Contents

Introduction

It is recommended that newcomers to **carma**, first read carma's documentation.

History

Version 0.2

New Features

1. **grcarma** can be launched with a directory path as a single argument. This directory should ideally be a directory created during a previous grcarma run or at the very least include the linked

PSF/DCD files from that run. The directory can be specified through the command line (`grcarma /path/to/dir`) or the graphical interface.

2. **grcarma** now supports 5 dimensional clustering in addition to the standard 3 dimensional clustering performed by **carma**. This is performed through the external program **cluster5D**. The cluster5D executable needs to be in the PATH in order for the 5 dimensional clustering analysis to be enabled.

3. Good-Turing convergence analysis is now available for linux/macOSX. This is performed through the external program **GoodTuringMD**. The source code of the program is included in the grcarma source code, so the user only needs to install the **R statistical package** and some required modules.

Modifications of existing features

1. The width of the cluster plot has been changed so that it takes advantage of the entire horizontal resolution of the screen.
2. The 'RMSD matrix' analysis window now includes a preview of the size of the matrix to be created(in megabytes and matrix dimensions).
3. The checks about the file/directory permissions are now more extensive.
4. The native structure in the 'Fraction of native contacts' analysis can now be defined with a PDF file.

Bug fixes

1. Running **grcarma** with 'carma.fitted.dcd' now produces a warning instead of overwriting the original DCD file. This happens because 'carma.fitted.dcd' is a filename reserved by **carma** for internal use.
2. Fixed a plot that manifested when viewing a cluster plot after performing both a dPCA and a cPCA analysis.
3. Atom selection during dPCA analysis with the 'chi1' flag now uses the correct atom types.
4. Various minor bug fixes.

Description

The documentation of grcarma is intended to be more of a tutorial to the program and it's most commonly used features rather than a Unix style manual, listing the various flags and possible parameters(the main reason for that being that grcarma does not accept any flags/parameters). In this section the main features of the program will be described, based on the order they appear in the program's main window. However, before the main window is created the files to be used must be specified.

File selection

Despite the fact that grcarma does not accept any flags, it can still be run from the command line with or without arguments. Two forms of arguments are accepted :

1. The first is a PSF/DCD pair. If the program detects that you have specified only a PSF and not a DCD file, or vice versa, a file that is neither a .psf or a .dcd, or a non matching pair it will terminate with an error message printed to the terminal.
2. The second is a path to a directory. This can be an absolute or relative path. The directory itself should contain the linked PSF/DCD files from a previous grcarma run, otherwise the program will

terminate with a relevant error message.

If the program is run without any files/directories as arguments, a window will pop up prompting the user to select a PSF/DCD pair or a directory, through a graphical interface. In case of non matching files the program will display an error message in a new window and then terminate. Alternatively, the program can be launched by running the executable, or a shortcut pointing to it, through the window manager. Upon successful file selection the main window of the program will come into view. If a directory has been specified the 'View Results' window will be available immediately.

Textbox

The status of the latest analysis is reported in the textbox in colored text. Successful reports are colored green, error reports are colored red, status reports are colored cyan and quick reports of the results of some analyses are colored white. The success reports prompt the user to use a relevant submenu to view the results, and error reports prompt the user to view the latest output of carma, which is stored in a file named **last_carma_run.log** located in the results subfolder. The results subfolder is a folder named **carma_results_DAY.MONTH.YEAR_HOUR.MIN.SEC** which is created every time the user initiates an analysis and is used for all the subsequent analyses, until the program is terminated. It is located in the folder in which the DCD file is located.

Active files

The lower part of the program's main window is taken up by two labels indicating the currently active PSF and DCD files. Should the user decide to perform an analysis with the PSF/DCD files originally specified, a click on the *'Go back to original psf/dcd'* button, will result in abandoning the currently active PSF/DCD pair and using the original ones.

Atom, Chain and residue selection

An atom, chain and residue selection can be specified for most of the analyses that can be performed with grcarma. The buttons that allow the user to perform these selections are included in the upper part of every task window.

Atom type selection

The first menu from the top can be used to specify the selection of atoms for which the desired analysis should be performed :

- **CA** Analysis will be performed only for the CA atoms. This is the default for most tasks.
- **ALLID** Analysis will be performed for all the atoms.
- **HEAVY** Analysis will be performed for all the non-hydrogen atoms.
- **Backbone** Analysis will be performed on **'CA' 'C' 'N' and 'O'** atoms.
- **Custom** Analysis will be performed on the atoms that will be specified by the user.

Chain selection

The second is used for specifying a chain. There are no defaults in this menu as the various checkbuttons will be populated according to the number of chain identifierss read from the .psf file. A usual PSF would contain a chain that will be titled A or another letter, several water molecules and perhaps some ions. This will result in the creation of 3 checkbuttons in this sub-menu, each named after the chain id they

represent.

Residue selection

The last sub-menu can be used to select atoms based on their residue number. The sub-menu consists of two radiobuttons titled **'All'** and **'Change'**. By default the active is the first one. This means that only the atoms specified through the atom type and chain sub-menus are included in the analysis. When the change radiobutton is activated a new window is created prompting the user to specify the span (for example 45 - 234) for which residues will be selected, for a particular chain identifier (eg 'A'). In case more than one selection of residues is required, an additional selection can be specified by clicking the **'Add'** button which creates a bar identical to the first one.

Users should take note that the selections should be ascending. This means that if the required residues belong in two groups, 50-100 and 150-300, the first selection should be specified in the first bar, and the second in the one that is created through the 'Add' button.

Upon submission of the selection a new PSF file is created containing the atoms selected, and is made the active PSF (the PSF file to be used in all subsequent analyses). This can be reversed by clicking on the 'All' radiobutton.

Task selection

The left side of the main window contains the buttons of the various tasks that can be performed :

Fitting

1. The first of these analyses makes use of the `-fit` flag. In this mode global rotations-translations are removed through least-squares fitting of every frame onto the first, or the one specified in the *Use frame as ref* box, creating a new DCD file named carma.fittedN.dcd, where N is the number of fittings that have been performed and thusly the number of fitted DCD files that have been created. Upon completion of the process the user will be prompted to answer whether the newly created DCD file is going to be used in other analyses. If the answer is *yes* the fitted DCD is rendered the active DCD, otherwise no change is made. The file **rms_from_frame_N.dat** that is created in the process contains the rms deviations of each structure versus the reference structure, and can be plotted through the **'View Results'** menu.

2. The second analysis makes use of the flag `-fit` along with the flag `-index`. In this mode the fitting is only performed for a subset of the atoms specified through atom type selection and chain selection, but the resulting DCD contains all the frames determined by those parameters. This mode can be enabled by clicking on the *Use a subset of the residues for the fitting*. The window is split in two parts for this analysis. The upper part is identical to the window of the fitting analysis. This part determines the atoms that will be written out to the fitted .dcd, and the lower part the atoms that will be used for the fitting.

RMSD matrix

This analysis makes use of the `-cross` flag. The matrix which is produced contains the rms deviations (using CA atoms only) between all possible pairs of structures from the trajectory of the DCD file. The matrix is then colored via the command `carma -color - < carma.RMSD.matrix` which results in the creation of a colored postscript representation of the matrix titled **RMSD_matrix.DCD_NAME.ps**. The successful completion of the command results in a status message being posted to the text-box. The

postscipt file is available for viewing through the 'View Results' Button located at the bottom of the left side buttons.

If the R statistical package is installed the 'Determine convergence using Good-Turing statistics' checkbutton will be enabled, otherwise it will be greyed out. If this button is active the Good-Turing convergence analysis will be performed on the RMSD matrix. For more details about what sort of results you can obtain using this procedure, take a look at the github page of the program.

Dihedral PCA

This analysis makes use of the `-dpca` flag. Note that at least one segment identifier must be specified in order to perform this analysis, and that atom type selection has no meaning in this analysis. The primary result of this analysis is the creation of a number of postscript files (default 3) displaying the free energy landscapes on the principal component planes defined by the eigenvector pairs 1-2, 1-3 and 2-3.

By activating the checkbutton titled *'Automatically isolate max ... clusters'* and specifying a number, a cluster analysis will be performed based on the file **dPCA.clusters.dat**. In case the user has specified more clusters than the ones contained in the file, the maximum number of clusters will be extracted. The flag used for this analysis is `-v -w -col -cov -dot -norm -super` and it results in the creation of three .pdb files for each cluster

1. The first is **average.cluster_N.pdb** N being the number of the cluster this .pdb file represents. In this file the average structure of the frames contained in the Nth cluster is presented.

2. The second is **superposition.cluster_N.pdb** and contains superpositions of the frames of the Nth cluster.

3. The third is **representative.cluster_N.pdb** and contains the natural structure with the smallest deviation from the average.

If the program **cluster5D** is included in the PATH, the 'Perform five dimensional clustering' checkbutton will be enabled. If it is activated then in addition to the 3 dimensional clustering procedure performed by **carma** a 5 dimensional clustering procedure will be performed by cluster5D. The files produced by cluster5D can be distinguished from the ones produced by carma by their '5-D' prefix. For more info on the cluster5D program and it's parameters take a look at it's github page.

Cartesian PCA

This analysis makes use of the flag `-proj`. Most of the above apply to the cPCA as well, except for the fact that atom type selections are meaningful for this analysis and that chain selection is not required.

Covariance, average and representative structures

This analysis makes use of the `-cov` and `-super` flags. After the analysis is complete and provided that it is successful the file **carma.rms-average,dat** is read in order to determine which dcd frame contains the natural structure that is closest to the average.

Secondary structure

This analysis is available only for linux/macosx platforms.

This analysis makes use of the flags `-stride` and `-pbd`. If the VMD program *stride* is found in the path then this task is enabled and *stride* is used to perform secondary structure analysis.

Fraction of Native contacts

This analysis makes use of the `-qfraction` flag. In this analysis the native contacts are calculated for residues whose distance (in Angstrom) is smaller than the number specified in *'Distance cutoff'* (default is 8), and which are at least as many residues apart as the number specified in *'Residue separation'* (default 2). The results are in a file named **Qfraction.dcd_name.dat** which can be plotted through the **'View results'** menu.

Distance maps

This analysis makes use of the flag `-rms`. In this analysis two files of interest are produced: **distance_map.average.dcd_name.ps** and **distance_map.RMSD.dcd_name.ps**. The first contains a map of the average values of all specified atom distances for all frames contained in the .dcd file and the second the corresponding root-mean-square deviations from the aforementioned averages. These two .ps files are then merged along their diagonal and a new .ps file named **distance_map.combined.dcd_name.ps** is created.

Solute entropy calculation

The only parameter required for this analysis is the temperature in which the simulation will be performed. In order to study the entropy of the molecule in a meaningful way the number of frames included in the analysis is increased by the number specified in the box *'Step'*. The default value is the number of frames contained in the .dcd header divided by 10. So if the .dcd file contained 2000000 frames and the analysis were to be run with the default values, the first run would be performed for the frames 1-200000, the second for the frames 1-400000 .. 1-2000000. Every time a run is completed successfully the entropy that was calculated by the program is reported in the textbox, and after all the runs have been completed the results are stored in the file **entropy.dat** that can be plotted through the **'View Results'** menu.

Extract PDB

This analysis makes use of the flag `-pdb`. In this analysis a step is specified (default .dcd frames / 10) and every step frames a .pdb file is written out. This .pdb file contains the structure corresponding to the particular frame.

Surface area

Flag `-surface`. In this mode the program will calculate a metric related to the total surface area of the atoms selected, and write the results in a file named **surface.dcd_name.dat**, which can be plotted. A segid selection is required for this analysis.

Radius of gyration, Distances, Bending angles and torsion angles (general)

The flags for these analyses are `-rg`, `-dist`, `-bend` and `-tor` respectively. In these analyses a particular metric of the molecule or a part of it is calculated. Specifically:

1. In the radius of gyration analysis the mass-weighted radius of gyration of the selected atoms is calculated. At least one segid is required for this analysis. The results of this analysis are contained

in the file **Rgyration.dcd_name.dat**.

2. In the distances analysis the program will calculate the distance of two atoms. The results are placed in a file named **distances.dcd_name.dat**.

3. In the bending angles analysis the bending angle between three specified atoms is calculated. The results are contained in the file **bendangles.dcd_name.dat**.

4. In the torsion angles analysis the torsion angles between four specified atoms are calculated, and the results are in the file **torsions.dcd_name.dat**.

All of the result files of the above analyses can be plotted through the **'View Results'** menu.

phi/psi dihedral angles

This task is a particular case of torsion angles calculation meaning that the torsion angles that will be calculated correspond to the phi/psi angles of the molecule and are suitable for viewing in a ramachandran plot.

Results

The results of most of the analyses can be viewed through the *'View Results'* button in the bottom left corner of the main window. Initially the button is grayed out as no files have been produced. After the successful completion of an analysis the user is prompted to use this button to examine the results. Generally the results that can be examined through this menu belong to one of three categories.

1. The first is postscript files ending in .ps. These files are (usually colored) representations of a matrix produced by carma, or an external program such as weblogo. In order to examine these files a postscript viewer must be installed on the system.

   For Linux/MACOSX users a search of the environmental variables of the system is performed and if the variable *GRCARMA_PS_VIEWER* has been set to a program in the path, then that program is used as the postscript viewer, else a search of the path is performed for one of the following programs:
   - evince
   - gv
   - gs
   - display

   in that order, and if one of them is found it is used as the postscript viewer, and a status message is printed in the textbox consisting of *The program selected for .ps viewing is evince/gs/display*. If none of these executables is found, a relevant warning message is printed informing the user that postscript viewing has been disabled.

   For Windows users no search of any folder is performed and no message is printed, due to the fact that windows uses file associations stored in the registry. However, if a program has been associated with the .ps extension then that program will automatically open the postscript file the user has selected through the graphical interface.

2. The second is ASCII formatted text files such as **distances.dcd_name.dat** or **rms_from_average.dat**. These files can be plotted and no external programs are used for viewing

the plots. In order to reduce computation time and speed up the plotting process, if the number of frames included in the DCD header is greater than the width (in pixels) of the screen that the program is being displayed on, a sampling of the values in the various text files will be performed.

3. The last type of files that can be viewed through this menu are .pdb files. Similarly to the .ps files, an external program is required in order to view these files.

   For Linux/MACOSX users a search of the environmental variables of the system is performed and if the variable *GRCARMA_PDB_VIEWER* has been set to a program in the path, then that program is used as the pdb viewer, else a search of the path is performed for one of the following programs:
   - rasmol
   - jmol
   - pymol
   - vmd

   in that order, and if one of them is found it is used as the PDB viewer, and a status message is printed in the textbox consisting of *The program selected for PDB viewing is rasmol*. If none of these executables is found, a relevant warning message is printed informing the user that pdb file viewing has been disabled.

   For Windows users no search of any folder is performed and no message is printed, due to the fact that windows uses file associations stored in the registry. However, if a program has been associated with the PDB extension then that program will automatically open the PDB file the user has selected through the graphical interface.

Version

Version 0.2 Oct 2014

Author

Panagiotis I. Koukos