

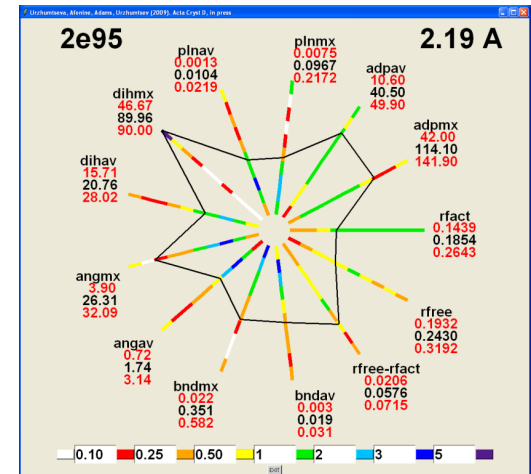
The Tcl/tk version 3.2 of the POLYGON program

17-01-2009

L.Urzhumtseva, P.V.Afonine, P.D.Adams, A.Urzhumtsev

The purpose of the program :

- show qualitatively if some models parameters are over- or under-refined
- compare the model characteristics with those for previously refined models; the full PDB or different personalized selections can be used



The set of files includes :

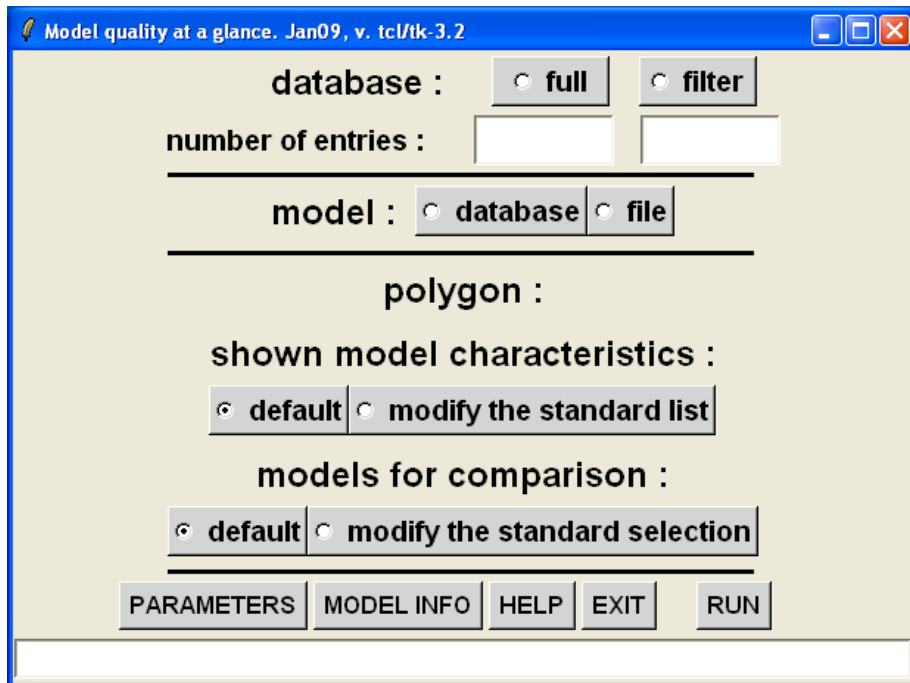
- Tcl source of the POLYGON program
- The database (file obtained initially by P.Afonine from 12.10.2008 using *phenix.model_vs_data* applied for all PDB models when possible and re-written in form of a table; the latter accelerates loading the program)
- Two examples of a *.pdb file, one refined by PHENIX, another one by CNS
- An example of a *.mvd file (output of *phenix.model_vs_data*)

How to run :

- Windows: make sure Tcl/tk is installed (see next page otherwise). To run POLYGON just double-click on its icon.
- LINUX/UNIX: one may need to correct the FIRST line of the source code: this is a comment line, but it should indicate the path to the program 'wish' in these libraries.
- I do not know if there are specific tricks to run Tck/tl programs for Macintosh
- Some comments about features of the program are available from Help button in program's GUI.

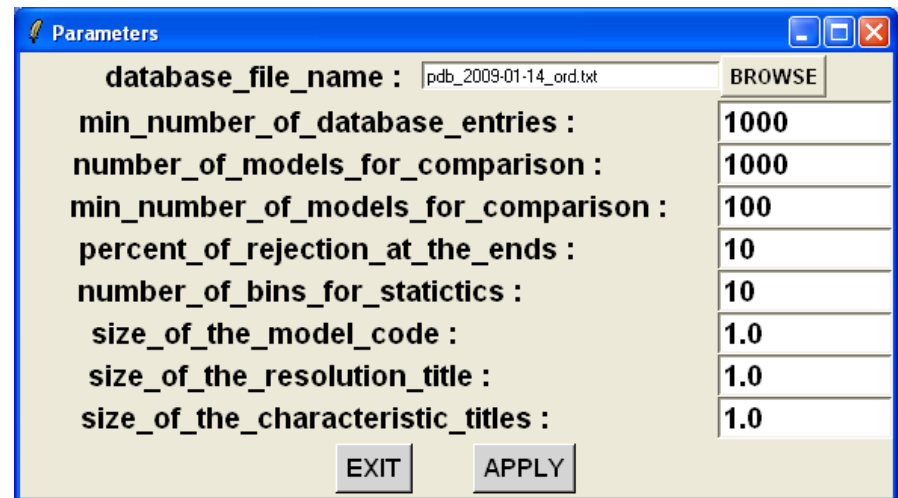
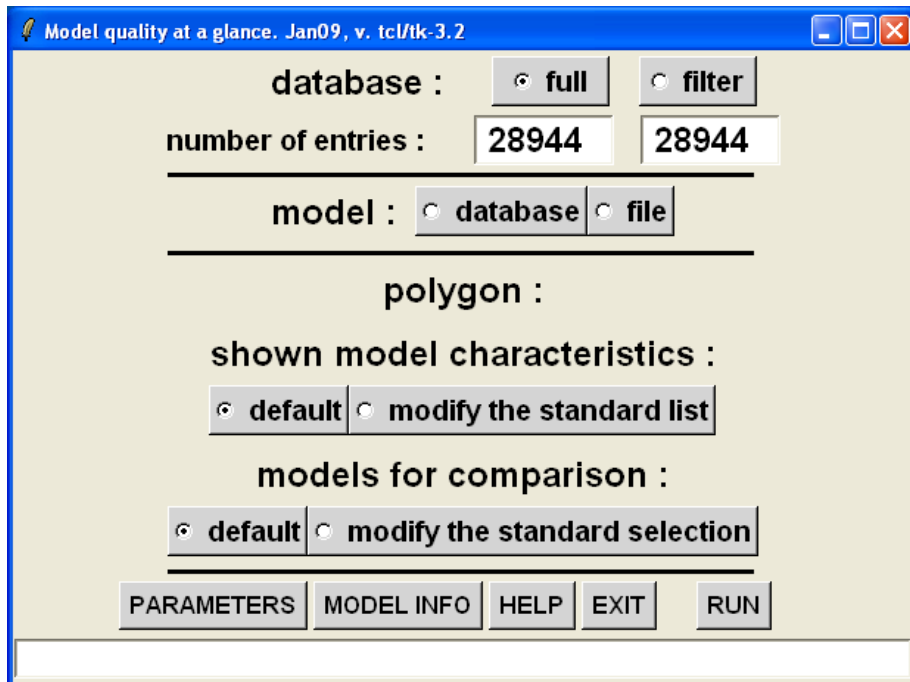
If Tcl/tk is not yet installed :

- get the installation files from, for example,
<http://www.tcl.tk/software/tcltk/>
<http://www.activestate.com/Products/activetcl/index.mhtml>
- follow the instructions
- do not forget to restart your computer after installation.



Choose if you filter the database or take it as a whole. If necessary, filtering can be done later.

By default the database is read in from the standard file but the choice of the file (for example, for a previously filtrated and saved database) can be made using the button 'Parameters' (at the beginning of the session or later).



Model quality at a glance. Jan09, v. tc/ltk-3.2

database : full filter

number of entries :

model : database file

enter a PDB code or click on the "Browse" button:

polygon :

shown model characteristics :

default modify the standard list

models for comparison :

default modify the standard selection

The model information for comparison can be selected from the database of entered from a file either in the pdb format or in the model_vs_data format.

For selection from the database, either type directly the PDB code or press "browse" and select the model from the list. The list can be ordered by the PDB code, by the high resolution limit of the year of the model deposition.

Model quality at a glance. Jan09, v. tc/ltk-3.2

database : full filter

number of entries :

model : database file

enter a PDB code or click on the "Browse" button:

polygon :

shown model characteristics :

default modify the standard list

models for comparison :

default modify the standard selection

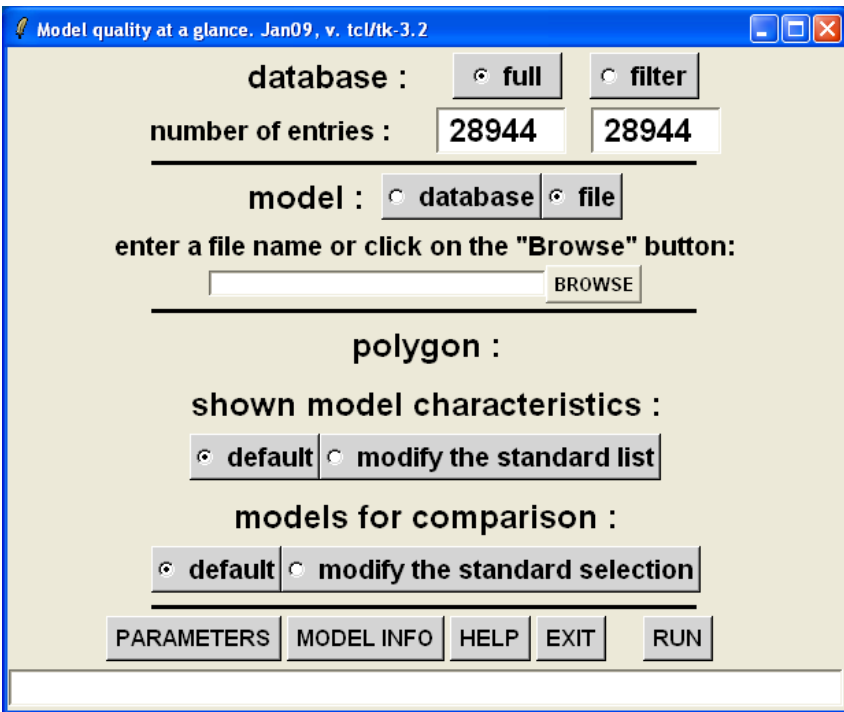
Model selection

select a model from the list

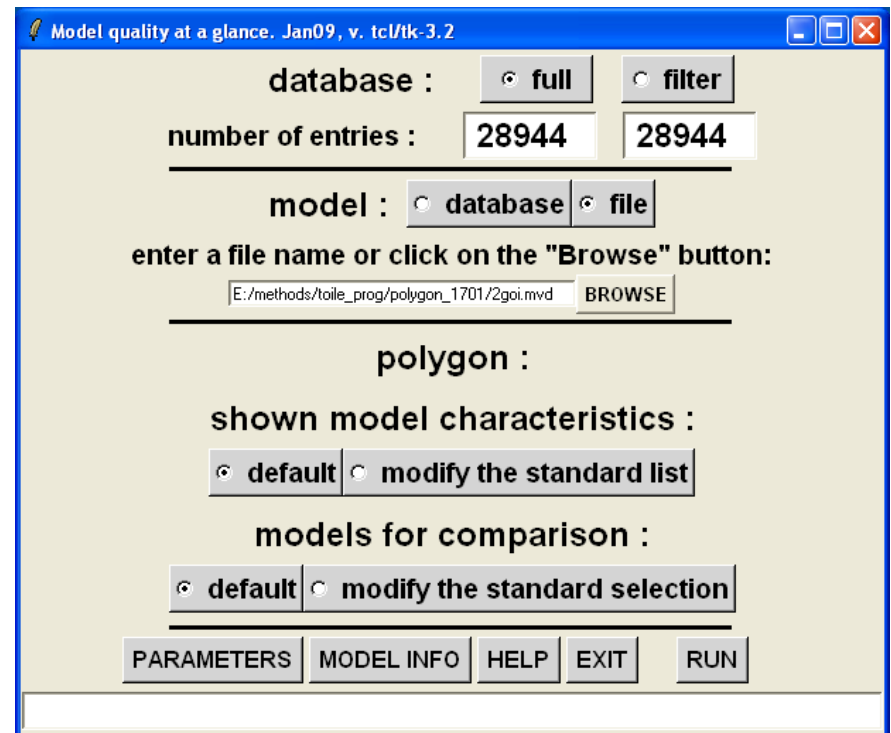
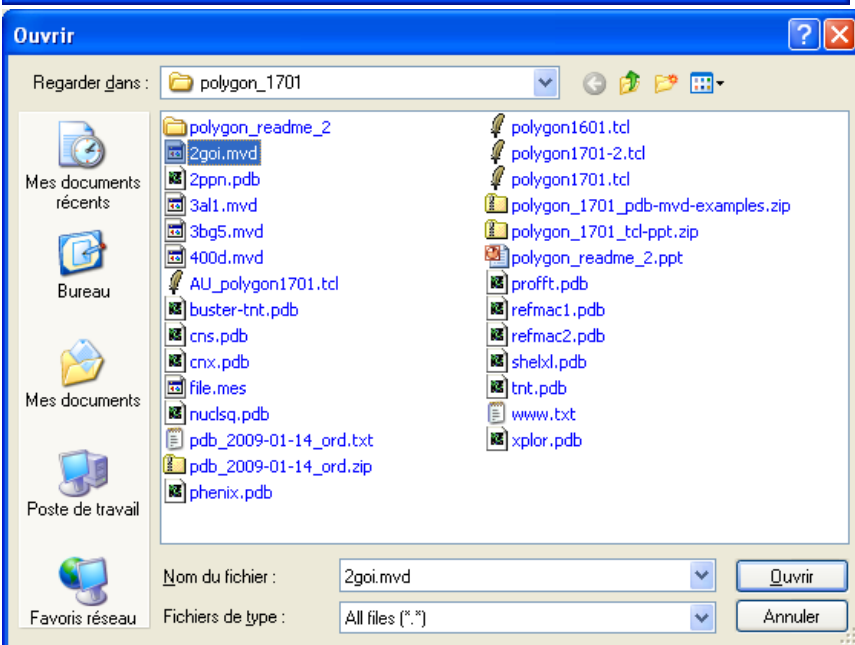
order by:

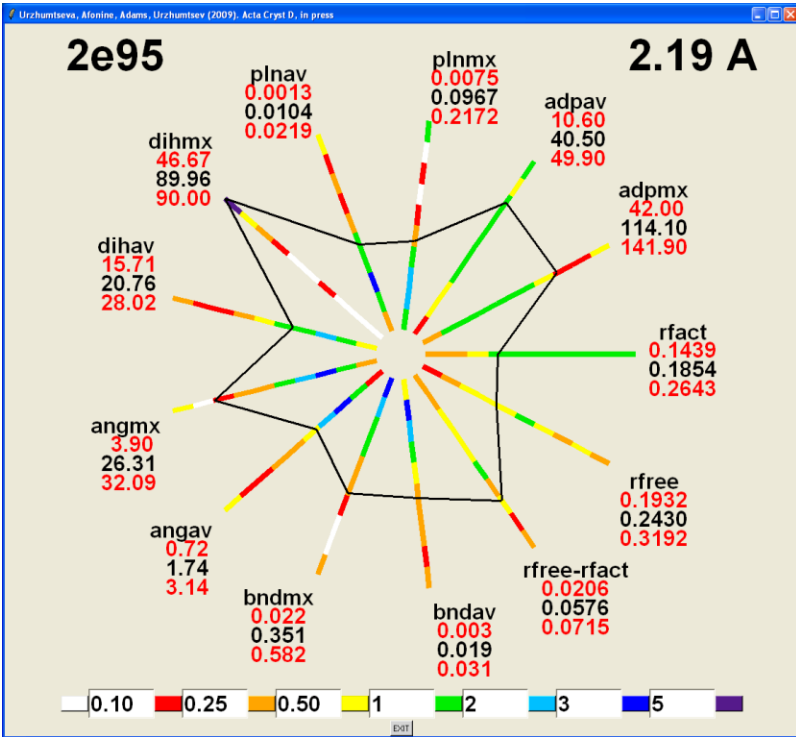
code dhigh year

1twf	2.19	2004
1w5a	2.19	2004
1rsv	2.19	2003
1kxi	2.19	1996
2e95	2.19	2007
1trn	2.19	1995
1rz0	2.19	2003
2i39	2.19	2006
2q1w	2.19	2007
1deh	2.19	1995
1del	2.19	1996
1m8w	2.19	2002



For selection from the file, type the filename in the window below. Otherwise, press "browse", select the directory and then the file.





Click 'Enter' or 'run' button to get a polygon calculated with default values of the characteristics shown and the models chosen for comparison by default conditions.

POLYGON - model quality at a glance

PURPOSE
 The program gives a simple graphic presentation of comparing it with that for other models deposited pre information is shown at the axes of a polygon so that a polygon at a middle of axes corresponds to a "mean" a polygon compressed to the center means a better r dilated along some axis indicates problems with the characteristic.

IMPORTANT
 The program does not perform any analysis of a mo graphically represents the model and data statistics t extracted from either input PDB file header or taken f of phenix.model_vs_data program.

THE DATABASE OF DATA AND MODEL CHARACTERI
 The set of key model and data characteristics (such deviations from ideal stereochemistry, etc.) is obtain structure factors from the PDB (July 2008) by running tool for each PDB entry.

DATABASE FILTERING
 If necessary, some models (selected by the user) ca the analysis. Those, for example, could be the neutro when the analysis is being done for a structure obtain To apply a filter, enter the interval limits for numeric c or choose one or several possibilities for non-number then activate corresponding filter. The counter show the models selected. The filtered set of models can b be used in subsequent runs (see PARAMETERS).

SOURCE OF THE MODEL INFORMATION
 The characteristics for the analyzed model can be t different sources: PDB file header, output of phenix.n internal database (in this case the PDB code should t

CHARACTERISTICS SHOWN

```

Model information
Program: model_vs_data (PHENIX)
Unit cell: (46.39, 116.26, 128.70, 90.00, 90.00, 90.00)
Space group: 19
Unit cell volume: 694.1
Model:
Atoms:
  atom_number          : 5508
  ADP_(min,max,mean)   : 14.90 114.10
  40.50
  occupancies_(min,max,mean) : 1.00 1.00
  1.00
  number_of_anisotropic : 0
  number_of_non_positive_definite : 0
Stereochemistry statistics (mean, max):
  bonds      : 0.019 0.351
  angles     : 1.74 26.31
  dihedrals  : 20.76 89.96
  chirality  : 0.131 0.573
  planarity  : 0.0104 0.0967
  non-bonded (min) : 1.83
  Ramachandran plot, number of:
  outliers   : 1
  general (%) : 86.28
  favored (%) : 95.85
Data:
  data_label      : FOBS
  high_resolution : 2.19
  low_resolution  : 33.20
  completeness_in_range : 0.95
  completeness_(d_min-inf) : 0.95
  completeness_(6A-inf) : 0.98
  wilson_b       : 29.1
  number_of_reflections : 3473
  test_set_size   : 0.050
  test_flag_value : 1
  number_of_Fobs_outliers : 12
  anomalous_flag  : False
Model_vs_Data:
  r_work(re-computed) : 0.185
  r_free(re-computed) : 0.241
  bulk_solvent_(k_sol,b_sol) : 0.3
Information extracted from PDB file:
  program_name : CNS
  year         : 2007
  r_work       : 0.185
  r_free       : 0.241
  high_resolution : 2.2
  low_resolution : 34.9
  sigma_cutoff  : 0.0
  
```

Program parameters that can be modified, model information in the text form and Help information are available when clicking at corresponding buttons

```

Parameters
database_file_name : pdb_2009-01-14_ord.txt BROWSE
min_number_of_database_entries : 1000
number_of_models_for_comparison : 1000
min_number_of_models_for_comparison : 100
percent_of_rejection_at_the_ends : 10
number_of_bins_for_statistics : 10
size_of_the_model_code : 1.0
size_of_the_resolution_title : 1.0
size_of_the_characteristic_titles : 1.0
EXIT APPLY
  
```

Select the models satisfying the following conditions

<input type="checkbox"/>	<= rfact <=	
<input type="checkbox"/>	<= rfree <=	
0.01	<input checked="" type="checkbox"/> <= rfree-rfact <=	1.0
<input type="checkbox"/>	<= rpdb <=	
<input type="checkbox"/>	<= rfpdb <=	
<input type="checkbox"/>	<= rfpdb-rpdb <=	
-0.05	<input checked="" type="checkbox"/> <= rfact-rpdb <=	0.05
<input type="checkbox"/>	<= adpmn <=	
<input type="checkbox"/>	<= adpav <=	
<input type="checkbox"/>	<= adpmx <=	
<input type="checkbox"/>	<= bulkk <=	
<input type="checkbox"/>	<= bulkb <=	
<input type="checkbox"/>	<= bndav <=	
<input type="checkbox"/>	<= bndmx <=	
<input type="checkbox"/>	<= angav <=	
<input type="checkbox"/>	<= angmx <=	
<input type="checkbox"/>	<= dihav <=	
<input type="checkbox"/>	<= dihmX <=	
<input type="checkbox"/>	<= plnav <=	
<input type="checkbox"/>	<= plnmX <=	
<input type="checkbox"/>	<= chrav <=	
<input type="checkbox"/>	<= chrmX <=	
<input type="checkbox"/>	<= noutl <=	
<input type="checkbox"/>	<= nonpos <=	
<input type="checkbox"/>	<= nbdis <=	
<input type="checkbox"/>	<= compl <=	
<input type="checkbox"/>	<= cmp-t <=	
<input type="checkbox"/>	<= cmp-l <=	

<input type="checkbox"/>	<= nrefl <=	
<input type="checkbox"/>	<= natoms <=	
<input type="checkbox"/>	<= nrefl / natoms <=	
<input type="checkbox"/>	<= natiso <=	
<input type="checkbox"/>	<= nmod <=	
<input type="checkbox"/>	<= bwils <=	
<input type="checkbox"/>	<= year <=	
<input type="checkbox"/>	<= t-set <=	
<input type="checkbox"/>	<= sigma <=	
<input type="checkbox"/>	<= rmout <=	
<input type="checkbox"/>	<= rmfav <=	
<input type="checkbox"/>	<= rmgen <=	
<input type="checkbox"/>	<= wat/res <=	

space group number - or - or -

molecule type prot nucl.acid prot+nucl.acid prot+nucl.acid+other
 prot+other nucl.acid+other other

data type FOBS_X FOBS_N IOBS_X IOBS_N OBS FOBS IOBS

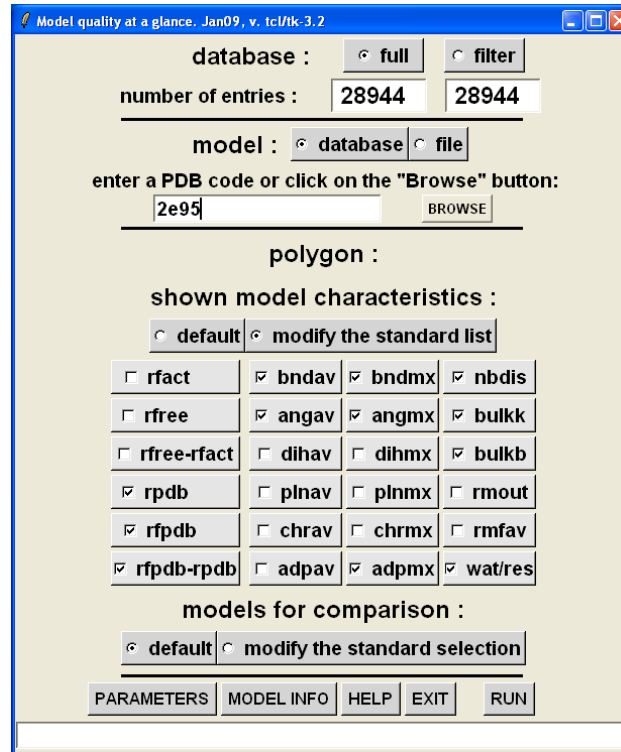
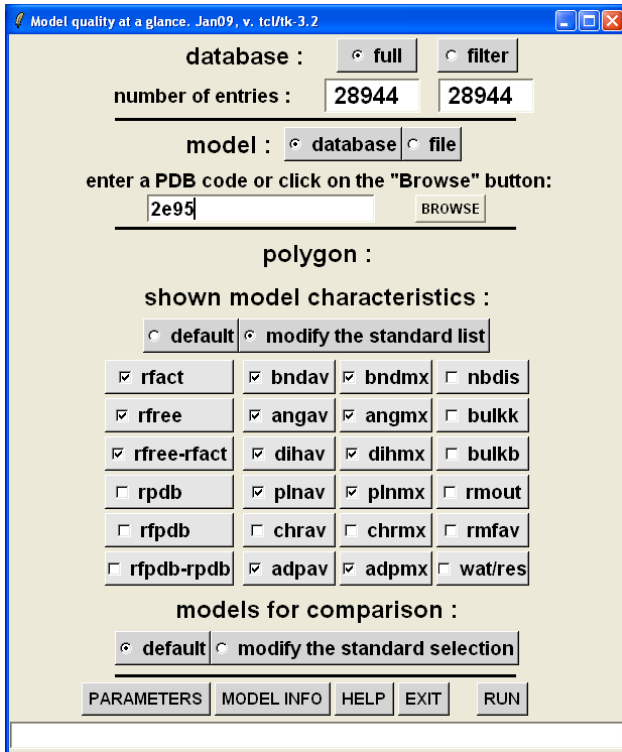
anomalous False True

program ARP BUSTER CCP4 CNS CNX CORELS
 EREF GPRLSA GSAS JACK_A MAIN MOLLY
 NMREF NUCLSQ PHENIX PROFFT PROLSQ REFMAC
 RESTRA SHELX SOLVE TNT X-PLOR unkn

number of selected entries: 14839

EXIT APPLY SAVE SELECTION : BROWSE

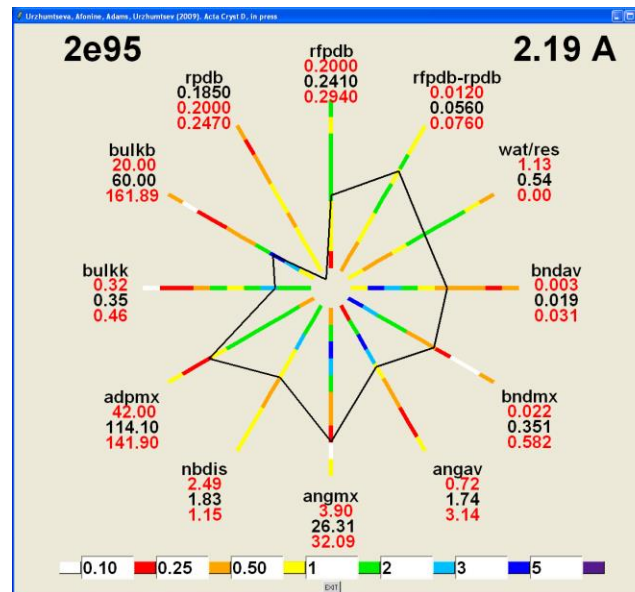
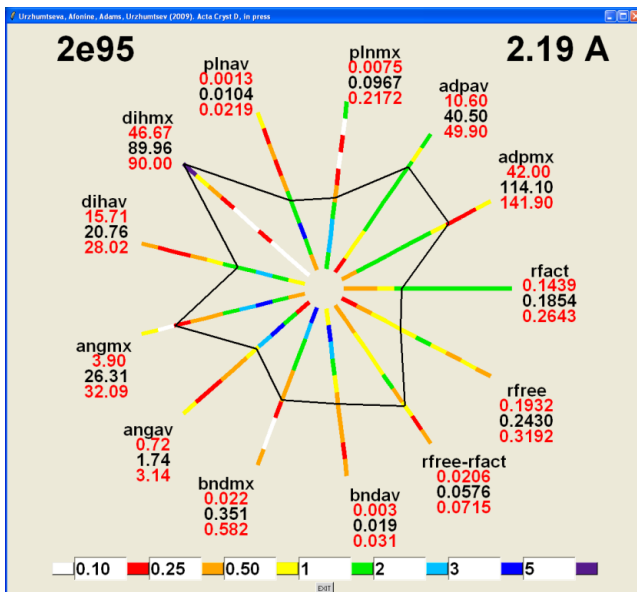
Filtering excludes from further analysis all models in the database that are considered to be wrong or inappropriate for comparison. Type first the values and then activate or deactivate the condition(s). The total number of selected models is shown. The filtered database can be saved for further comparisons. Choose then if you accept this selection or not to continue the session. Return to the full database is possible during the session.



The default list of characteristics can be modified.

Minimal number of activated (displayed) characteristics is 3.

The modification of the list is done automatically if the input PDB file miss some of them.



If phenix-calculated R-factors (shown as **rfact** and **rfree**) are missed in the file, the R-factors calculated by other programs are shown (as **rpdb** and **rfpdb !!**)

By default, about 1000 models (from the filtered database if filtration applied) closest in resolution to the current model, are chosen for comparison. This value can be changed using 'Parameters'.

In the advanced mode, the selection of similar models can be tuned manually. The values for the current model are communicated to simplify the choice.

Model quality at a glance. Jan09, v. tcl/tk-3.2

database : full filter

number of entries :

model : database file

enter a PDB code or click on the "Browse" button:

polygon :

shown model characteristics :

default modify the standard list

models for comparison :

default modify the standard selection

Parameters

database_file_name :

min_number_of_database_entries :	1000
number_of_models_for_comparison :	1500
min_number_of_models_for_comparison :	100
percent_of_rejection_at_the_ends :	10
number_of_bins_for_statistics :	10
size_of_the_model_code :	0.0
size_of_the_resolution_title :	1.5
size_of_the_characteristic_titles :	1.0

Model quality at a glance. Jan09, v. tcl/tk-3.2

database : full filter

number of entries :

model : database file

enter a PDB code or click on the "Browse" button:

polygon :

shown model characteristics :

default modify the standard list

models for comparison :

default modify the standard selection

current model:

<input type="text" value="2.19"/>	<input type="text"/>	<input type="checkbox"/> <= dhhigh <=	<input type="text"/>
<input type="text" value="69.0"/>	<input type="text"/>	<input type="checkbox"/> <= mw(kda) <=	<input type="text"/>
<input type="text" value="694.1"/>	<input type="text"/>	<input type="checkbox"/> <= vol/1000 <=	<input type="text"/>
<input type="text" value="2.52"/>	<input type="text"/>	<input type="checkbox"/> <= cmatthw <=	<input type="text"/>

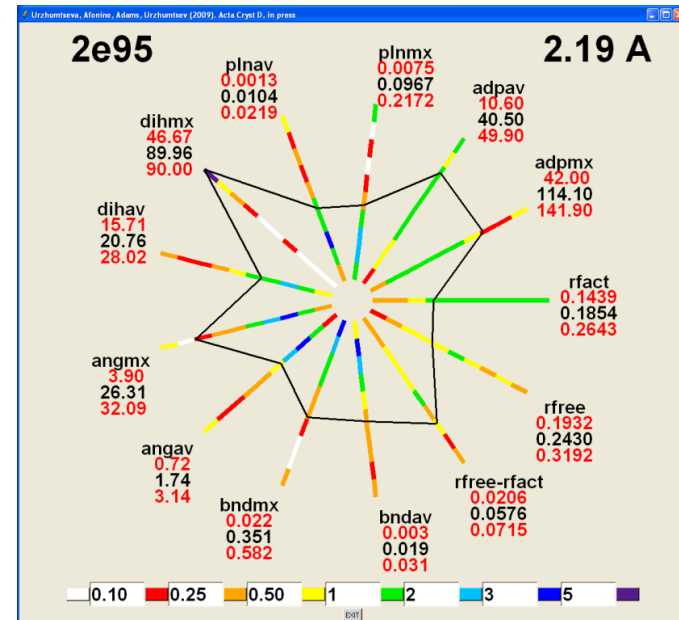
number of selected models

The scale at each ray is linear. The min and max limits are defined from the values of this parameters for the models selected for comparison.

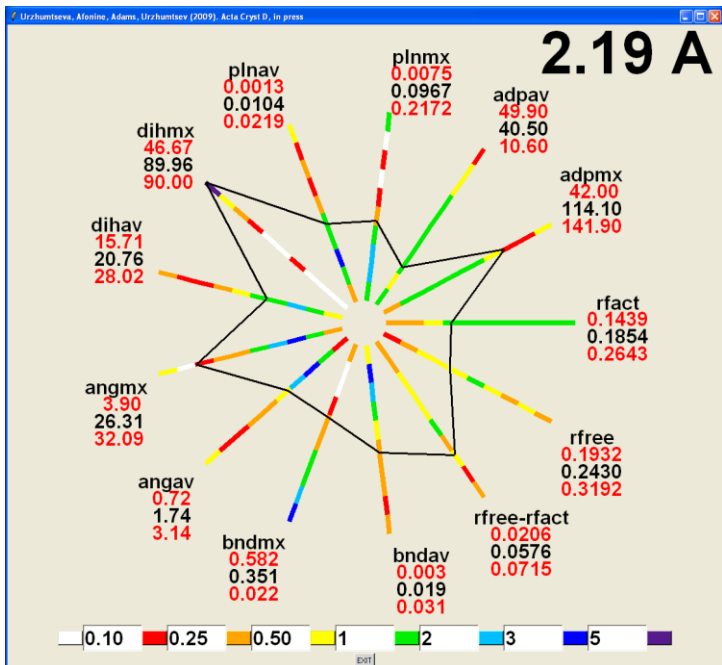
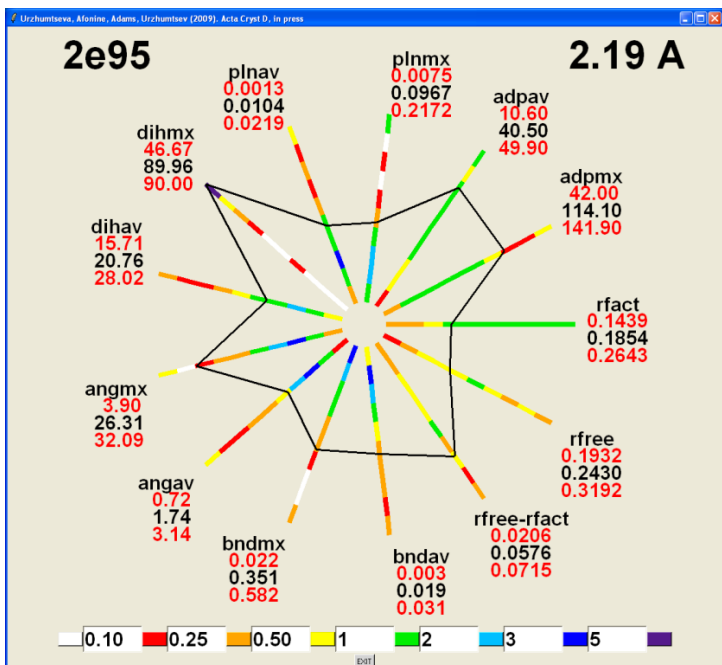
Initial values Xmin and Xmax are defined, and the interval is divided in 3 equal bins. If one of the terminal bins (or a terminal and the central bins) contains too few variants, it is removed from further analysis and the interval is shrunk. Then the procedure is repeated. The shrinking is terminated when either each of bin contains enough models or if already too many models are removed. The default values (10%) can be changed using 'Parameters'. This shrinking is very important to avoid quasi-empty ends and too sharp and non-informative peaks.

For all rays Xmin is close to the centre, Xmax is at the periphery EXCEPT the non-bonded distance 'nbdis' and the %residues in the favorable zones of the Ramachandran plot; for them the direction is inverted.

There is a space between the center and the point for Xmin, and between the periferic circle and Xmax, where the point for the model is positionned if the model parameters are outside the interval (Xmin,Xmax)



For each ray Xmin and Xmax are given, as well as the value for the current model (the position of this number is given with respect to Xmin and Xmax).



Each ray is colored accordingly to the frequency with which the corresponding values are found for the selected models. Each interval (Xmin,Xmax) is divided in 10 bins (this number can be modified using 'parameters') and the number of models N_k , $k=1,\dots,10$, for each bin is calculated. The 'mean' value is also calculated as $N_{uniform}=N_{models}/N_{bins}$.

For each interval the ratio $N_k/N_{uniform}$ is calculated and the sub-interval for this bin is colored accordingly to this value. The color for the corresponding value is chosen from the bottom of the image. The limit values can be updated interactively retyping the new values in the windows.

The fonts are slightly system-computer-Tcl-version dependent. The size of letters can be modified using appropriate scale through 'Parameters'. Definition of a scale equal to 0 hides completely the corresponding text (for example, the PDB code).

FUTURE VERSIONS should include :

- a button «PRINT» near to «EXIT» at the polygon window to make its PDF-JPEG-GIF copy (currently, under Tcl/tk , we can do only a screen copy; if somebody knows how to do it - the advice is welcome).
- when clicking at any ray, a more complet information about the distribution of this characteristic should appear : initial min/max values; initial distribution in form of a curve; the same for Xmin/Xmax shrinked; names for the PDB codes with extreme values, etc etc...
- a possibility to order the database list by other parameters