Research Article

Introducing Thetis: a comprehensive suite for event detection in molecular dynamics

Eleni Picasi[1#], Athanasios Tartas[1#], Vasileios Megalooikonomou[2] and Dimitrios Vlachakis[1,2]

[1]Genetics and Computational Biology Group, Laboratory of Genetics, Department of Biotechnology, Agricultural University of Athens, 75 Iera Odos, 11855, Athens, Greece
[2]Computer Engineering and Informatics Department, School of Engineering, University of Patras, 26500 Patras, Greece
[#]Equal contributions

Received on August 30, 2018; Accepted on September 22, 2018; Published on October 3, 2018

Correspondence should be addressed to Dimitrios Vlachakis; Email: dimvl@aua.gr

Abstract

A suite of computer programs has been developed under the general name Thetis, for monitoring structural changes during molecular dynamics (MD) simulations on proteins. Conformational analysis includes estimation of structural similarities during the simulation and analysis of the secondary structure with emphasis on helices. In contrast to available freeware dealing with MD snapshots, Thetis can be used on a series of consecutive MD structures, thus allowing a detailed conformational analysis over the time course of the simulation.

Introduction

Molecular dynamics (MD) simulations permit the study of complex, dynamic processes that occur in biological systems (Dalkas *et al*. 2013). This computational method calculates the time dependent behaviour of a molecular system. MD simulations can be applied, for example, for the evaluation of protein stability, conformational changes, protein folding, molecular recognition in proteins, DNA, membranes or complexes and ion transport in biological systems (Kabsh *et al*. 1983). Moreover, molecular dynamics provide the mean of carrying out structure refinement studies (X-ray), determination studies (NMR) or even *in silico* drug design experiments (Papageorgiou *et al*. 2016, Vlachakis *et al*. 2013, 2014).

Given that molecular dynamics trajectories are very complicated functions of the proteins and the environment, comparing different trajectories, even under the same conditions, is not straightforward (Vlachakis *et al*. 2013b). Several methods have been suggested that attempt to set the criteria for the evaluation and the comparison of MD trajectories at different levels of complexity (Vlachakis *et al*. 2013b). The simpler methods are geometry based and make use of the root-mean squared deviations between structures, while the more complicated methods are based on the time variation of the various properties of the system during the MD simulation (Vlachakis *et al*. 2014b). A great number of computer programs that focus on the trajectory analysis are available, whereas software for the actual monitoring of conformational changes over the course of the MD simulation in detail is not available (Vlachakis *et al*. 2015).

In this report, we present a new suite of computer programs for monitoring structural changes during molecular dynamics (MD) simulations on proteins with emphasis on helices (Vlachakis *et al*. 2015). Thetis is freeware and offers extensive possibilities for detailed conformational analysis using a series of consecutive coordinate files (MD structures) that are produced by taking successive time-snapshots throughout the process of molecular dynamics (MD) simulations (Vlachakis *et al*. 2017).

Description of the program

Thetis is written in optimised Basic (Liberty Basic v4.0) and was developed initially on a Pentium 4 machine running Windows. The program has also been tested in various UNIX systems running Linux (i386) or IRIX (SGi) with the aid of freeware emulators. There is a multitude of freeware windows emulators for Linux capable of providing an adequate platform for our programs.

The Windows-based version of Thetis

supports multiple processor systems (hyperthreading, dual core or dual CPU).

The programs are provided in standalone versions, compressed in zip format. All the essential libraries and accompanying sub-routines are provided in compressed zip format too. The full size of the suite does not exceed 5 MB in total size.

Thetis is controlled via simple pre-formatted text files through a set of keywords. Parameters used by the program, such as hydrogen bonding distance cut -offs, residue range for analysis and many more can be fully customized in order to meet the needs of the experiment. Thetis is flexible in the type and quantity of output that it generates and can read protein structure coordinate data in PDB or ENT format.

The program can operate in either of two modes, either reading, analyzing, and visualizing structures on an individual basis or automatically batch -processing sets of structures for the large-scale analysis of multiple proteins.

Because, long MD calculations generate enormous amounts of data (e.g. big matrices of data) that need to be loaded onto the computer's RAM, our programs are capable of operating in two modes:

a) Fast implementation: large computer memory is required (where, each coordinate file is accessed once and all information is stored permanently in the computer's memory).

b) Slow implementation: small memory is adequate (where, each coordinate file is accessed many times, and as a result each batch of information is read when needed and straight afterwards dumped from the computer's memory).

Having identified hydrogen bonds, Thetis proceeds to assign individual residues to a secondary structure type by matching observed hydrogen-bonding patterns to those characteristic of ideal secondary structures. The three main types of helical secondary structure analyzed by the program are the a, pi and 3_{10} helices. Following Kabsch-Sander and/or the Ramachandran definitions, these patterns can be described by simple logical conditions expressed in terms of the hydrogen bond connectivity matrix. As soon as the protein system has been loaded and analysed the most representative calculations that can be done with Thetis are:

- Recognition of 3_{10}-, α- and pi- helices using as criterion the H bonding pattern of the polypeptide chain. Contrary to the algorithm of DSSP2 (http:// swift.cmbi.ru.nl/gv/dssp), Thetis presents analytically the score each residue receives (3, 4 or 5 depending on the type of helix) per entry judging whether it belongs in helical formation or not. This is achieved by checking for repetition of overlapping residue windows throughout the sequence. This way it becomes possible to monitor the stability of helices during the course of MD simulations. The output for each MD snapshot entry is recorded in a single txt file.
- Generation of a txt file containing all φ/ψ and ω conformational angles for each amino acid of each MD snapshot entry.

Structural analysis of the α-helical conformation of each residue, according to the criteria of Ramachandran. The α-helical area is restricted to the most favoured regions on the plot as defined in the Procheck program 3, 4 (http:// www.biochem.ucl.ac.uk/~roman/procheck/ procheck.html). The following parameters are calculated for each residue, of each file-snapshot of the molecular dynamics simulation:

→ DaHC (Degree of α-helical Conformation) = $\frac{\sum_{i=1}^{n}(-1)^{a_i}\cdot d_i}{n}$

→ RMS of angular distances from the α-region of Ramachandran = $\sqrt{\frac{\sum_{i=1}^{n} d_i^{\,2}}{n}}$

→ Standard deviation from mean angular distance = $\sqrt{\frac{\sum_{i=1}^{n}\left(d_i-\bar{d}\right)^2}{n}}$

Where,

n = the numbers of files in question

d_i = the angular distance of each residue point from the borders of the nominal α-helical Ramachandran area in entry file i

a_i = odd, if the residue point is located within the Ramachandran α-helical area

a_i = even, if the residue point is located outside the Ramachandran α-helical area

- Ramachandran plots can be custom-drawn for a user-defined set of file entries. The plots can be coloured both in monochrome or colour, according to either the residue IDs or MD snapshot entries. This way, the user can either follow the course of conformational changes for each residue or the overall conformational changes of the protein during the course of the simulation.
- RMSd of the C-alpha can be generated for a user-defined set of residues and time course of the MD simulation. The difference between two structures of the same protein can be evaluated by measuring the relative distance between the corresponding atoms. When contributions from translation and rotation are

subtracted and the absolute values are scaled by the total number of atoms, these distances become the familiar root-mean-squared deviations (RMSD). Generally, the Ca RMSD from the starting structure is reported for MD simulations. For native simulations (typically at 300 K), the Ca RMSD should vary little over time and should not deviate more than a couple of Angstrom units from the starting structure.

- As soon as the Ca RMS distances between the given residue-set have been calculated, an all-by-all pair-wise C^{α} RMSD matrix is subsequently constructed in order to compare the various MD structures between them. A color scale from blue to red was used to depict the variation of the C^{α} RMSD values. In such matrices, clusters of similar structures should appear as squares of approximately the same color, about the diagonal.

Using the above, even the smallest changes that may occur in the α-helical conformation during the course of MD can be detected. All output files produced by the software suite are in common-read TXT file-format, except picture files that are output in versatile .pov file-format.

Thetis is a complete suite for the analysis of the helical conformation of a molecular system over the course of molecular dynamics (MD) simulations. Through Thetis it is now possible to:

- Evaluate the α - / $^{3}/_{10}$ - / pi - helical conformation of a set of multiple consecutive snapshots from molecular dynamics simulations, in the universally established .pdb or .ent file format.
- The evaluation of the phi and psi dihedral angles of each residue from each coordinate file over the full MD course.
- The "mobility" of each residue over the MD course. Here, the term mobility refers to the positional coordinate fluctuations of a given residue during MD, estimated from an interactive Ramachandran plot analysis.
- The α - helical or non α - helical conformation of a given (set of) residue(s) from the "core" region of the Ramachandran plot as defined by Procheck. This calculation can either be performed on a file-by-file basis or over a given set of consecutive snapshot-files outputted from molecular dynamics simulations. The tendency of a given residue to acquire α - helical conformation during the course of a molecular dynamics simulation, is also evaluated by determining the distance of each residue's phi/psi angles from the α – helical "core" region of the Ramachandran plot.
- Graphical representation of the conformational fluctuations of a given set of residues on a Ramachandran plot over the selected course of MD simulation. Colored data points on the Ramachandran plot can either refer to selected residues or vary over time.

The Thetis suite is available as freeware (open -source software). The program's distribution includes documentation, example scripts, and standalone executables for windows (9x, NT, 2k and XP). All copyrights are held by their by their respective owners, unless specifically noted otherwise. All files provided by the Thetis team have been thoroughly tested and examined for viruses using a set of anti-viral applications. It can be downloaded from http:// dimitrislab.com

Installation and Parameterization of Thetis

The Thetis suite comes in a single compressed file in .zip format. For Windows XP® and VISTA® running systems uncompressing and running the executables is automatically supported by the operating system. For earlier versions of windows an uncompressing utility such as WinZIP® or WinRAR® must be pre-installed (links to shareware and freeware versions can be found in our website). All programs are supplied in separate folders accompanied by their parameter files. All downloaded files are provided in RON (read only) format, to prevent alterations of the original files. Users are STRONGLY encouraged to copy and re-attribute the Thetis suite in their current working directories.

MODULE #1: "Helixanalysis"

BINARY: Helixanalysis.exe (binary file)
PARAMETER FILE: elixanalysisparameters.txt (parameter file)
USAGE: <path>/Helixanalysis.exe (param file in same folder)
DESCRIPTION: Determines the helical conformation of a given set of residues within the selected time-frame, and performs a set of statistical calculations.

→ INPUT FILE FORMATTING (remarks showing in *italics*):
 # Load file set: C:\Helix\AlaninTest\
Location of the files to be analysed
 # Filename conserved part: >AlaninHelix_<
Common filename string among candidate files
 # Starting serial number of file data set: 100
The first serial number of the input file set
 # Last serial number of file data set: 130
The last serial number of the input file set
 # Sampling step: 3
The sampling step factor. A sampling factor of 1 will

include all files in the calculation, same way a sampling factor of 2 will take file1, file3, file5, a sampling factor of 5 will include file1, file5,file10 etc.
 # Chain identifier: > <
Applies to .pdb files incorporating more than one chain. Here the chain of interest may be selected, otherwise (if left blank) all chains will be included in the calculation.
 # Output files for each snapshot seperately (Y/N)?: Y
Output format can be either in the form of a single file containing all calculations (choose N) or output per input file format (depending on the number of input files, choose Y).
 # Accepted bond patterns to the $^{3}/_{10}$ helixes: 001, 010, 100, 110, 101, 011, #
All hydrogen bonding combinations for $^{3}/_{10}$ helices. See "Hydrogen-helix scoring" section below.
 # Accepted bond patterns to the α helixes: 0001, 0010, 0100, 1000, 1100, 1010, 1001, 0110, 0101,0011, 1110, 1101, 1011, 0111, #
All hydrogen bonding combinations for α helices.
 # Accepted bond patterns to the p helixes: 10000, 01000, 00100, 00010, 00001, 11000, 10100, 10010, 10001, 01001, 01100, 01010, 00110, 00101, 00011, 11100, 11010, 11001, 10101, 10110, 10011, 01110, 01101, 01011, 00111, 11110 ,11101, 11011, 10111, 01111
All hydrogen bonding combinations for pi helices.

→ OUTPUT FILE FORMATTING:

A. Summary output file.

Filename format:
<path>/conserved filename part_helixR_1-50ST2_sums.txt, where:
conserved filename part is the common filename string among input files.
1 is the number of the first input file (taken from the filename).
50 is the number of the last input file (taken from the filename).
2 is the sampling step selected in the parameter file.
File format:
The following columns will appear in the output .txt file:
AA RN 310 3Su aHel aSu piH pSu Kink Rems, where:

AA: is the serial number of the current residue
RN: 3-letter code of the current residue
310: H-bonds satisfying a $^{3}/_{10}$ helix. 000 means no H-bond, while 111 means all H-bonds are present.
3Su: an H will appear here for those residues that satisfy the criteria used in the parameter file.
aHel: H-bonds satisfying an α helix. 0000 means no H-bond, while 1111 means all H-bonds are present.
aSu: an H will appear here for those residues that satisfy the criteria used in the parameter file.
piH: H-bonds satisfying an α helix. 00000 means no H-bond, while 11111 means all H-bonds are present.
pSu: an H will appear here for those residues that satisfy the criteria used in the parameter file.
Kink: Will identify kinks in the polypeptide chain.
Rems: Remarks

B. Per file analysis output file.

Filename format: <path>/conserved filename part_helixR_1.txt ,where conserved filename part is the common filename string among input files. 1 is the number of the file, being analyzed.
File format: The following columns will appear in the output .txt file:
AA 3H-bond filter% aH-bond filter% pH-bond filter%, where:
AA: is the serial number of the current residue
3H-bond: is the mean of the H-bonds satisfying a $^{3}/_{10}$ helix from all files.
filtro%: % of residues in $^{3}/_{10}$ helical conformation
aH-bond: is the mean of the H-bonds satisfying an α helix from all files.
filtro%: % of residues in α helical conformation
pH-bond: is the mean of the H-bonds satisfying a pi helix from all files.
filtro%: % of residues in pi helical conformation

C. Per residue analysis output file.

Filename format:
<path>/conserved filename part_helixR_1XXX.txt ,where:
conserved filename part is the common filename string among input files.
1 is the number of the file, being analyzed.
XXX is the three-letter code of the residue being analysed (e.g. ALA)
File format:
The following columns will appear in the output .txt file:
File 310 3Su aHel aSu piH pSu Kink Rems, where:

File: is the serial number of the file, being analysed
310: H-bonds satisfying a $^{3}/_{10}$ helix. 000 means no H-bond, while 111 means all H-bonds are present.
3Su: an H will appear here for those residues that satisfy the criteria used in the parameter file.
aHel: H-bonds satisfying an α helix. 0000 means no

H-bond, while 1111 means all H-bonds are present.
aSu: an H will appear here for those residues that satisfy the criteria used in the parameter file.
piH: H-bonds satisfying an α helix. 00000 means no H-bond, while 11111 means all H-bonds are present.
pSu: an H will appear here for those residues that satisfy the criteria used in the parameter file.
Kink: Will identify kinks in the polypeptide chain.
Rems: Remarks

MODULE #2: "Renamer"

BINARY: Renamer.exe (binary file)
PARAMETER FILE: Renamerparameters.txt (parameter file)
USAGE: <path>/ Renamer.exe (param file in same folder)
DESCRIPTION: Batch renaming of a set of files for subsequent analysis with Thetis.

→ INPUT FILE FORMATTING (remarks showing in *italics*):
 # Load file set: C:\Helix\AlaninTest\
Location of the files to be analysed.
 # Filename conserved part: >AlaninHelix_<
Common filename string among candidate files.
 # Filename Variable part: >001<
Variable filename string among candidate files.
 # First File number ID: 001
The serial number of the first file to rename.
 # Last File number ID: 100
The serial number of the last file to rename.
 # Leading Zeros: 3
If the variable filename string contains leading zeroes, enter the full length of the numerical part or else enter "N".
 # Output Filename String: >output_<
A small string that will be added to all filenames upon renaming.
 # Renumber from: 051
For renumbering the output files, enter the number of the first output file (here 051 instead of 001)

MODULE #3: "Dihedralcalc"

BINARY: Dihedralcalc.exe (binary file)
PARAMETER FILE: Dihedralcalcparameters.txt (parameter file)
USAGE: <path>/ Dihedralcalc.exe (param file in same folder)
DESCRIPTION: Program that calculates the Phi, Psi and Omega dihedral angles for a given residue range from multiple coordinate files.

→ INPUT FILE FORMATTING (remarks showing in *italics*):
 # Load file set: C:\Helix\AlaninTest\
Location of the files to be analysed.
 # Filename conserved part: >AlaninHelix_<
Common filename string among candidate files.
 # First File number ID: 001
The serial number of the first file to be analysed.
 # Population: 999
Number of files to be analysed (here 001,003,…,999).

MODULE #4: "Dihedralstats"

BINARY: Dihedralstats.exe (binary file)
PARAMETER FILE: Dihedralstatsparameters.txt (parameter file)
USAGE: <path>/ Dihedralstats.exe (parameter file in same folder)
DESCRIPTION: Program analyses output from Dihedralcalc and calculates the mean distance from Ramachandran plot.
INFO: The Dihedralstats will analyse the output from Dihedralcalc and will calculate the mean distance from the Ramachandran plot "core α-helical" region. A weight file can be used to generate a score of a-helical conformational significance along the MD course. An indicative multiplication factor is the current energy of the system. So, for example: the Dihedralcalc returns a value or 3 Angstroms for ALA001 at time 10 picoseconds when the system energy is 55000 Kcal/ mole and 3 Angstroms for the same residue at time 10 nanoseconds when the system energy has dropped to -25000 Kcal/mole. Since the system's overall energy has dropped the RMSd of that residue should have dropped as well in the 10 ns snapshot. So, ALA001 is conformationally unstable. By applying the formula for a specific snapshot**: $\mathbf{RMSd_{res} * E_{total}^{-1} = CS\ score}$**, where by CS score, we refer to the Conformational Stability of that specific residue at the given time snapshot.

→ INPUT FILE FORMATTING (remarks showing in *italics*):
 # Load file set: C:\Helix\AlaninTest\

Location of the files to be analysed.
 # Filename conserved part: >AlaninHelix_<
Common filename string among candidate files.
 # First file number ID: 001
The serial number of the first file to be analysed.
 # Population: 999
Number of files to be analysed (here 001,003,…,999).
 # Iterative mode (Y/N): N
Select Y (Yes) to produce separate outputs for each residue involved in the calculation, rather than the mean (default: N).
 # Use weight file (Y/N): N
The user can define in a separate file a residue importance weight factor to be used in a scoring function (default: N).
 # Load weight file: C:\Helix\AlaninTest\weight.txt
Location of the weight file.
 # Weight file order (as in weight file: N reverse: Y):N
The weight factor can be either used in the order found in the weight file or reversed.

MODULE #5: "Ramawalk"

BINARY: Ramawalk.exe (binary file)
PARAMETER FILE: Ramawalkparameters.txt (parameter file)
USAGE: <path>/ Ramawalk.exe (parameter file must be in the same folder)
DESCRIPTION: Program that calculates the distance covered by each residue on a Ramachandran plot during the course of MD simulations. The Ramawalk values are indicative of the residue's tendency to acquire α-helical conformation.

→ INPUT FILE FORMATTING (remarks showing in *italics*):
 # Load file set: C:\Helix\AlaninTest\
Location of the files to be analysed.
 # Filename conserved part: >AlaninHelix_<
Common filename string among candidate files.
 # First file number ID: 001
The serial number of the first file to be analysed.
 # Population: 999
Number of files to be analysed (here 001,003,…,999).
 # Use weight file (Y/N): N
The user can define in a separate file a residue importance weight factor to be used in a scoring function (default: N).
 # Load weight file: C:\Helix\AlaninTest\weight.txt
Location of the weight file.
 # Weight file order (as in weight file: N reverse: Y): N
The weight factor can be either used in the order found in the weight file or reversed.

MODULE #6: "Ramaplotter"

BINARY: Ramaplotter.exe (binary file)
PARAMETER FILE: Ramaplotterparameters.txt (parameter file)
USAGE: <path>/ Ramaplotter.exe (param file in same folder)
DESCRIPTION: Program that generates multiple residue or multiple positions (of same residue) on a Ramachandran plot with enhanced display capabilities.

→ INPUT FILE FORMATTING (remarks showing in *italics*):
 # Load file set: C:\Helix\AlaninTest\
Location of the files to be analysed.
 # Filename conserved part: >AlaninHelix_<
Common filename string among candidate files.
 # Consecutive files (S) or random (R): S
If the filenames of the files are in consecutive order choose S, otherwise R.
 # First file number ID: 001
The serial number of the first file to be analysed.
 # Last file number ID: 999
The serial number of the last file to be analysed.
 # Sampling step: 1
The sampling step factor. A sampling factor of 1 will include all files in the calculation, same way a sampling factor of 2 will take file1, file3, file5, a sampling factor of 5 will include file1, file5,file10 etc.
 # Determine the filenames of the Random files: 1, 10, 100
If the filenames of the files to be analysed are not in consecutive order give the variable numerical string of each one separated by a comma (,).
 # Chain identifier of Random files: > <
When using Random rather than consecutive filenames the ID chain must be selected.
 # Consecutive residues (S) or random (R): S
If the residues of interest within each file are in consecutive order choose S (ALA001, GLU002, ARG003), otherwise R (ALA001, ARG003, MET009).
 # First residue number: 001
The serial number of the first residue to be analysed.
 # Last residue number: 100
The serial number of the last residue to be analysed.
 # Determine the residue numbers when using: 1, 10, 100
If the residue numbers of the amino-acids to be analysed are not in consecutive order give the numerical string of each one separated by a comma (,).
 # Residue spot size: 1.5

The size of the spot of each residue on the Ramachandran plot (minimum is 0.8).
 # Background color: W
The background color of the Ramachandran plot. Two choices: White (W) and black (B).
 # Palette color: W
For a palette from red à purple enter (P), whereas for a palette from red à blue (B).
 # Scale up color: P
To scale up from purple or blue to red enter (R), whereas to scale up from red to purple or blue enter (P).
 # Coloring manner: T
To color dots depending on time progress enter (T), whereas to color dots per residue enter (R).
 # Coloring manner scale: A
A for absolute coloring and R for relative coloring.
 # Relative coloring lower value: 12
The number of the file or residue that will be set to have the lower value in the Relative coloring mode.
 # Relative coloring higher value: 1280
The number of the file or residue that will be set to have the higher value in the Relative coloring mode.

Conclusions

All in all, Thetis is a versatile, fast and flexible suite of programs mainly designed not only to evaluate the helical protein conformation, but also to provide the scientist with an extended and comprehensive overview-analysis of a molecular system during the course of MD simulations. Thetis is written in BASIC and is compatible with all native Windows® or Windows® – emulator equipped computers.

Conflicts of interest

The authors declare no conflicts of interest.

References

Kabsch W & Sander C 1983 Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen -Bonded and Geometrical Features. *Biopolymers* **22** 2577

Dalkas GA, Vlachakis D, Tsagkrasoulis D, Kastania A & Kossida S 2013 State-of-the-art technology in modern computer-aided drug design. *Brief Bioinform,* **14** 745-752

Vlachakis D, Armaos A & Kossida S 2017 Advanced Protein Alignments Based on Sequence, Structure and Hydropathy Profiles; The Paradigm of the Viral Polymerase Enzyme. *Mathematics in Computer Science* **11** 197-208

Vlachakis D, Bencurova E, Papangelopoulos N & Kossida S 2014b Current state-of-the-art molecular dynamics methods and applications. *Adv Protein Chem Struct Biol* **94** 269-313

Vlachakis D, Champeris Tsaniras S, Feidakis C & Kossida S 2013 Molecular modelling study of the 3D structure of the biglycan core protein, using homology modelling techniques. *J Mol Biochem* **2** 85-93

Vlachakis D, Champeris Tsaniras S, Ioannidou K, Papageorgiou L, Baumann M & Kossida S 2014 A series of Notch3 mutations in CADASIL; insights from 3D molecular modelling and evolutionary analyses. *J Mol Biochem* **3** 97-105

Vlachakis D, Fakourelis P, Megalooikonomou V, Makris C & Kossida S 2015 DrugOn: a fully integrated pharmacophore modeling and structure optimization toolkit. *PeerJ* **3** e725

Vlachakis D, Tsagrasoulis D, Megalooikonomou V & Kossida S 2013b Introducing Drugster: a comprehensive and fully integrated drug design, lead and structure optimization toolkit. *Bioinformatics* **29** 126-128

Papageorgiou L, Loukatou S, Sofia K, Maroulis D & Vlachakis D 2016 An updated evolutionary study of Flaviviridae NS3 helicase and NS5 RNA-dependent RNA polymerase reveals novel invariable motifs as potential pharmacological targets. *Mol Biosyst* **12** 2080 -2093