Biophysical Journal, Volume 98

Supporting Material

Predicting Interaction Sites from the Energetics of Isolated Proteins: A New Approach to Epitope Mapping

Guido Scarabelli, Giulia Morra, and Giorgio Colombo

Predicting Interaction Sites from the Energetics of Isolated Proteins: A New Approach to Epitope Mapping

Guido Scarabelli, Giulia Morra, and Giorgio Colombo*

Istituto di Chimica del Riconoscimento Molecolare, CNR, Via Mario Bianco 9, 20131 Milano, Italy

Molecular Dynamics Simulation Runs.

All the starting structures of the proteins were downloaded from the Protein Data Bank, their respective codes are reported in Table I and subjected to explicit water MD simulations. The structures of the isolated antigens were solvated with explicit water, using the SPC water model {Berendsen, 1987 #309} in triclinic or cubic boxes (depending on the protein shape) large enough to contain the whole protein and 1.4 nm of solvent from the protein. The total charge of each system was neutralized with suitable counter ions. The systems were subsequently optimized through a molecular mechanics process with the steepest descent algorithm (2000 steps).

The MD simulations were started from the minimized systems and carried out using GROMACS {van der Spoel, 2004 #2708}, with GROMOSS96 43A1 force field {Scott, 1999 #2357}. The LINCS algorithm {Hess, 1997 #1425} was used to constrain the bond lengths for the all atoms. The electrostatic interactions were calculated through PME implementation of Ewald summation method. The temperature was set to 300K and kept constant with the Berendsen thermostat {Berendsen, 1984 #157} with a coupling constant of 0.1 ps. The timestep used was 2 fs. 5 simulations of 30 ns each with different random initial velocities were run for each protein, to check the dependence of the results on the simulation conditions. The first 5 ns of each simulation were discarded from the final analysis. The average simulation time using a parallel calculation on 32 Intel Xeon 3.166GHz cores for a protein of 200 residues is around 24 hours.

References.

1. Berendsen, H. J. C., J. R. Grigera, and P. R. Straatsma. 1987. The missing term in effective pair potentials. J. Phys. Chem. 91:6269-6271.
2. van der Spoel, D., E. Lindahl, B. Hess, A. R. van Buuren, E. Apol, P. J. Meulenhoff, D. P. Tieleman, A. L. T. M. Sijbers, K. A. Feenstra, R. van Drunen, and H. J. C. Berendsen. 2004. Gromacs User Manual version 3.2. www.gromacs.org.
3. Scott, W. R. P., P. H. Hunenberger, I. G. Tironi, A. E. Mark, S. R. Billeter, J. Fennen, A. E. Torda, T. Huber, P. Kruger, and W. F. V. Gunsteren. 1999. TheGROMOS biomolecular simulation program package. J.Phys.Chem.A 103:3596-3607.
4. Hess, B., H. Bekker, J. G. E. M. Fraaije, and H. J. C. Berendsen. 1997. A linear constraint solver for molecular simulations. J.Comp.Chem. 18:1463-1472.
5. Berendsen, H. J. C., J. P. M. Postma, W. F. van Gunsteren, A. Di Nola, and J. R. Haak. 1984. Molecular dynamics with coupling to an external bath. J. Chem. Phys. 81:3684-3690.

Table S1. PDB codes of the isolated antigen, of its complex with the antibody and the biological role. In parentheses, we have reported the resolution of the crystals from which the antigens structures were obtained.

Antigen	Antigen-Antibody complexes	Biological Role
1AO3 (2.2)	1FE8 (2.03); 2ADF (2.90)	von Willerbrand factor domain A3
1AUQ (2.3)	1OAK (2.20)	von Willerbrand factor domain A1
1BV1 (2.0)	1FSK (2.90)	Major pollen allergen Bet V 1-A
1CK4 (2.2)	1MHP (2.80)	alpha I beta 1 Integrin I-domain
1CMW (2.6)	1BGX (2.30)	Taq DNA polymerase I
1D7P (1.5)	1IQD (2.00)	Coagulation factor VIII precursor
1GWP (NMR)	1AFV (3.70)	Gag polyprotein
1HCN (2.6)	1QFW (1.36)	Human chronic gonadotropin
1K59 (1.8)	1H0D (2.0)	Angiogenin
1KDC (2.00)	1NSN (2.8)	Staphilococcal nuclease
1KZQ (1.7)	1YNT (3.1)	Major Surface Antigen SAG1
1P4P (2.0)	1RJL (2.6)	Outer surface protein B
1PKO (1.45)	1PKQ (3.0)	Myelin olygodendrocyte glycoprotein
1POH (2.0)	2JEL (2.5)	Histidine containing phosphocarrier protein
1TFH (2.4)	1AHW (2.5)	Human tissue factor
1UW3 (2.05)	1TPX (2.56)	Prion protein
2VPF (2.5)	1BJ1 (2.4), 2FJG (2.8)	Vascular endothelial grow factor
3LZT [2] (0.93)	1FDL (2.5), 1YQV (1.7), 1MLC (2.5), 1IC4 (2.5), 1NDG (1.9), 1DQJ (2.00), 1NDM (2.1),	Lysozyme
7NN9 (2.0)	1NCA (2.5)	Neuroamidase N9

Table S2. Distance between the first and second eigenvectors, average distance between all other eigenvectors for simulated proteins, percentage of the stabilization energy accounted for by the first eigenvalue and corresponding eigenvector.

Protein	$\Delta\lambda_{1\text{-}2}$	$\Delta\lambda_{average}$	$\lambda_1 w_i^1 w_j^1$ fraction of E_{nb}(%)
1AO3	2.412	0.045	74.7
1AUQ	1.597	0.060	89.3
1BV1	2.550	0.069	68.0
1CK4	2.709	0.054	58.0
1CMW	2.07	0.051	51.0
1D7P	2.617	0.070	65.8
1GWP	3.088	0.042	73.4
1HCN	1.409	0.031	47.0
1K59	2.981	0.077	70.0
1KDC	3.685	0.083	69.1
1KZQ	2.851	0.067	66.0
1P4P	3.038	0.056	56.7
1PKO	2.665	0.085	61.2
1POH	4.893	0.089	64.6
1TFH	1.187	0.055	44.6
1UW3	1.893	0.116	65.1
2VPF	1.358	0.111	73.7
3LZT	2.971	0.071	72.1
7NN9	1.987	0.022	62.6

Table S3: Sequences of the predicted epitopes for each of the proteins.

1AO3	
Sequences	97-TSEMHG-102 104-RPG-106 124-DAAADAARSNRV-135 137-V 152-RILAGPAGDSNV-163 171-D 173-PTMVT-177

1AUQ	
Sequences	559-HDG-561 585-KYAGSQV-591 602-LFQ-604 607-SKIDRPE-613 626-EPQRMSRN-633 636-RYVQGLKKKKV-646 652-GIG-654 668-QAPEN-672

1BV1	
Sequences	14-PA-15 17-R 31-PKVAPQ-36 46-GNGGPG-51 58-F 60-EGLPFKYV-67 88-GGP-90 107-TPDGG-111 113-IL-114 124-GDH-126

1CK4	
Sequences	156-SNSIYPWESVIA-167 191-GEN-193 212-AANK-215 218-RQGGLQT-224 260-SHDNY-264 286-LGHYNRGNLSTEK-298 319-SDELALVT-326

1CMW	
Sequences	30-LKGLTTSRGEP-40 87-PTP-89 193-LPGVKGIGEKTARK-206 221-LDRLKPAI-228 230-EKI-232 234-AH-235 237-DDL-239 264-PDRERLRA-271 273-LERLEFGSL-281 335-APEPYKA-341 364-GLGLP-368 381-DPSNTTPEGVARRYGG-396 421-EGEER-425 447-TGVR-450 452-DVAYLRALSLEVA-464 546-IDPLPD-551 585-PVR-587 648-GVPREAVDPLMRR-660 674-SAHRLSQELAIPYEEAQAF-692 713-EGRRRG-718 729-YVPDLEARVKSVREAAE-745

1D7P	
Sequences	2197-TNMFA-2201 2221-PQVNNP-2226 2236-KT-2237 2239-K 2241-T 2249-KSLLT-2253 2274-FFQNGKVKV-2282 2288-D 2290-FTP-2292 2294-V 2296-CLDPPL-2301

1GWP	
Sequences	84-HPVHAGPIAPGQMREPRG-101 119-THNPPI-1124

1HCN	
Sequences	<u>Chain A</u> 15-NPFFSQPGAPI-25 69-TVMGGFK-75 <u>Chain B</u> 22-GCPVC-26 49-LPA-51 67-IRLPGCP-73 75-GVNPVV-80

1K59	
Sequences	36-TSPC-39 63-NPHRENL-69 83-LHGGSPWPPCQ-93 108-ENGLP-112

1KDC	
Sequences	26-M 28-KGQPM-32 41-TPETKHPKKGVEKYGPEASA-60 84-KYG-86

1KZQ

Sequences	13-PDKKST-18 34-CPKTA-38 55-PAGT-58 61-SCTS-64 102F 104-VT-105 115-KGDDAQ-120 122-C 129-QAR-131 135-VVN-137 144-YGA-146 156-SAEGPT-161 167-GKDG-171 199-LPKLTEN-205 214-DKG-216 222-KKEAF-226 239-GGSPEK-244

1P4P	
Sequences	228-DTAGSNKK-235 250-ADSKK-254 269-QQYNTAGTSLEGSA-282

1PKO	
Sequences	27-SPGKNATG-34 76-IGEGKV-81 102-DHSY-105 107-E

1POH	
Sequences	11-PNG-13 15-HTRP18 55-LTQG-58

1TFH	
Sequences	81-G 83-VA-84 111-PTI-113 116-FEQVGTK-122 154-T 156-YYWKSSSSGKKTAKT-170 179-VDKGENYCFS-188

1UW3	
Sequences	162-Y 164-RP-165 167-DQ-168 188-TVTTTTKG-195

2VPF	
Sequences	23-RSYCH-27 59-GGCCNDEGLE-67

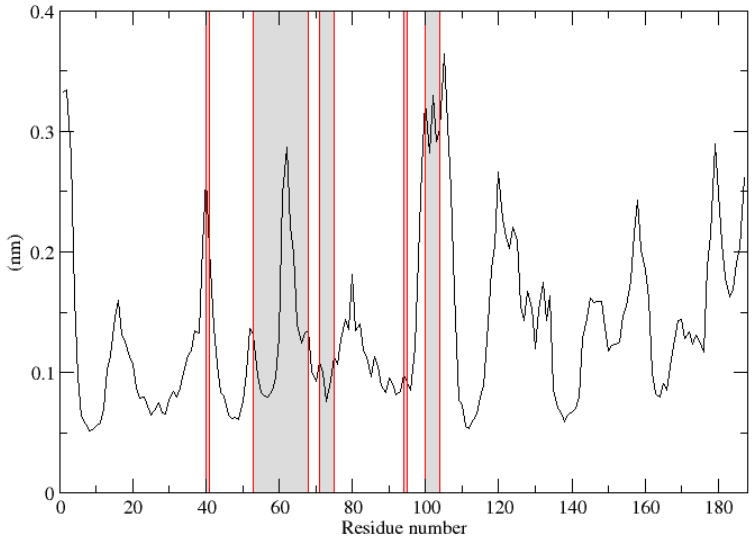
3LZT	
Sequences	11-A 13-KRHG-16 66-DGRTPGS-72 114-RCKGTD-119

7NN9	
Sequences	105-AVRIGEDSDV-114 138-TTIRGKHSNGTI-149 165-SPPT-168 247-TGP-249 269-AGT-271 307-PVA-309 318-CSPV-321 327-RPNDPTVGKCNDPYPG-343 356-DGVN-359 377-L 379-VPNALTDDKSKPTQ-392 451-TEFLGQWDW-459

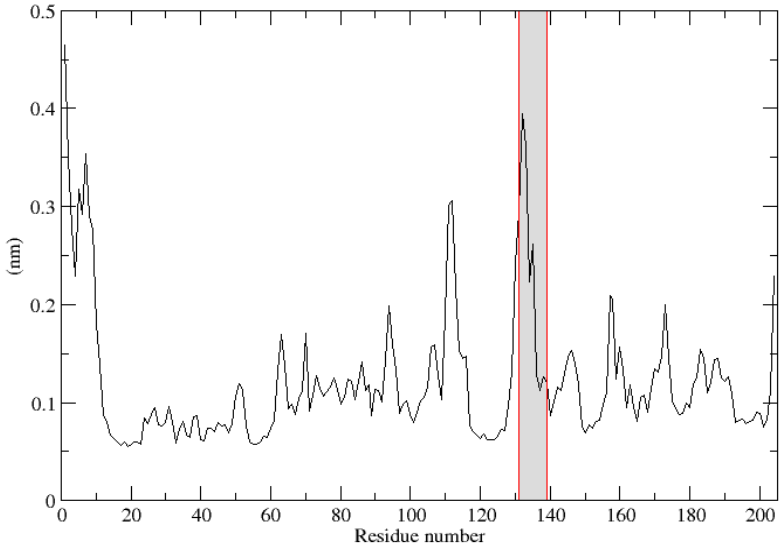
Flexibility of the proteins analyzed from MD simulations.

Protein flexibility was characterized in terms of Root Mean Square Fluctuations (RMSF) around the average structure obtained from MD simulations. Epitope residues are highlighted by gray shading. It is immediately apparent that flexibility alone is not a sufficient criterion to discriminate epitopes or protein interaction sites from other regions.

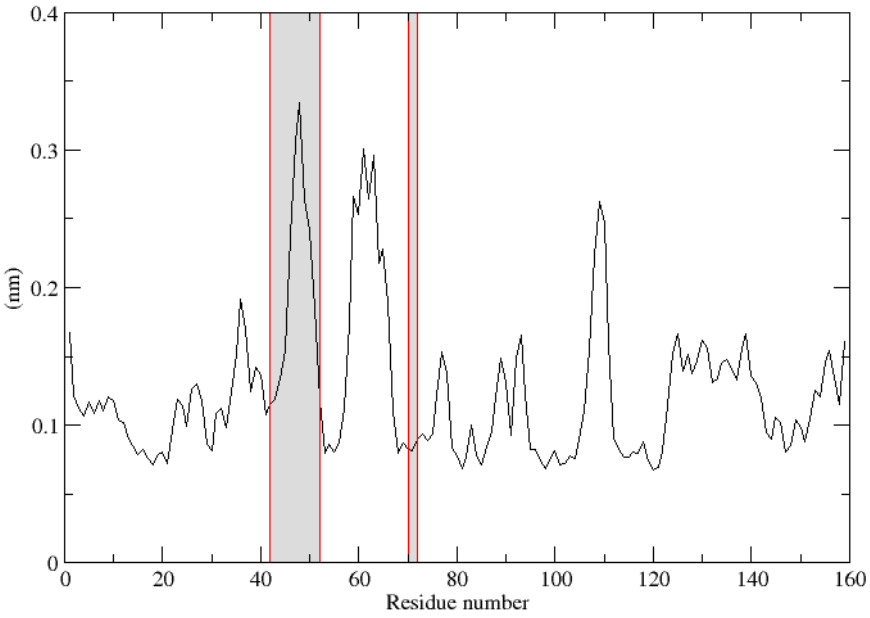
1AO3



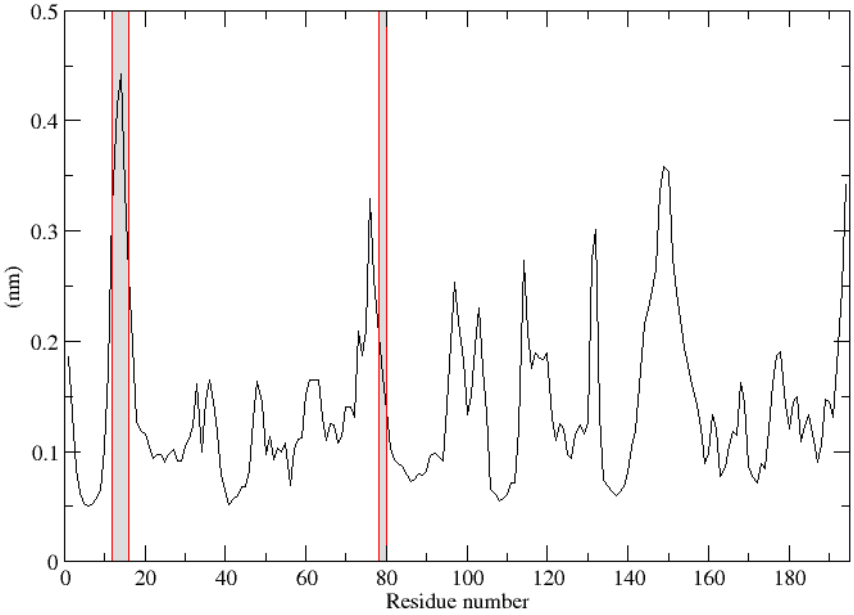
1AUQ



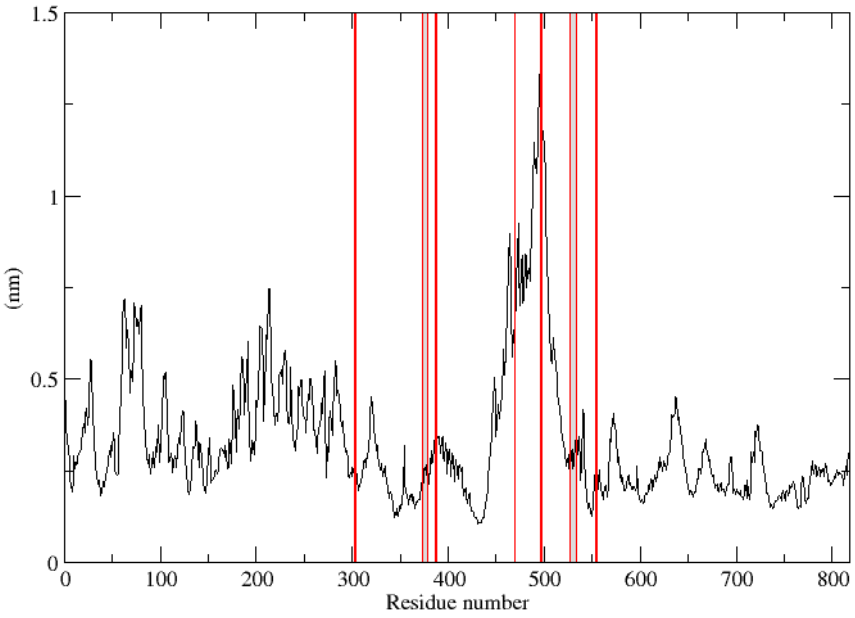
1BV1


0.4
0.3
0.2
0.1
0
(nm)
0
20
40
60
80
100
120
140
160
Residue number

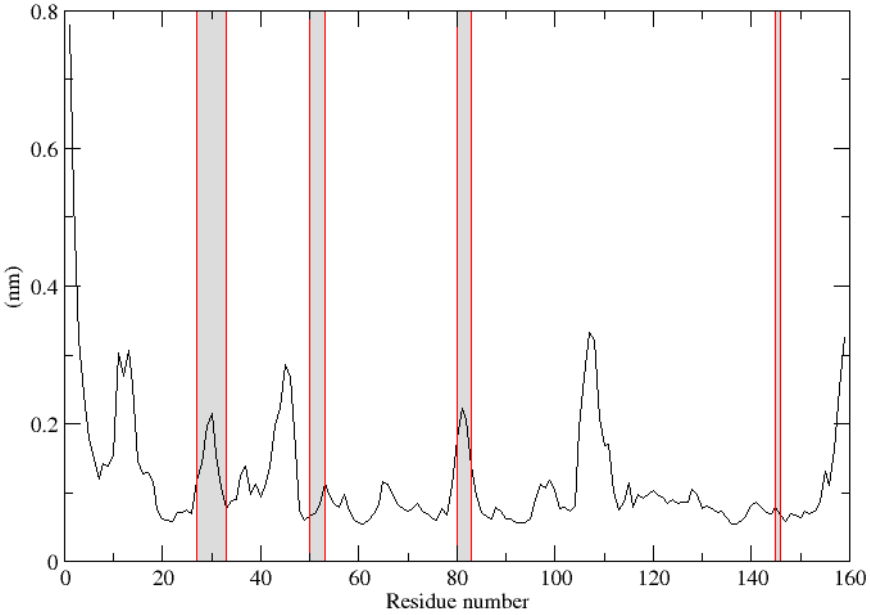
1CK4


0.5
0.4
0.3
0.2
0.1
0
(nm)
0
20
40
60
80
100
120
140
160
180
Residue number

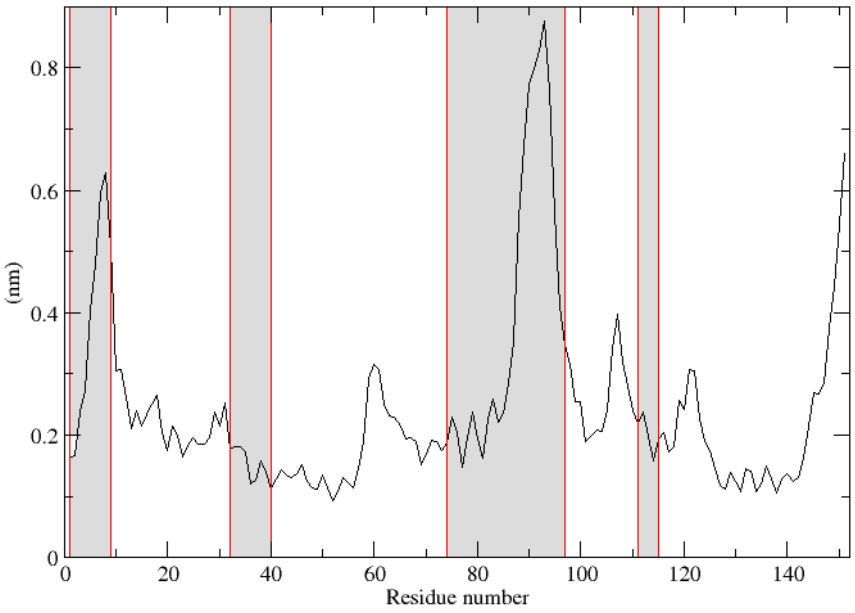
1CMW


1.5
1
0.5
0
(nm)
0
100
200
300
400
500
600
700
800
Residue number

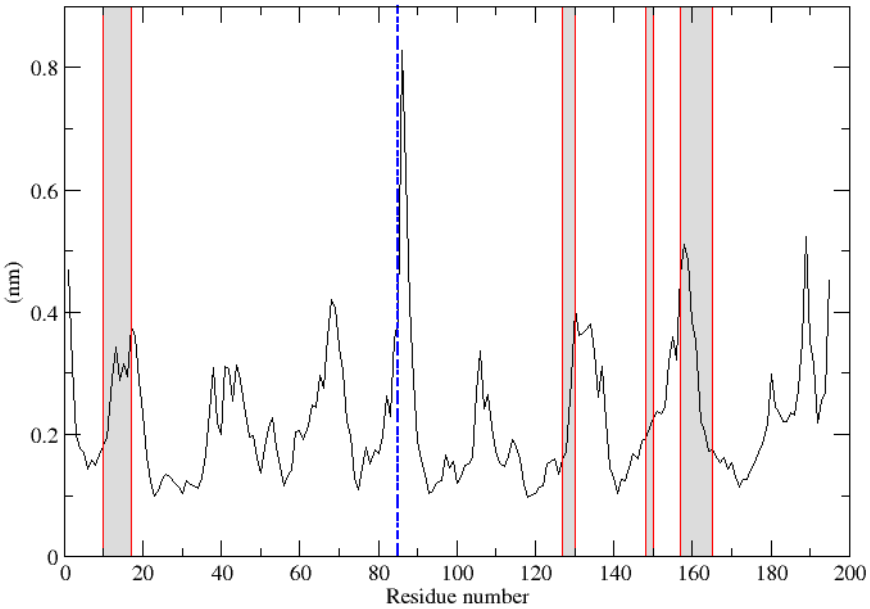
1D7P


0.8
0.6
0.4
0.2
0
(nm)
0
20
40
60
80
100
120
140
160
Residue number

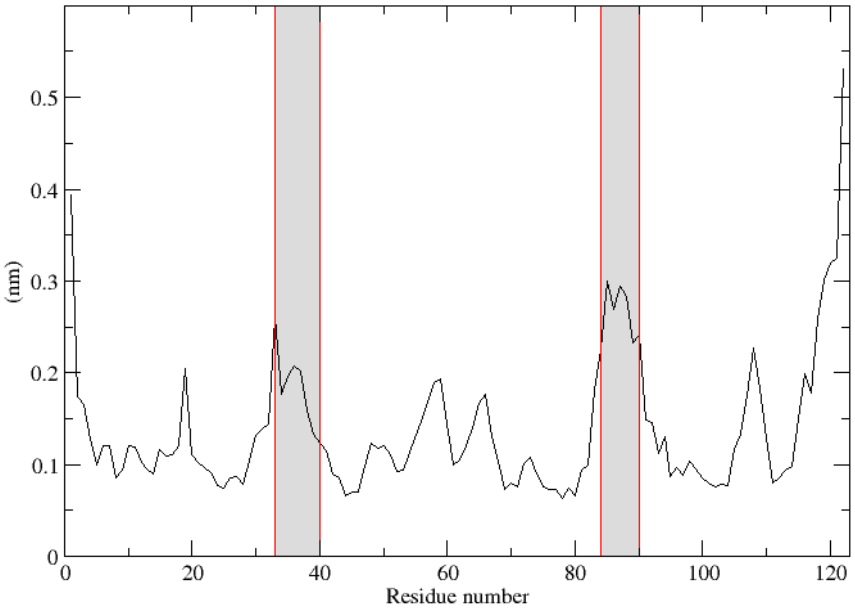
1GWP


0.8
0.6
0.4
0.2
0
(nm)
0
20
40
60
80
100
120
140
Residue number

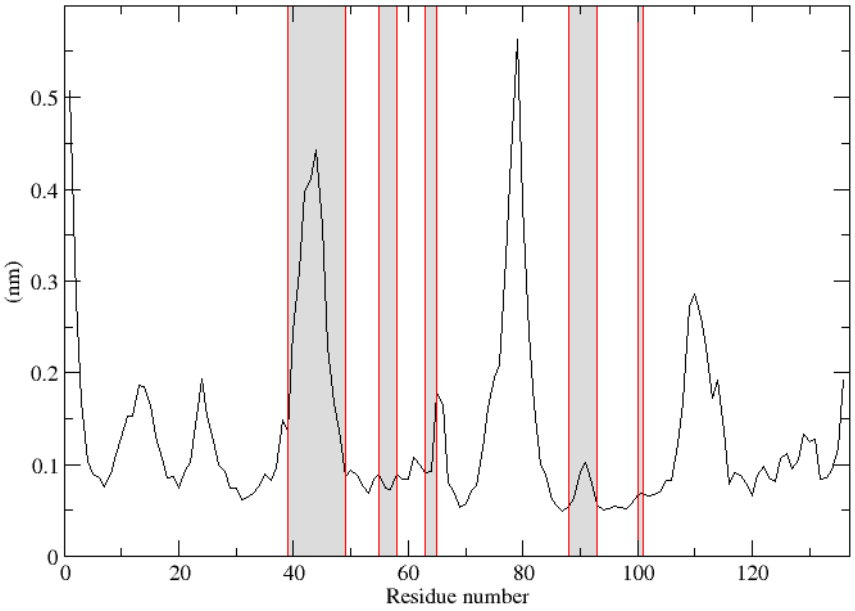
1HCN


0.8
0.6
0.4
0.2
0
(nm)
0
20
40
60
80
100
120
140
160
180
200
Residue number

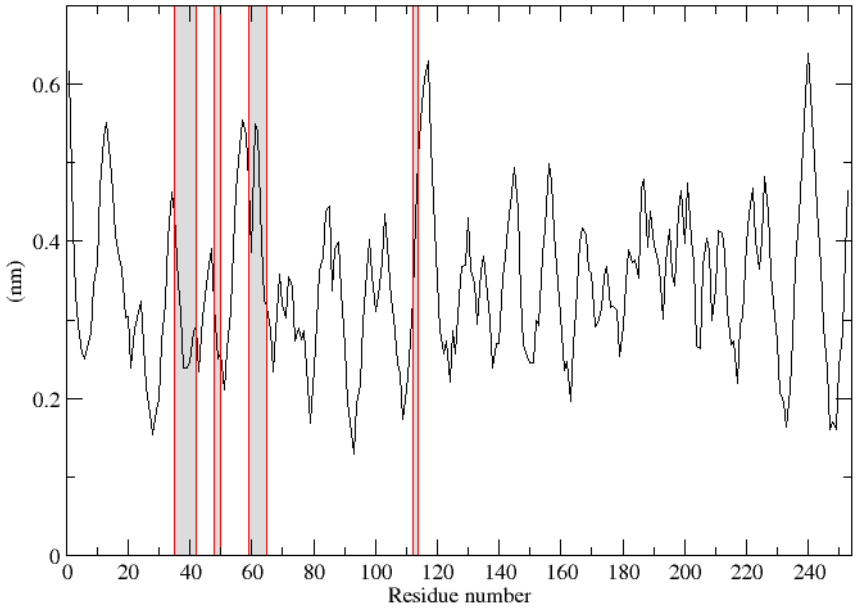
1K59


(nm)
0
0.1
0.2
0.3
0.4
0.5
0
20
40
60
80
100
120
Residue number

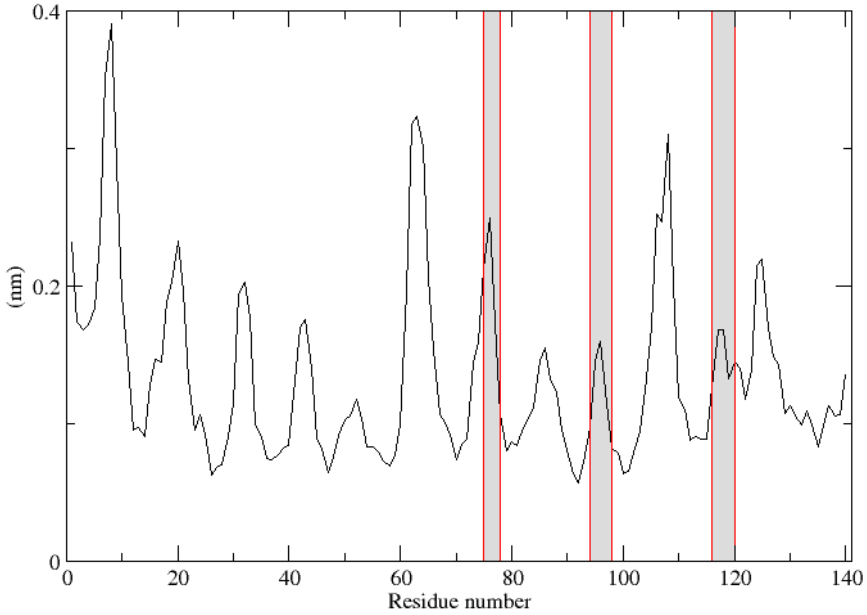
1KDC


(nm)
0
0.1
0.2
0.3
0.4
0.5
0
20
40
60
80
100
120
Residue number

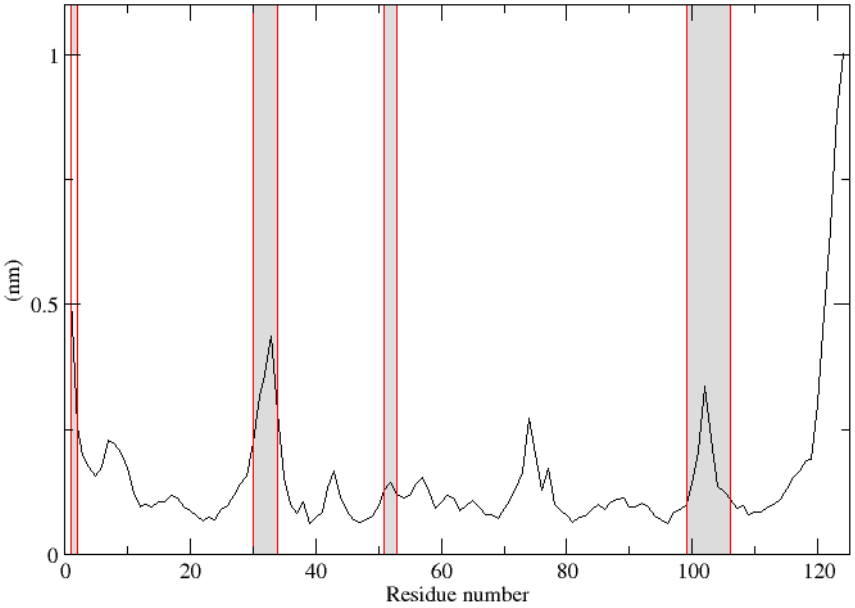
1KZQ


0.6
0.4
0.2
0
(nm)
0
20
40
60
80
100
120
140
160
180
200
220
240
Residue number

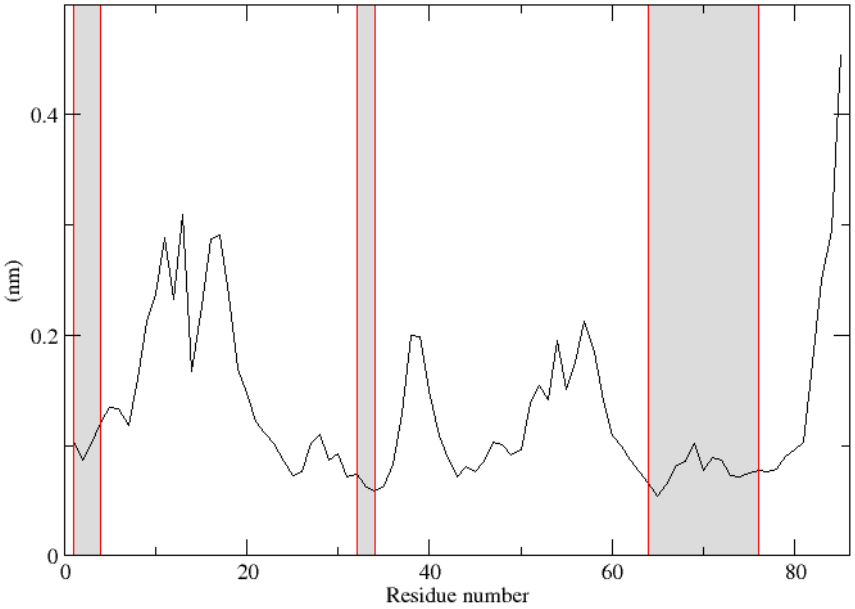
1P4P


0.4
0.2
0
(nm)
0
20
40
60
80
100
120
140
Residue number

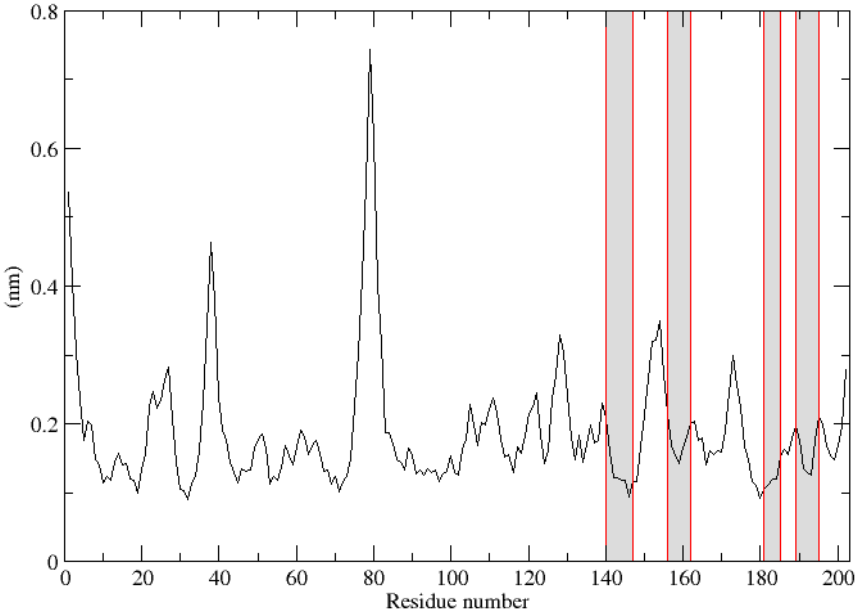
1PKO


1
0.5
0
(nm)
0
20
40
60
80
100
120
Residue number

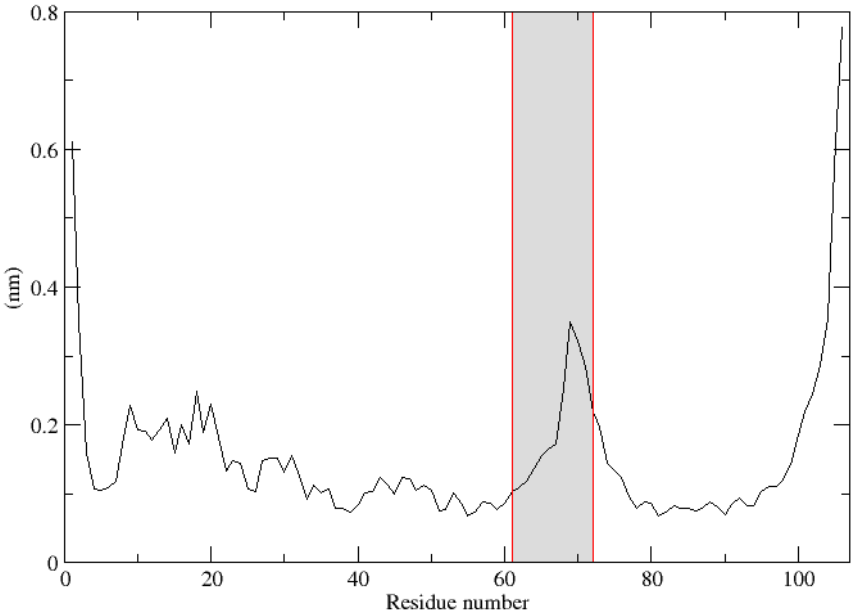
1POH


0.4
0.2
0
(nm)
0
20
40
60
80
Residue number

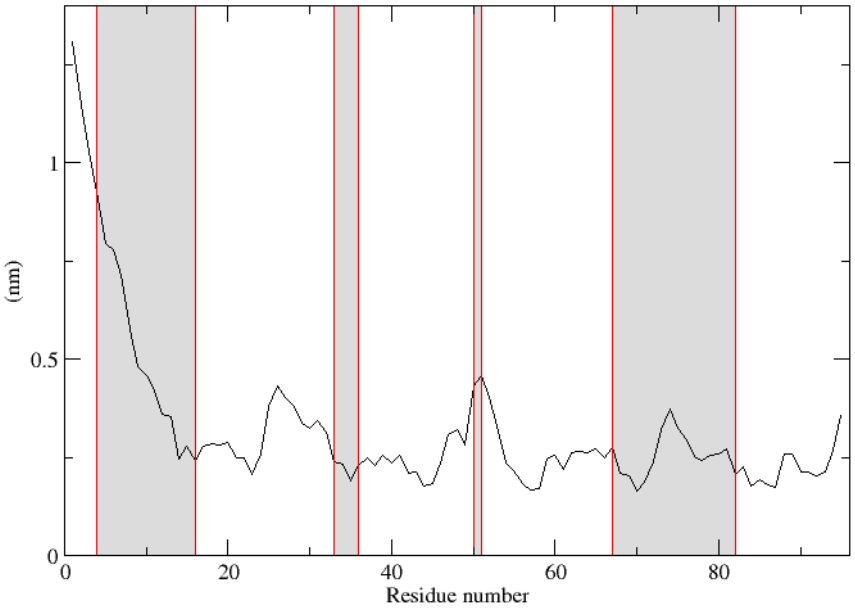
1TFH


0.8
0.6
0.4
0.2
0
(nm)
0
20
40
60
80
100
120
140
160
180
200
Residue number

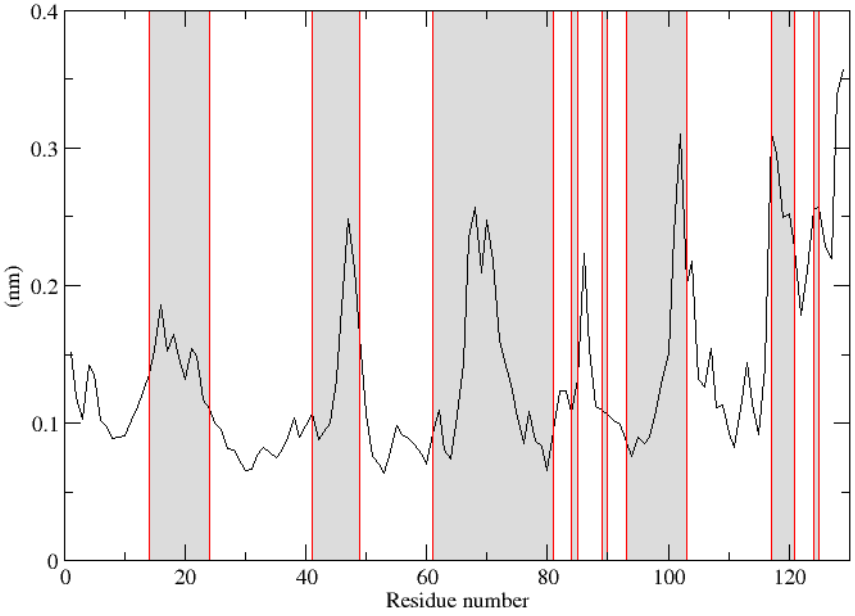
1UW3


0.8
0.6
0.4
0.2
0
(nm)
0
20
40
60
80
100
Residue number

2VPF


(nm)
0
0.5
1
0
20
40
60
80
Residue number

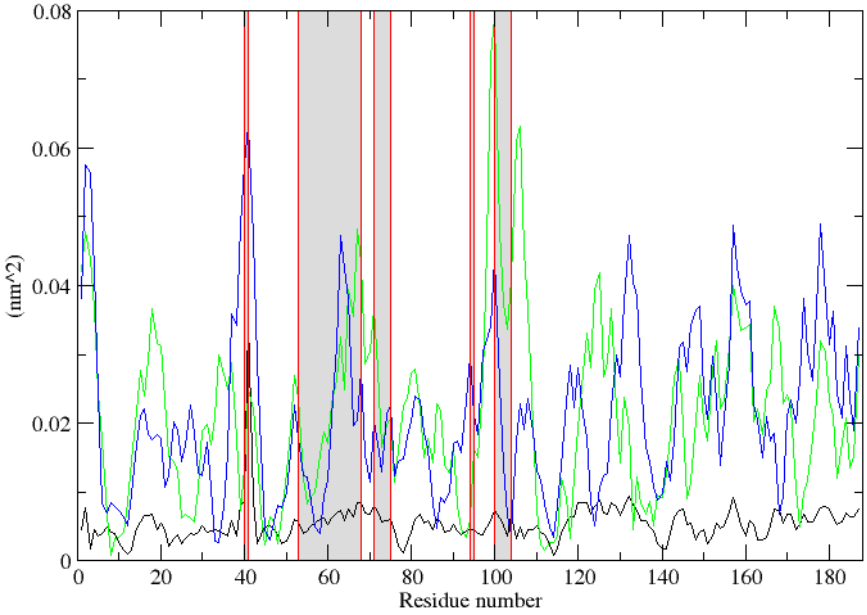
3LZT


(nm)
0
0.1
0.2
0.3
0.4
0
20
40
60
80
100
120
Residue number

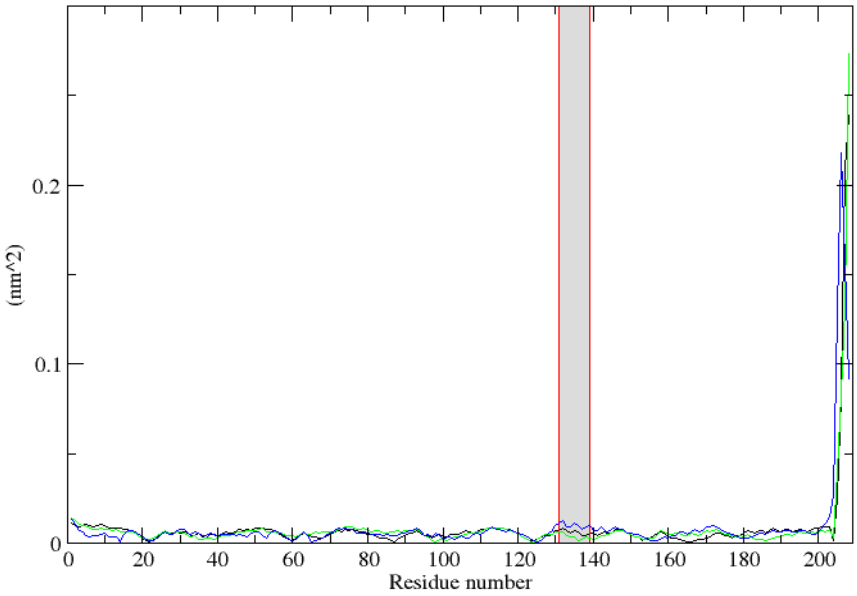
Normal Mode Analysis.
Fluctuations of the proteins along the first three lowest energy normal modes.
Normal modes were obtained both by using an Elastic Network Model to represent the structure of the protein, as implemented in Suhre K, Sanejouand YH. Nucleic Acids Res. 2004 Jul 1;32; and by an all atom representation of the protein as implemented in the Amber program. The results are highly consistent. Herein we are showing the data from the ENM calculations.
Gray Shadings indicate the locations of the epitopes.

Black: First normal mode;
Green: Second normal mode;
Blue: Third normal mode.

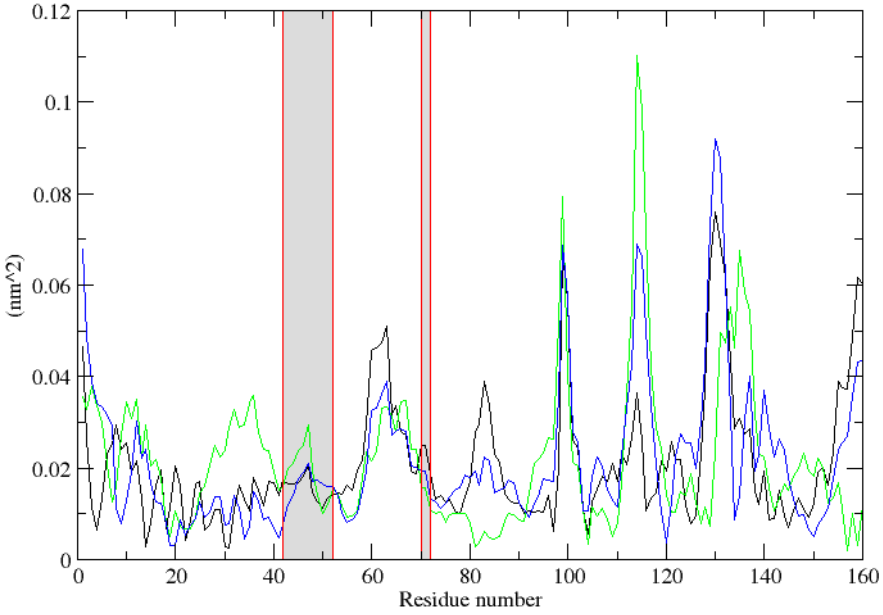
1AO3



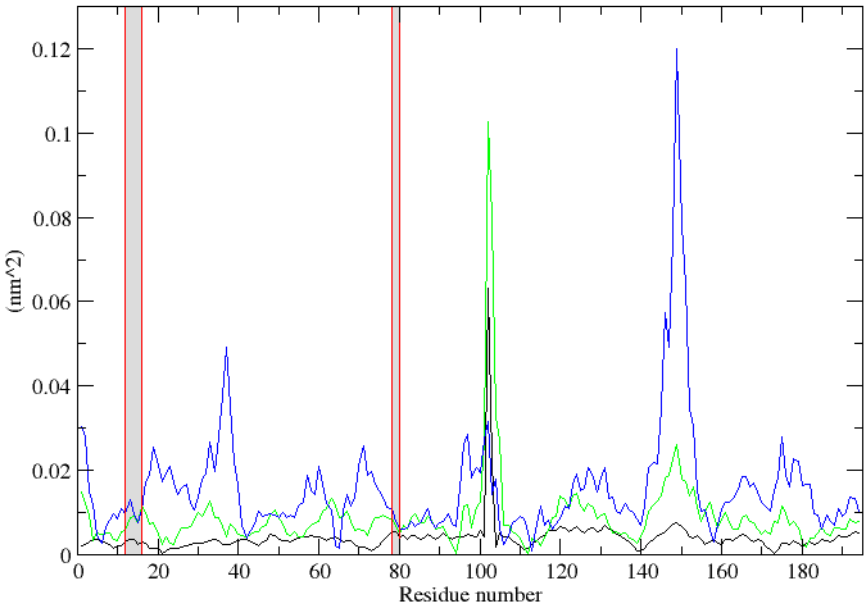
1AUQ


0.2
0.1
0
(nm^2)
0
20
40
60
80
100
120
140
160
180
200
Residue number

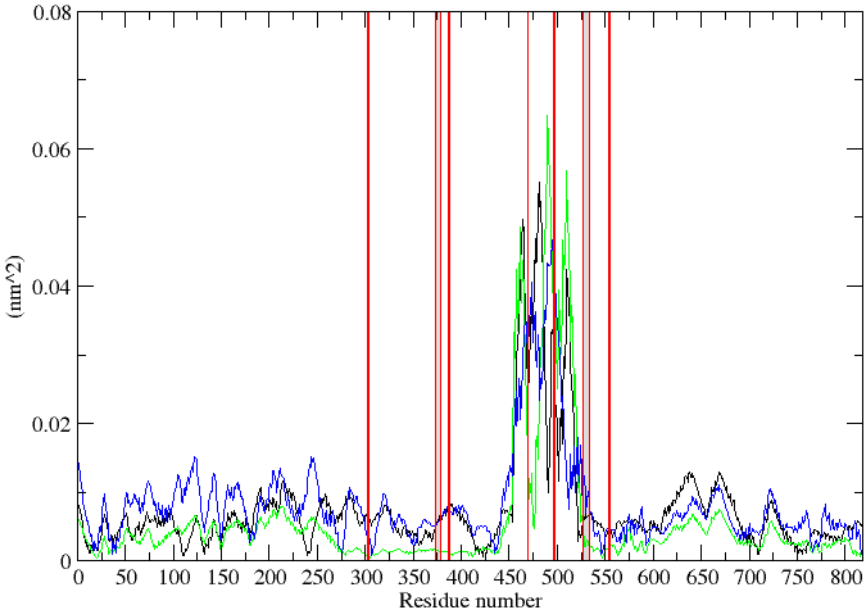
1BV1


0.12
0.1
0.08
0.06
0.04
0.02
0
(nm^2)
0
20
40
60
80
100
120
140
160
Residue number

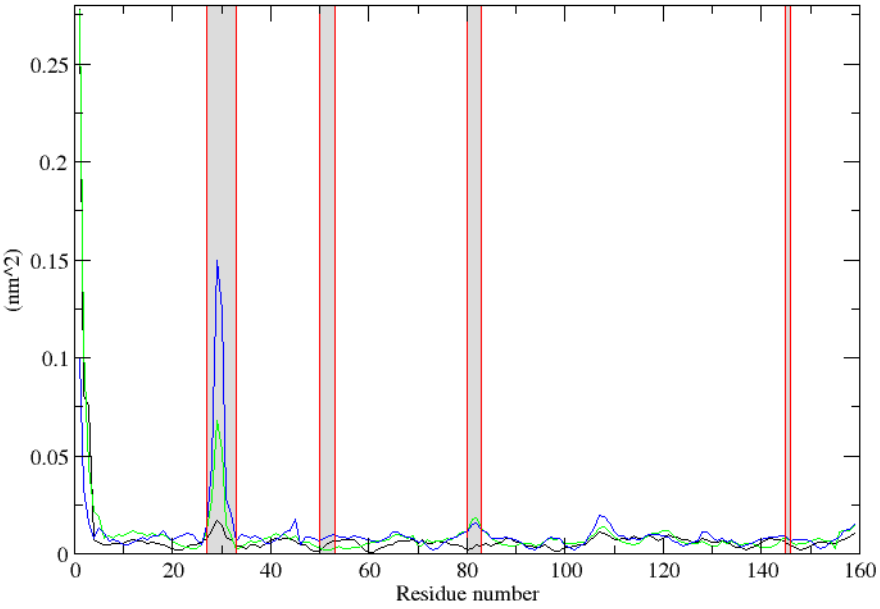
1CK4


0.12
0.1
0.08
0.06
0.04
0.02
0
(nm^2)
0
20
40
60
80
100
120
140
160
180
Residue number

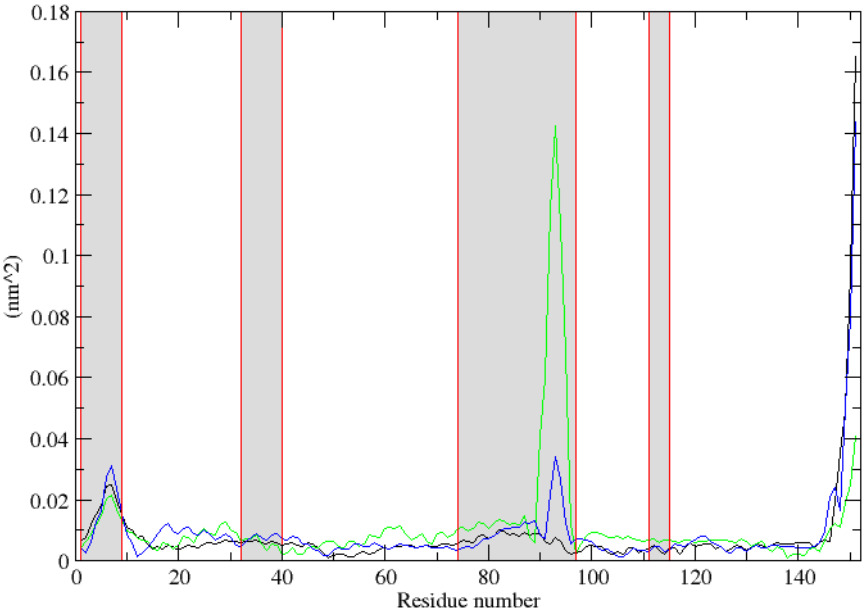
1CMW


0.08
0.06
0.04
0.02
0
(nm^2)
0
50
100
150
200
250
300
350
400
450
500
550
600
650
700
750
800
Residue number

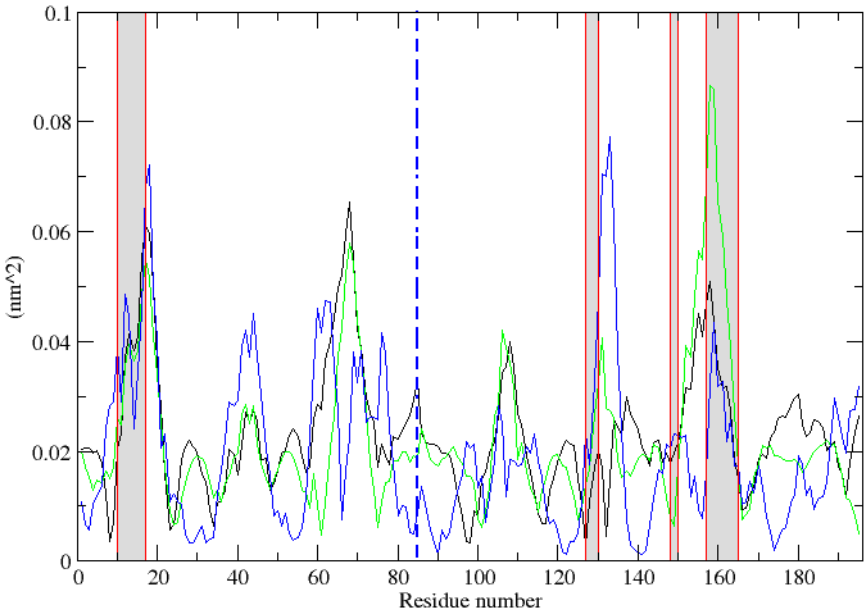
1D7P


0.25
0.2
0.15
0.1
0.05
0
(nm^2)
0
20
40
60
80
100
120
140
160
Residue number

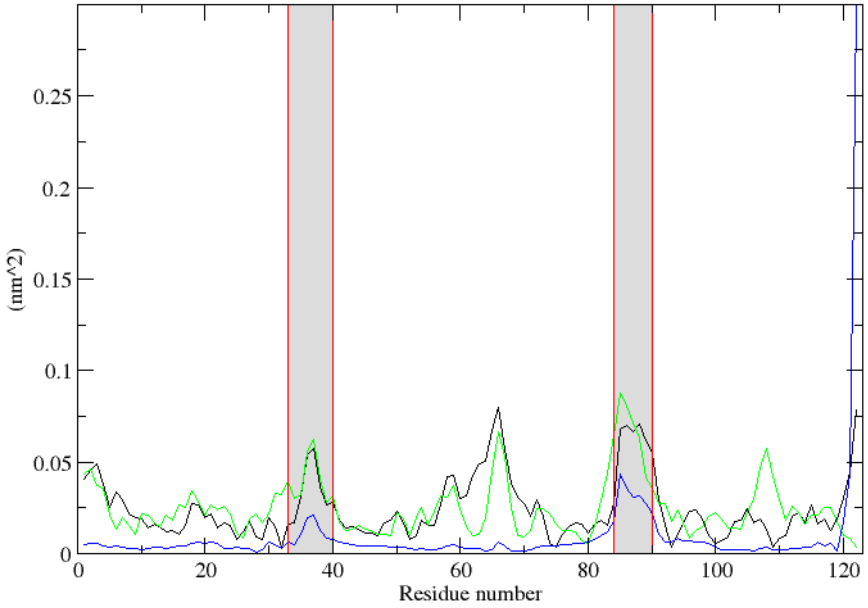
1GWP


0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
(nm^2)
0
20
40
60
80
100
120
140
Residue number

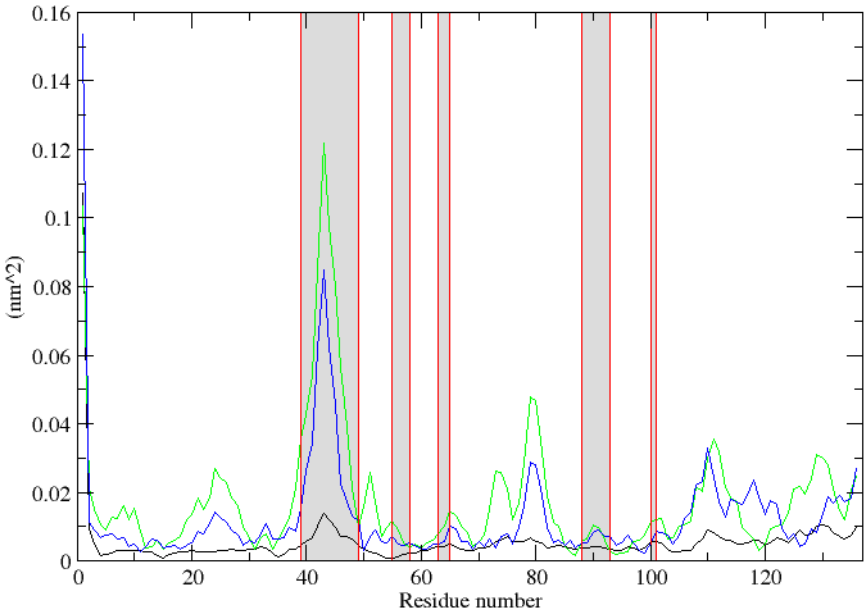
1HCN


0.1
0.08
0.06
0.04
0.02
0
(nm^2)
0
20
40
60
80
100
120
140
160
180
Residue number

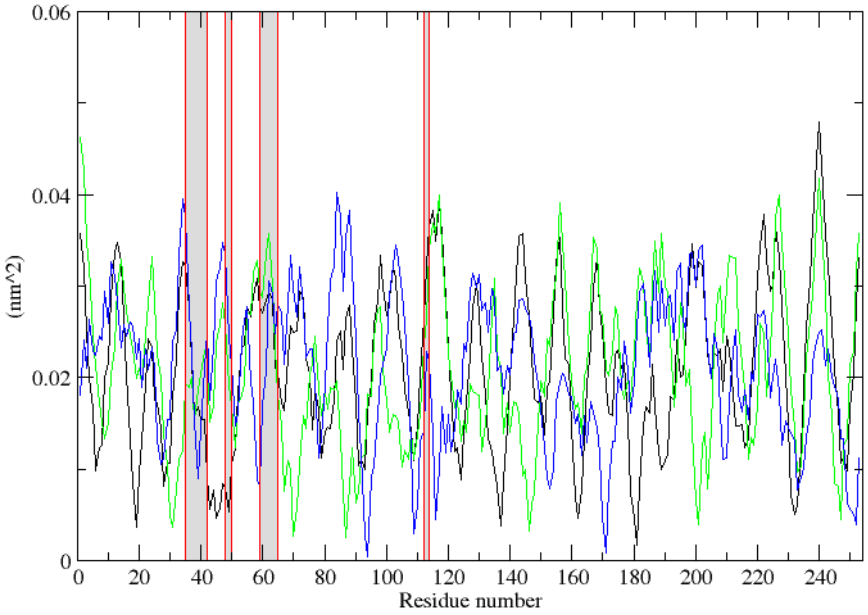
1K59


0.25
0.2
0.15
0.1
0.05
0
(nm^2)
0
20
40
60
80
100
120
Residue number

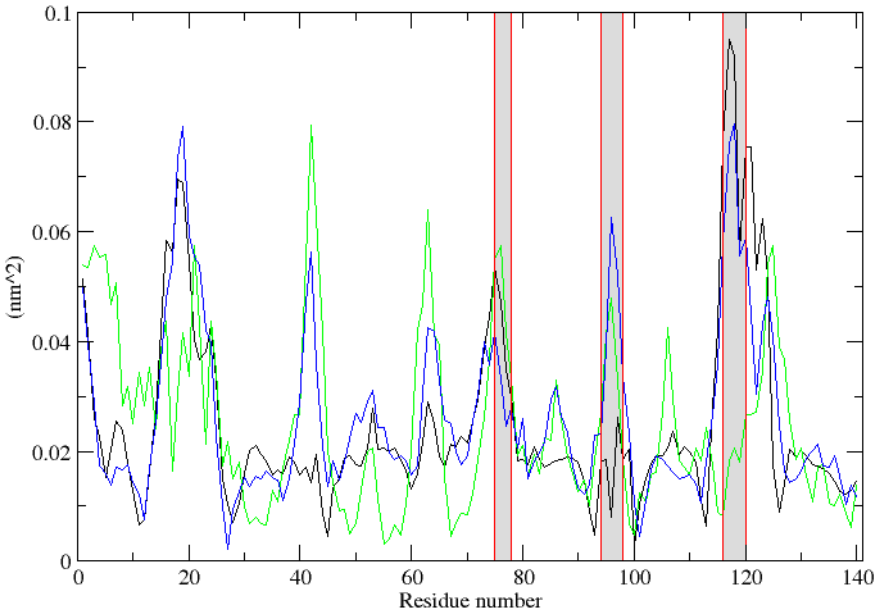
1KDC


0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
(nm^2)
0
20
40
60
80
100
120
Residue number

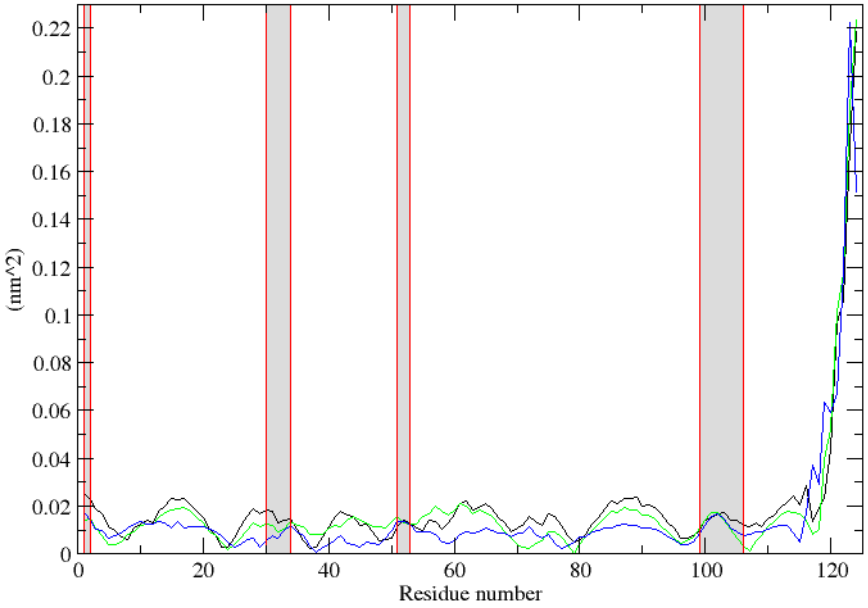
1KZQ


0.06
0.04
0.02
0
(nm^2)
0
20
40
60
80
100
120
140
160
180
200
220
240
Residue number

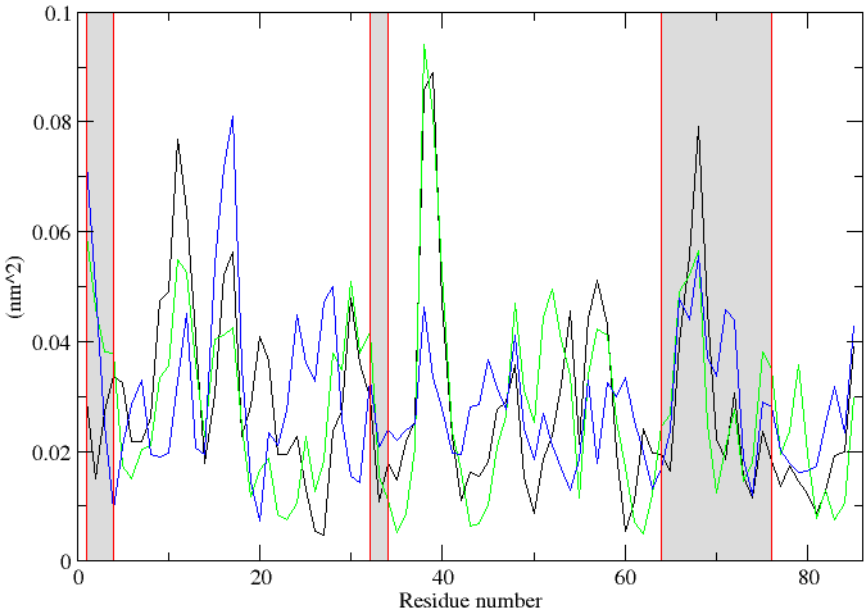
1P4P


0.1
0.08
0.06
0.04
0.02
0
(nm^2)
0
20
40
60
80
100
120
140
Residue number

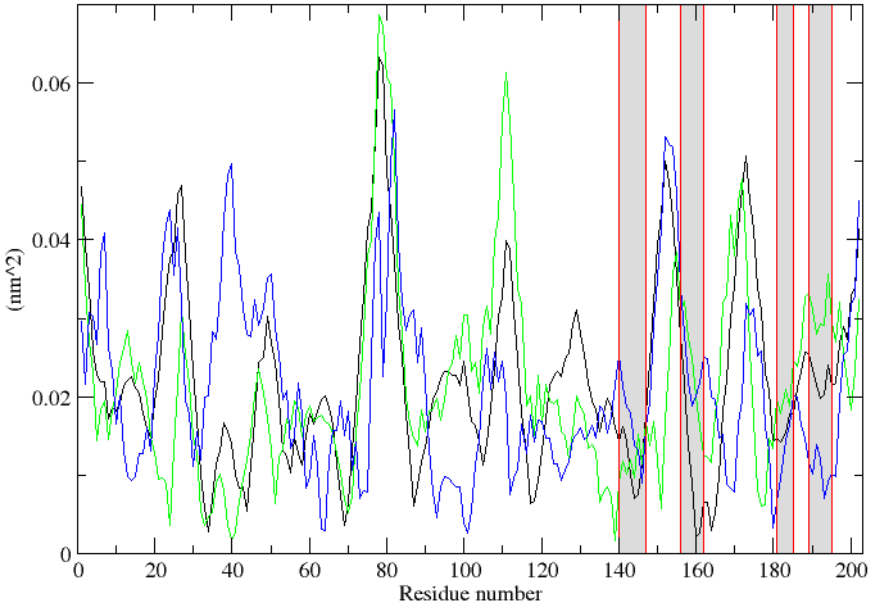
1PKO


0.22
0.2
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
(nm^2)
0
20
40
60
80
100
120
Residue number

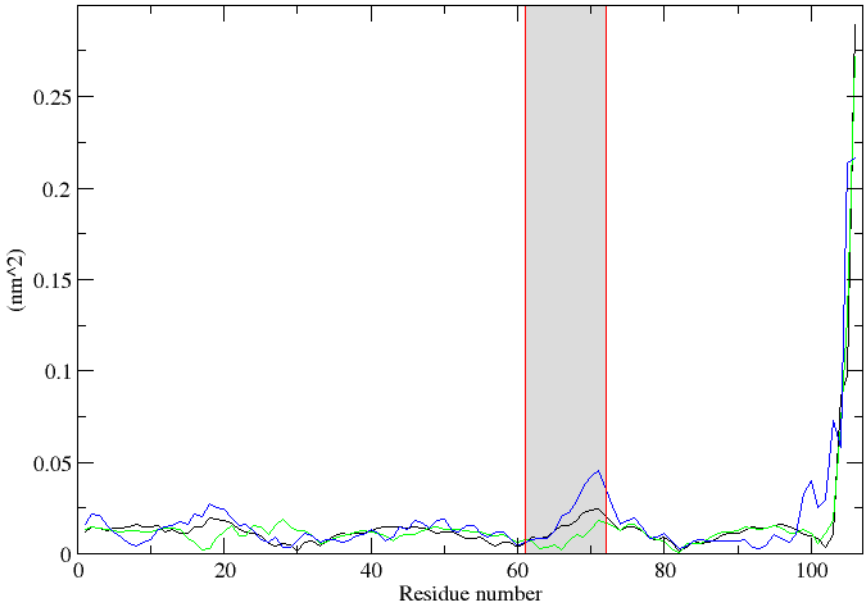
1POH


0.1
0.08
0.06
0.04
0.02
0
(nm^2)
0
20
40
60
80
Residue number

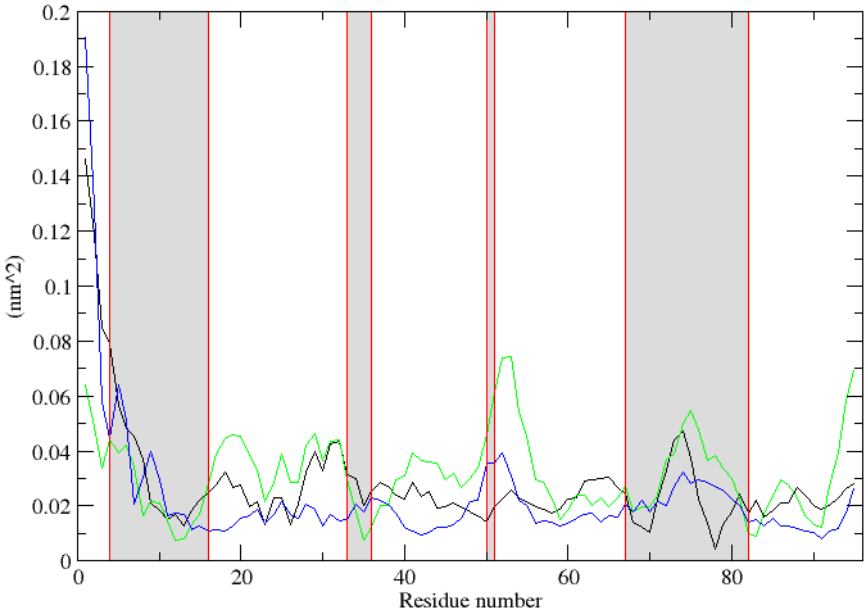
1TFH


0.06
0.04
0.02
0
(nm^2)
0
20
40
60
80
100
120
140
160
180
200
Residue number

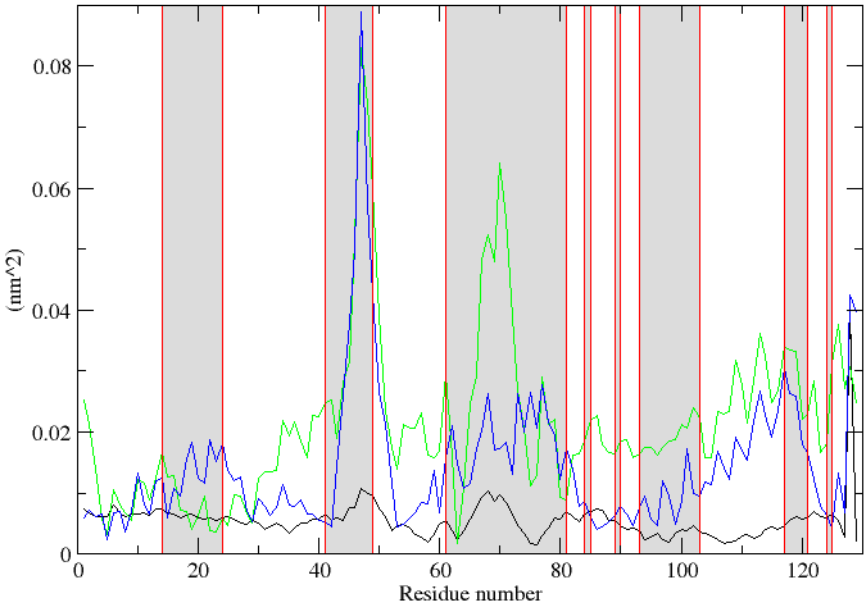
1UW3


0.25
0.2
0.15
0.1
0.05
0
(nm^2)
0
20
40
60
80
100
Residue number

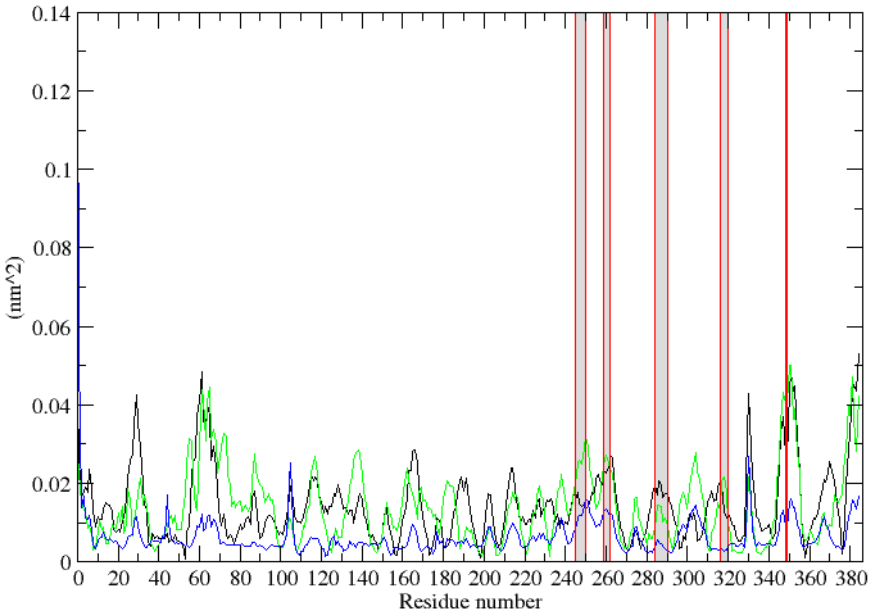
2VPF


0.2
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
(nm^2)
0
20
40
60
80
Residue number

3LZT


0.08
0.06
0.04
0.02
0
(nm^2)
0
20
40
60
80
100
120
Residue number

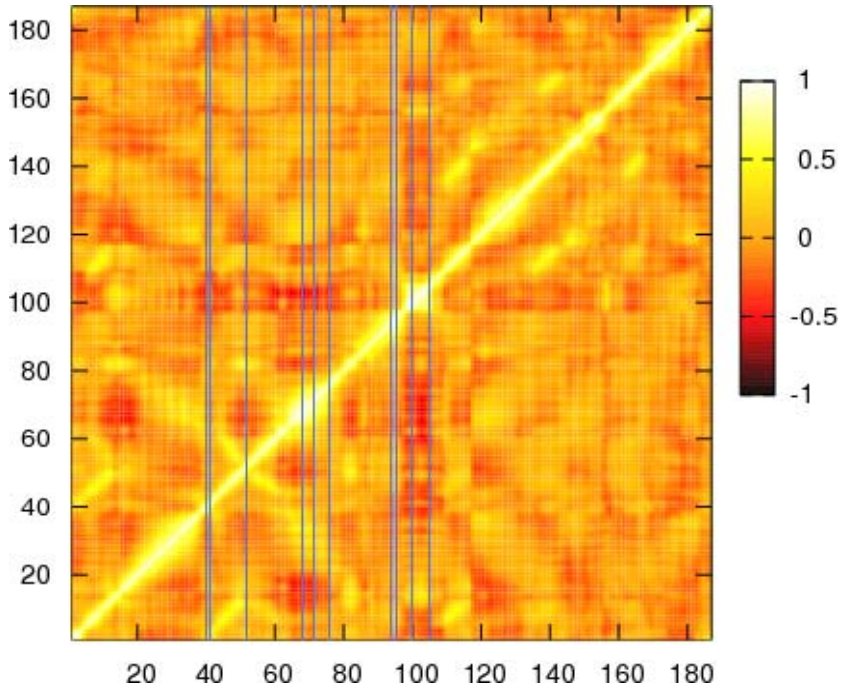
7NN9

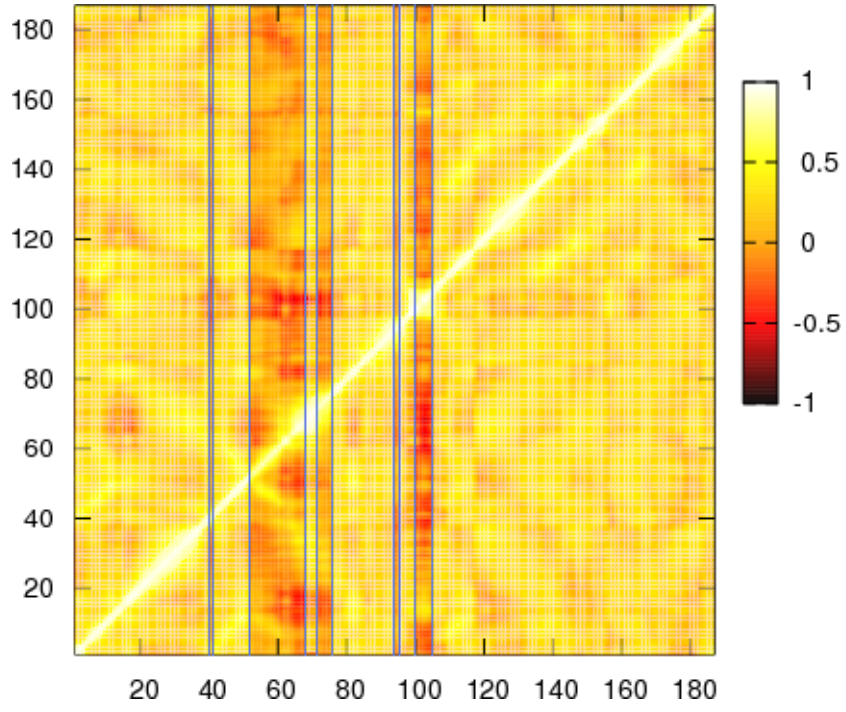

0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
(nm^2)
0 20 40 60 80 100 120 140 160 180 200 220 240 260 280 300 320 340 360 380
Residue number

Residue-Residue Cross-Correlation Matrices

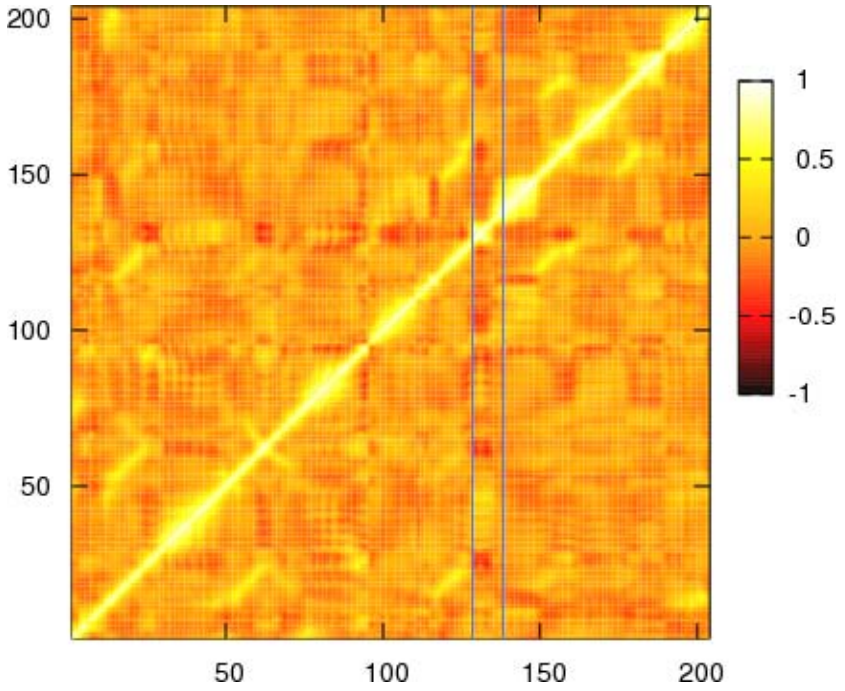
The matrices reported in the following are color-coded according to the degree of cross-correlation between any two residues in the sequence. Blue lines highlight epitope regions. For each antigen, we have reported two representation of the matrix. The top one represents the whole protein. The bottom one is all blurred except for the cross-correlations involving epitope regions. This representation has been added for clarity. It is apparent that no specific dynamic signal differentiates epitope from non-epitope sequences.

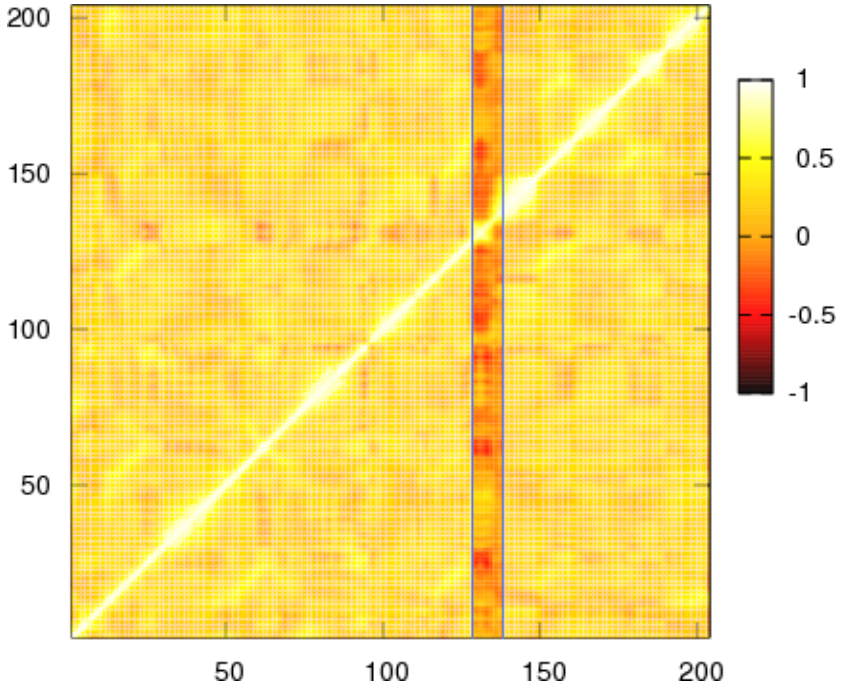
1AO3



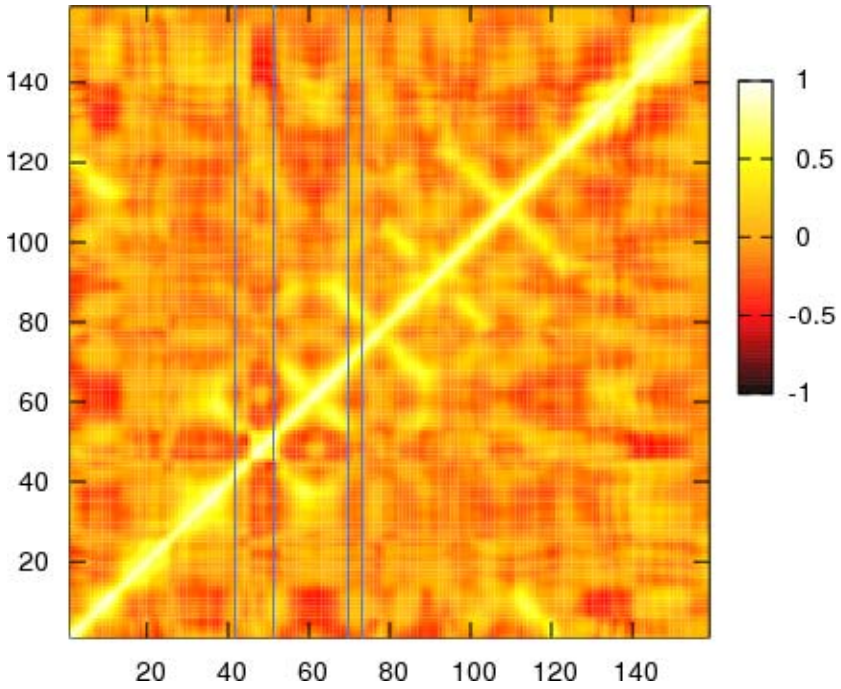


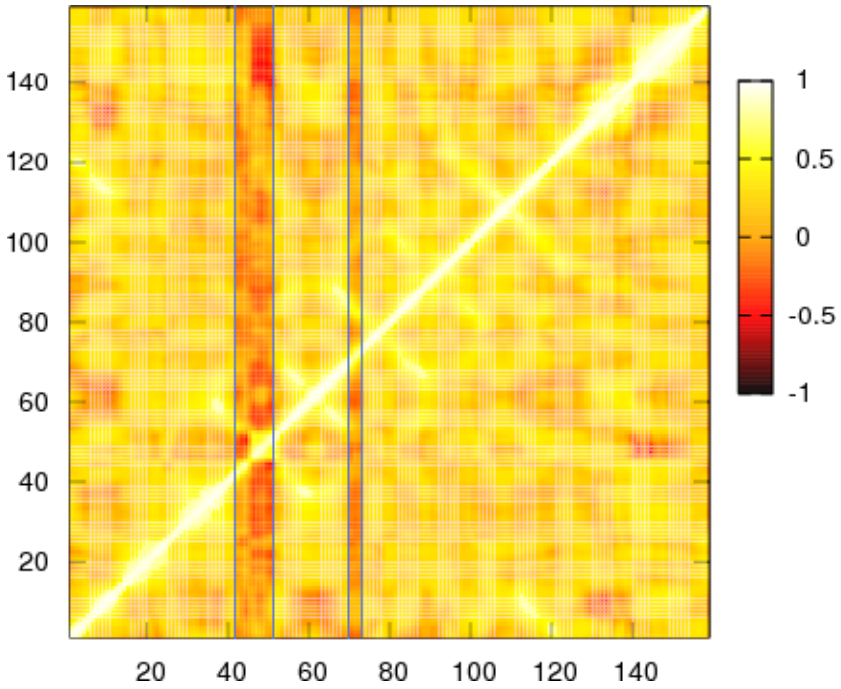
1AUQ


200
150
100
50
50
100
150
200
1
0.5
0
-0.5
-1

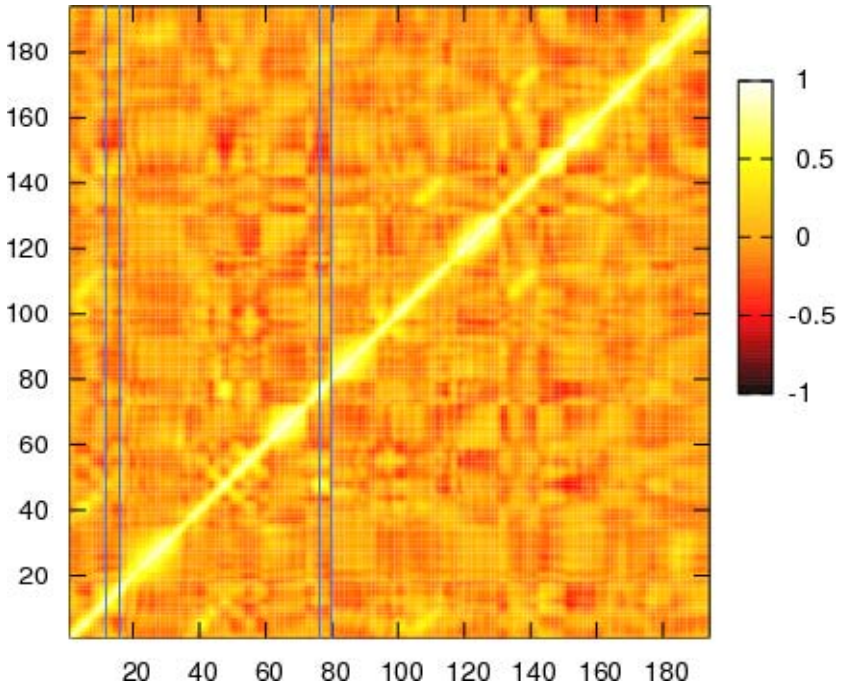

200
150
100
50
50
100
150
200
1
0.5
0
-0.5
-1

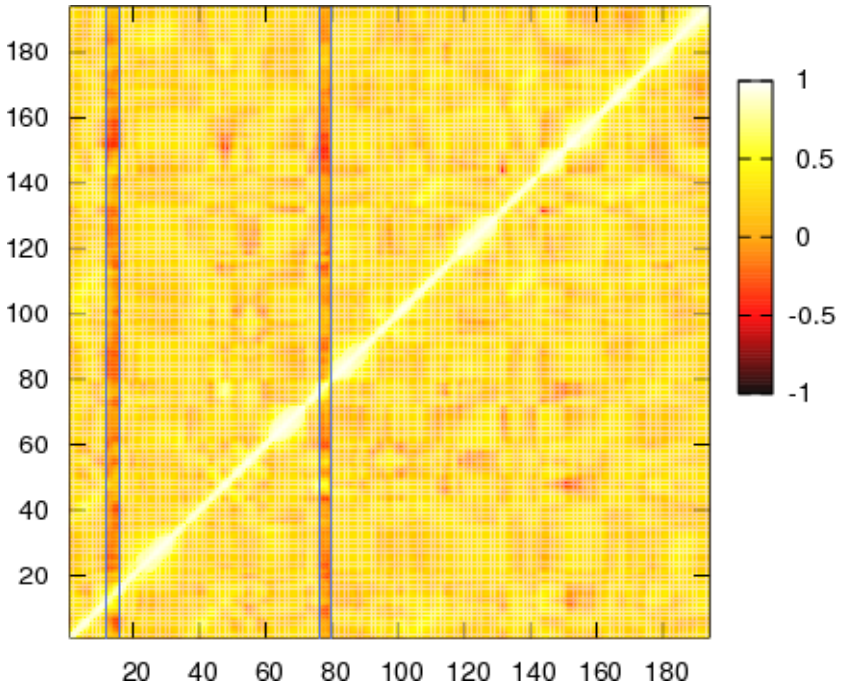
1BV1


140
120
100
80
60
40
20
20 40 60 80 100 120 140
1
0.5
0
-0.5
-1

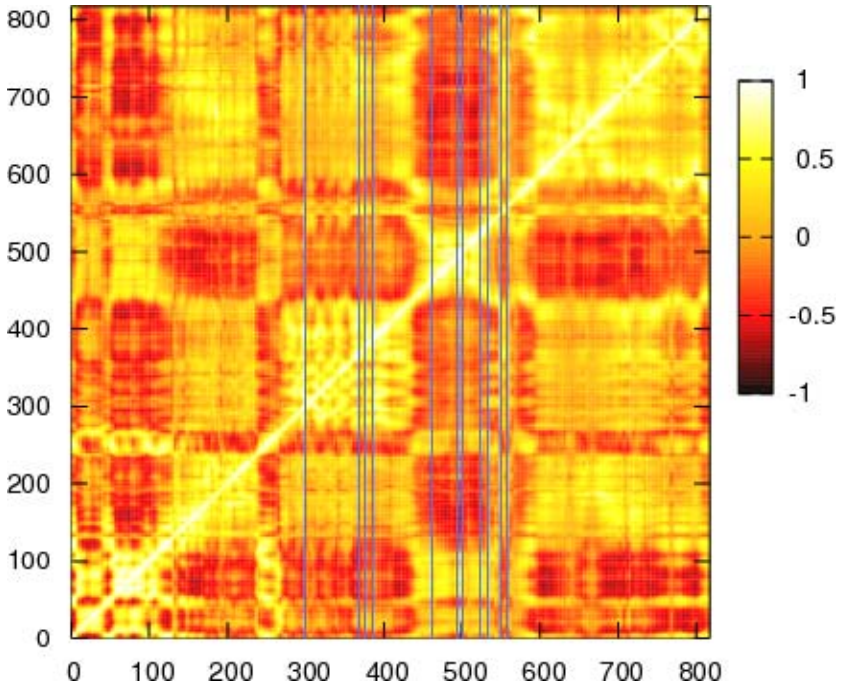

140
120
100
80
60
40
20
20 40 60 80 100 120 140
1
0.5
0
-0.5
-1

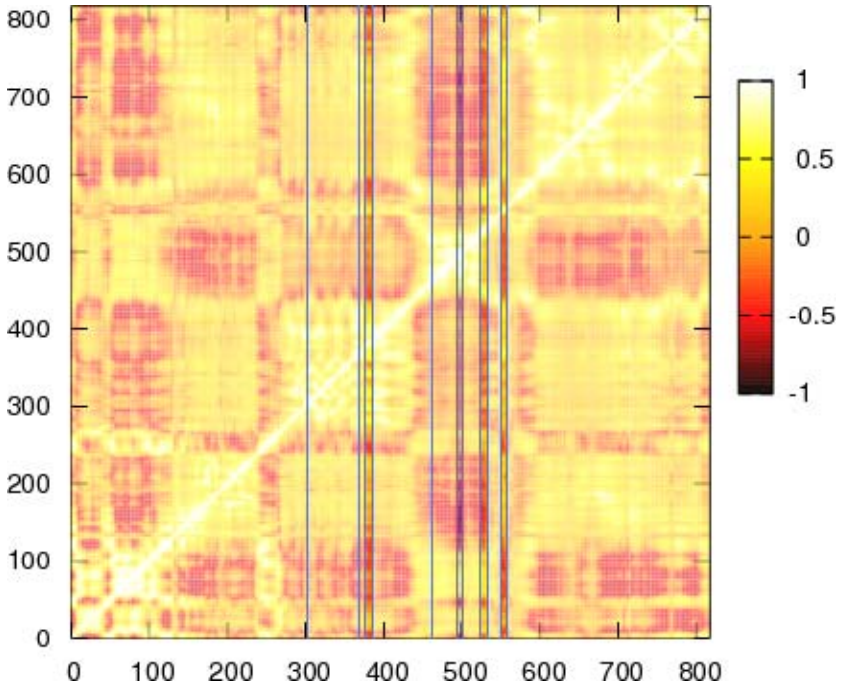
1CK4


180
160
140
120
100
80
60
40
20
20 40 60 80 100 120 140 160 180
1
0.5
0
-0.5
-1

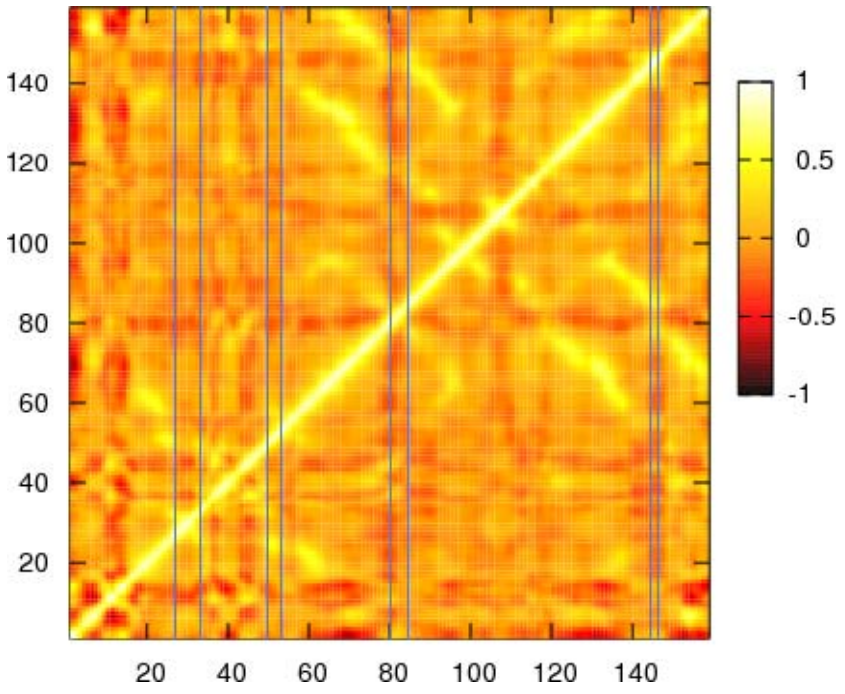

180
160
140
120
100
80
60
40
20
20 40 60 80 100 120 140 160 180
1
0.5
0
-0.5
-1

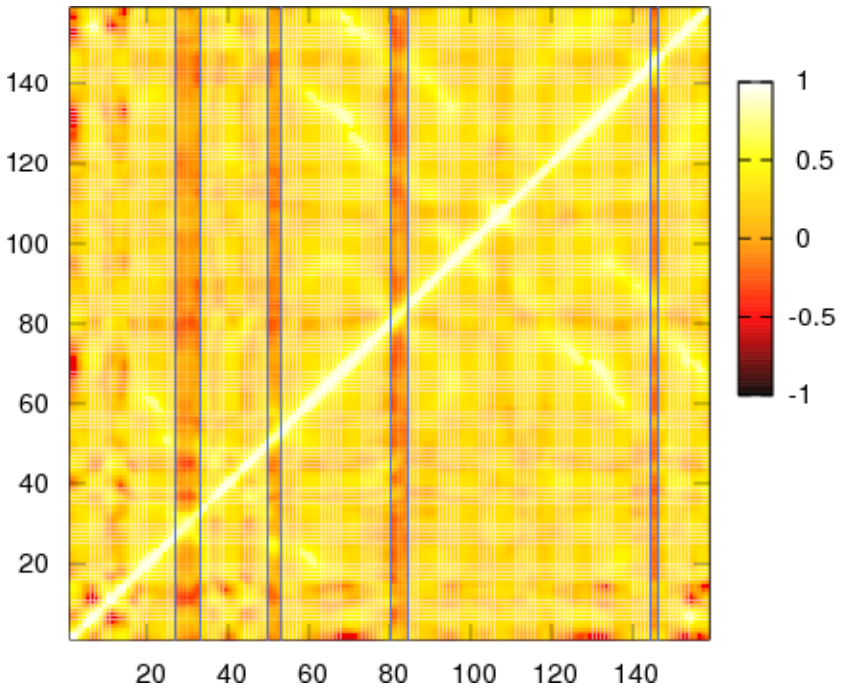
1CMW


800
700
600
500
400
300
200
100
0
0 100 200 300 400 500 600 700 800
1
0.5
0
-0.5
-1

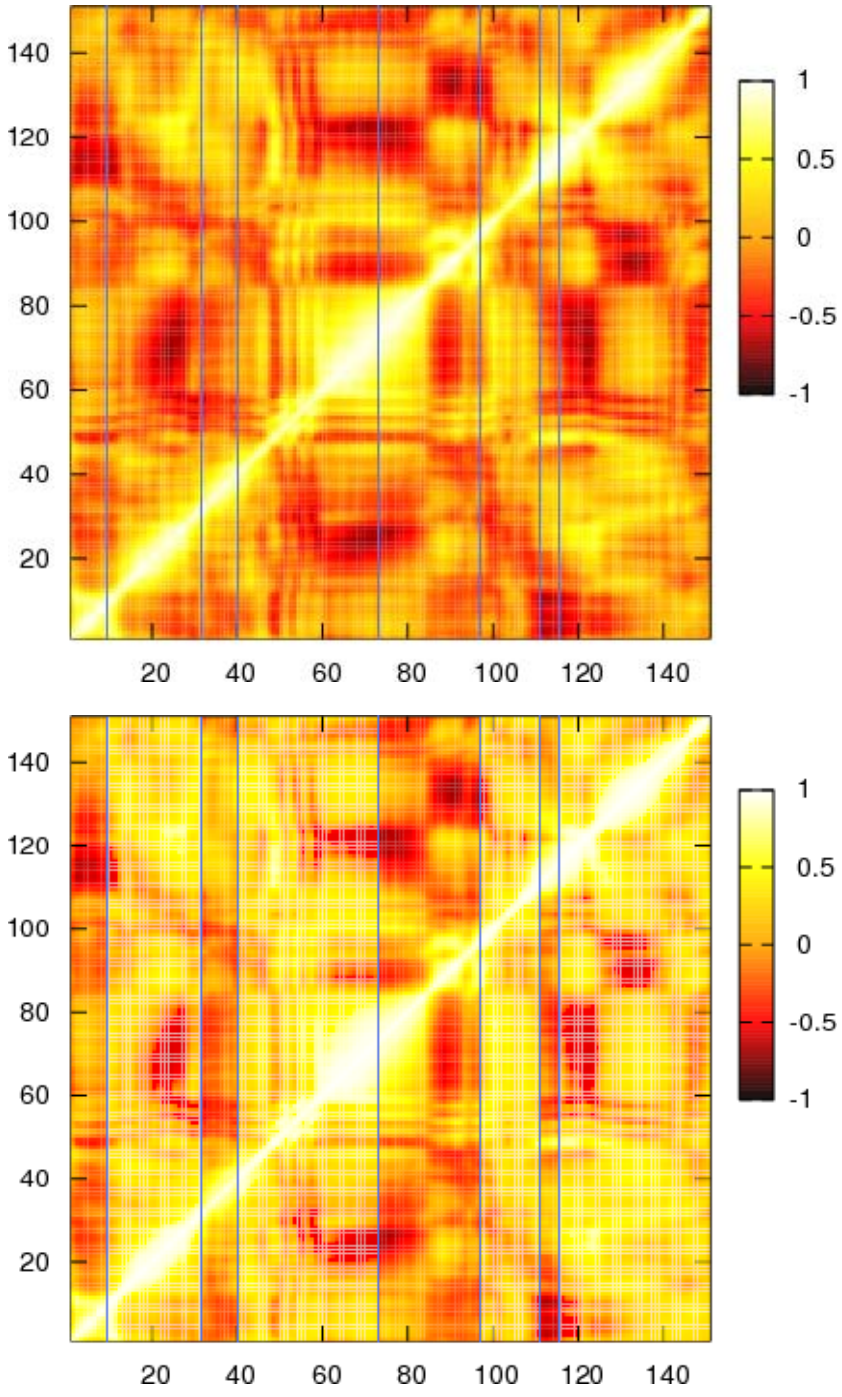

800
700
600
500
400
300
200
100
0
0 100 200 300 400 500 600 700 800
1
0.5
0
-0.5
-1

1D7P

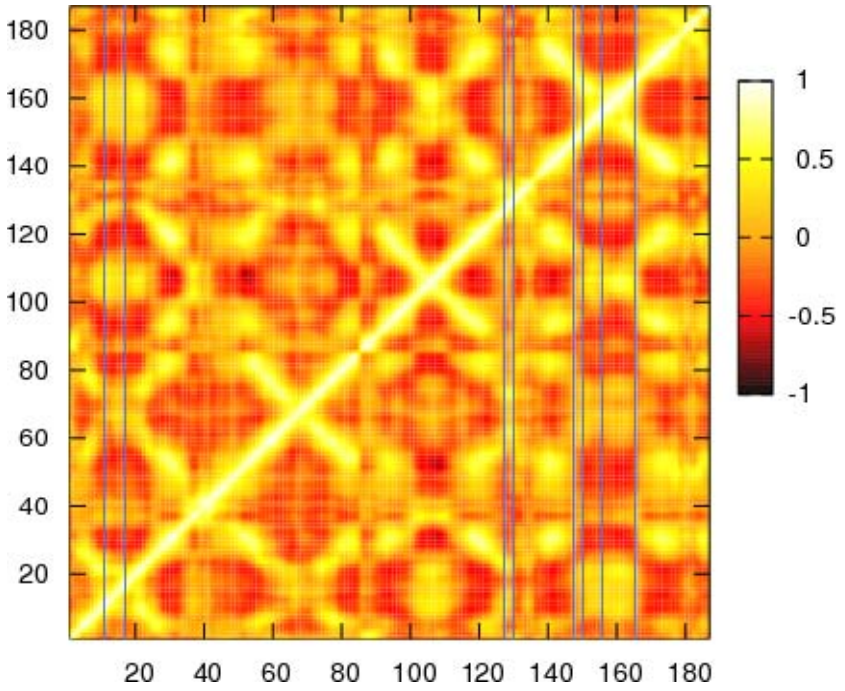

140
120
100
80
60
40
20
20
40
60
80
100
120
140
1
0.5
0
-0.5
-1

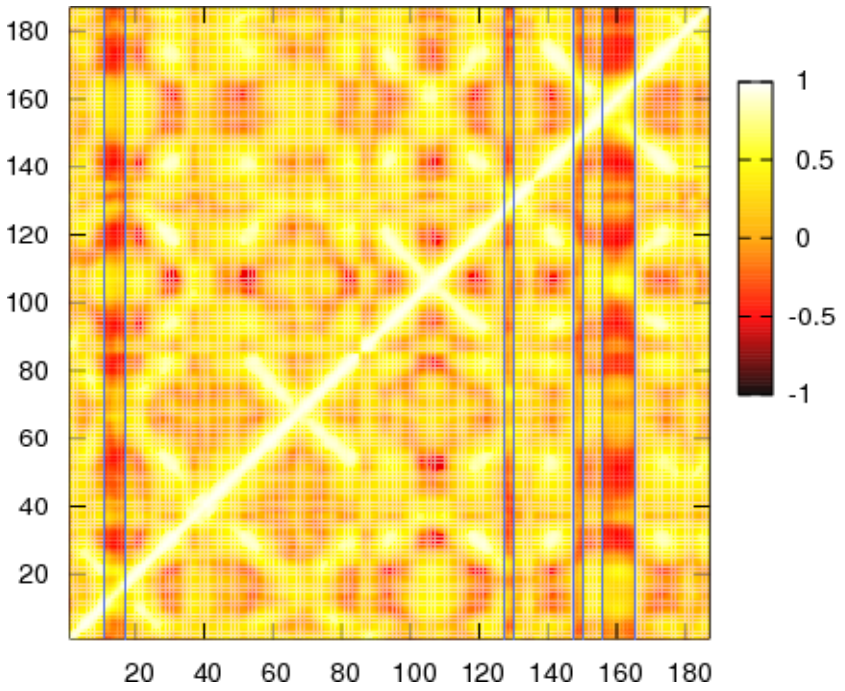

140
120
100
80
60
40
20
20
40
60
80
100
120
140
1
0.5
0
-0.5
-1

1GWP

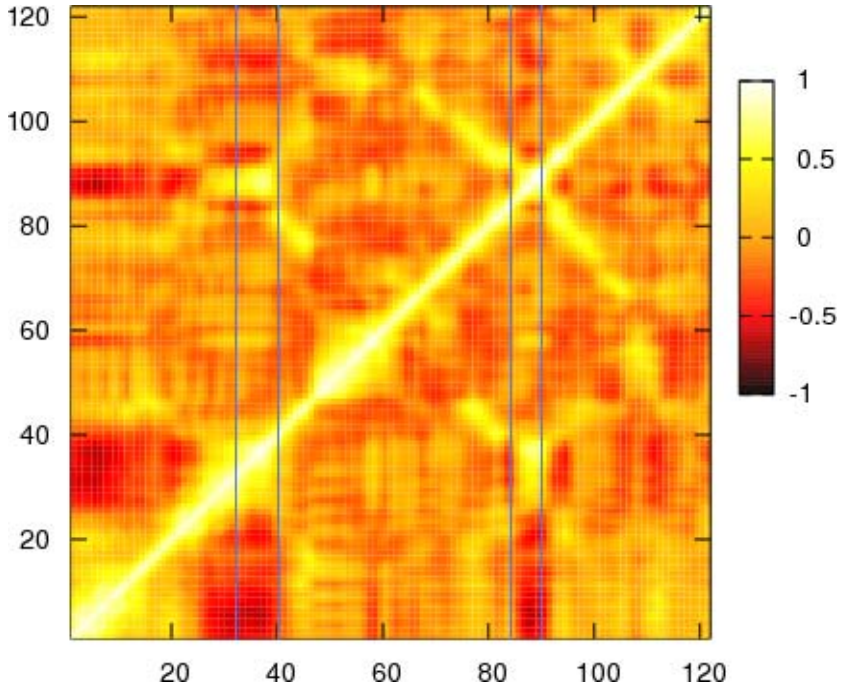

140
120
100
80
60
40
20
20
40
60
80
100
120
140
1
0.5
0
-0.5
-1
140
120
100
80
60
40
20
20
40
60
80
100
120
140
1
0.5
0
-0.5
-1

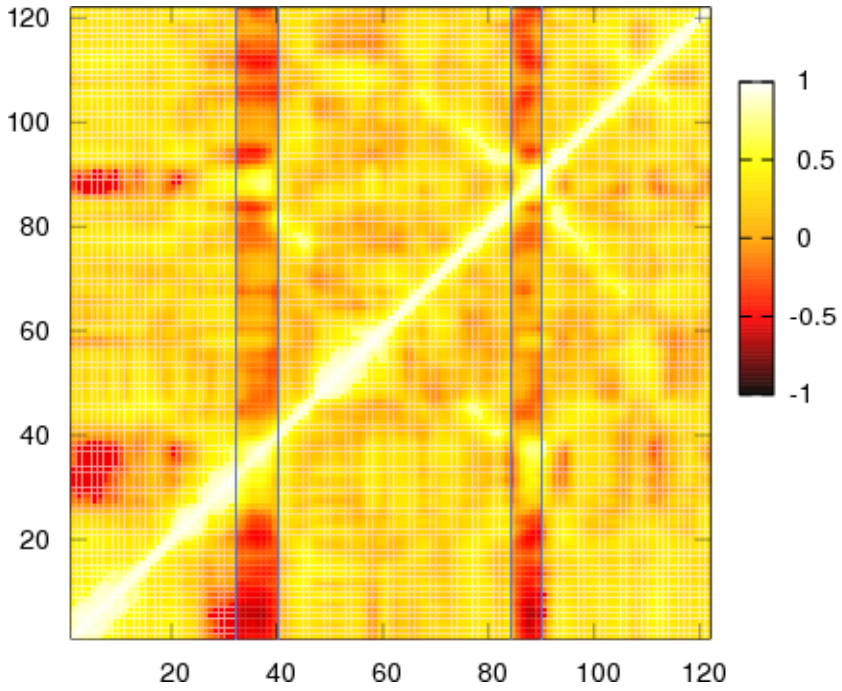
1HCN


180
160
140
120
100
80
60
40
20
20 40 60 80 100 120 140 160 180
1
0.5
0
-0.5
-1

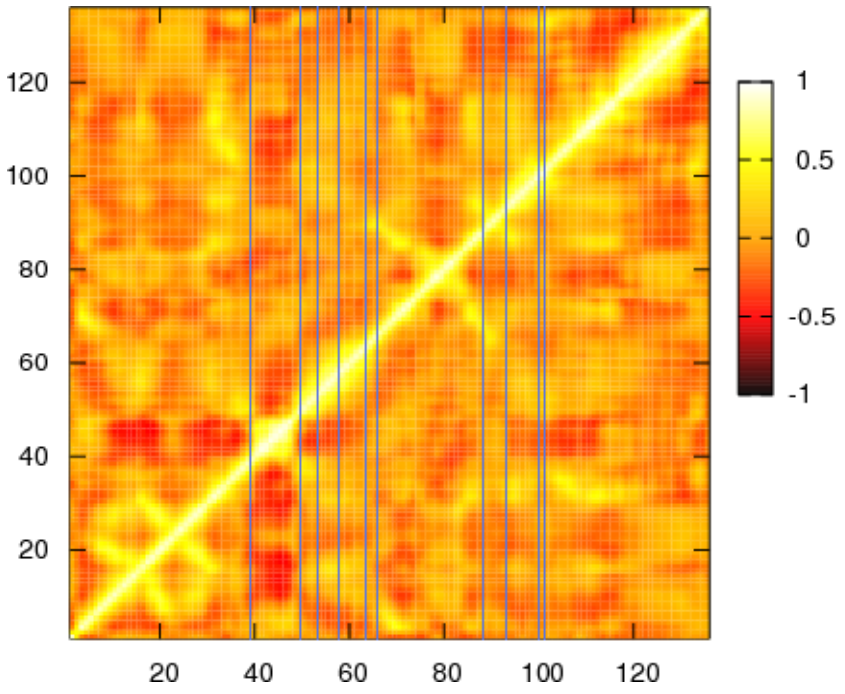

180
160
140
120
100
80
60
40
20
20 40 60 80 100 120 140 160 180
1
0.5
0
-0.5
-1

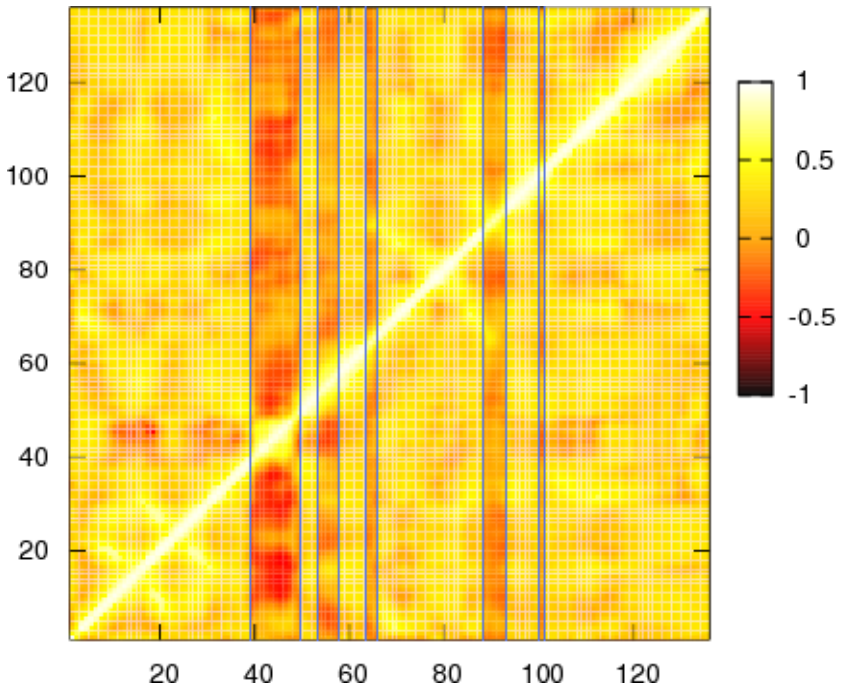
1K59


120
100
80
60
40
20
20
40
60
80
100
120
1
0.5
0
-0.5
-1

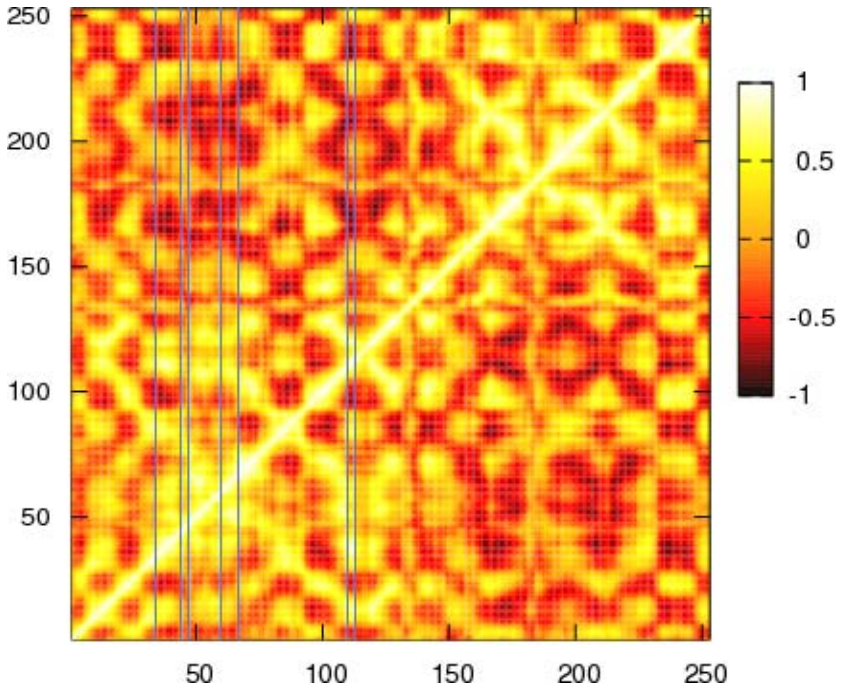

120
100
80
60
40
20
20
40
60
80
100
120
1
0.5
0
-0.5
-1

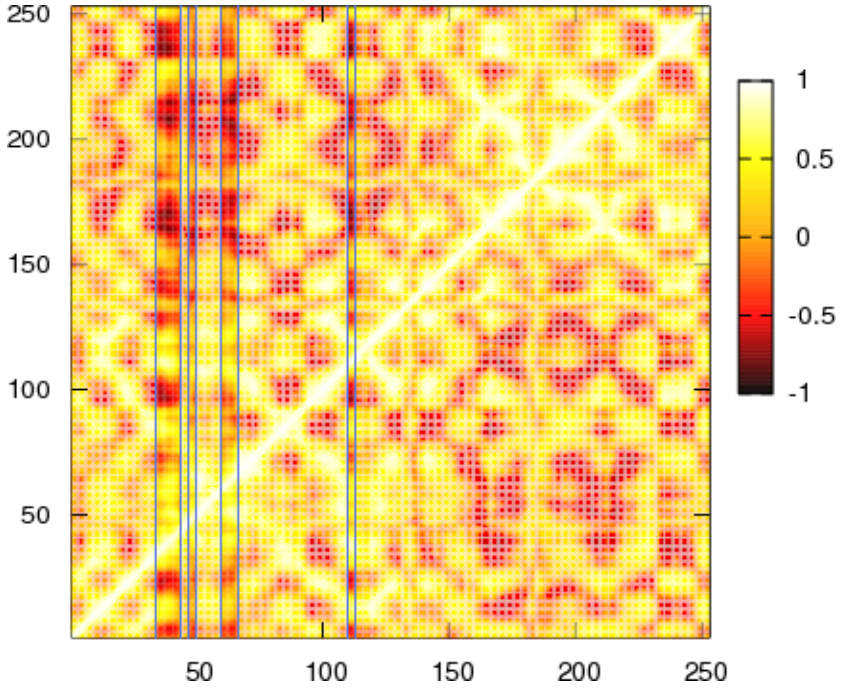
1KDC


120
100
80
60
40
20
20
40
60
80
100
120
1
0.5
0
-0.5
-1

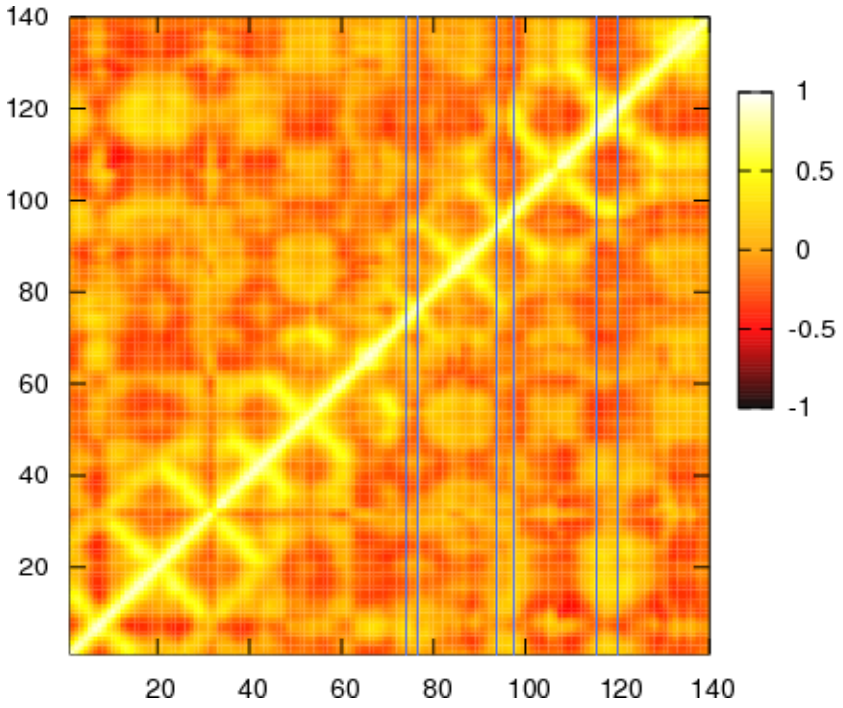

120
100
80
60
40
20
20
40
60
80
100
120
1
0.5
0
-0.5
-1

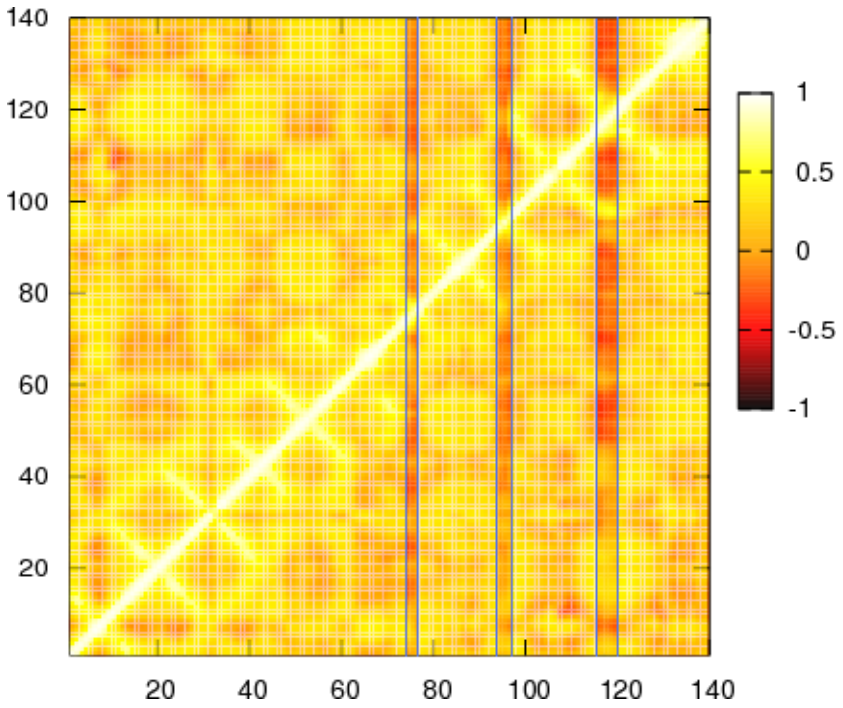
1KZQ


250
200
150
100
50
50
100
150
200
250
1
0.5
0
-0.5
-1

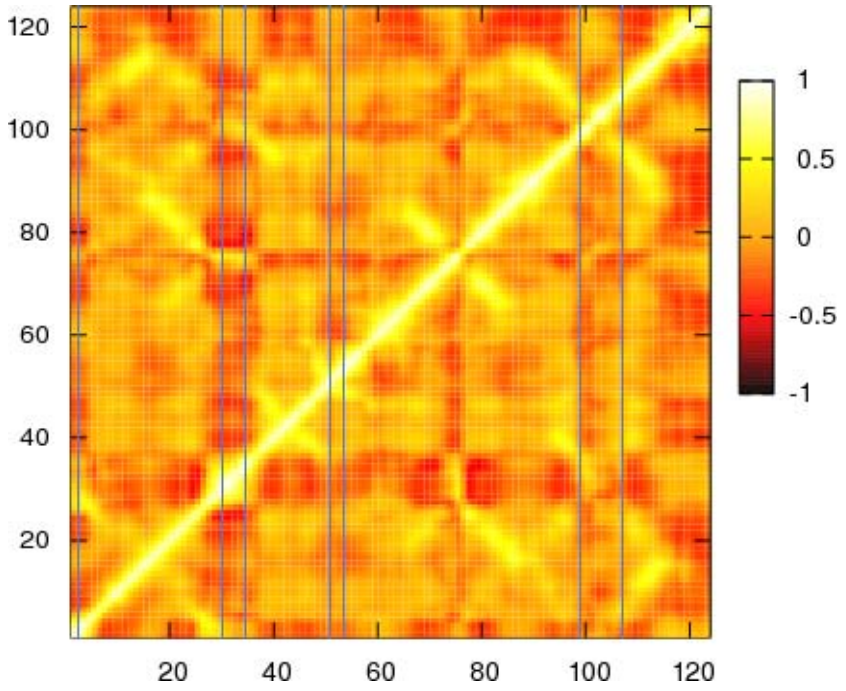

250
200
150
100
50
50
100
150
200
250
1
0.5
0
-0.5
-1

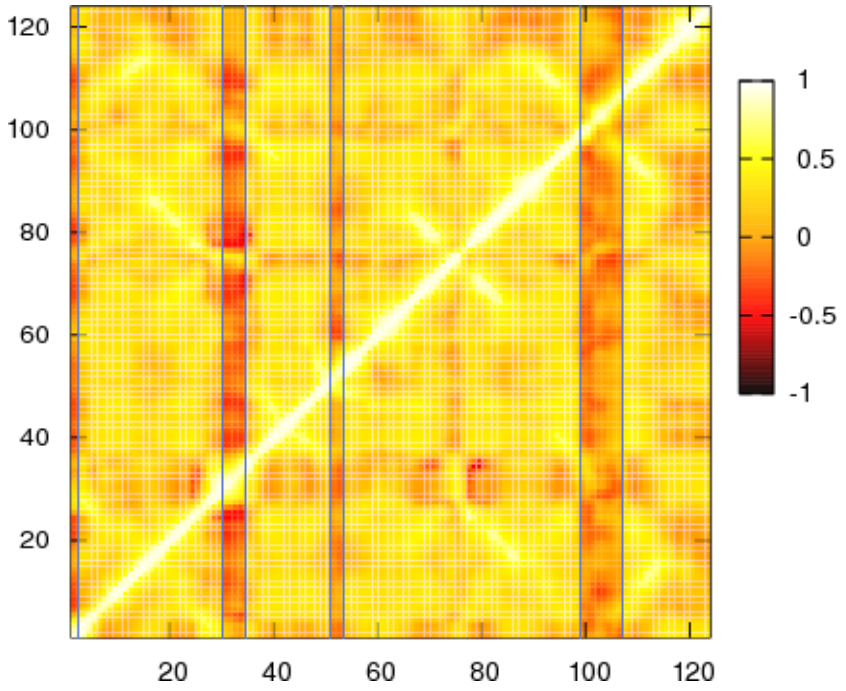
1P4P


140
120
100
80
60
40
20
20
40
60
80
100
120
140
1
0.5
0
-0.5
-1

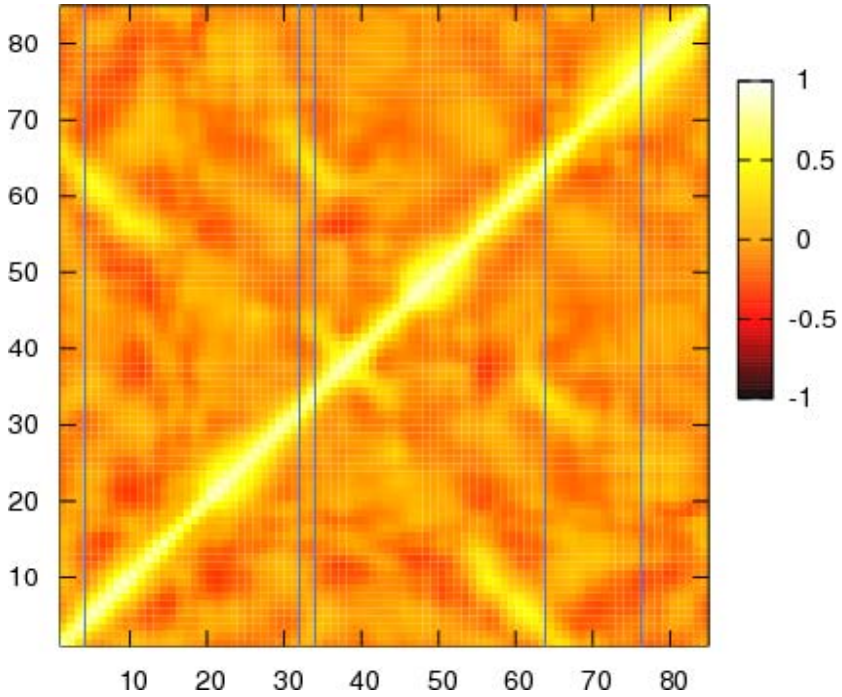

140
120
100
80
60
40
20
20
40
60
80
100
120
140
1
0.5
0
-0.5
-1

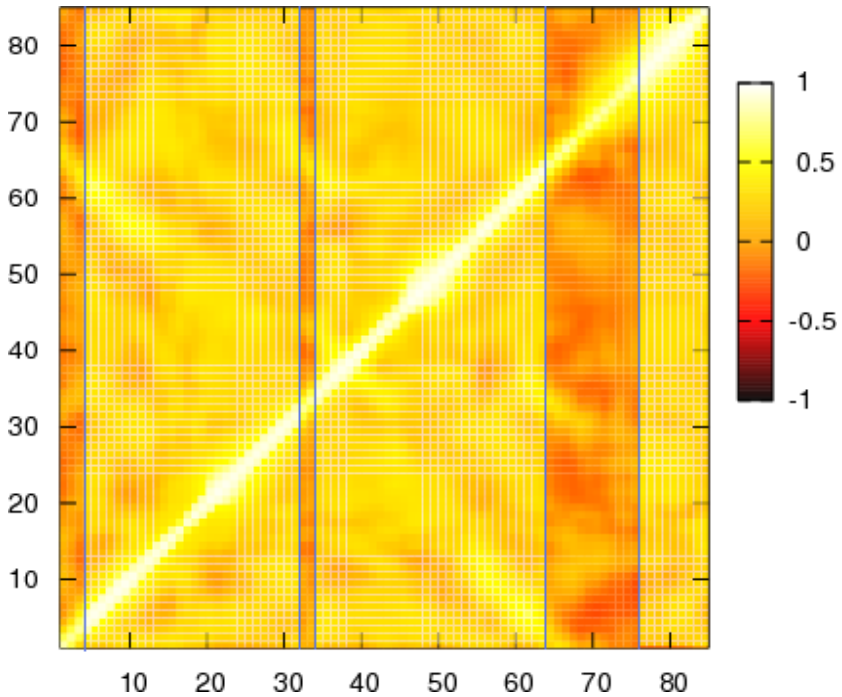
1PKO


120
100
80
60
40
20
20
40
60
80
100
120
1
0.5
0
-0.5
-1

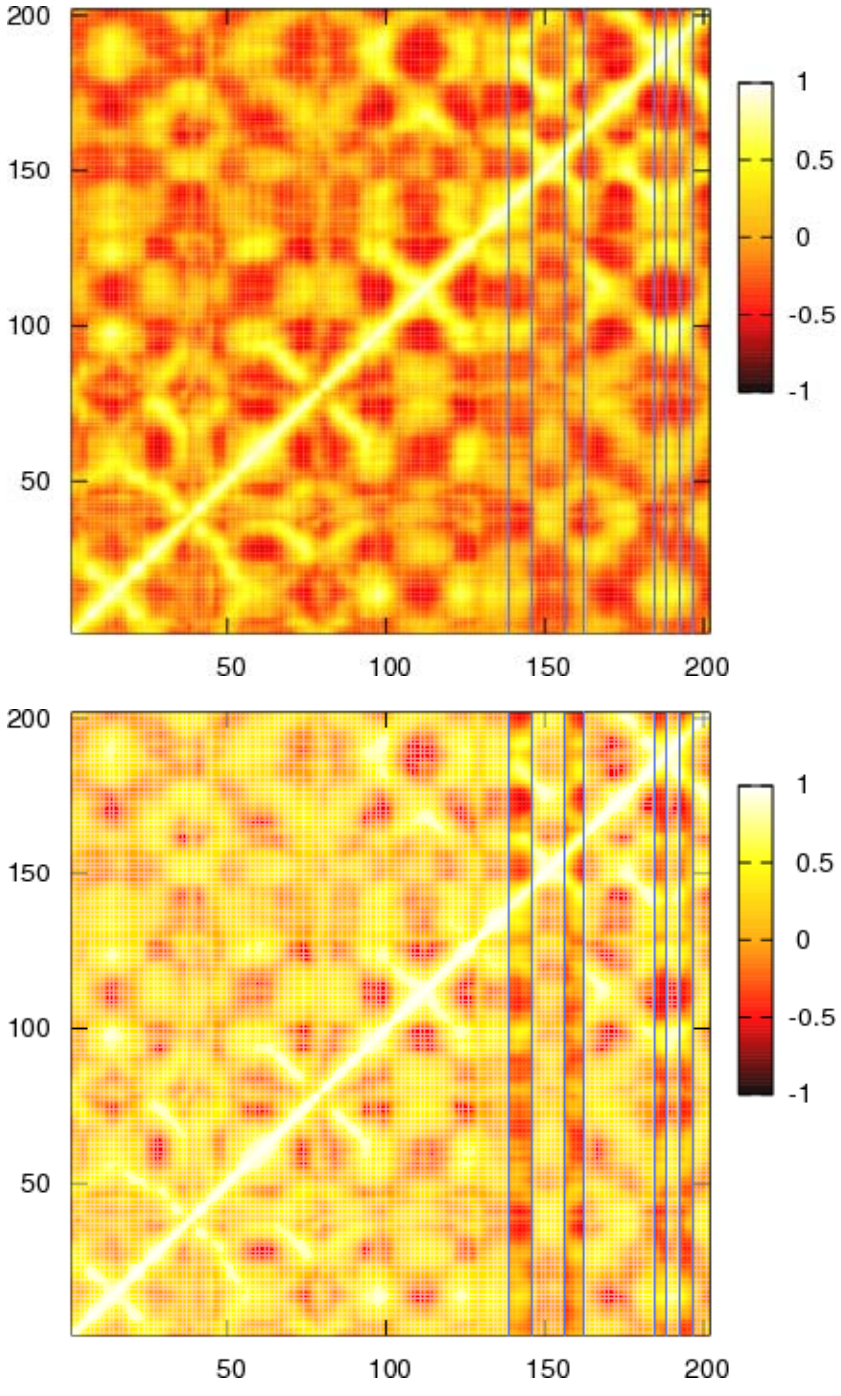

120
100
80
60
40
20
20
40
60
80
100
120
1
0.5
0
-0.5
-1

1POH

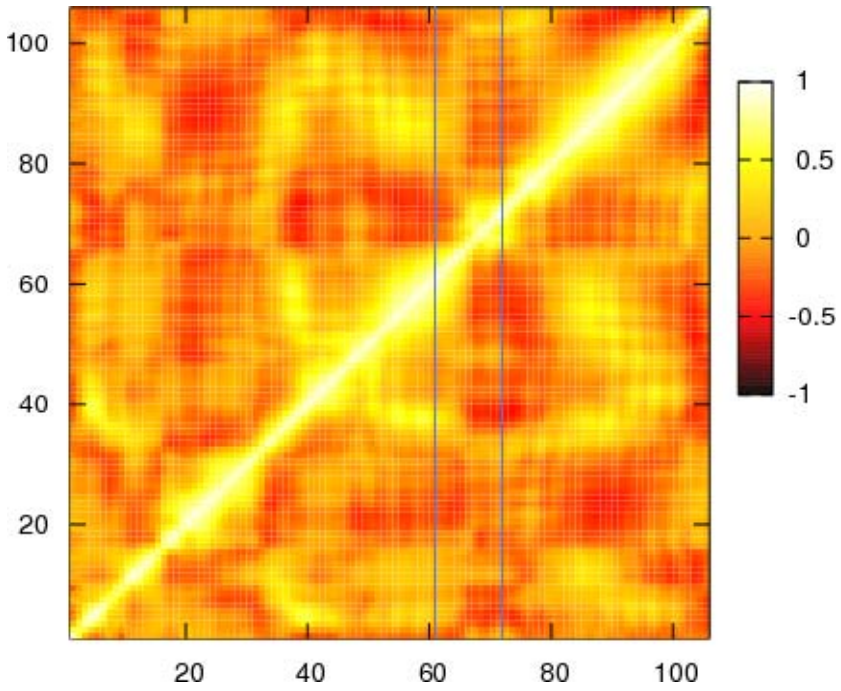

80
70
60
50
40
30
20
10
10
20
30
40
50
60
70
80
1
0.5
0
-0.5
-1

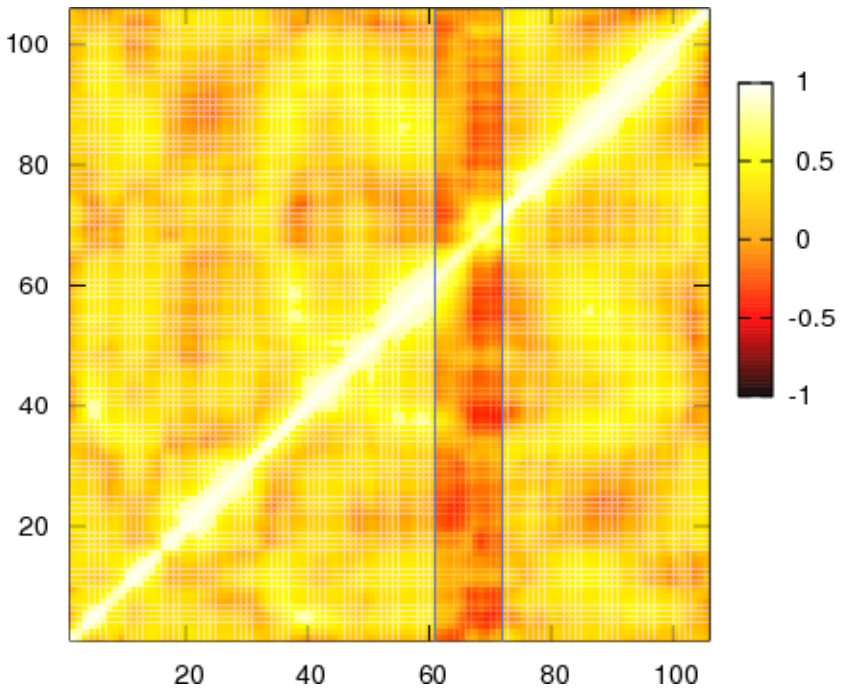

80
70
60
50
40
30
20
10
10
20
30
40
50
60
70
80
1
0.5
0
-0.5
-1

1TFH

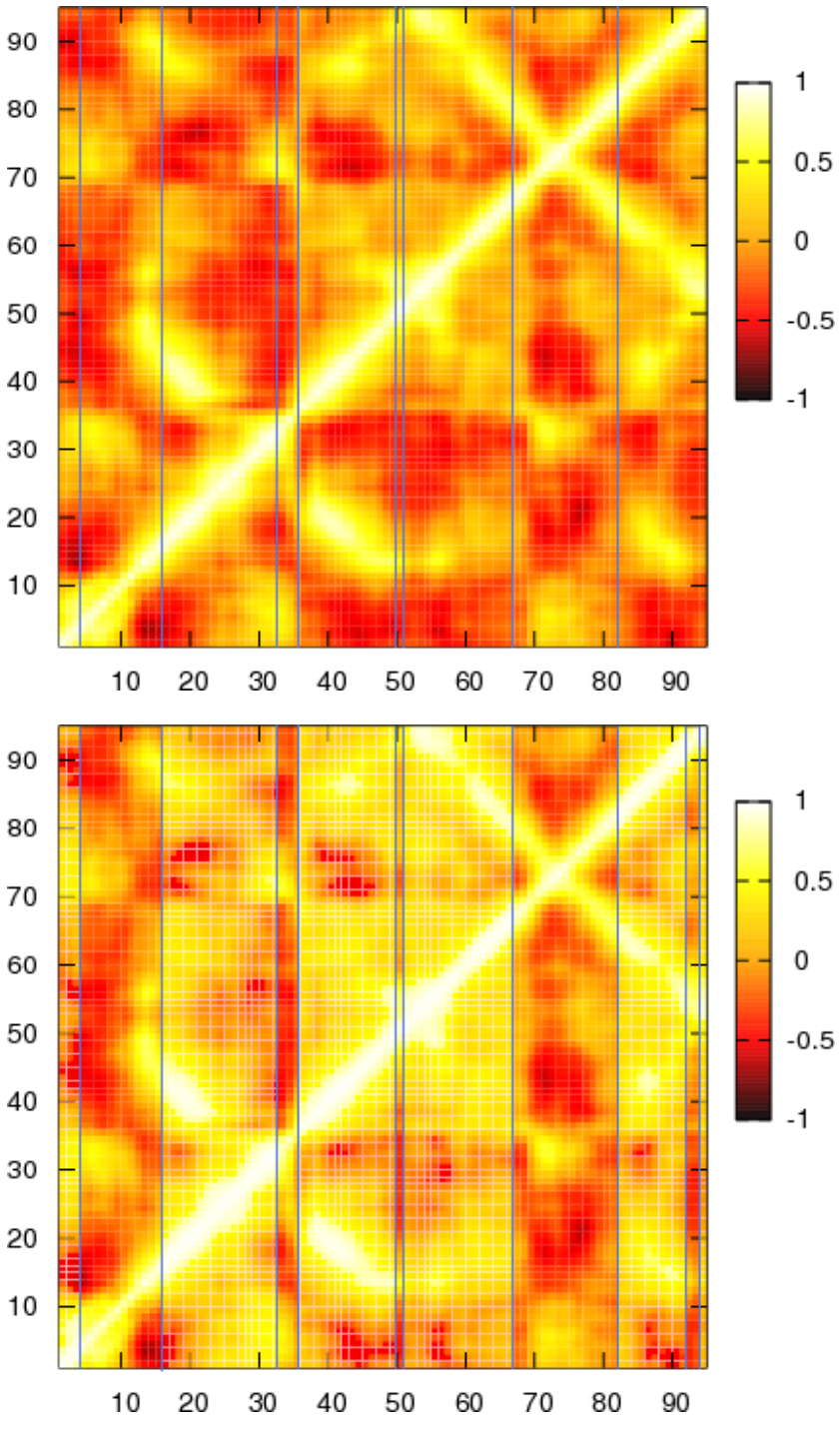

200
150
100
50
50
100
150
200
1
0.5
0
-0.5
-1
200
150
100
50
50
100
150
200
1
0.5
0
-0.5
-1

1UW3

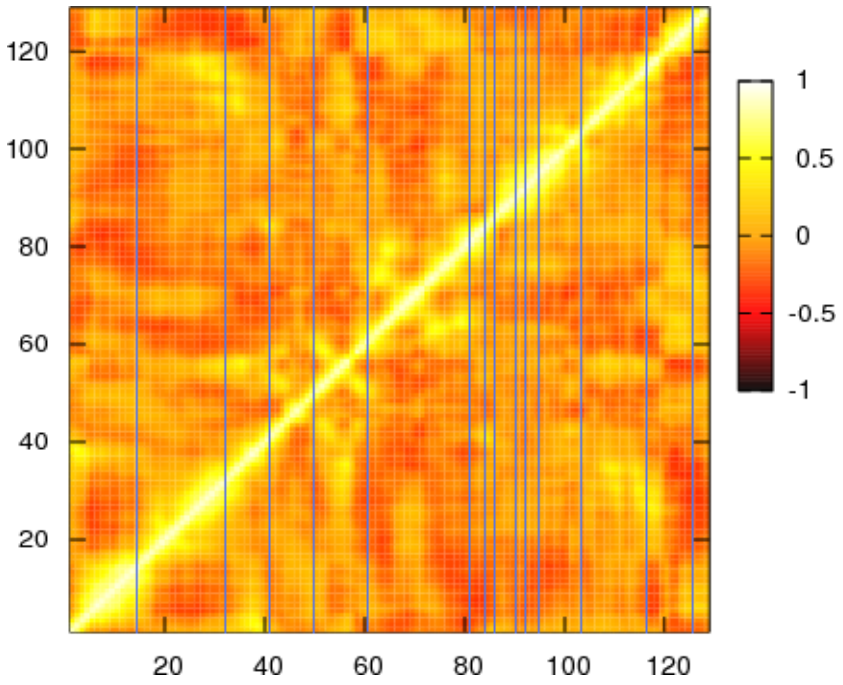

100
80
60
40
20
20
40
60
80
100
1
0.5
0
-0.5
-1

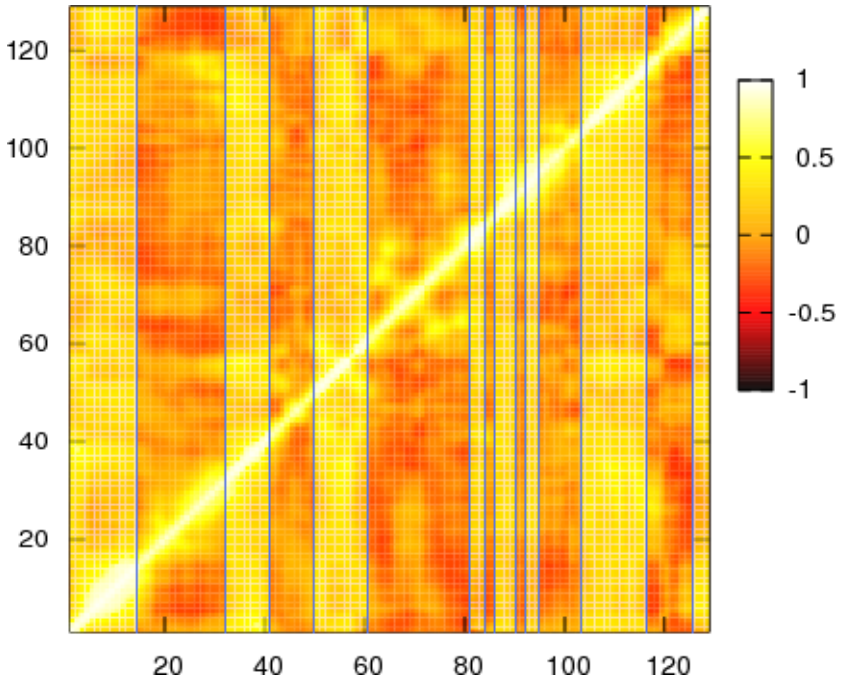

100
80
60
40
20
20
40
60
80
100
1
0.5
0
-0.5
-1

2VPF

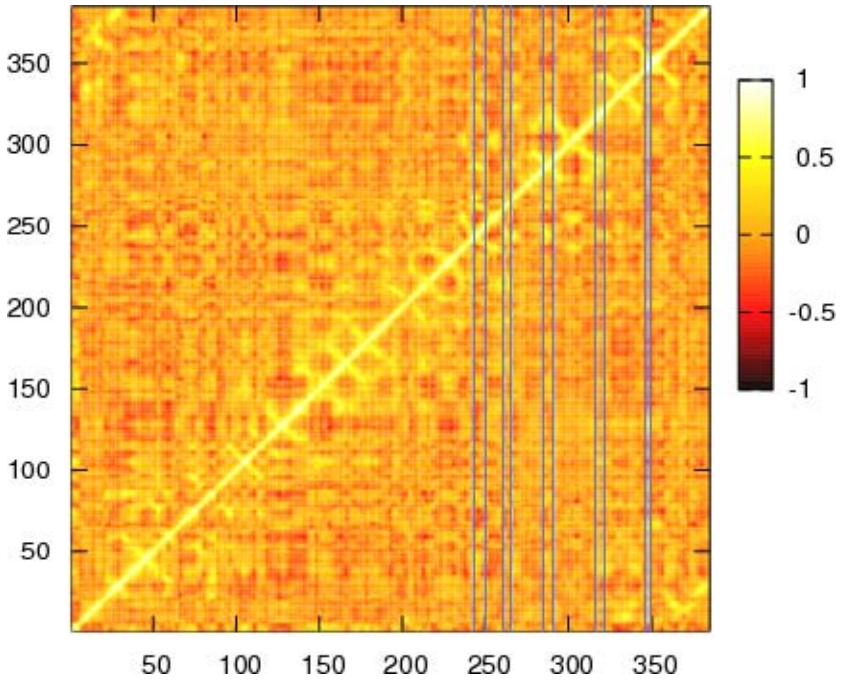

90
80
70
60
50
40
30
20
10
10 20 30 40 50 60 70 80 90
1
0.5
0
-0.5
-1
90
80
70
60
50
40
30
20
10
10 20 30 40 50 60 70 80 90
1
0.5
0
-0.5
-1

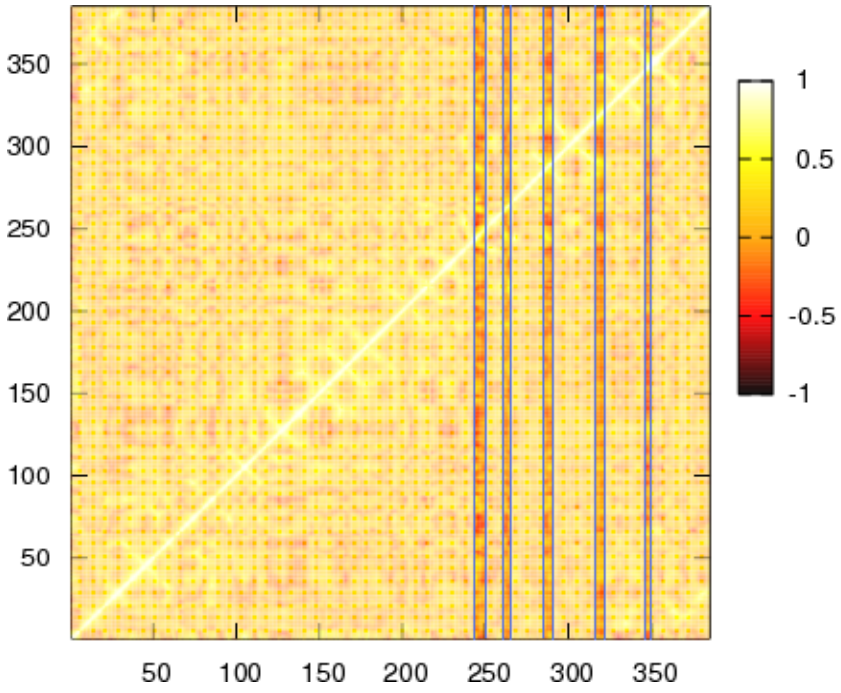
3LZT


120
100
80
60
40
20
20
40
60
80
100
120
1
0.5
0
-0.5
-1


120
100
80
60
40
20
20
40
60
80
100
120
1
0.5
0
-0.5
-1

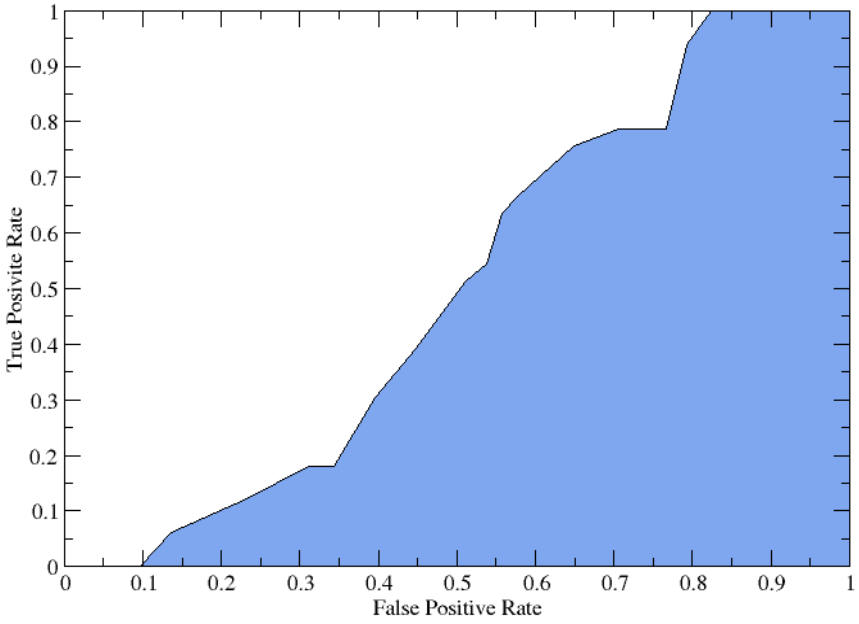
7NN9


350
300
250
200
150
100
50
50 100 150 200 250 300 350
1
0.5
0
-0.5
-1

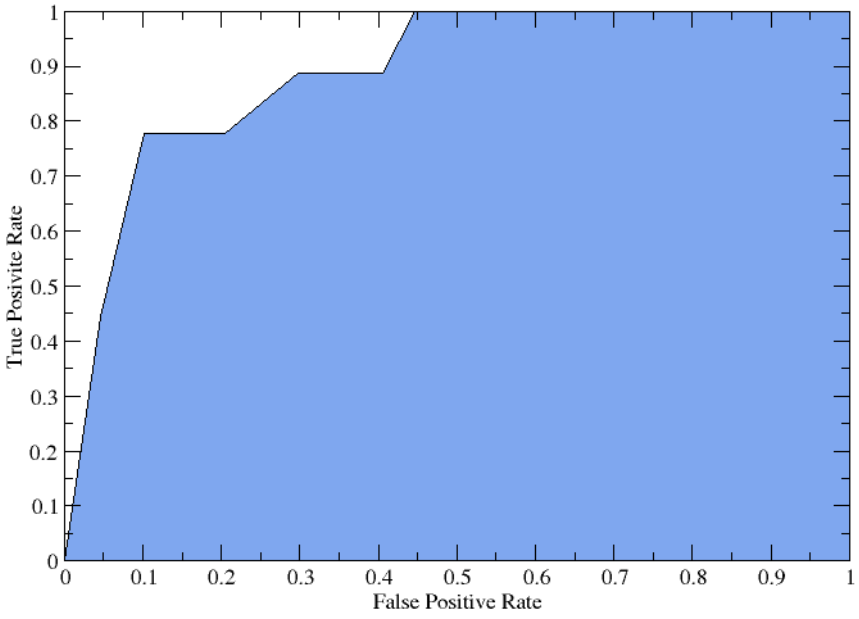

350
300
250
200
150
100
50
50 100 150 200 250 300 350
1
0.5
0
-0.5
-1

ROC curves

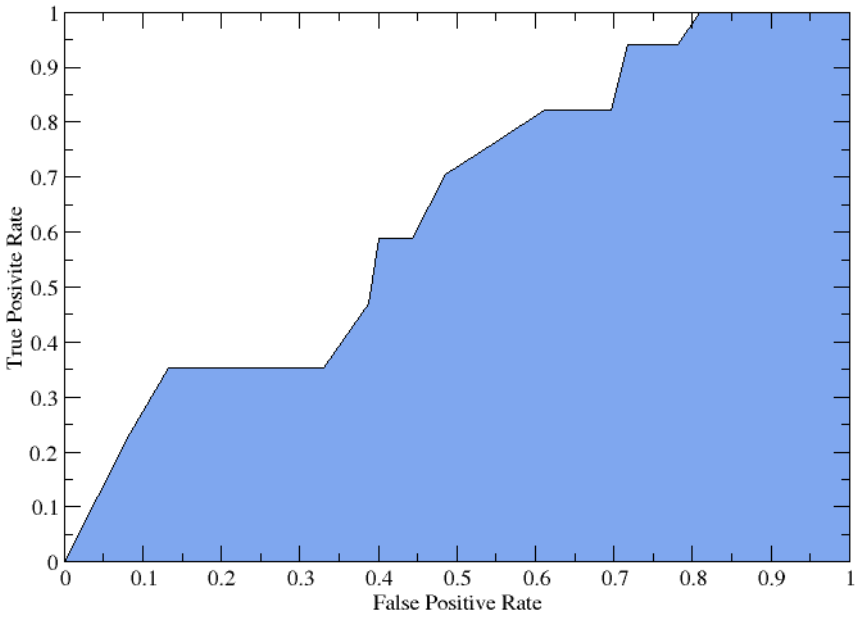
1AO3



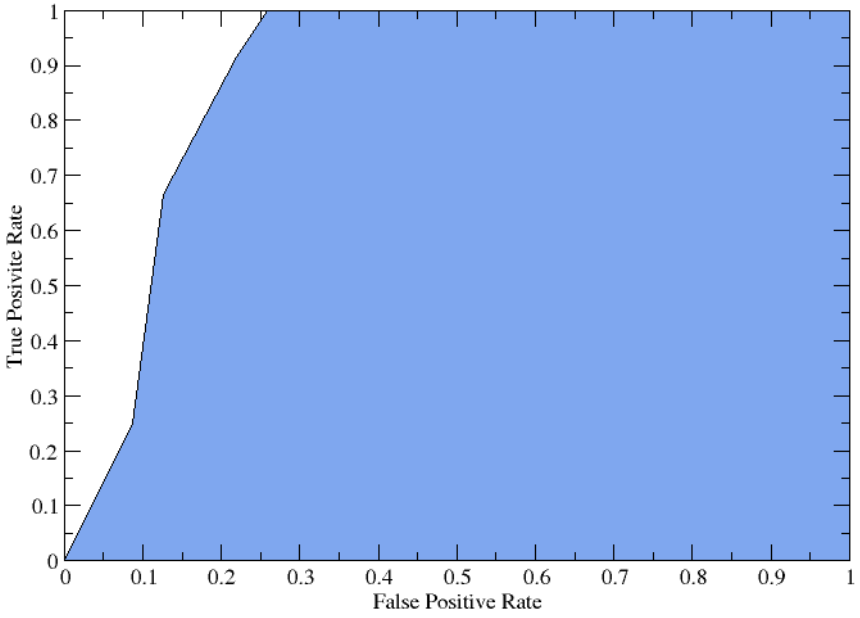
1AUQ



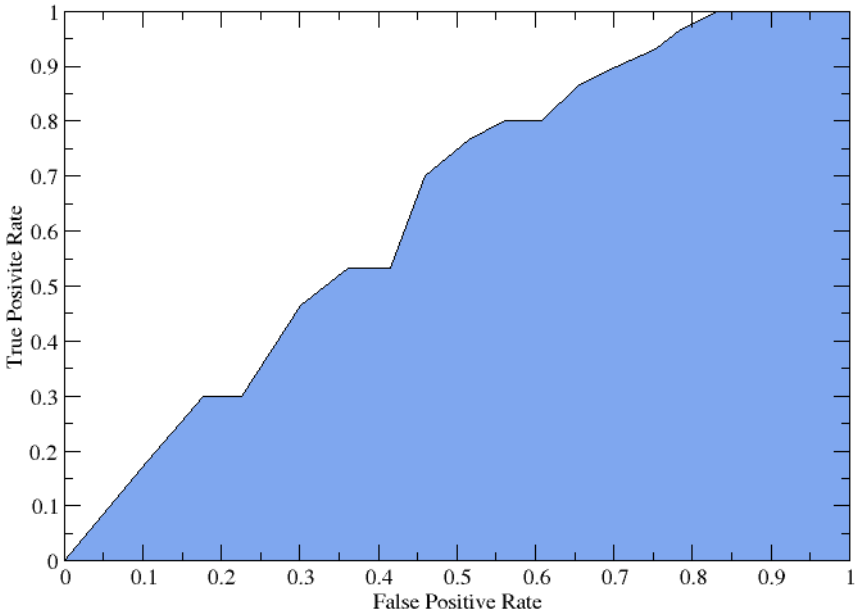
1BV1


True Posivite Rate
False Positive Rate
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

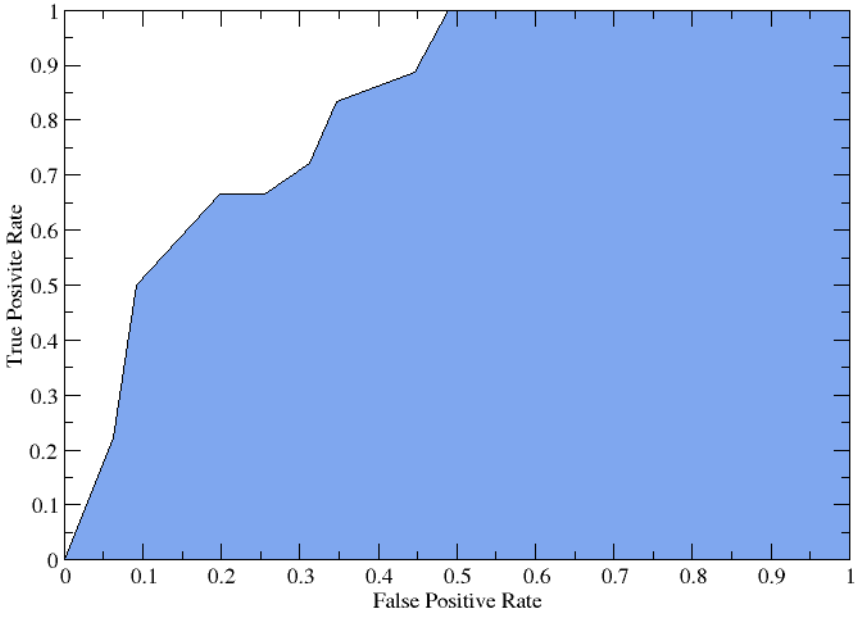
1CK4


True Posivite Rate
False Positive Rate
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

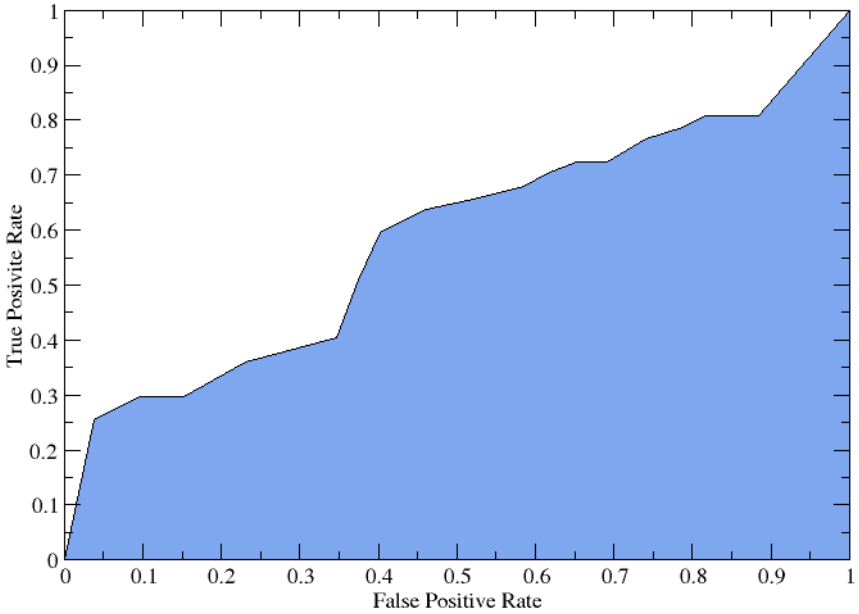
1CMW


True Posivite Rate
False Positive Rate
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

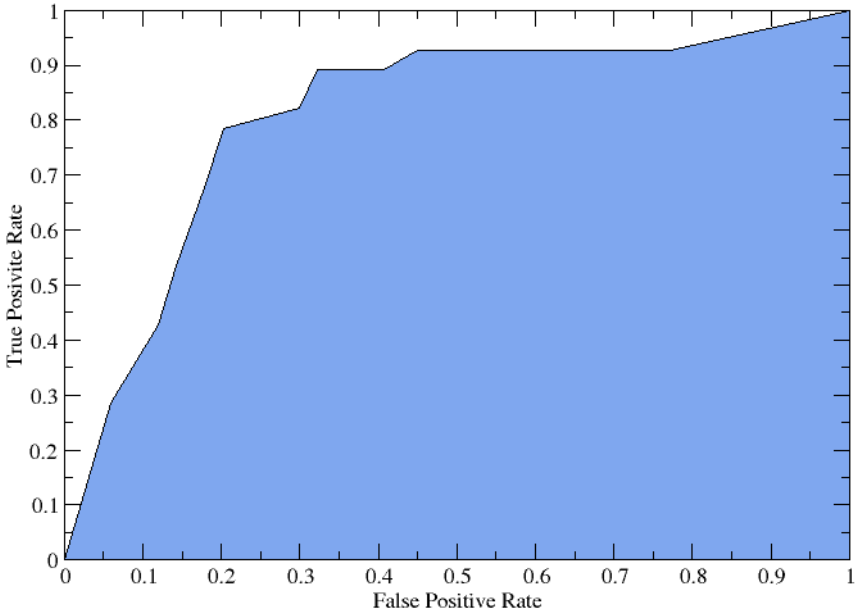
1D7P


True Posivite Rate
False Positive Rate
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

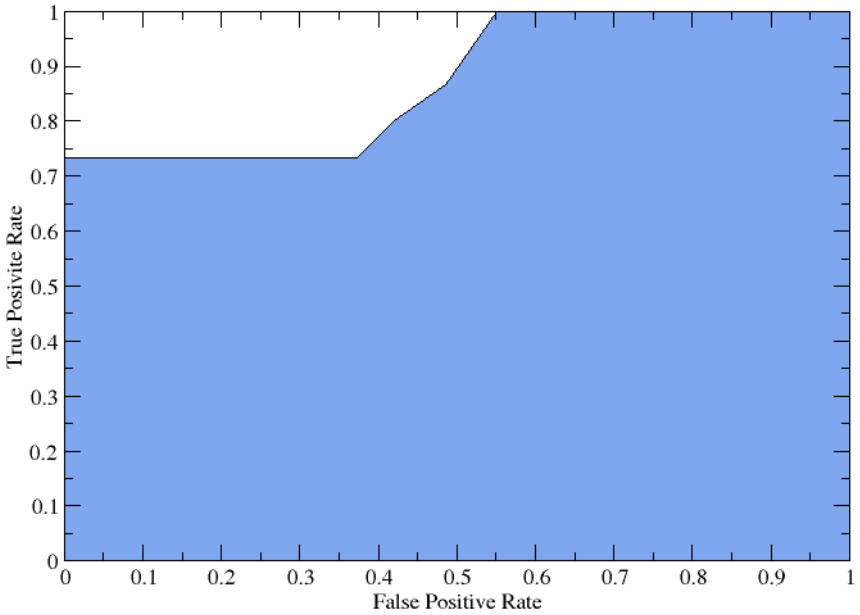
1GWP


True Posivite Rate
False Positive Rate
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

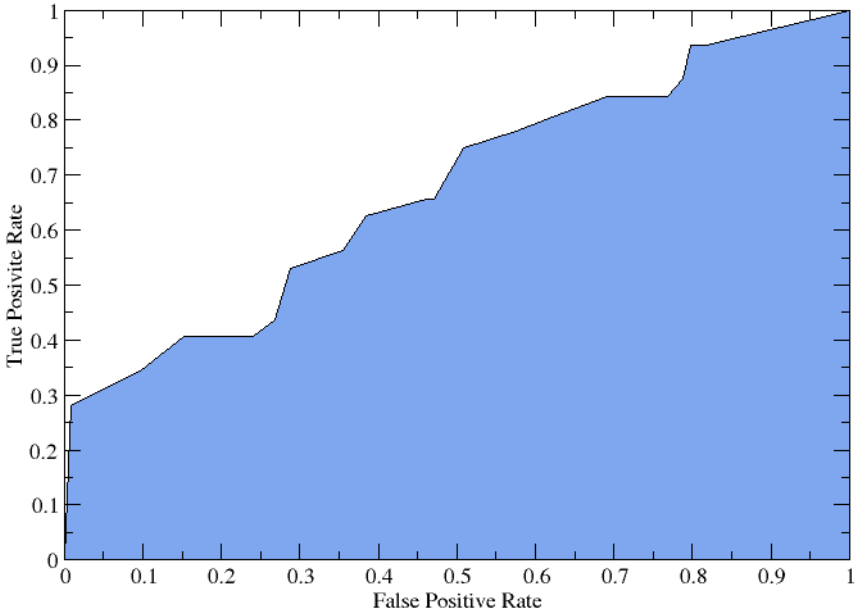
1HCN


True Posivite Rate
False Positive Rate
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

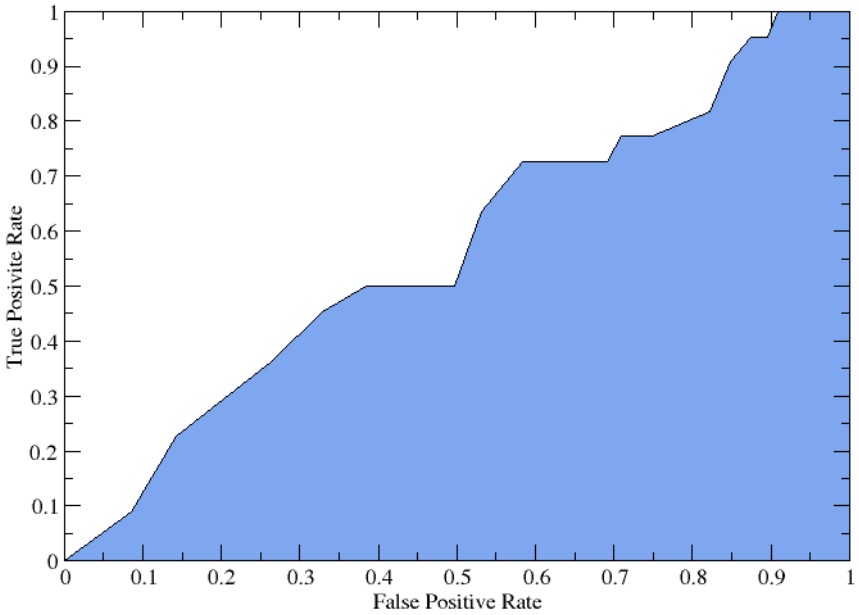
1K59


True Posivite Rate
False Positive Rate
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

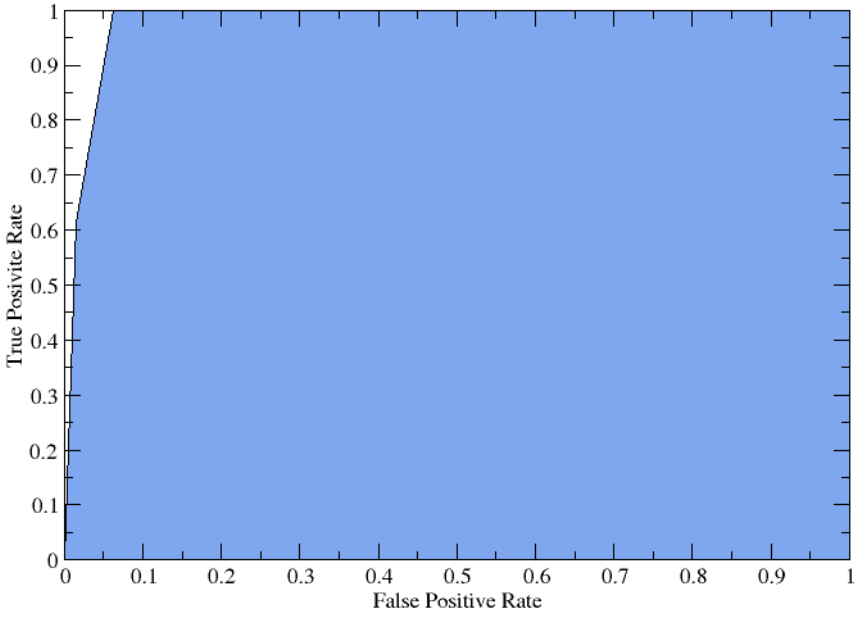
1KDC


True Posivite Rate
False Positive Rate
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

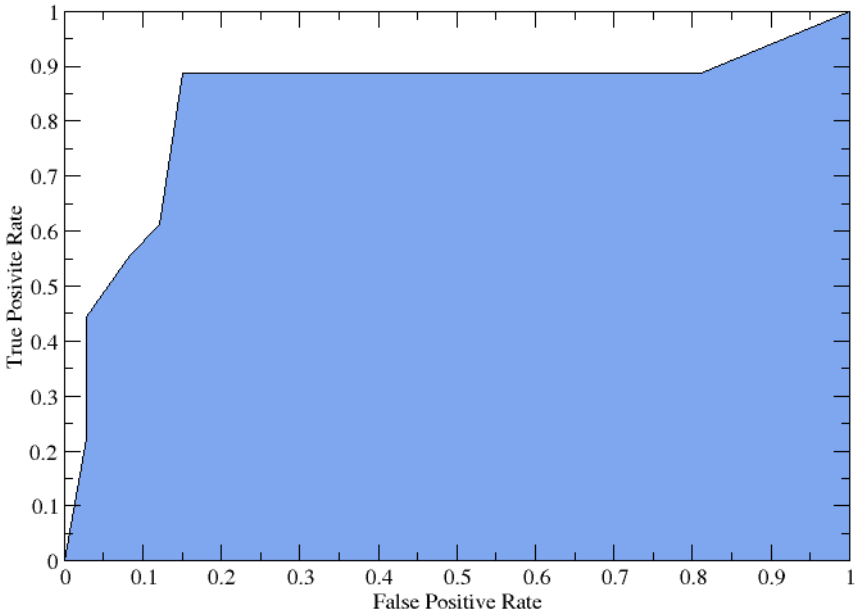
1KZQ


True Posivite Rate
False Positive Rate
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

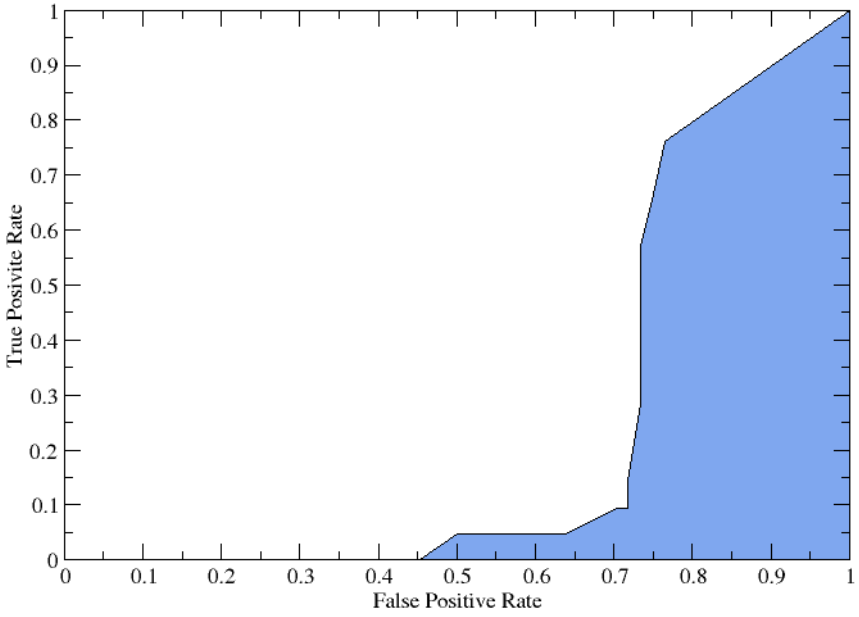
1P4P


True Posivite Rate
False Positive Rate
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

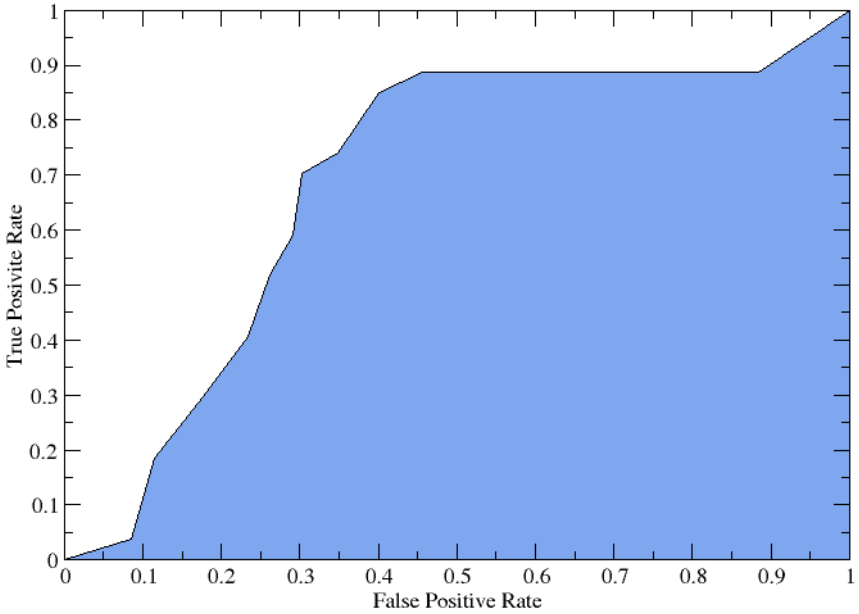
1PKO


True Posivite Rate
False Positive Rate
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

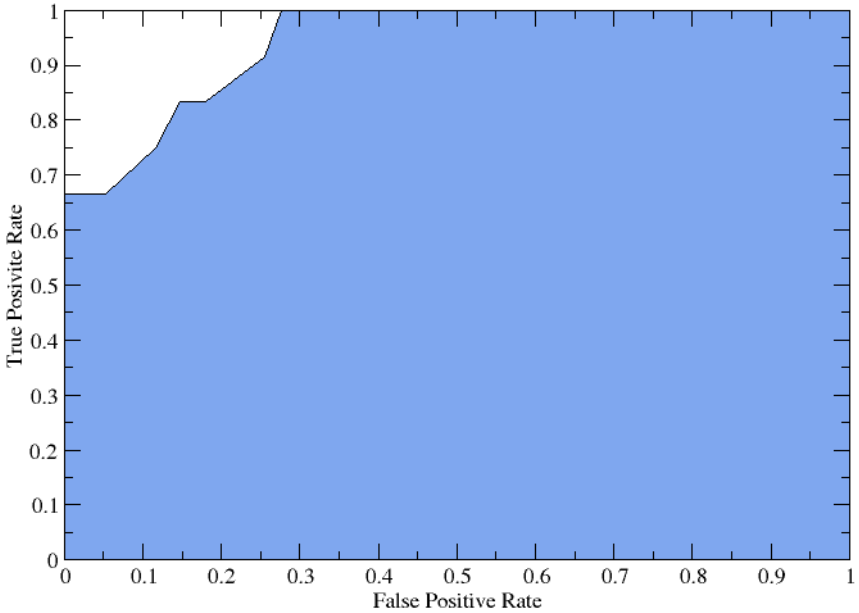
1POH


True Posivite Rate
False Positive Rate
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

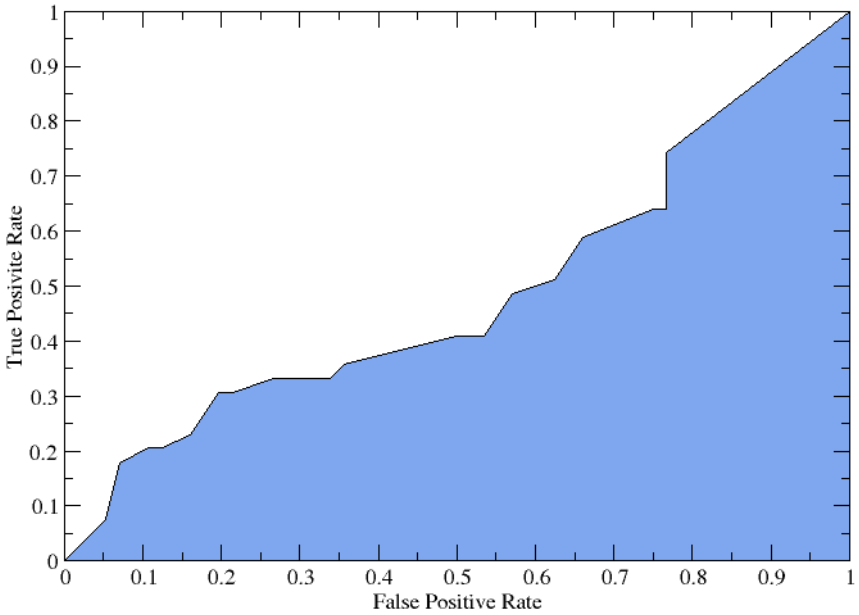
1TFH


True Posivite Rate
False Positive Rate
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

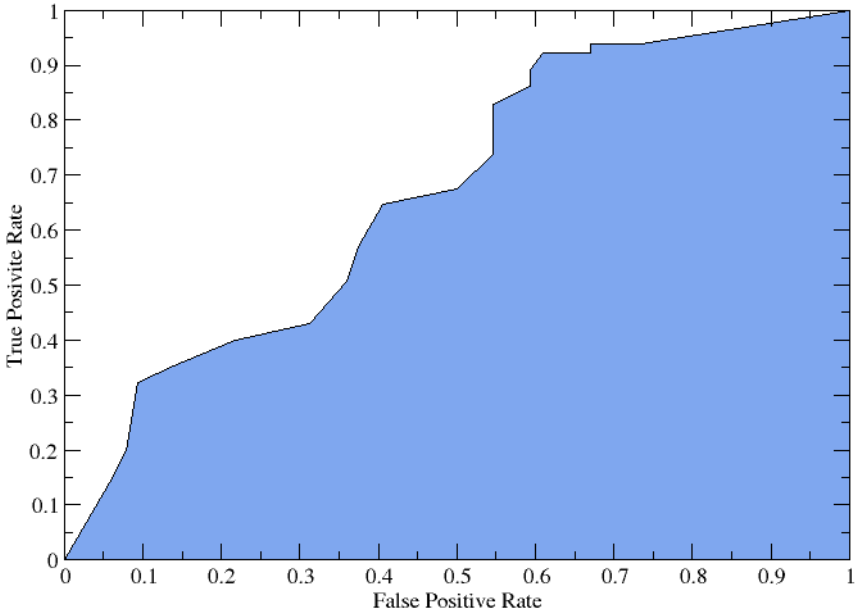
1UW3


True Posivite Rate
False Positive Rate
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

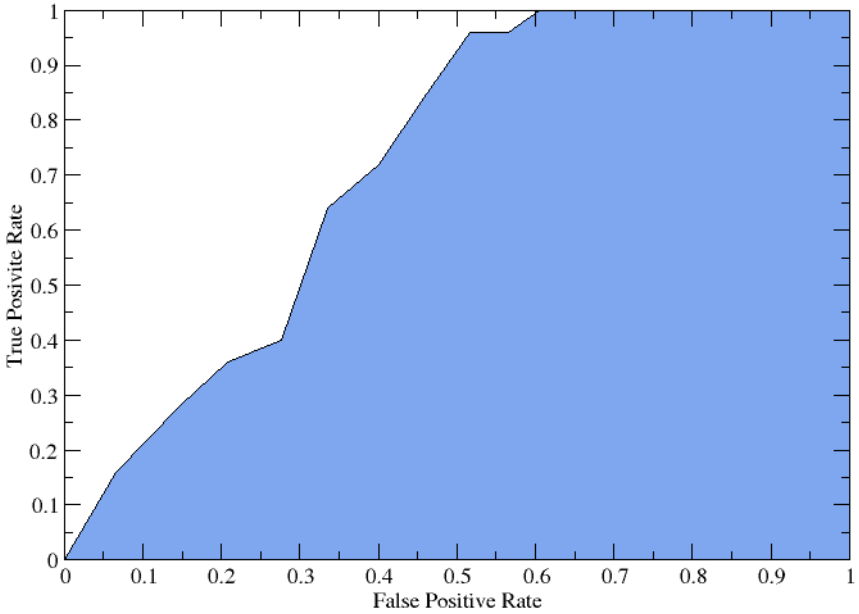
2VPF


True Posivite Rate
False Positive Rate
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

3LZT

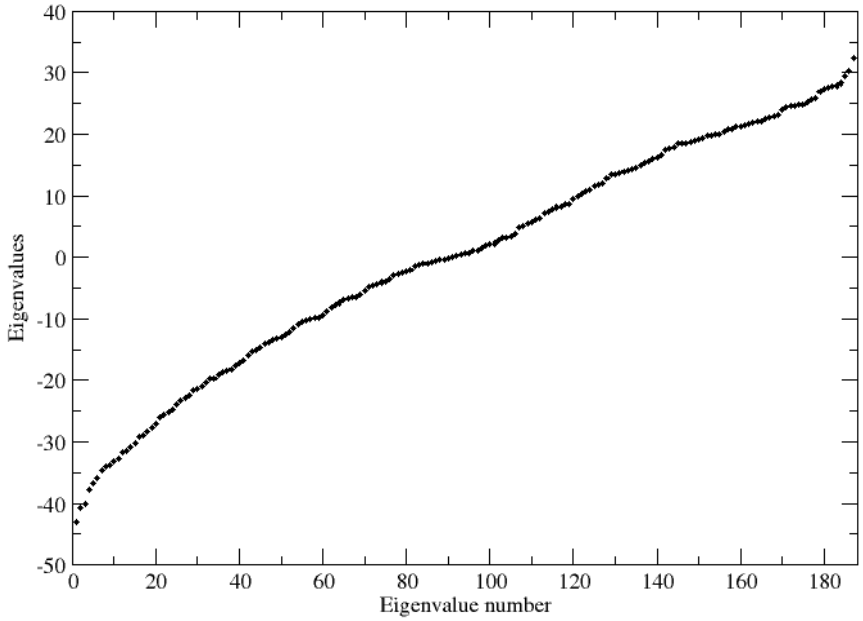

True Posivite Rate
False Positive Rate
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

7NN9

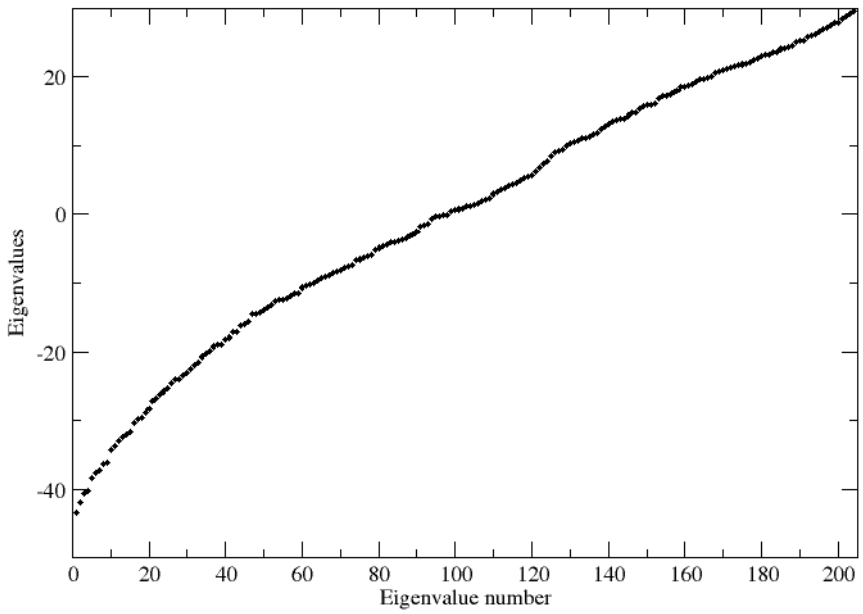

True Posivite Rate
False Positive Rate
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

Distribution of Eigenvalues Calculated with the Energy Decomposition approach.

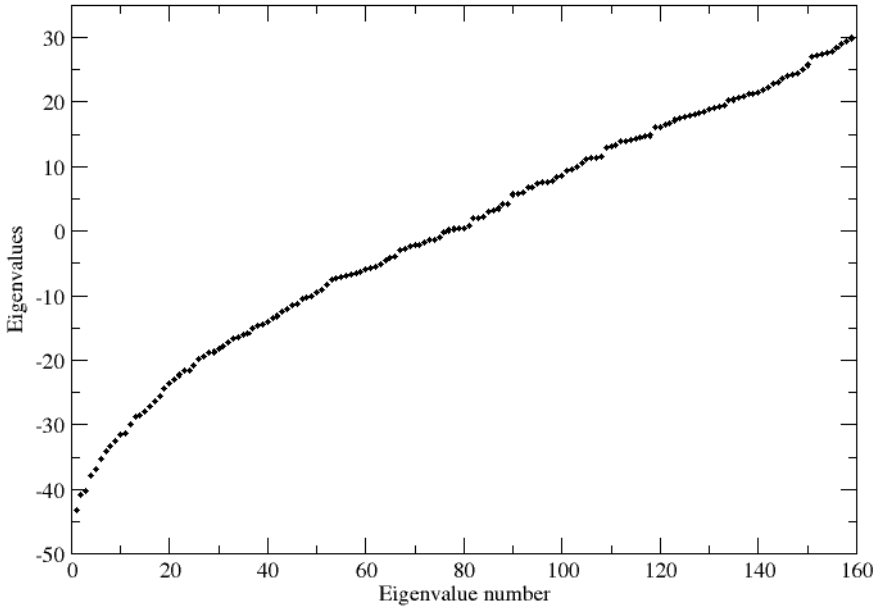
1AO3



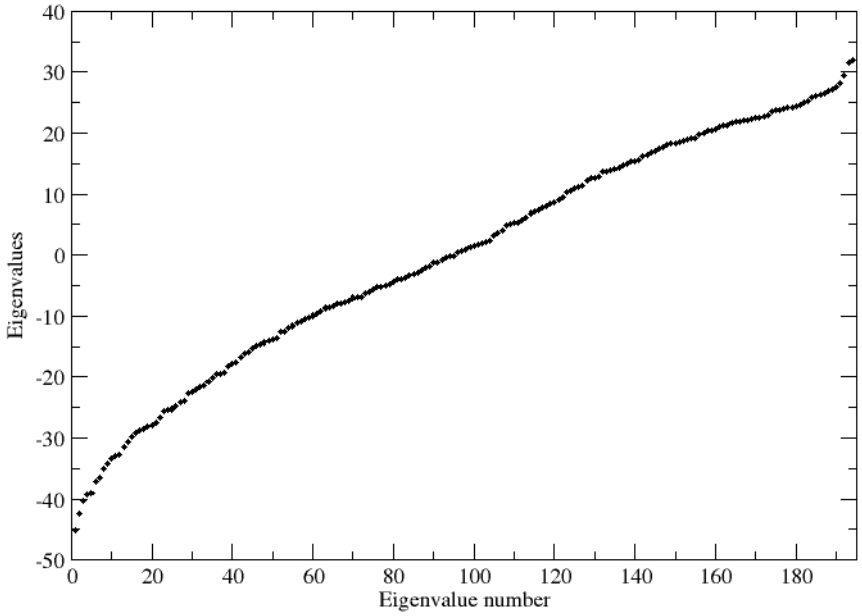
1AUQ



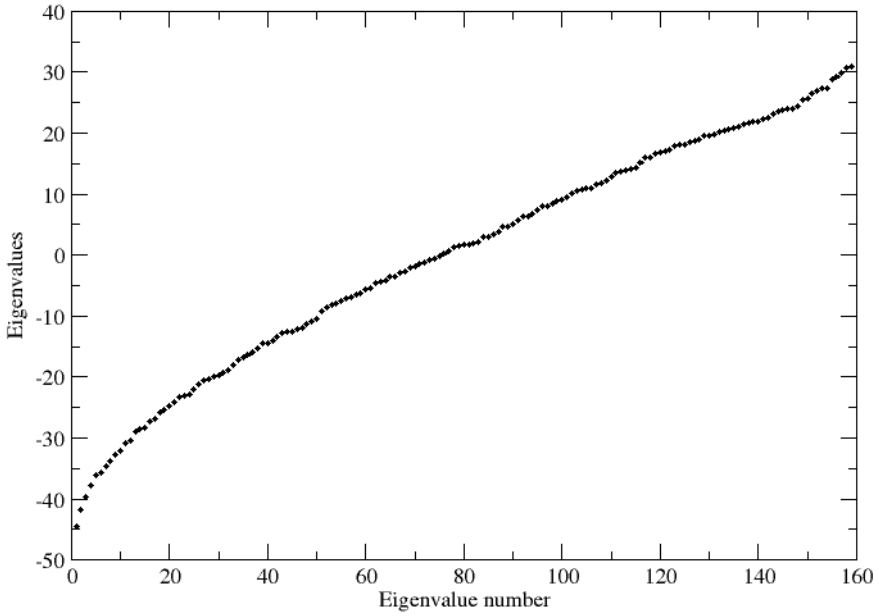
1BV1


Eigenvalues
Eigenvalue number
30
20
10
0
-10
-20
-30
-40
-50
0
20
40
60
80
100
120
140
160

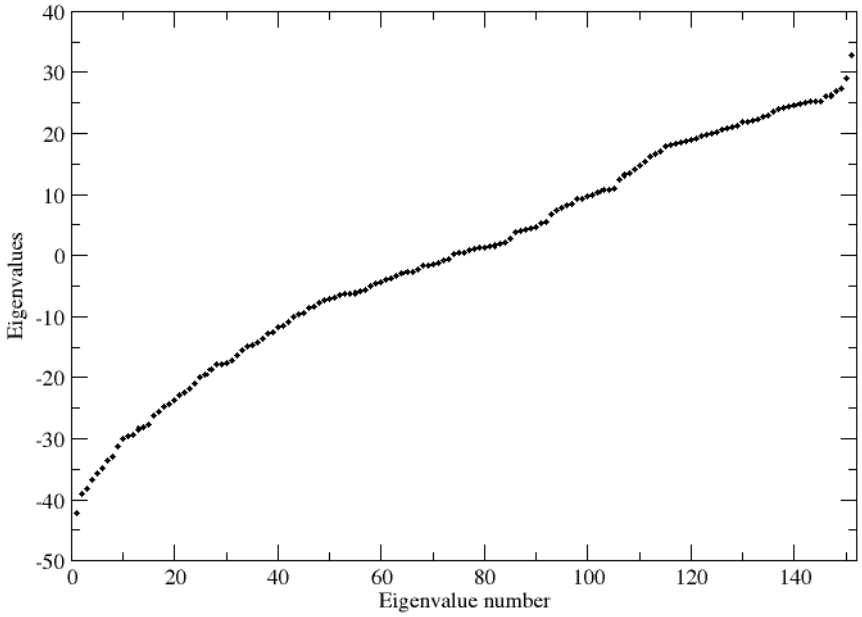
1CK4


Eigenvalues
Eigenvalue number
40
30
20
10
0
-10
-20
-30
-40
-50
0
20
40
60
80
100
120
140
160
180

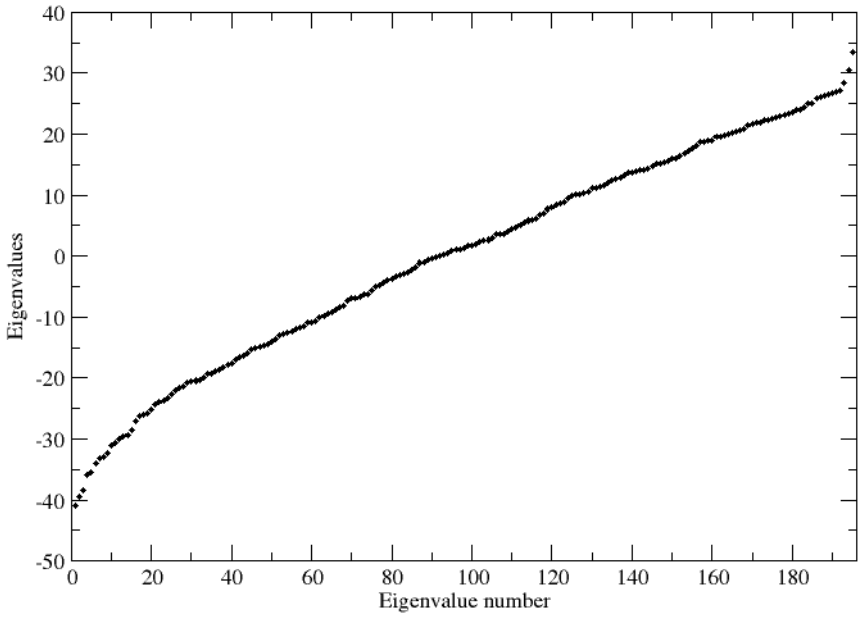
1D7P


Eigenvalues
Eigenvalue number
40
30
20
10
0
-10
-20
-30
-40
-50
0
20
40
60
80
100
120
140
160

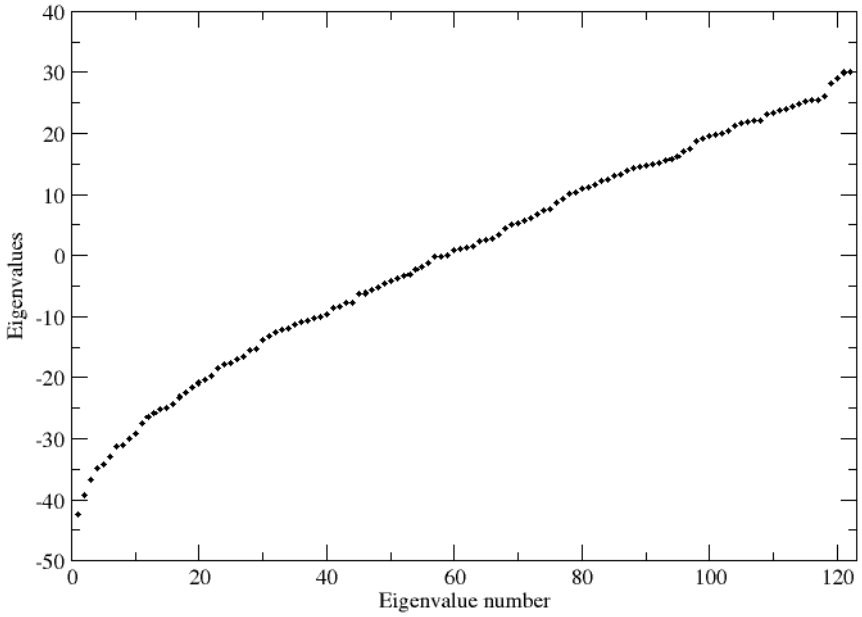
1GWP


Eigenvalues
Eigenvalue number
40
30
20
10
0
-10
-20
-30
-40
-50
0
20
40
60
80
100
120
140

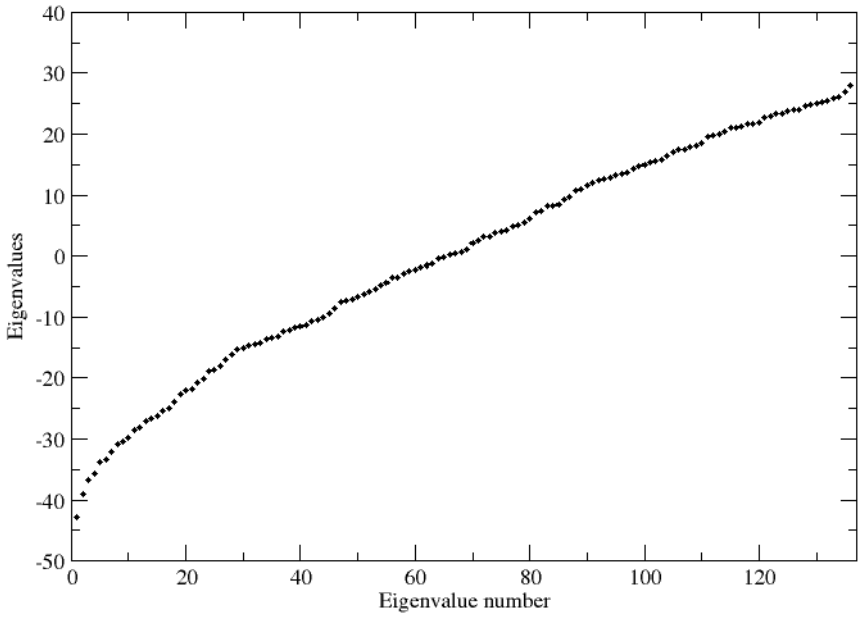
1HCN


Eigenvalues
Eigenvalue number
40
30
20
10
0
-10
-20
-30
-40
-50
0
20
40
60
80
100
120
140
160
180

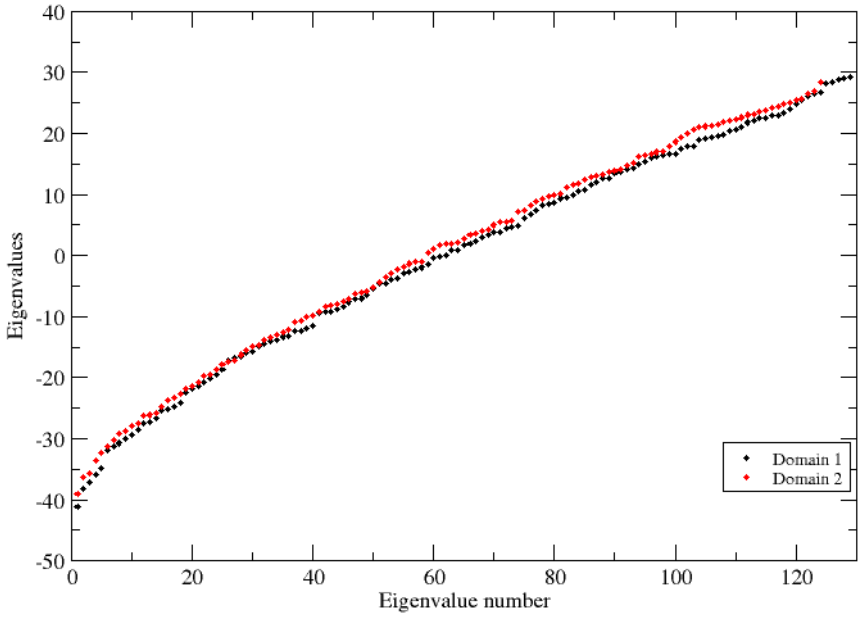
1K59


Eigenvalues
Eigenvalue number
40
30
20
10
0
-10
-20
-30
-40
-50
0
20
40
60
80
100
120

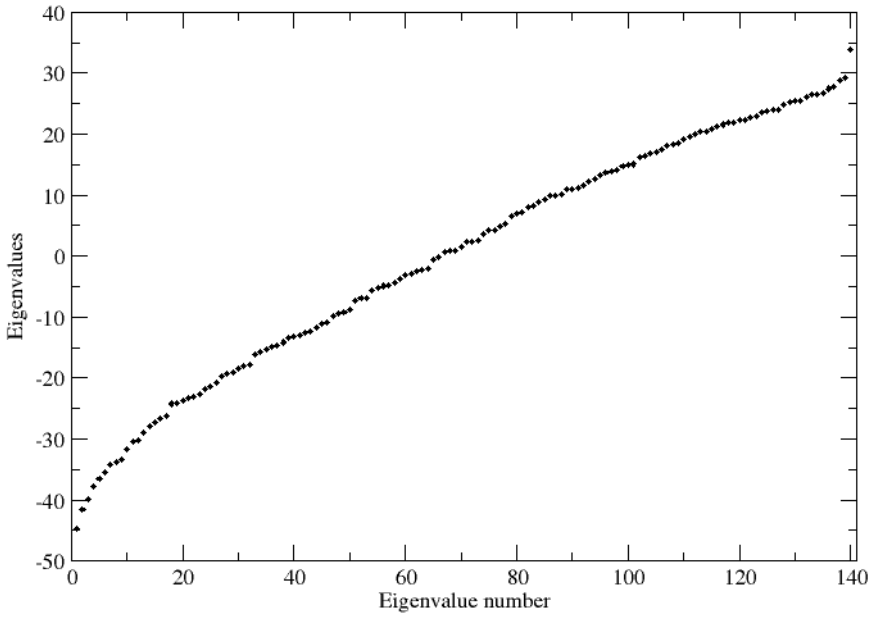
1KDC


Eigenvalues
Eigenvalue number
40
30
20
10
0
-10
-20
-30
-40
-50
0
20
40
60
80
100
120

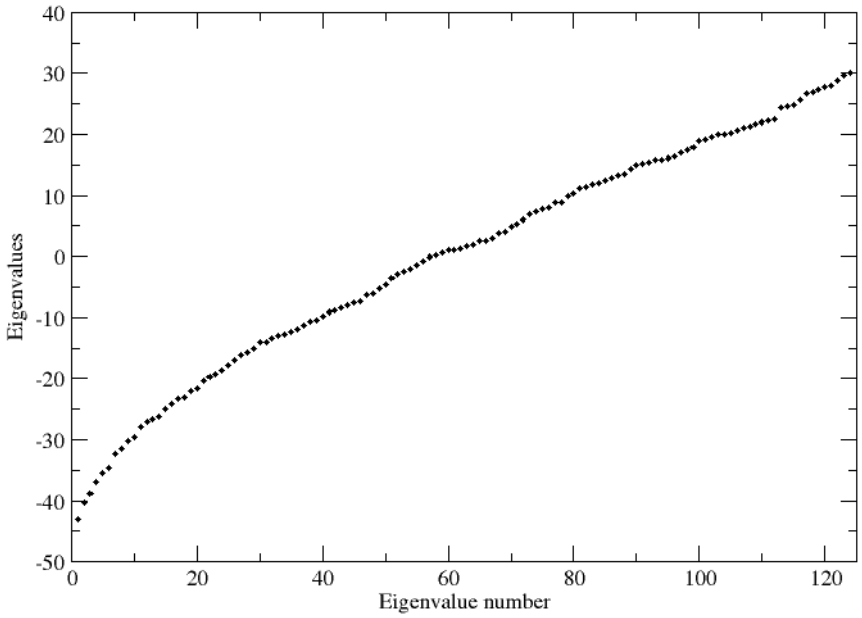
1KZQ


Eigenvalues
Eigenvalue number
Domain 1
Domain 2
40
30
20
10
0
-10
-20
-30
-40
-50
0
20
40
60
80
100
120

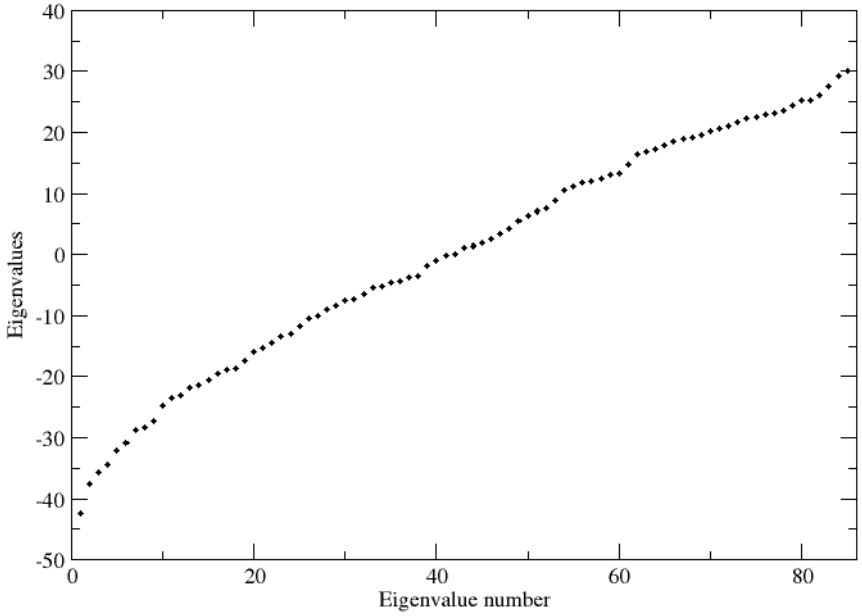
1P4P



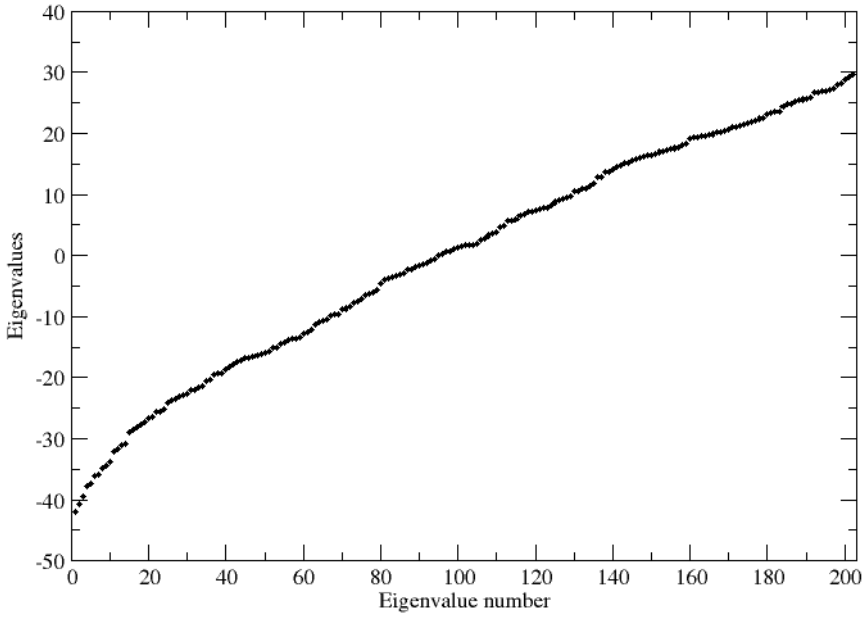
1PKO



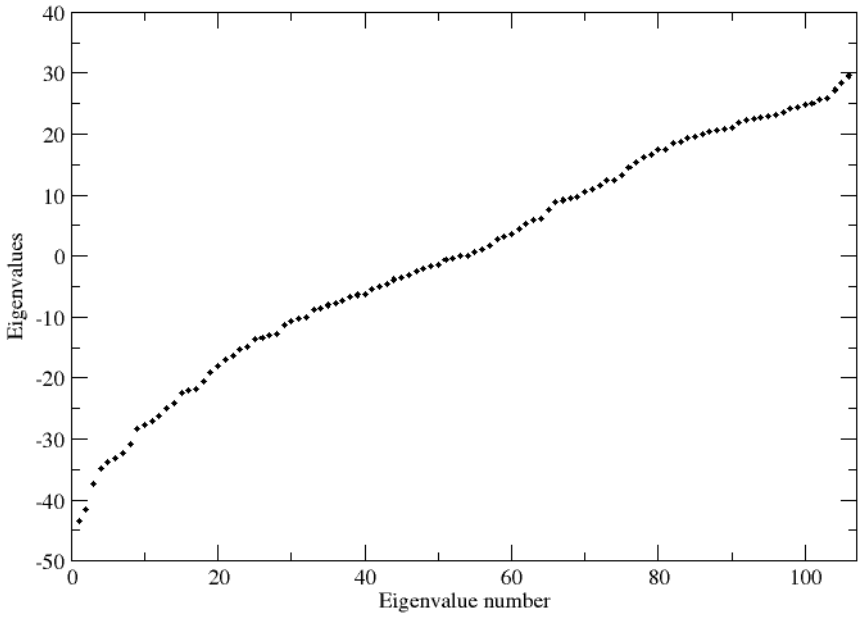
1POH


Eigenvalues
Eigenvalue number
40
30
20
10
0
-10
-20
-30
-40
-50
0
20
40
60
80

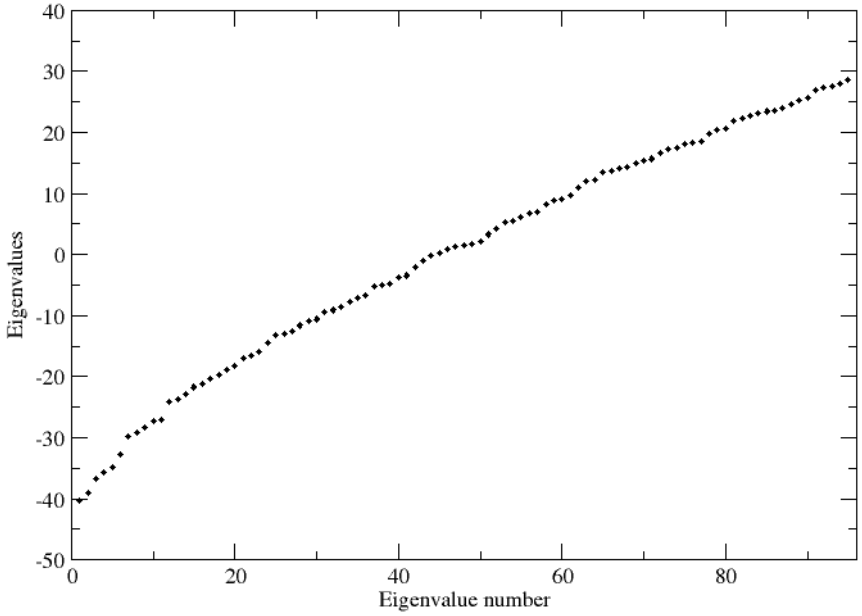
1TFH


Eigenvalues
Eigenvalue number
40
30
20
10
0
-10
-20
-30
-40
-50
0
20
40
60
80
100
120
140
160
180
200

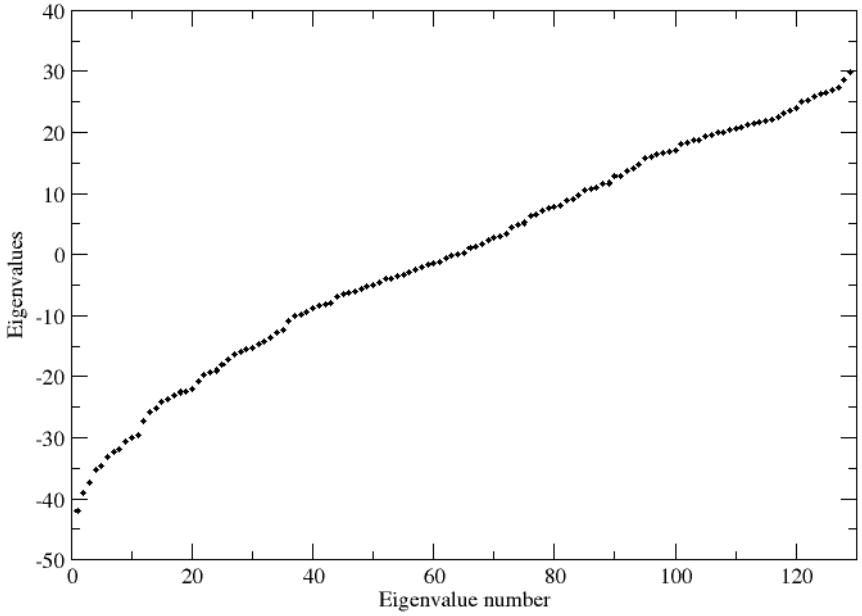
1UW3



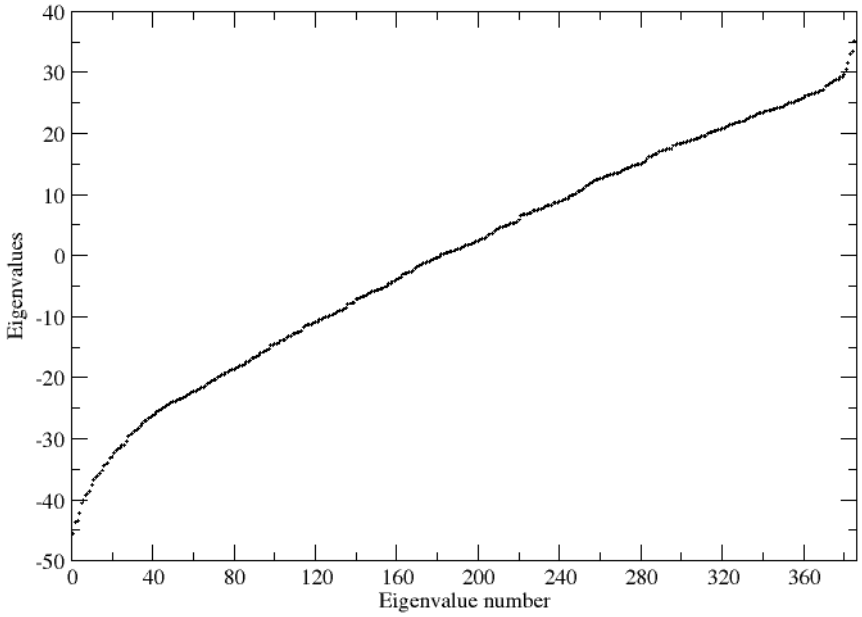
2VPF



3LZT

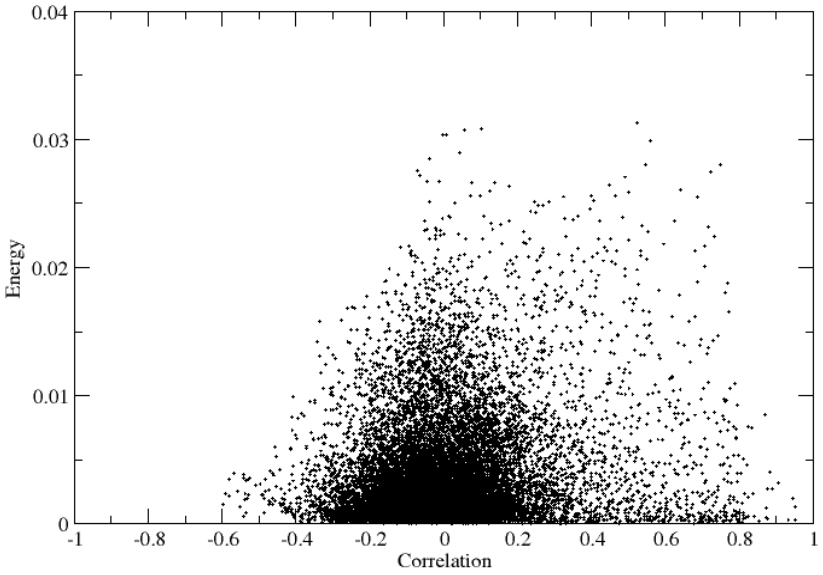

Eigenvalues
Eigenvalue number
40
30
20
10
0
-10
-20
-30
-40
-50
0
20
40
60
80
100
120

7NN9


Eigenvalues
Eigenvalue number
40
30
20
10
0
-10
-20
-30
-40
-50
0
40
80
120
160
200
240
280
320
360

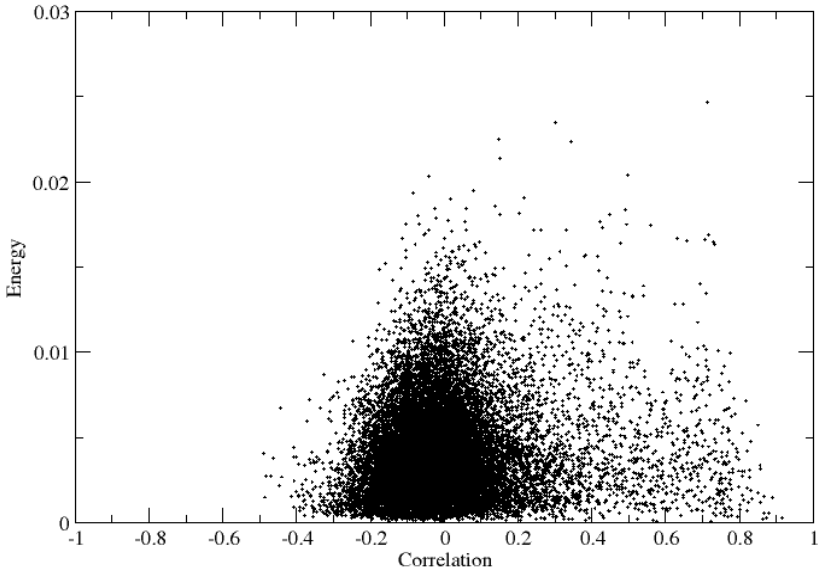
Comparison of the simplified energy matrices for each protein with the respective residue-residue cross correlation matrix.

1AO3



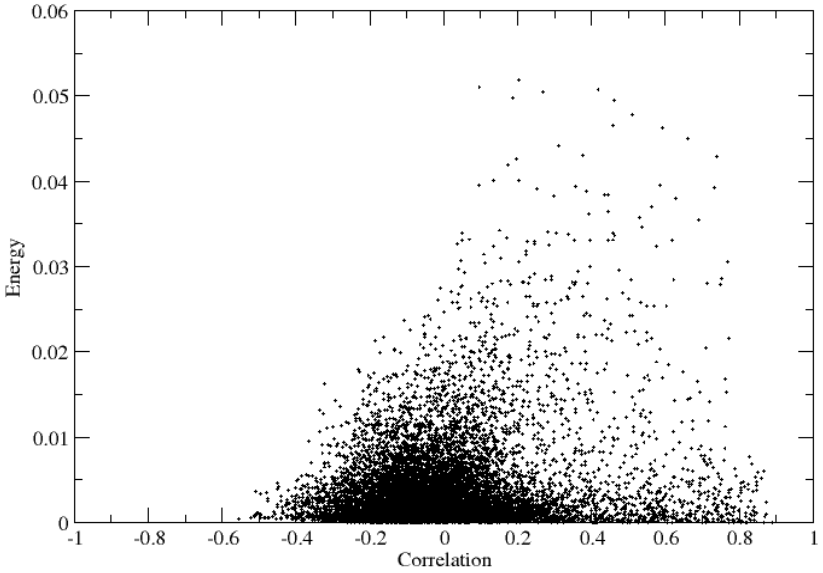
Correlation coefficient = 0.16

1AUQ



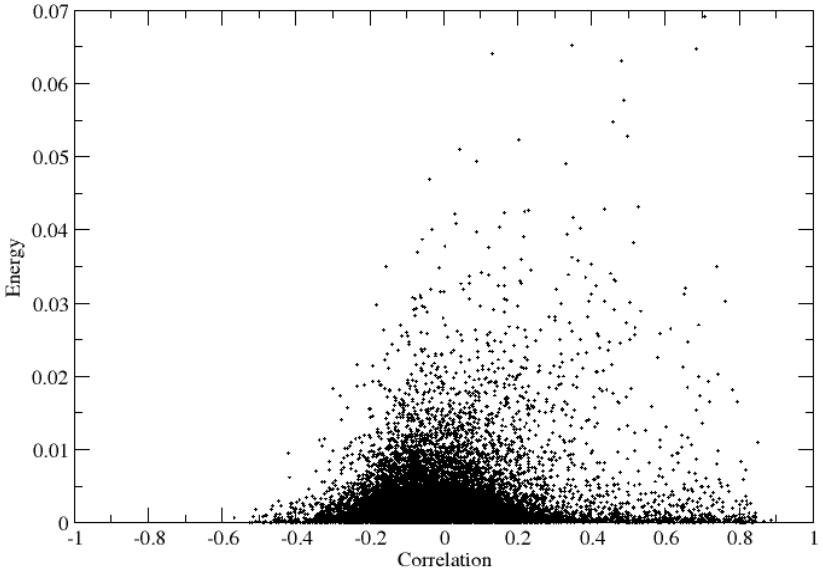
Correlation coefficient = 0.09

1BV1



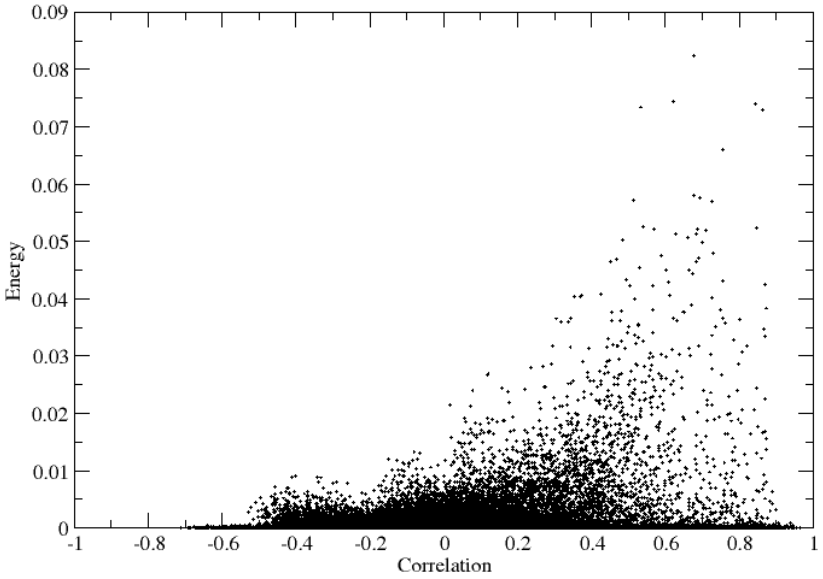
Correlation coefficient = 0.18

1CK4



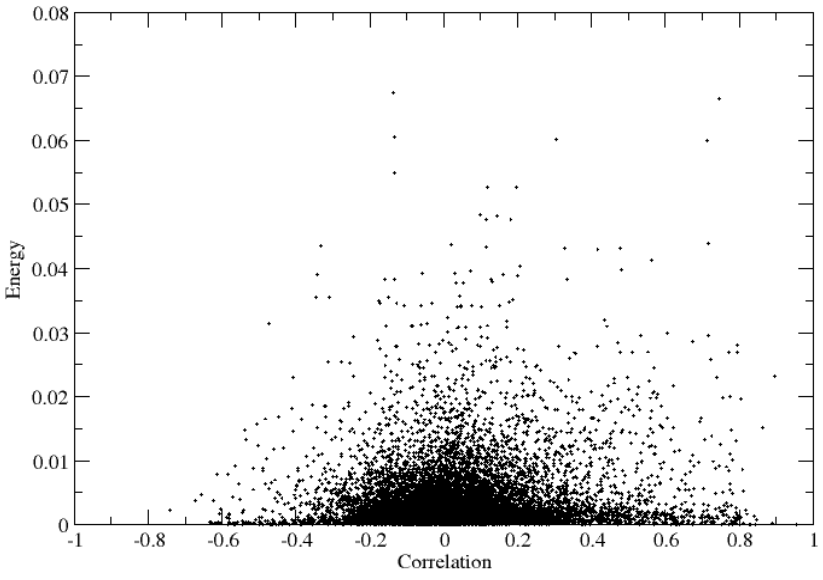
Correlation coefficient = 0.13

1CMW



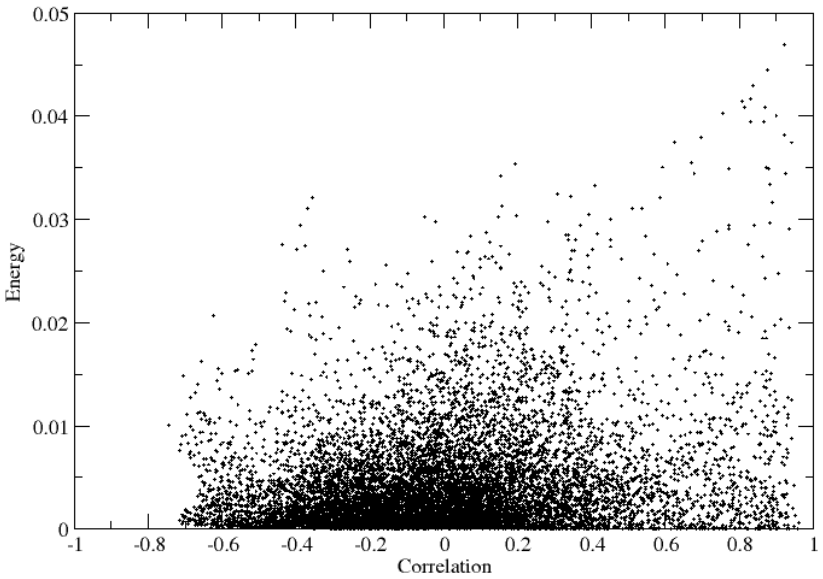
Correlation coefficient = 0.27

1D7P



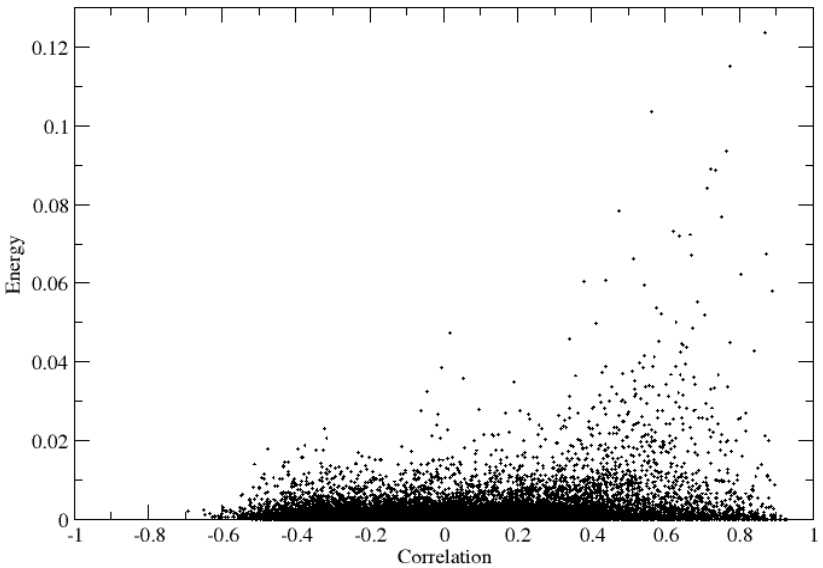
Correlation coefficient = 0.13

1GWP



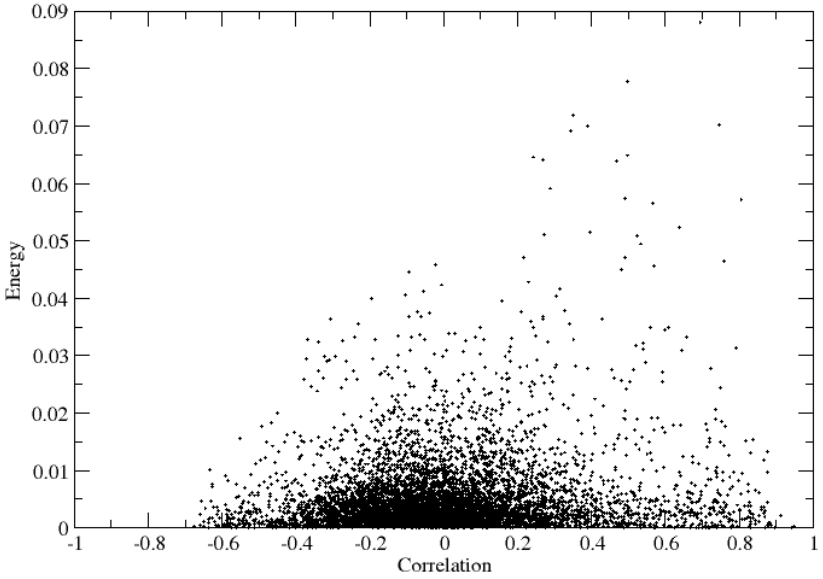
Correlation coefficient = 0.20

1HCN



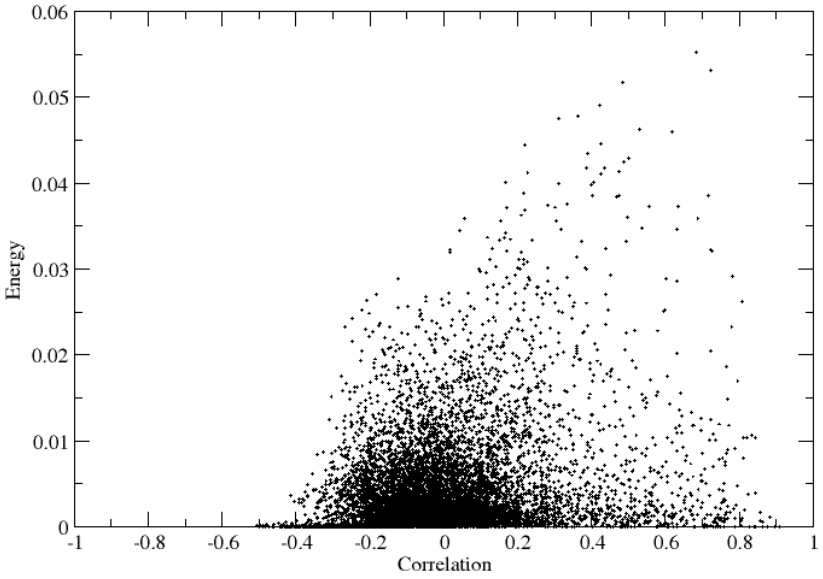
Correlation coefficient = 0.27

1K59



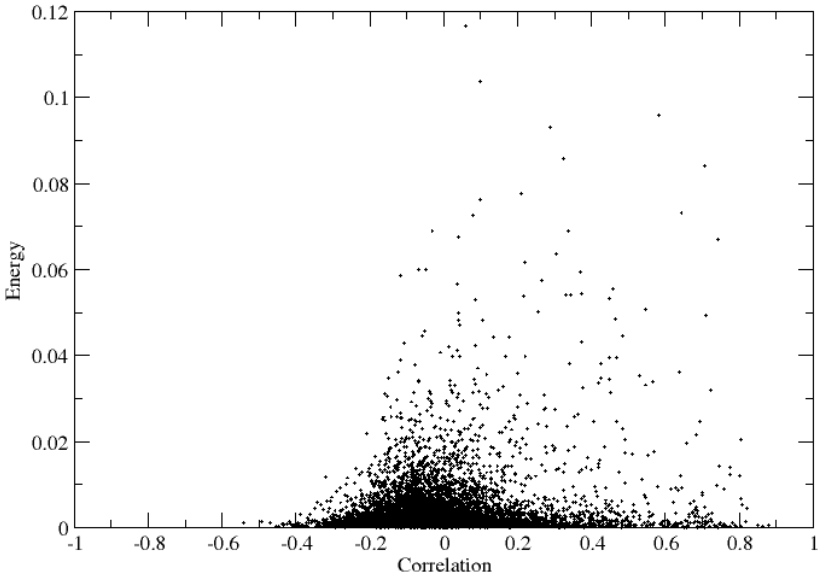
Correlation coefficient = 0.14

1KDC



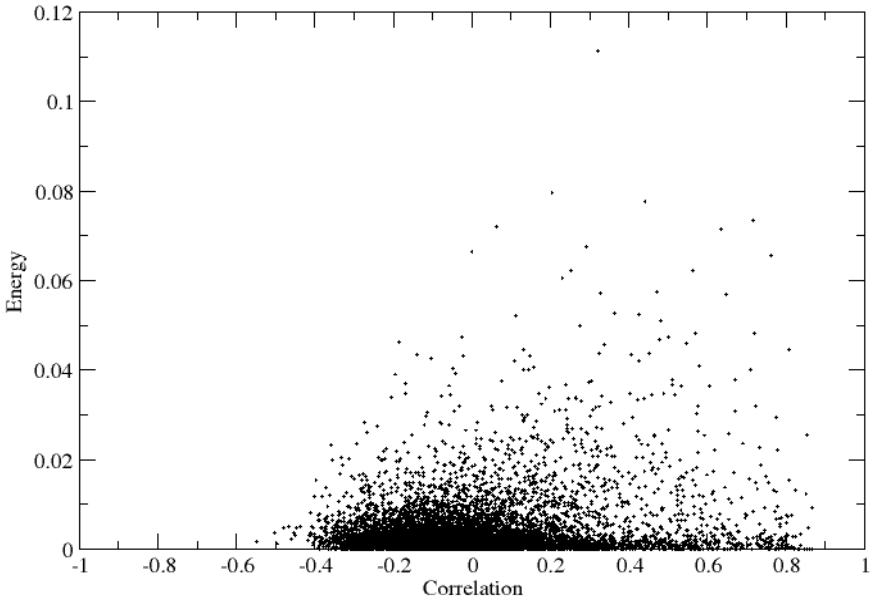
Correlation coefficient = 0.21

1KZQ



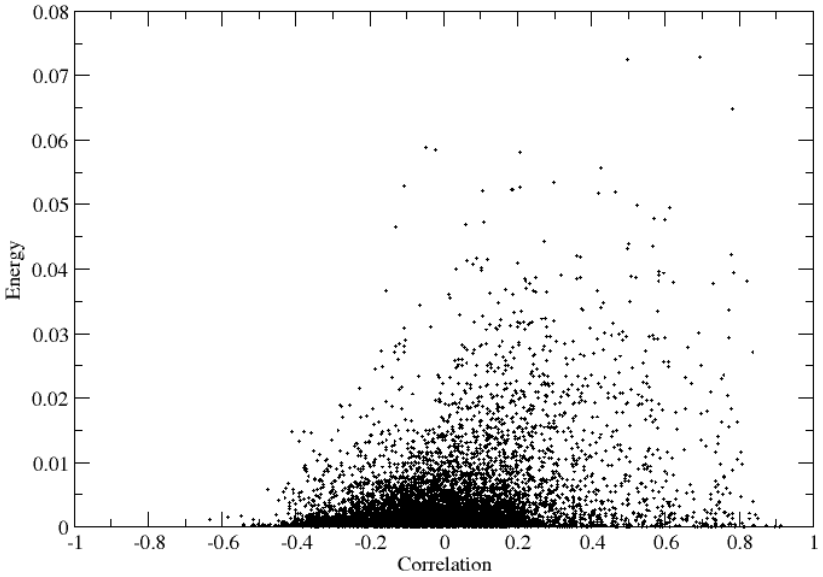
Correlation coefficient = 0.12

1P4P



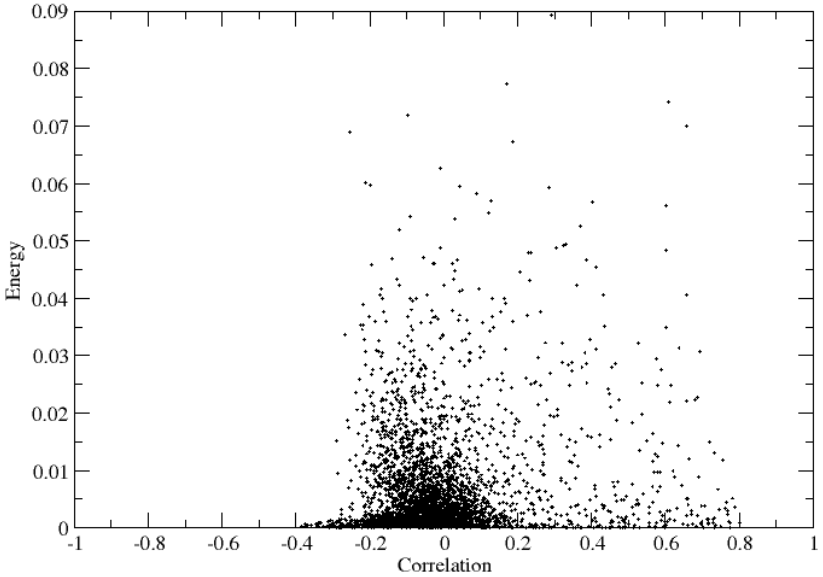
Correlation coefficient = 0.13

1PKO



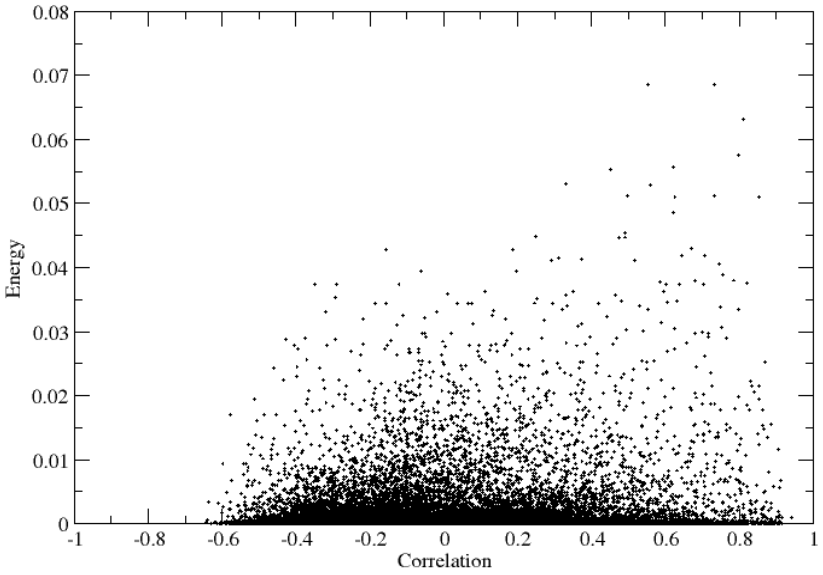
Correlation coefficient = 0.31

1POH



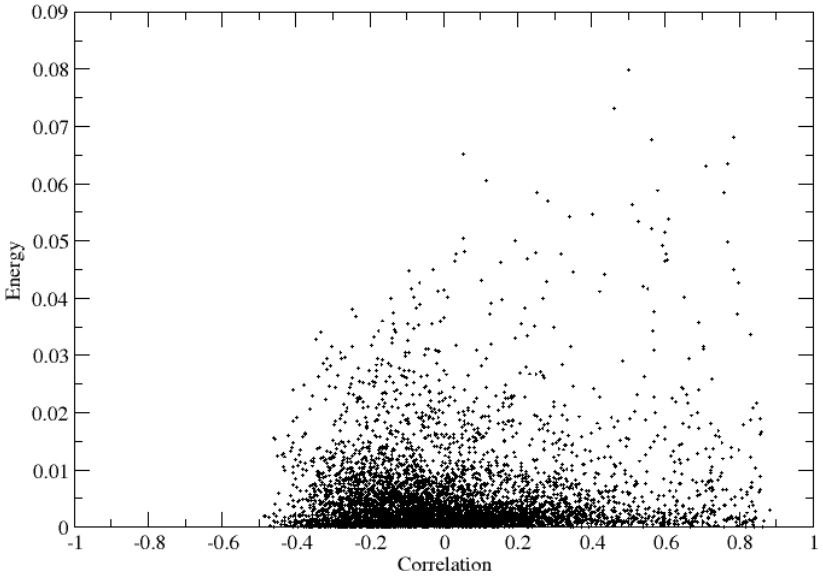
Correlation coefficient = 0.12

1TFH



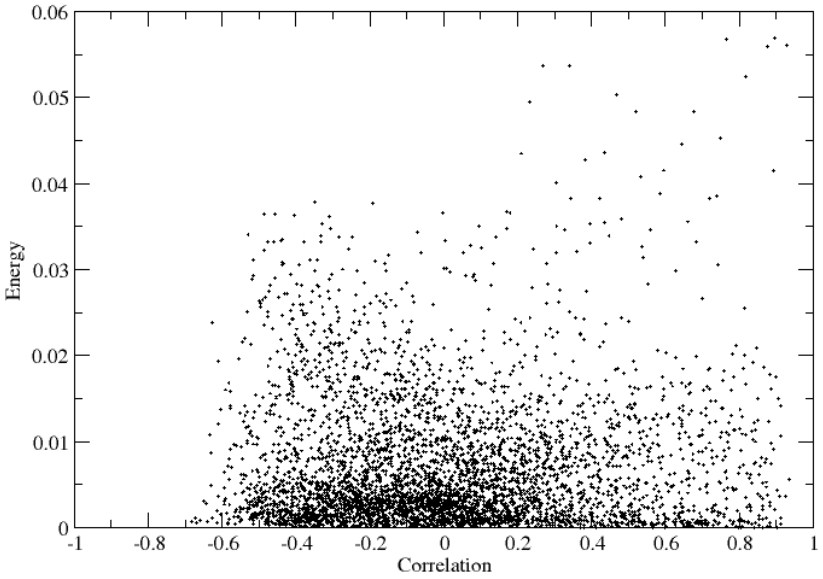
Correlation coefficient = 0.18

1UW3



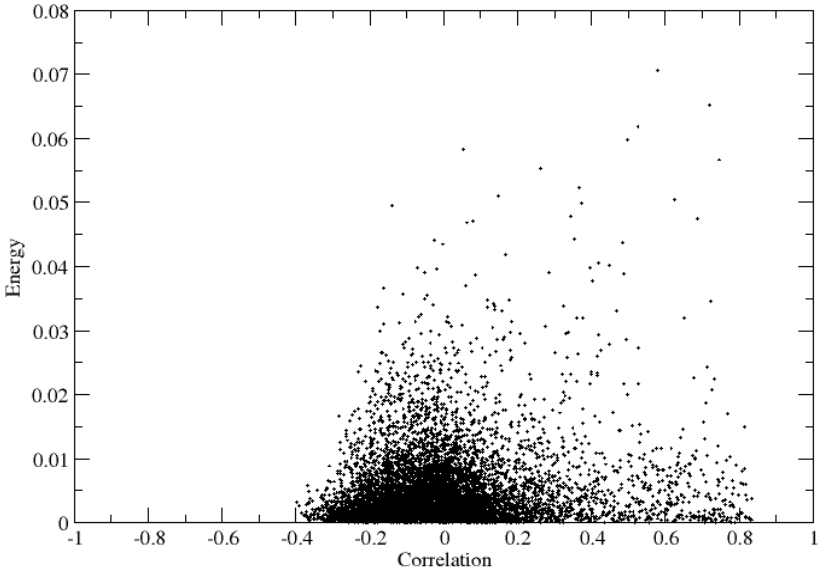
Correlation coefficient = 0.09

2VPF



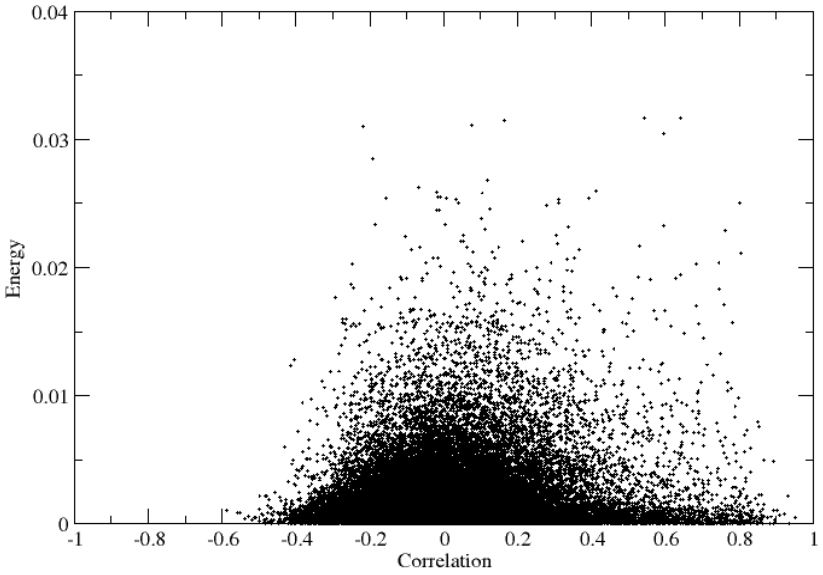
Correlation coefficient = 0.02

3LZT



Correlation coefficient = 0.12

7NN9



Correlation coefficient = 0.11

SUPPLEMENTAR Y MATERIAL 2: ALIGNMENT OF ALL ANTIGENS

```
CLUSTAL 2.0.12 multiple sequence alignment
1AO3            ------------------------------------------------------------
1CK4            ------------------------------------------------------------
1AUQ            ------------------------------------------------------------
1HCN_B          ------------------------------------------------------------
2VPF            ------------------------------------------------------------
1TFH            ------------------------------------------------------------
1POH            ------------------------------------------------------------
1P4P            ------------------------------------------------------------
1CMW            MRGMLPLFEPKGRVLLVDGHHLAYRTFHALKGLTTSRGEPVQAVYGFAKSLLKALKEDGD 60
1KDC            ------------------------------------------------------------
1D7P            ------------------------------------------------------------
1UW3            ------------------------------------------------------------
1K59            ------------------------------------------------------------
1KZQ            ------------------------------------------------------------
3LZT            ------------------------------------------------------------
7NN9            ------------------------------------------------------------
1HCN_A          ------------------------------------------------------------
1PKO            ------------------------------------------------------------
1BV1            ------------------------------------------------------------
1GWP            ------------------------------------------------------------

1AO3            ------------------------------------------------------------
1CK4            ------------------------------------------------------------
1AUQ            ------------------------------------------------------------
1HCN_B          ------------------------------------------------------------
2VPF            ------------------------------------------------------------
1TFH            ------------------------------------------------------------
1POH            ------------------------------------------------------------
1P4P            ------------------------------------------------------------
1CMW            AVIVVFDAKAPSFRHEAYGGYKAGRAPTPEDFPRQLALIKELVDLLGLARLEVPGYEADD 120
1KDC            ------------------------------------------------------------
1D7P            ------------------------------------------------------------
1UW3            ------------------------------------------------------------
1K59            ------------------------------------------------------------
1KZQ            ------------------------------------------------------------
3LZT            ------------------------------------------------------------
7NN9            ------------------------------------------------------------
1HCN_A          ------------------------------------------------------------
1PKO            ------------------------------------------------------------
1BV1            ------------------------------------------------------------
1GWP            ------------------------------------------------------------

1AO3            ------------------------------------------------------------
1CK4            ------------------------------------------------------------
1AUQ            ------------------------------------------------------------
1HCN_B          ------------------------------------------------------------
2VPF            ------------------------------------------------------------
1TFH            ------------------------------------------------------------
1POH            ------------------------------------------------------------
1P4P            ------------------------------------------------------------
1CMW            VLASLAKKAEKEGYEVRILTADKDLYQLLSDRIHVLHPEGYLITPAWLWEKYGLRPDQWA 180
1KDC            ------------------------------------------------------------
1D7P            ------------------------------------------------------------
1UW3            ------------------------------------------------------------
1K59            ------------------------------------------------------------
1KZQ            ------------------------------------------------------------
3LZT            ------------------------------------------------------------
7NN9            ------------------------------------------------------------
1HCN_A          ------------------------------------------------------------
1PKO            ------------------------------------------------------------
1BV1            ------------------------------------------------------------
1GWP            ------------------------------------------------------------

1AO3            ------------------------------------------------------------
1CK4            ------------------------------------------------------------
1AUQ            ------------------------------------------------------------
1HCN_B          ------------------------------------------------------------
2VPF            ------------------------------------------------------------
1TFH            ------------------------------------------------------------
1POH            ------------------------------------------------------------
1P4P            ------------------------------------------------------------
1CMW            DYRALTGDESDNLPGVKGIGEKTARKLLEEWGSLEALLKNLDRLKPAIREKILAHMDDLK 240
1KDC            ------------------------------------------------------------
1D7P            ------------------------------------------------------------
1UW3            ------------------------------------------------------------
1K59            ------------------------------------------------------------
1KZQ            ------------------------------------------------------------
3LZT            ------------------------------------------------------------
7NN9            ------------------------------------------------------------
1HCN_A          ------------------------------------------------------------
1PKO            ------------------------------------------------------------
1BV1            ------------------------------------------------------------
1GWP            ------------------------------------------------------------

1AO3            ------------------------------------------------------------
1CK4            ------------------------------------------------------------
1AUQ            ------------------------------------------------------------
1HCN_B          ------------------------------------------------------------
2VPF            ------------------------------------------------------------
1TFH            ------------------------------------------------------------
1POH            ------------------------------------------------------------
1P4P            ------------------------------------------------------------
1CMW            LSWDLAKVRTDLPLEVDFAKRREPDRERLRAFLERLEFGSLLHEFGLLESPKALEEAPWP 300
1KDC            ------------------------------------------------------------
1D7P            ------------------------------------------------------------
1UW3            ------------------------------------------------------------
1K59            ------------------------------------------------------------
1KZQ            ------------------------------------------------------------
3LZT            ------------------------------------------------------------
7NN9            ------------------------------------------------------------
1HCN_A          ------------------------------------------------------------
1PKO            ------------------------------------------------------------
```

```
1BV1            ------------------------------------------------------------
1GWP            ------------------------------------------------------------

1AO3            ------------------------------------------------------------
1CK4            ------------------------------------------------------------
1AUQ            ------------------------------------------------------------
1HCN_B          ------------------------------------------------------------
2VPF            ------------------------------------------------------------
1TFH            ------------------------------------------------------------
1POH            ------------------------------------------------------------
1P4P            ------------------------------------------------------------
1CMW            PPEGAFVGFVLSRKEPMWADLLALAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLA 360
1KDC            ------------------------------------------------------------
1D7P            ------------------------------------------------------------
1UW3            ------------------------------------------------------------
1K59            ------------------------------------------------------------
1KZQ            ------------------------------------------------------------
3LZT            ------------------------------------------------------------
7NN9            -------------------------------------------------RDFNNLTKGLCT 12
1HCN_A          ------------------------------------------------------------
1PKO            ------------------------------------------------------------
1BV1            ------------------------------------------------------------
1GWP            ------------------------------------------------------------

1AO3            ------------------------------------------------------------
1CK4            ------------------------------------------------------------
1AUQ            ------------------------------------------------------------
1HCN_B          ------------------------------------------------------------
2VPF            ------------------------------------------------------------
1TFH            ----------------------------------------------SGTTNTVAAYNLTWKS 16
1POH            ------------------------------------------------------------
1P4P            ------------------------------------------------------------
1CMW            LREGLGLPPGDDPMLLAYLLDPSNTTPEGVARRYGGEWTEEAGERAALSERLFANLWGRL 420
1KDC            ------------------------------------------------------------
1D7P            ------------------------------------------------------------
1UW3            ------------------------------------------------------------
1K59            ------------------------------------------------------------
1KZQ            -------------------SDPPLVANQVVTCPDKKSTAAVILTPTENHFTLKCPKTALT 41
3LZT            ------------------------------------------------------------
7NN9            INSWHIYGKDNAVRIGEDSDVLVTREPYVSCDPDECRFYALSQGTTIRGKHSNGTIHDRS 72
1HCN_A          ------------------------------------------------------------
1PKO            ------------------------------------------------------------
1BV1            ------------------------------------------------------------
1GWP            ------------------------------------------------------------

1AO3            ------------CSQPLDVILLLDGSSSFPASYFDEMKSFAKAFISKANIGPRLTQVSVL 48
1CK4            ------------CSTQLDIVIVLDGSNSIYP--WESVIAFLNDLLKRMDIGPKQTQVGIV 46
1AUQ            -DISEPPLHDFYCSRLLDLVFLLDGSSRLSEAEFEVLKAFVVDMMERLRISQKWVRVAVV 59
1HCN_B          ------------------------------------------------------------
2VPF            ------------------------------------------------------------
1TFH            TNFKTILEWEPKPVNQVYTVQISTKSGDWKSKCFYTTDTECDLTDEIVKDVKQTYLARVF 76
1POH            ------------------------------------------------------------
1P4P            ------------------------------------------------ANKLDSKKLTRSNGTTL 17
1CMW            EGEERLLWLYREVERPLSAVLAHMEATGVRLDVAYLRALSLEVAEEIARLEAEVFRLAGH 480
1KDC            --------------------------------ATSTKKLHKEPATLIKAIDGDTVKLMYK 28
1D7P            ---------------------LNSCSMPLGMESKAISDAQITASSYFTNMFATWSPSKAR 39
1UW3            ------------------------------------------------------------
1K59            ------------------------------------------------------------
1KZQ            EPPTLAYSPNRQICPAGTTSSCTSKAVTLSSLIPEAEDSWWTGDSASLDTAGIKLTVPIE 101
3LZT            ------------------------------------------------------------
7NN9            QYRALISWPLSSPPTVNSRVECIGWSSTSCHDGKTRMSICISGPNNNASAVIWYNRRPVT 132
1HCN_A          ----------------------------------------SKEPLRPRCRPINATLAVE 19
1PKO            ------------------------------------------------MRGSGQFRVIG-- 11
1BV1            ---------------------------------GVFNYETETTSVIPAARLFKAFILD 25
1GWP            ----------------------------------------PIVQNLQGQMVHQAISPR 18


1AO3            QYGSITT--------IDVPWNVVPEKAHLLSLVDVMQREGGPSQIGDA-LGFAVRYLTSE 99
1CK4            QYGENVT--------HEFNLNKYSSTEEVLVAANKIGRQGGLQTMTALGIDTARKEAFTE 98
1AUQ            EYHDGSH--------AYIGLKDRKRPSELRRIASQVKYAGSQVASTSEVLKYTLFQIFSK 111
1HCN_B          ------------------APDVQDCPECTLQENPFFSQPGAP-------ILQCMGCCFSR 35
2VPF            ------------------GQNHHEVVKFMDVYQRSYCHPIET--------LVDIFQEYPD 34
1TFH            SYPAGNV--------ESTGSAGEPLYENSPEFTPYLETNLGQPTIQSFEQVGTKVNVTVE 128
1POH            ---------------------------MFQQEVTITAPNGL-----------HTRPAAQ 21
1P4P            EYSQITD--------ADNATKAVETLKNSIKLEGSLVVGKTTVEIKEGTVTLKREIEKDG 69
1CMW            PFNLNSRDQLERVLFDELGLPAIGKTEKTGKRSTSAAVLEALREAHPIVEKILQYRELTK 540
1KDC            GQPMTFR----------LLLVDTPETKHPKKGVEKYGPEASAFTKKMVENAKKIEVEFDK 78
1D7P            LHLQGRS---------NAWRPQVNNPKEWLQVDFQKTMKVTGVTTQGVKSLLTSMYVKEF 90
1UW3            -----------------LGGYMLGSAMSRPLIHFGNDYEDCYYRENMHRYPNQVYYRPVD 43
1K59            ------------------EDNSRYTHFLTQHYDAKPQGRDDRYCESIMRRRGLTSPCKD 41
1KZQ            KFPVTTQTFVVGCIKGDDAQSCMVTVTVQARASSVVNNVARCSYGADSTLGPVKLSAEGP 161
3LZT            -----------------KVFGRCELAAAMKRHGLDNYRGYSLGNWVCAAKFESNFNTQA 42
7NN9            EINTWARNILRTQESECVCHNGVCPVVFTDGSATGPAETRIYYFKEGKILKWEPLAGTAK 192
1HCN_A          KEG------------CPVCITVNTTICAGYCPTMTRVLQGVLPALPQVVCNYRDVRFES 66
1PKO            -----------------PGHPIRALVGDEAELPCRISPGKNATGMEVGWYRSPFSRVVH 53
1BV1            GDNLFPK---------VAPQAISSVENIEGNGGPGTIKKISFPEGLPFKYVKDRVDEVDH 76
1GWP            TLNAWVK----------VVEEKAFSPEVIPMFSALSEGATPQDLNTMLNTVGGHQAAMQ 67

1AO3            MHGAR--PGASKAVVILVTDVSVD--SVDAAAD-AARSNRVTVFPIGIGDRYDAAQLR-- 152
1CK4            ARGAR--RGVKKVMVIVTDGESHDNYRLKQVIQ-DCEDENIQRFSIAILGHYNRGNLSTE 155
1AUQ            IDRPE--ASRIALLLMASQEPQRMSRNFVRYVQ-GLKKKKVIVIPVGIGPHANLKQIR-- 166
1HCN_B          AYPTP--LRSKKTMLVQKN--------VTSEST-CCVAKSYNRVTV-MGG-FKVENHT-- 80
2VPF            EIEYI--FKPSCVPLMRCGG------CCNDEGL-ECVPTEESNITMQIMRIKPHQGQH-- 83
1TFH            DERTL--VRRNNTFLSLRDVFGKDLIYTLYYWK-SSSSGKKTAKTNTNEFLIDVDKGE-- 183
1POH            FVKEA--KGFTSEITVTSNG-------KSASAK-SLFKLQTLGLTQGTVVTISAEGED-- 69
1P4P            KVKVF--LNDTAGSNKKTGKWEDSTSTLTISAD-SKKTKDLVFLTDGTITVQQYNTAG-- 124
1CMW            LKSTY--IDPLPDLIHPRTGRLHTRFNQTATATGRLSSSDPNLQNIPVRTPLGQRIRRAF 598
1KDC            GQRTD--KYGRGLAYIYADG------KMVNEALVRQGLAKVAYVYNPNNTHEQHLRKS-- 128
1D7P            LISSS--QDGHQWTLFFQNGK-----VKVFQGNQDSFTPVVNCLDPPLLTRYLRIHPQS- 142
1UW3            QYSNQ--NNFVHDCVNITVKQH----TVTTTTKGENFTETDIKIMERVVEQMCITQYQ-- 95
1K59            INTFI--HGNKRSIKAICENKNGNPHRENLRISKSSFQVTTCKLHGGSPWPPCQYRAT-- 97
1KZQ            TTMTL--VCGKDGVKVPQDNNQYCSGTTLTGCNEKSFKDILPKLTENPWQGNASSDKGAT 219
3LZT            TNRNT--DGSTDYGILQINSRWWCNDGRTPGSRNLCNIPCSALLSSDITASVNCAKKIVS 100
7NN9            HIEECSCYGERAEITCTCRDNWQGSNRPVIRIDPVAMTHTSQYICSPVLTDNPRPNDPTV 252
1HCN_A          IRLPGCPRGVNPVVSYAVALSCQCALCRRSTTDCGGPKDHPLTCDDPRFQDSSSSKAP-- 124
```

```
1PKO            LYRNG--KDQDAEQAPEYRGRTELLKESIGEGKVALRIQNVRFSDEGGYTCFFRDHSY-- 109
1BV1            TNFKYNYSVIEGGPIGDTLEKISNEIKIVATPDGGSILKISNKYHTKGDHEVKAEQVK-- 134
1GWP            MLKETINEEAAEWDRLHPVHAGPIAPGQMREPRGSDIAGTTSTLQEQIGWMTHNPPIPVG 127

1AO3            ----ILAGPAGDSNVVKLQRIEDLPTMVTLGNSFLHKLC---------------------- 187
1CK4            KFVEEIKSIASEPTEKHFFNVSDELALVTIVKALGERIFALEA----------------- 198
1AUQ            ----LIEKQAPENKAFVLSSVDELEQQRDEIVSYLCDLAPEAPPPT-------------- 208
1HCN_B          ---------ACHCSTCYYHKS------------------------------------- 92
2VPF            ---------IGEMSFLQHNKCECRPKKD------------------------------ 102
1TFH            -------NYCFSVQAVIPSRTVNRKSTDSPVECMGQEKGEFRE---------------- 219
1POH            ----------EQKAVEHLVKLMAELE-------------------------------- 85
1P4P            --------TSLEGSASEIKNLSELKNALK---------------------------- 145
1CMW            IAEEGWLLVALDYSQIELRVLAHLSGDENLIRVFQEGRDIHTETASWMFGVPREAVDPLM 658
1KDC            ---------EAQAKKEKLNIWSENDADSGQ----------------------------- 149
1D7P            ---------WVHQIALRMEVLGCEAQ-------------------------------- 159
1UW3            -----------RESQAYYQRGA------------------------------------ 106
1K59            --------AGFRNVVVACENGLPVHLDGSIFRRP------------------------ 123
1KZQ            LTIKKEAFPAESKSVIIGCTGGSPEKHHCTVKLEFAGAAGSAKSAAGTASHVSIFAMVIG 279
3LZT            ------DGNGMNAWVAWRNRCKGTDVQAWIRGCRL------------------------ 129
7NN9            GKCNDPYPGNNNNGVKGFSYLDGVNTWLGRTISIASRSGYEMLKVPNALTDDKSKPTQGQ 312
1HCN_A          -----------PPSLPSPSRLPGPSDTPILPQ---------------------------- 145
1PKO            -----------QEEAAVELKVEDPFYWINPGRSR------------------------- 132
1BV1            ---------ASKEMGETLLRAVESYLLAHSDAYN------------------------- 159
1GWP            -------EIYKRWIILGLNKIVRMYSPTSIL---------------------------- 151

1AO3            ------------------------------------------------------------
1CK4            ------------------------------------------------------------
1AUQ            ------------------------------------------------------------
1HCN_B          ------------------------------------------------------------
2VPF            ------------------------------------------------------------
1TFH            ------------------------------------------------------------
1POH            ------------------------------------------------------------
1P4P            ------------------------------------------------------------
1CMW            RRAAKTINFGVLYGMSAHRLSQELAIPYEEAQAFIERYFQSFPKVRAWIEKTLEEGRRRG 718
1KDC            ------------------------------------------------------------
1D7P            ------------------------------------------------------------
1UW3            ------------------------------------------------------------
1K59            ------------------------------------------------------------
1KZQ            LIGSIAACVA------------------------------------------------- 289
3LZT            ------------------------------------------------------------
7NN9            TIVLNTDWSGYSGSFMDAEGECYRACFYVELIRGRPKEDKVWWTSNSIVSMCSSTEFLGQ 372
1HCN_A          ------------------------------------------------------------
1PKO            ------------------------------------------------------------
1BV1            ------------------------------------------------------------
1GWP            ------------------------------------------------------------


1AO3            ------------------------------------------------------------
1CK4            ------------------------------------------------------------
1AUQ            ------------------------------------------------------------
1HCN_B          ------------------------------------------------------------
2VPF            ------------------------------------------------------------
1TFH            ------------------------------------------------------------
1POH            ------------------------------------------------------------
1P4P            ------------------------------------------------------------
1CMW            YVETLFGRRRYVPDLEARVKSVREAAERMAFNMPVQGTAADLMKLAMVKLFPRLEEMGAR 778
1KDC            ------------------------------------------------------------
1D7P            ------------------------------------------------------------
1UW3            ------------------------------------------------------------
1K59            ------------------------------------------------------------
1KZQ            ------------------------------------------------------------
3LZT            ------------------------------------------------------------
7NN9            WDWPDGAKIEYFL---------------------------------------------- 385
1HCN_A          ------------------------------------------------------------
1PKO            ------------------------------------------------------------
1BV1            ------------------------------------------------------------
1GWP            ------------------------------------------------------------

1AO3            ----------------------------------------------------
1CK4            ----------------------------------------------------
1AUQ            ----------------------------------------------------
1HCN_B          ----------------------------------------------------
2VPF            ----------------------------------------------------
1TFH            ----------------------------------------------------
1POH            ----------------------------------------------------
1P4P            ----------------------------------------------------
1CMW            MLLQVHDELVLEAPKERAEAVARLAKEVMEGVYPLAVPLEVEVGIGEDWLSAKE 832
1KDC            ----------------------------------------------------
1D7P            ----------------------------------------------------
1UW3            ----------------------------------------------------
1K59            ----------------------------------------------------
1KZQ            ----------------------------------------------------
3LZT            ----------------------------------------------------
7NN9            ----------------------------------------------------
1HCN_A          ----------------------------------------------------
1PKO            ----------------------------------------------------
1BV1            ----------------------------------------------------
1GWP            ----------------------------------------------------
```

SUPPLEMENTARY MATERIAL 3: ALIGNMENT OF EPITOPE SEQUENCES

```
1CK4            ----------SNSIY------------------------------ 5
7NN9            --PRPNISIASRSLNTD-------W---------------------- 16
1AUQ            ----QRMS--RNFVR------------------------------- 9
3LZT            ---RHGLDNYRGYSQATNRNTDGRWWCNDGRTPGSRNLSDGNGTDVQ 44
1P4P            ---GSNADSKKNTAGT------------------------------ 13
1CMW            ---SNTTPEKSTYIDP------------------------------ 13
1AO3            ---ITTIDVPWNVVPEKAHSLVDVMHGAR----------------- 26
1KZQ            ---TALTEPPTSCTSKAV---------------------------- 15
1TFH            ---KDLIYTLYKKTAKTNQAVIPVNRKSTD---------------- 27
1UW3            ----KQHTVTTTTKGE------------------------------ 12
2VPF            ---KFMDVYQRSYCHPYIFKQIMRIKPHQGQHIGEM---------- 33
1BV1            ----ENIEGNGGPGT------------------------------- 11
1GWP            PIVQNLQGQFSPEVIPMFNEEAAEWDRLHPVHAGPIAPGQMRLQEQI 47
1POH            -MFQQSAEGEDEQKAVEH---------------------------- 17
1K59            ----GLTSPCKDGGSPWPP--------------------------- 15
1KDC            --TKHPKKGVEKYAFTKYADGKM----------------------- 21
1PKO            ---KNATGRDHSYQEE------------------------------ 13
1HCN            ---ENPFFSQPRVLQPRGVNPVVS---------------------- 21
1D7P            ---TNMFATWRPQVSLLT---------------------------- 15
>1AO3
ITTIDVPWNVVPEKAHSLVDVMHGAR
>1AUQ
QRMSRNFVR
>1BV1
ENIEGNGGPGT
>1CK4
SNSIY
>1CMW
SNTTPEKSTYIDP
>1D7P
TNMFATWRPQVSLLT
>1GWP
PIVQNLQGQFSPEVIPMFNEEAAEWDRLHPVHAGPIAPGQMRLQEQI
>1HCN
ENPFFSQPRVLQPRGVNPVVS
>1K59
GLTSPCKDGGSPWPP
>1KDC
TKHPKKGVEKYAFTKYADGKM
>1KZQ
TALTEPPTSCTSKAV
>1P4P
GSNADSKKNTAGT
>1PKO
KNATGRDHSYQEE
>1POH
MFQQSAEGEDEQKAVEH
>1TFH
KDLIYTLYKKTAKTNQAVIPVNRKSTD
>1UW3
KQHTVTTTTKGE
>2VPF
KFMDVYQRSYCHPYIFKQIMRIKPHQGQHIGEM
>3LZT
RHGLDNYRGYSQATNRNTDGRWWCNDGRTPGSRNLSDGNGTDVQ
>7NN9
PRPNISIASRSLNTDW
```