

Supporting Information

This Supporting Information contains 1 figure (Figure S1) and an appendix. The appendix contains the source code of the Python script to produce Figure S1.

Extracting Dynamical Correlations and Identifying Key Residues for Allosteric Communication in Proteins by *correlationplus*

Mustafa Tekpinar^{1*}, *Bertrand Neron*², *Marc Delarue*^{1*}

¹Unit of Architecture and Dynamics of Biological Macromolecules, Pasteur Institute, UMR 3528 CNRS, 25 Rue du Dr. Roux, 75015 Paris, France.

²Bioinformatics and Biostatistics Hub, Computational Biology Department, 28 Rue du Dr. Roux, 75015 Paris, France.

KEYWORDS. Allostery, coevolution, coupling, dynamical cross-correlation, elastic network model, linear mutual information, molecular dynamics simulations.

SUPPORTING FIGURES.

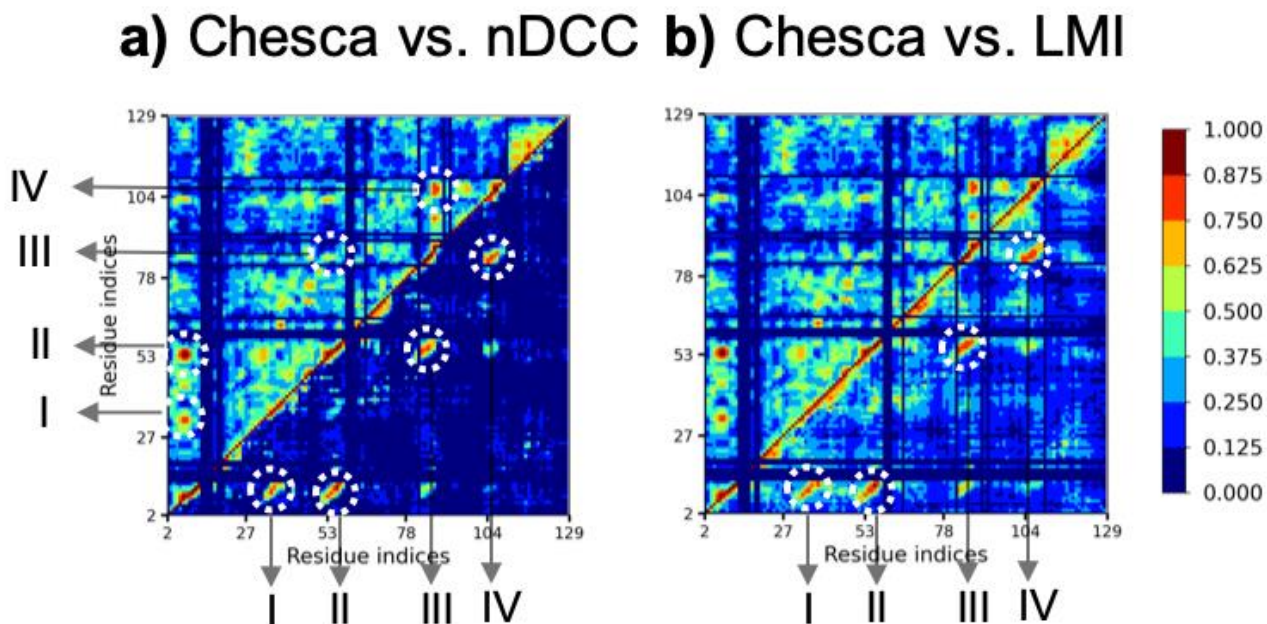


Figure S1. Absolute values of Chesca NMR correlations (upper triangle) vs computed correlations (lower triangle) for Chemotactic Y protein (PDB ID: 1f4v¹ Chain A). **a)** Absolute values of Chesca NMR vs. absolute values of nDCC. **b)** Absolute values of Chesca NMR vs. normalized LMI. Three out of four major peaks in LMI and nDCC matrices (red and dark red regions labeled with Roman numerals from I to IV) have corresponding high correlations in the experimental data.

Details and analysis of Figure S1

Chesca NMR values were obtained from Supplementary Data 2 of Wang et. al.² and it was processed as described by Boulton et. al.³. The Chesca NMR values given in the figure are Gaussian filtered as recommended by Wang et al.². Residues 1, 2, 13, 14, 15, 16, 18, 19, 59, 60, 61, 65, 82, 90, 92, and 110 were missing in the original data set. Therefore, they have been removed from the computed data. Normalized nDCC and LMI matrices were obtained with anisotropic network model⁴. Cutoff distance was 15 Å and only the first 100 normal modes were used for the computations of the correlation matrices⁵.

Computed high correlation regions I, II and IV have corresponding high correlations in Chesca data. Region III has high correlations in nDCC and LMI matrices. However, the Chesca matrix show moderate correlations in this region. Altogether, three out of the four computed high correlations regions have a corresponding region in the Chesca experimental map.

APPENDIX: Source code of the Python script to produce Figure S1

```
import numpy as np
from scipy.ndimage import gaussian_filter
from prody import parsePDB
from correlationplus.visualize import *
from correlationplus.calculate import calcENM_LMI
from correlationplus.calculate import calcENMnDCC

#Parse Chesca data
chescaMatrix = np.loadtxt('dm.dat')

#Parse the CheY pdb file
selectedAtoms = parsePDB('1f4v_chA.pdb', subset='ca')

#Calculate LMI matrix
lmiMatrix = calcENM_LMI(selectedAtoms, 15, method="ANM", \
                        nmodes=100, normalized=True, \
                        saveMatrix=False)

#Zero lmiMatrix data points that doesn't exist in the experimental data.
lmiMatrix[0,:] = lmiMatrix[:,0] = 0.0
lmiMatrix[11,:] = lmiMatrix[:,11] = 0.0
lmiMatrix[12,:] = lmiMatrix[:,12] = 0.0
lmiMatrix[13,:] = lmiMatrix[:,13] = 0.0
lmiMatrix[14,:] = lmiMatrix[:,14] = 0.0
lmiMatrix[16,:] = lmiMatrix[:,16] = 0.0
lmiMatrix[17,:] = lmiMatrix[:,17] = 0.0
lmiMatrix[57,:] = lmiMatrix[:,57] = 0.0
lmiMatrix[58,:] = lmiMatrix[:,58] = 0.0
lmiMatrix[59,:] = lmiMatrix[:,59] = 0.0
lmiMatrix[63,:] = lmiMatrix[:,63] = 0.0
lmiMatrix[80,:] = lmiMatrix[:,80] = 0.0
lmiMatrix[88,:] = lmiMatrix[:,88] = 0.0
lmiMatrix[90,:] = lmiMatrix[:,90] = 0.0
lmiMatrix[108,:] = lmiMatrix[:,108] = 0.0

#Calculate absolute values of normalized DCC matrix
dccMatrix = np.absolute(calcENMnDCC(selectedAtoms, 15, \
                                   method="ANM", \
                                   nmodes=100, \
                                   normalized=True, \
                                   saveMatrix=False))

#Zero dccMatrix data points that doesn't exist in the experimental data.
dccMatrix[0,:] = dccMatrix[:,0] = 0.0
dccMatrix[11,:] = dccMatrix[:,11] = 0.0
```

```

dccMatrix[12,:] = dccMatrix[:,12] = 0.0
dccMatrix[13,:] = dccMatrix[:,13] = 0.0
dccMatrix[14,:] = dccMatrix[:,14] = 0.0
dccMatrix[16,:] = dccMatrix[:,16] = 0.0
dccMatrix[17,:] = dccMatrix[:,17] = 0.0
dccMatrix[57,:] = dccMatrix[:,57] = 0.0
dccMatrix[58,:] = dccMatrix[:,58] = 0.0
dccMatrix[59,:] = dccMatrix[:,59] = 0.0
dccMatrix[63,:] = dccMatrix[:,63] = 0.0
dccMatrix[80,:] = dccMatrix[:,80] = 0.0
dccMatrix[88,:] = dccMatrix[:,88] = 0.0
dccMatrix[90,:] = dccMatrix[:,90] = 0.0
dccMatrix[108,:] = dccMatrix[:,108] = 0.0

```

#Apply a Gaussian filter to the experimental data

#as recommended in <https://doi.org/10.1038/s41467-020-17618-2>

```
chescaFiltered = gaussian_filter(np.absolute(chescaMatrix), sigma=0.75)
```

#Gaussian filtering causes non-zero values at points where they must be zero.

#Let's correct it!

```

chescaFiltered[0,:] = chescaFiltered[:,0] = 0.0
chescaFiltered[11,:] = chescaFiltered[:,11] = 0.0
chescaFiltered[12,:] = chescaFiltered[:,12] = 0.0
chescaFiltered[13,:] = chescaFiltered[:,13] = 0.0
chescaFiltered[14,:] = chescaFiltered[:,14] = 0.0
chescaFiltered[16,:] = chescaFiltered[:,16] = 0.0
chescaFiltered[17,:] = chescaFiltered[:,17] = 0.0
chescaFiltered[57,:] = chescaFiltered[:,57] = 0.0
chescaFiltered[58,:] = chescaFiltered[:,58] = 0.0
chescaFiltered[59,:] = chescaFiltered[:,59] = 0.0
chescaFiltered[63,:] = chescaFiltered[:,63] = 0.0
chescaFiltered[80,:] = chescaFiltered[:,80] = 0.0
chescaFiltered[88,:] = chescaFiltered[:,88] = 0.0
chescaFiltered[90,:] = chescaFiltered[:,90] = 0.0
chescaFiltered[108,:] = chescaFiltered[:,108] = 0.0

```

```
ImiMatrixCombined = np.triu(ImiMatrix, k=0) + np.tril(chescaFiltered, k=0)
```

```
dccMatrixCombined = np.triu(dccMatrix, k=0) + np.tril(chescaFiltered, k=0)
```

#Use correlationplus to produce the figures as png files

```

intraChainCorrelationMaps(ImiMatrixCombined, 0.0, 1.0, \
    "chesca_vs_Imi", "Chesca vs LMI", \
    selectedAtoms, False)

```

```

intraChainCorrelationMaps(dccMatrixCombined, 0.0, 1.0, \
    "chesca_vs_dcc", "Chesca vs nDCC", \
    selectedAtoms, False)

```

AUTHOR INFORMATION

Corresponding Authors

Mustafa Tekpinar, Email: tekpinar@buffalo.edu; Marc Delarue, Email: delarue@pasteur.fr.

Phone: +33 01 40 61 31 11

ACKNOWLEDGMENT

MT thanks to *Programme PAUSE* of *College de France* for their support.

ABBREVIATIONS

Anisotropic network model: ANM, Coarse-grained: CG, Dynamical Cross-Correlations: DCC, Elastic Network Model: ENM, Gaussian network model: GNM, Linear Mutual Information: LMI.

REFERENCES

- (1) Lee, S. Y.; Cho, H. S.; Pelton, J. G.; Yan, D. L.; Henderson, R. K.; King, D. S.; Huang, L. S.; Kustu, S.; Berry, E. A.; Wemmer, D. E. Crystal structure of an activated response regulator bound to its target. *Nat. Struct. Biol.* **2001**, *8* (1), 52.
- (2) Wang, J.; Jain, A.; McDonald, L. R.; Gambogi, C.; Lee, A. L.; Dokholyan, N. V. Mapping allosteric communications within individual proteins. *Nat Commun* **2020**, *11* (1), 3862.
- (3) Boulton, S.; Selvaratnam, R.; Ahmed, R.; Melacini, G. Implementation of the NMR CHEMical Shift Covariance Analysis (CHESCA): A Chemical Biologist's Approach to Allostery. *Methods Mol Biol* **2018**, *1688*, 391.
- (4) Atilgan, A. R.; Durell, S. R.; Jernigan, R. L.; Demirel, M. C.; Keskin, O.; Bahar, I. Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys J* **2001**, *80* (1), 505.
- (5) Tekpinar, M.; Yildirim, A. Only a Subset of Normal Modes is Sufficient to Identify Linear Correlations in Proteins. *J Chem Inf Model* **2018**, *58* (9), 1947.