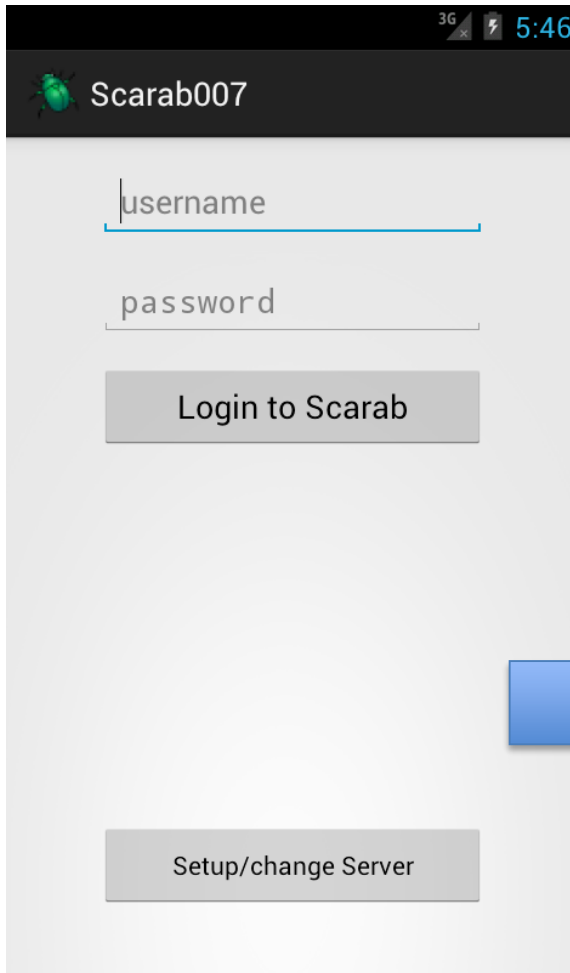
Text Image Excel Attach Reference Resultset ICB Session



SLC2A3A-x0137 RHS used for run2 spin-on-spot



UNSTRUCTURED DATA - MOBILE





[\[edit title...\]](#)

PAGE16-00005

Author: **administrator**

Created: 2016-Feb-09

Last updated: 2016-Mar-02 17:19:09

Experiment started: [\[edit date...\]](#)

Projects: Reader;Writer [\[edit projects...\]](#)

Tags: [\[edit tags...\]](#)

Related Pages: [\[edit related pages...\]](#)

Referenced by:

Empty Item - double click to edit this title



- You can import an image:
- (1) Via the clipboard (right-click and select)
 - (2) By dragging and dropping an image file
 - (3) By clicking here to import an image from the ICM works
 - (4) Dragging an image from the ICM works

- Open New Image...
- Browse Mobile App Images...**
- Remove Image
- Save With Annotations As...
- Save As (no annotations)...
- Copy With Annotations to Clipboard
- Copy to Clipboard (no annotations)
- Open External Viewer
- Paste
- Invert Image
- Edit Image...



[\[edit title...\]](#)

PAGE16-00005

Author: **administrator**

Created: 2016-Feb-09

Last updated: 2016-Mar-02 17:19:09

Experiment started: [\[edit date...\]](#)

Projects: Reader;Writer [\[edit projects...\]](#)

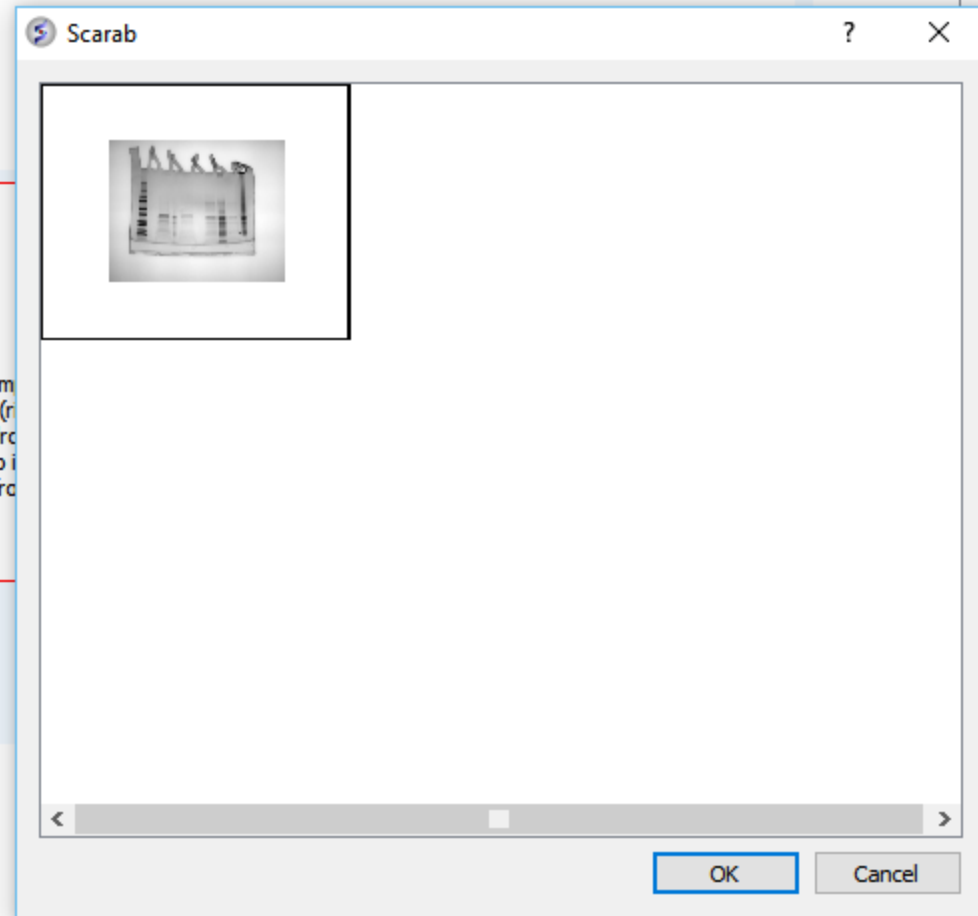
Tags: [\[edit tags...\]](#)

Related Pages: [\[edit related pages...\]](#)

Referenced by:

Empty Item - double click to edit this title

- You can im
- (1) Via the clipboard (r
 - (2) By dragging and dro
 - (3) By clicking here to i
 - (4) Dragging an image fro



UNSTRUCTURED DATA - PAGES



New Molsoft ICMdb 3.8-5 [NewProject *] 0

File Edit View Tools Windows Help Wizards Helpers

Workspace Panel

Use this box to filter the tree

My Stuff (bmarsden)

BRD4 compounds

Page Title	Page ID/Experiment No	Author	Date Started	Last Edited	Experiment Started	Completed
SPOT Assay: Membrane 30 & C-Terminal BET BRDs	PAGE15-01602	abrosig	2015-04-16	2015-05-01 18:20:13	2015-04-01	no
SPOT Assay: Membrane 30 & N-Terminal BET BRDs	PAGE15-01603	abrosig	2015-04-16	2015-05-01 18:19:23	2015-03-25	no
BRDTBD1 (without tag) Expression & Purification	PAGE15-01745	abrosig	2015-04-26	2015-05-01 13:16:32	2015-03-11	no
BRDTBD2 (without tag) Expression & Purification	PAGE15-01746	abrosig	2015-04-26	2015-04-29 16:42:30	2015-03-11	no
BRD3BD2 (without tag) Expression & Purification	PAGE15-01744	abrosig	2015-04-26	2015-04-29 16:16:23	2015-03-11	no
BRDTBD2-His Expression & Purification	PAGE15-01714	abrosig	2015-04-23	2015-04-26 18:41:12	2015-03-18	no
BRDTBD1-His Expression & Purification	PAGE15-01675	abrosig	2015-04-21	2015-04-26 18:34:51	2015-03-18	no
VORLAGE - EXPRESSION PURIFICATION CLEAVAGE	PAGE15-01747	abrosig	2015-04-26	2015-04-26 18:26:21	2015-03-03	no
BRD3BD2-His Expression & Purification	PAGE15-01694	abrosig	2015-04-22	2015-04-26 17:25:23	2015-03-11	no
BRD3BD1 (without tag) Expression & Purification	PAGE15-01743	abrosig	2015-04-26	2015-04-26 16:35:13	2015-03-03	no
BRD2BD2 (without tag) Expression & Purification	PAGE15-01740	abrosig	2015-04-24	2015-04-26 16:12:51	2015-03-03	no
BRD4BD2 (without tag) Expression & Purification	PAGE15-01735	abrosig	2015-04-24	2015-04-26 15:58:01	2015-03-11	no
BRD3BD1-His Expression & Purification	PAGE15-01672	abrosig	2015-04-21	2015-04-26 15:20:16	2015-03-03	no
BRD4BD1 (without tag) Expression & Purification	PAGE15-01719	abrosig	2015-04-24	2015-04-26 14:54:20	2015-03-03	no
BRD2BD2-His Expression & Purification	PAGE15-01647	abrosig	2015-04-20	2015-04-26 13:17:46	2015-03-11	no
BRD2BD1 (without tag) Expression & Purification	PAGE15-01727	abrosig	2015-04-24	2015-04-26 12:43:00	2015-03-03	no
BRD4BD2-His Expression & Purification	PAGE15-01710	abrosig	2015-04-23	2015-04-24 18:14:26	2015-03-11	no
BRD4BD1-His Expression & Purification	PAGE15-01379	abrosig	2015-03-31	2015-04-24 15:57:03	2015-03-02	no
BRD2BD1-His Expression & Purification	PAGE15-01519	abrosig	2015-04-11	2015-04-24 15:56:06	2015-03-03	no
-	PAGE15-01131	abrosig	2015-03-17	2015-04-21 17:29:15		no

2015

- All
- Activity Assay
- Expression
- Purification

Templates

Queries (4 items)

- abullock Alexander Bullock
- acarpenter Anna Carpenter
- acarter Alison Carter

Pages

Welcome to ICM

STRUCTURED DATA MANAGEMENT



Displaying all the targets for a given compound (aspirin)

New Molsoft ICMDb 3.8-5 [NewProject '1] 0

File Edit View Tools Windows Help

Workspace Panel

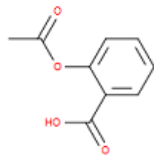
Use this box to filter the tree

- Testing (4 items)
- ChEMBL 18 (5 items)
- ChEMBL 21 (5 items)
- Target (11 items)
 - action_type ~29 rows, View & Refer
 - protein_family_classification ~857 rows, View & Refer
 - organism_class ~3608 rows, View & Refer
 - target_type ~27 rows, View & Refer
 - target_relations ~4786 rows, View & Refer
 - cell_dictionary ~1612 rows, View & Refer
 - protein_class_synonyms ~7539 rows, View & Refer
 - protein_classification ~858 rows, View & Refer
 - target_dictionary ~11019 rows, View & Refer
- Binding Sites (4 items)
- Target Components (4 items)
- Compound (11 items)
 - molecule_hierarchy ~1516066 rows, View & Refer
 - compound_properties ~1587883 rows, View & Refer
 - compound_structures ~1583897 rows, View & Refer
 - molecule_synonyms ~143493 rows, View & Refer
 - molecule_dictionary ~1592191 rows, View & Refer
 - research_companies ~812 rows, View & Refer
 - research_stem ~670 rows, View & Refer
 - usan_stems ~556 rows, View & Refer
 - compound_records ~1929473 rows, View & Refer
- Bio Therapeutics (3 items)
- Drug (5 items)
- Experimental Data (9 items)
 - source ~32 rows, View & Refer
 - docs ~62502 rows, View & Refer
 - assay_type ~6 rows, View & Refer
 - assays ~1212831 rows, View & Refer
 - assay_parameters ~3672553 rows, View & Refer
 - activities ~13968617 rows, View & Refer
 - relationship_type ~6 rows, View & Refer
 - parameter_type ~51 rows, View & Refer
 - Activity-related information (4 items)
- Mechanism/Binding Annotation (2 items)
- Misc (17 items)
- Users
- Pages (0 items)

My Stuff (arman) Targets for Aspirin

Run Save Save As Clear All Keep disabled constraints

Max number of rows to fetch: 1000

Loc	Column	Relation	Value	Logical
1	molfile	Substructure		x

Loc	Column	Type	Grouping	Sort Order	Functions
1	full_mwt	Real	no grouping	no sorting	x
2	standard_type	String	no grouping	no sorting	x
3	description	String	no grouping	no sorting	x
4	pref_name	String	no grouping	no sorting	x
5	target_desc	String	no grouping	no sorting	x
6	assay_desc	String	no grouping	no sorting	x
7	src_description	String	no grouping	no sorting	x

Create List Save Format

	full_mwt	standard_type	description	pref_name
7	180.16	Inhibition	Percentage inhibition against platelet activating factor (2 ng/mL) indu...	Unchecked
8	180.16	IC50	Inhibitory concentration in DMSO with purified human Prostaglandin ...	Cyclooxygenase-2
9	180.16	IC50	Inhibitory concentration in DMSO with purified ovine Prostaglandin G/...	Cyclooxygenase-1
10	180.16	Dose	Evaluated for dose to inhibit Prostaglandin G/H synthase of platelets ...	Cyclooxygenase
11	180.16	ED50	Analgesic activity was assessed from the ability to inhibit phenylquin...	Mus musculus
12	180.16	ED50	Effective dose required for writhing response in mouse after 20 minut...	Mus musculus
13	180.16	EC50	In vitro inhibition of collagen-induced human platelet aggregation.	Homo sapiens
14	180.16	Relative activity	Relative activity (compared to acetylsalicylic acid) for in vitro inhibi...	Homo sapiens
15	180.16	ED33	Ex vivo inhibition of collagen-induced rat platelet aggregation after per...	Rattus norvegicus
16	180.16	Relative activity	Activity relative to dihydralazine for in vitro hypotensive activity in an...	Rattus norvegicus
17	180.16	Ratio	In vitro inhibition of collagen induced aggregation of human plateles a...	Unchecked
18	180.16	Inhibition	Analgesic activity was measured on phenylquinone writhing in mice a...	Mus musculus
19	180.16	MIC	Compound was measured as concentration of compound required for	Trichomonas vaginalis

Showing row 1 of 1000 rows (more to be fetched, but limit is 1000)

RESULT QUERY SPLIT



Fetch list of compounds for given target(s)

New Molsoft ICMDb 3.8-5 [NewProject *] 0

File Edit View Tools Windows Help

Workspace Panel

Use this box to filter the tree

- acd_most_apka
- acd_most_bpka
- acd_logp
- acd_logd
- molecular_species
- full_mwt
- aromatic_rings
- heavy_atoms
- num_alerts
- qed_weighted
- mw_monoisotopic
- full_molformula
- hba_lipinski
- hbd_lipinski
- num_lipinski_ro5_violations
- compound_structures ~1583897 rows, View & Refer
- molecule_synonyms ~143493 rows, View & Refer
- molecule_dictionary ~1592191 rows, View & Refer
- research_companies ~812 rows, View & Refer
- research_stem ~670 rows, View & Refer
- usan_stems ~556 rows, View & Refer
- compound_records ~1929473 rows, View & Refer
- Bio Therapeutics (3 items)
- Drug (5 items)
- Experimental Data (9 items)
- source ~32 rows, View & Refer
- src_id
- src_description
- src_short_name
- docs ~62502 rows, View & Refer
- assay_type ~6 rows, View & Refer
- assays ~1212831 rows, View & Refer
- assay_id Key
- doc_id
- description
- assay_type
- assay_test_type
- assay_category
- assay_organism
- assay_tax_id
- assay_strain

Edit: Activity Assay Constraint: <Insert empty row> Value: Edit Data

My Stuff (arman) Targets for Aspirin target_type target_dictionary

Run Save Save As Clear All Keep disabled constraints Max number of rows to fetch

Loc	Column	Relation	Value	Logical
1	pref_name	Starts With	Estrogen receptor	AND
2	organism	Equals	Homo sapiens	

Loc	Column	Type	Grouping	Sort Order	Functions
1	molfile	Chemical	no grouping	N/A	
2	standard_type	String	no grouping	no sorting	
3	description	String	no grouping	no sorting	
4	src_description	String	no grouping	no sorting	
5	full_mwt	Real	no grouping	no sorting	
6	assay_type	String	no grouping	no sorting	

Create List Save Format

	molfile	standard_type	description	src_description	full_mwt	assay_type
51		Relative activation	Activation of the estrogen receptors of MCF-7-2a cells at 10e-6 M.	Scientific Literature	367.27 B	
52		RBA	Percent relative binding affinity for estrogen receptor ligand binding domain expressed in BL21 cells at 2C	Scientific Literature	432.55 B	
53		RBA	Percent relative binding affinity for estrogen receptor ligand binding domain expressed in BL21 cells at 2C	Scientific Literature	399.52 B	

Showing row 3 of 1001 rows (more to be fetched)

RESULT QUERY SPLIT

STRUCTURED DATA MANAGEMENT



New Molsoft ICMdb 3.8-5 [NewProject *] 0

File Edit View Tools Windows Help Wizards Helpers

Workspace Panel

Use this box to filter the tree

Databases (1 item)

- Scarab
 - bmarsden Brian Marsden
 - Target History
 - Last Entity IDs
 - Gene Projects
 - Check DataSource ddamerell
 - Shared Queries (19 items)
 - Shared Lists (2 items)
 - Institute Projects
 - Tags
 - Groups
 - Global Templates
 - OxHive (17 items)
 - Useful Lists (7 items)
 - Target Information (5 items)
 - Grant / Project (6 items)
 - Genes (18 items)
 - Target (55 items)
 - Target Tags (6 items)
 - Bioinformatics (22 items)
 - Cloning (13 items)
 - Test Expression, Scale-up and Purification (7 items)
 - Biophysics & Chemical Biology (16 items)
 - Chemical Probes (16 items)
 - Recombinant Antibodies: Phage Display (12 items)
 - NMR (5 items)
 - Protein Crystallisation (11 items)
 - X-ray Structure Determination (7 items)
 - Structure Deposition (37 items)
 - Structure Deposition ~597 rows
 - Useful Queries (1 item)
 - Structure Deposition PDB ID Update
 - Date PDB Deposition Occurred Update
 - Date PDB Released Update
 - Structure Deposition Type Update
 - Structure Deposition SGC Number Update
 - Structure Deposition Can Be Counted? Update
 - Counting Summary Update
 - Counted - ULTRA-DD Update
 - Counted - TEP Update

My Stuff (bmarsden) Structure Deposi... X

Create List Save Format

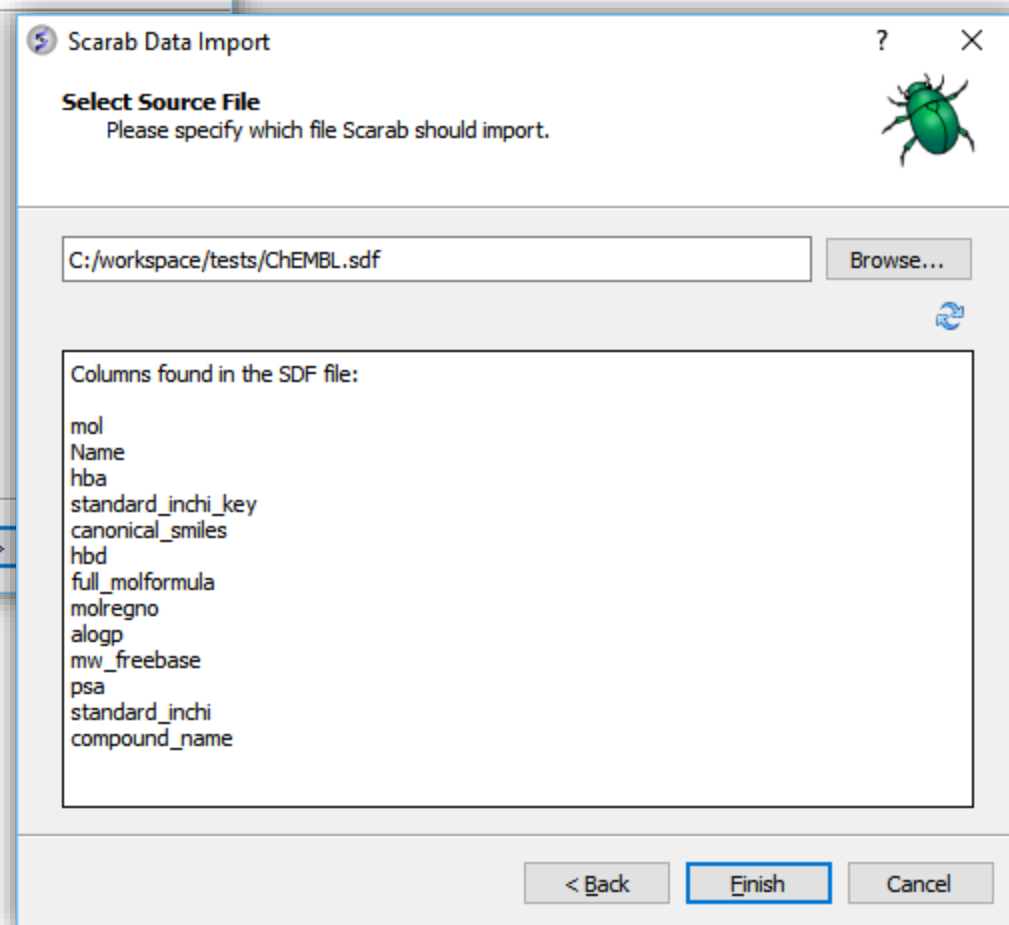
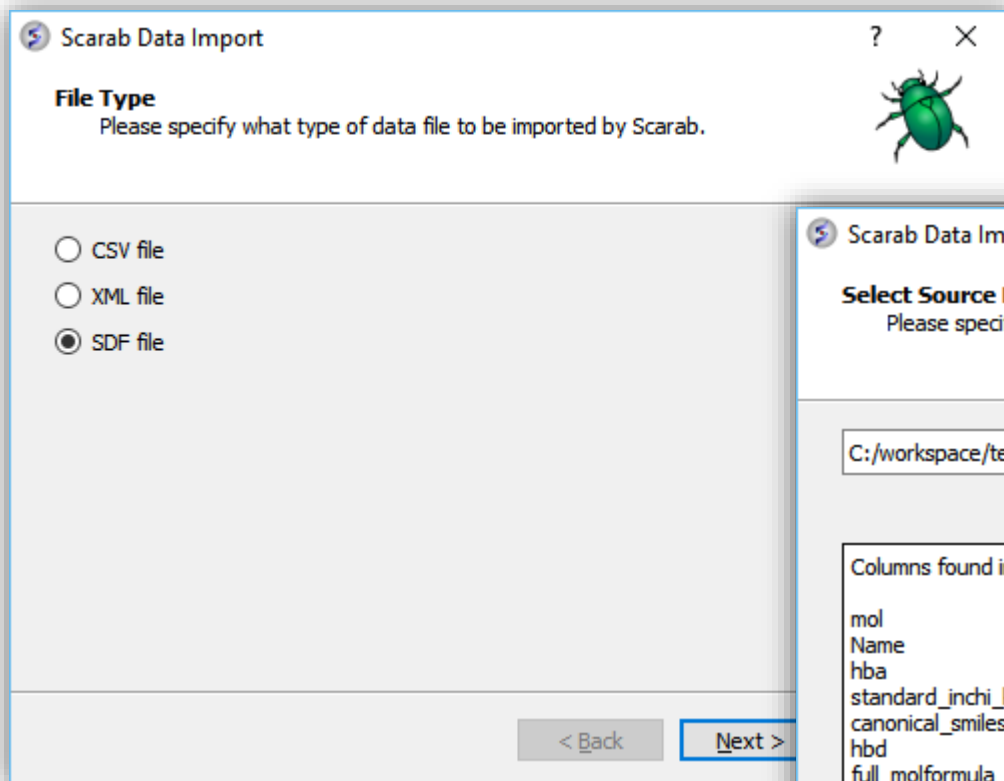
Structure Deposition PDB ID	Date PDB Deposition Occurred	NCBI Gene ID
1 5fwj	2016-02-17	8242
2 5IGN	2016-02-28	65980
3 5II2	2016-03-01	55193
4 5II1	2016-03-01	55193
5 5IID	2016-03-01	55193
6 5fxv	2016-03-03	7404
7 5fxw	2016-03-03	7404
8 5fxx	2016-03-03	7404
9 5fzx	2016-03-03	7404
10 5fy0	2016-03-03	7404
11 5fy1	2016-03-03	7404
12 5fy4	2016-03-04	10765
13 5fy9	2016-03-04	10765
14 5fyi	2016-03-07	9682
15 5fyu	2016-03-10	10765
16 5fyt	2016-03-10	10765
17 5fyv	2016-03-10	10765
18 5fys	2016-03-10	10765
19 5fz8	2016-03-11	10765
20 5FPV	2015-12-21	9682
21 5FUN	2016-01-28	10765
22 5fy7	2016-03-04	7404
23 5fy8	2016-03-04	9682
24 5fz6	2016-03-11	10765
25 5fz7	2016-03-11	10765
26 5fze	2016-03-14	10765
27 5fzg	2016-03-14	10765
28 5fzh	2016-03-14	10765
29 5FFY	2015-12-19	7862
30 5fy5	2016-03-04	10765
31 5fym	2016-03-08	7404
32 5fz9	2016-03-12	10765
33 5fyy	2016-03-10	10765
34 5fz1	2016-03-10	10765
35 5fz4	2016-03-10	10765

Showing row 1 of 64 rows

RESULT QUERY SPLIT

All Pages

Welcome to ICM





- Post-import of a snippet of ChEMBL:

The screenshot shows the Molsoft ICMdb 3.8-5 interface. The left sidebar (Workspace Panel) displays a tree view of the database structure, including folders for Projects, Tags, Groups, Global Templates, TEST (2 items), sharedQueries (0 items), Page Edit Queries (2 items), and ChEMBL (1 item). The ChEMBL folder is expanded to show a list of fields: mol, molregno, standard_inchi, standard_inchi_key, canonical_smiles, alogp, full_molformula, mw_freebase, hba, hbd, psa, compound_name, and ld. The main panel shows a 'My Stuff' tab with a search box and a table of saved queries.

Query Name	Save	Access	Referenced	Rows Fetched	Date Modified
<input type="checkbox"/> chembl 18 Pages	saved	public	Determine	n/a	2016-03-10 16:07
<input type="checkbox"/> chembl 18 Query	saved	public	Determine	2000	2016-03-10 16:00
<input type="checkbox"/> test_table dupl names	saved	public	Determine	600	2016-02-17 11:58
<input type="checkbox"/> test query	saved	public	Determine	8	2016-02-17 11:44
<input type="checkbox"/> keep const	saved	public	Determine	2000	2016-02-17 11:43
<input type="checkbox"/> AAA	saved	public	Determine	2000	2016-02-11 17:39



Query the data:

New Molsoft ICMdb 3.8-5 [NewProject *] 0

File Edit View Tools Windows Help

Workspace Panel

Use this box to filter the tree

My Stuff (adminst... CHEMBL Query

Run Save Save As Clear All Keep disabled constraints Max number of rows to fetch 1000

Loc	Column	Relation	Value	Logical
1	mol	Substructure <input type="checkbox"/> Match stereo <input type="checkbox"/> Ignore salt		X

Loc	Column	Type	Grouping	Sort Order	Functions
1	mol	Chemical	no grouping	N/A	
2	compound_name	String	no grouping	no sorting	
3	full_molformula	String	no grouping	no sorting	
4	mw_freebase	Real	no grouping	no sorting	
5	hba	Int	no grouping	no sorting	
6	hbd	Int	no grouping	no sorting	
7	psa	Real	no grouping	no sorting	

CHEMBL/CHEMBL/mol

Create List Save Format

	mol	compound_name	full_molformula	mw_freebase	hba	hbd	psa
27		Phenyl-[(E)-styryl]-quinazolin-4-yl-amine	C22H17N3	323.39	3	1	37.81
28		Benzo[1,3]dioxol-5-yl-(6,7,8-trimethoxy-quinazolin-4-yl)-amine	C18H17N3O5	355.34	8	1	83.96

Showing row 1 of 271 rows

RESULT QUERY SPLIT

All Pages

Welcome to ICM



- Capture your ICM session in a Page

The screenshot displays the ICM (Interchangeable Chemical Modeling) software interface. The main window shows a 3D molecular model of a streptavidin complex with biotin, rendered in a stick representation with atoms colored by element (carbon in grey, oxygen in red, nitrogen in blue). The interface includes a menu bar (File, Tools, Edit, View, Bioinfo, Homology, Chemistry, Docking, MollMechanics, Windows, Help), a toolbar with various manipulation tools, and a Workspace Panel on the left. The Workspace Panel shows a tree view of objects: 'objects (2 items)' containing 'biotin [1] ICM; 2.6A', 'biotin H btn', 'streptavidin complex with biotin', 'rec [2*] ICM; 1.8A; streptavidin c', and 'a 116A'. Below the Workspace Panel is a terminal window with the following text:

```
Info> a_biotin. "streptavidin complex with biotin"  
Info> a_rec. "streptavidin complexed with haba"  
Info> 127 colors and 18 graphic modes read from C:/MinGW/msys/1.0/home/arman/main_workspace/icmd/icm.clr  
Info> 29 shell objects read (skipped 1) from C:/Users/arman/AppData/Local/Temp/pageicb_8_23_in.icb  
icm/rec>
```

The terminal window also shows the status 'Welcome to ICM' and '2 ICM Obj'.



- Capture your ICM session in a PAGE

The screenshot shows a web application window titled "PAGE16-00007". The interface includes a top toolbar with options like "Export to PDF", "Export to Word", "Print", "Submit", "Preferences", "View Audit", "Authors", and "Refresh". Below the toolbar, the page content is displayed within a light blue border. The main content area has a title "ICB session test" and a subtitle "PAGE16-00007". It lists metadata: "Author: administrator", "Created: 2016-Feb-15", "Last updated: 2016-Mar-17 08:13:42", "Experiment started:", "Projects: Reader", "Tags:", "Related Pages:", and "Referenced by:". Below this, there is a section titled "Streptavidin complex with biotin" with a subtitle "Double click to open ICB session". The main visual element is a 3D ball-and-stick molecular model of the Streptavidin complex with biotin, rendered in grey, blue, and red.



- Enhanced web/mobile interfaces
- Further integration with ICM capabilities
- LDAP authentication
 - For enterprise environments
- Deep search (elasticsearch-based)
 - Google-like search to your data



- Scarab captures structured and unstructured data
- Scarab provides a means to present multiple databases in a unified, manner without changes to those databases
- Scarab provides a powerful and unique query/report builder

- A critical component of the SGC's success
- Ensures we have captured all aspects of our work
- Allows us to continue to be recognised as being able to deliver reproducible data

- Scarab is an ideal tool to capture the provenance of your work in ICM. Please try it!

ACKNOWLEDGEMENTS



SGC Oxford RI Group

Brian Marsden

David Damerell

Uka Osim

Wen Hwa Lee

Qi Liu

Julián Atienza-Herrero

Hong Harper

Minghua Wang

James Bray

MolSoft

Ruben Abagyan

Arman Sahakyan

Eugene Raush

Max Totrov

Rita Abagyan

Andy Orry

Slava Brover

SGC Toronto

Matthieu Schapira

Lihua Liu

FUNDING PARTNERS

The SGC is a registered charity (number 1097737) that receives funds from AbbVie, Bayer Pharma AG, Boehringer Ingelheim, Canada Foundation for Innovation, Eshelman Institute for Innovation, Genome Canada, Innovative Medicines Initiative (EU/EFPIA), Janssen, Merck & Co., Novartis Pharma AG, Ontario Ministry of Economic Development and Innovation, Pfizer, São Paulo Research Foundation-FAPESP, Takeda, and Wellcome Trust