Supplementary Material

Exploring linkage dependence of polyubiquitin conformations using molecular modeling

David Fushman & Olivier Walker

Table S1. Accessible (for selected residues) and buried (BSA) surface areas for the best structures for all generated di-ubiquitin chains.

	Distal domain					Proximal domain				
Chain linkage	V8 [a] (Å^2)	L44 [a] (Å^2)	I70 [a] (Å^2)	BSA [b] (Å^2)	% [c]	V8 [a] (Å^2)	L44 [a] (Å^2)	I70 [a] (Å^2)	BSA [b] (Å^2)	% [c]
K48	38.4	1.2	1.1	704	15	19.6	2.6	0.2	658	14
K63	8.5	13.3	27.6	451	10	133.5	41.1	39.8	445	10
K6	7.0	4.6	5.7	518	12	19.8	5.0	11.6	555	12
K33	41.9	9.5	1.5	411	9	142.0	33.6	24.7	409	9
K29	31.4	0.0	2.7	514	11	151.1	26.4	28.3	506	11
K27	53.5	27.9	8.4	590	12	79.7	6.9	4.8	549	12
K11	52.2	16.1	3.3	726	15	59.9	5.4	3.4	680	15
HT	48.6	1.4	0	552	13	146.6	36.8	37.1	521	13
K48*	66.2	21.8	23.1	120	3	116.1	26.8	19.8	195	4

[a] Accessible surface area for all atoms of the indicated residues
[b] Buried surface area for a given Ub domain computed with the ProFace server[1]
[c] Percentage of the buried surface area with respect to the total surface of a given domain

Table S2. Detailed analysis of the hydrophobic contacts for the best structure of all generated di-ubiquitin structures

Linkage	Distal Ub	Atom	Proximal Ub	Atom	Distance (Å)
K48					
	LEU 8	CG	LEU 108	CD1	3.6927
	LEU 8	CD1	LEU 108	CD1	3.6243
	LEU 8	CD1	LEU 108	CD2	3.5428
	LEU 8	CD1	ILE 144	CB	3.8321
	LEU 8	CD1	ILE 144	CG2	3.4398
	LEU 8	CD2	LEU 108	CD1	3.4706
	LEU 8	CD2	LEU 108	CD2	3.2821
	ILE 44	CG2	VAL 170	CG1	3.3951
	ILE 44	CD1	VAL 170	CB	3.4933
	ILE 44	CD1	VAL 170	CG1	3.3169
	ILE 44	CD1	VAL 170	CG2	3.4479
	ALA 46	C	THR 109	CG2	3.8304
	LYS 48	CD	LEU 173	CD1	3.8369
	GLN 49	CB	ARG 172	CD	3.4165
	GLN 49	CD	ARG 172	CD	3.8101
	GLU 51	CD	ARG 172	CZ	3.7977
	HIS 68	CB	LEU 108	CD1	3.7102
	HIS 68	CG	LEU 108	CD1	3.8432
	HIS 68	CD2	LEU 108	CB	3.8161
	HIS 68	CD2	LEU 108	CD1	3.8323
	VAL 70	CG1	ILE 144	CG1	3.7345
	VAL 70	CG1	ILE 144	CG2	3.3605
	VAL 70	CG1	ILE 144	CD1	3.2544
	ARG 72	CD	GLN 149	CB	3.6555
	ARG 72	CD	GLN 149	CD	3.6357
	GLY 76	C	LYS 148	CG	3.6998
K63					
	LEU 8	CD2	LEU 167	CA	3.8686
	LEU 8	CD2	HIS 168	CG	3.8698
	ILE 44	CG2	ALA 146	CB	3.4716
	ILE 44	CD1	ALA 146	CB	3.6041
	GLY 47	CA	LYS 148	CB	3.6563
	HIS 68	CD2	ALA 146	C	3.6138
	HIS 68	CD2	GLY 147	CA	3.6379
	ARG 72	CZ	GLN 162	CD	3.6860
	GLY 76	CA	LYS 163	CE	3.2881
	GLY 76	C	LYS 163	CD	3.8408
	GLY 76	C	LYS 163	CE	2.9687

K6

LEU 8	CB	VAL 170	CB	3.6134
LEU 8	CB	VAL 170	CG1	3.5834
LEU 8	CB	VAL 170	CG2	3.4358
LEU 8	CG	VAL 170	CG2	3.7952
LEU 8	CD1	ILE 144	CG1	3.2689
LEU 8	CD1	ILE 144	CD1	3.7985
LEU 8	CD1	VAL 170	CG1	3.5014
LEU 8	CD1	VAL 170	CG2	3.5130
LEU 8	CD2	ILE 144	CG1	3.7194
LEU 8	CD2	ILE 144	CD1	3.4261
VAL 70	CG1	LEU 108	CD1	3.8282
LEU 71	CB	LEU 108	CD1	3.6275
GLY 76	CA	LYS 106	CE	3.6235
GLY 76	C	LYS 106	CE	3.4208
GLY 76	C	HIS 168	CE1	3.6086

K33

LEU 8	CB	LYS 111	CD	3.7876
LEU 8	CD1	THR 107	CB	3.8609
LEU 8	CD1	THR 109	CB	3.5231
LEU 8	CD1	GLY 110	C	3.5986
LEU 8	CD1	LYS 111	CB	3.5215
LEU 8	CD2	THR 109	CB	3.8702
ILE 44	CD1	GLY 110	CA	3.3949
ILE 44	CD1	GLY 110	C	3.2441
HIS 68	CE1	THR 109	CG2	3.6187
VAL 70	CG1	LYS 111	CA	3.5818
VAL 70	CG1	LYS 111	CB	3.8524
VAL 70	CG1	LYS 111	CG	3.7282
VAL 70	CG2	GLY 110	C	3.8106
GLY 76	CA	THR 114	CB	3.5400
GLY 76	CA	LYS 133	CE	3.2218
GLY 76	C	LYS 133	CE	3.2562

K29

LEU 8	CB	GLU 124	CD	3.6287
LEU 8	CD1	PRO 138	C	3.6895
HIS 68	CD2	ASP 139	CG	3.5146
ARG 72	CD	ASP 132	CG	3.8713
GLY 76	CA	LYS 129	CG	3.3689
GLY 76	CA	LYS 129	CE	3.2681
GLY 76	C	LYS 129	CE	2.9820

K27

LEU 8	CB	VAL 170	CB	3.5925
LEU 8	CB	VAL 170	CG1	3.3477
LEU 8	CD1	VAL 170	CB	3.6579
LEU 8	CD1	VAL 170	CG2	3.5858
LEU 8	CD2	VAL 170	CB	3.8928
ARG 74	CZ	GLU 151	CD	3.7677
GLY 76	CA	ASP 152	CG	3.6636
GLY 76	C	ARG 142	CD	3.5304

K11

LEU 8	CB	VAL 170	CB	3.8168
LEU 8	CB	VAL 170	CG1	3.3952
LEU 8	CB	VAL 170	CG2	3.4215
LEU 8	CD1	ILE 144	CG1	3.7584
LEU 8	CD1	ILE 144	CG2	3.5840
LEU 8	CD1	ILE 144	CD1	3.5567
LEU 8	C	VAL 170	CG2	3.7993
THR 9	CB	ARG 172	CB	3.5523
THR 9	CG2	ARG 142	CB	3.6045
THR 9	CG2	ARG 142	CG	3.4507
THR 9	CG2	VAL 170	CG2	3.6381
THR 9	CG2	LEU 171	C	3.6172
THR 9	CG2	ARG 172	CA	3.8273
THR 9	CG2	ARG 172	CB	3.5893
ILE 36	CG1	ARG 174	C	3.6287
PRO 37	CD	GLY 175	CA	3.7910
PRO 37	CD	GLY 175	C	3.8833
ILE 44	CD1	LEU 108	CB	3.7526
ILE 44	CD1	LEU 108	CG	3.8855
ILE 44	CD1	LEU 108	CD1	3.7808
ILE 44	CD1	LEU 108	CD2	3.5742
VAL 70	CG1	LEU 108	CB	3.4238
VAL 70	CG2	LEU 108	CB	3.4030
VAL 70	CG2	LEU 108	CD1	3.5491
LEU 71	CD1	LEU 173	CB	3.6805
LEU 71	C	THR 109	CG2	3.8131
ARG 72	CA	THR 109	CG2	3.6567
ARG 72	CG	THR 109	CB	3.6985
ARG 72	CG	THR 109	CG2	3.6046
ARG 72	CD	THR 109	CB	3.4274
ARG 72	CD	THR 109	CG2	3.6889
GLY 76	CA	LYS 111	CB	3.7632
GLY 76	CA	LYS 111	CE	3.3298
GLY 76	C	LYS 111	CE	2.9809

HT

LEU 8	CD1	LEU 115	CD2	3.784
LEU 8	CD2	LYS 133	CB	3.8453
LEU 8	CD2	LYS 133	CG	3.8271
LEU 8	CD2	LYS 133	CD	3.7156
ILE 44	CD1	THR 114	CB	3.842
ILE 44	CD1	LYS 133	CE	3.6356
GLY 47	CA	LYS 111	CE	3.6155
VAL 70	CG2	LYS 133	CE	3.6768
ARG 72	CD	GLN 102	CD	3.5372
ARG 72	CZ	PHE 104	CE1	3.5885
ARG 72	CZ	PHE 104	CZ	3.8362
GLY 75	CA	LYS 163	CD	3.7694
GLY 75	C	LYS 163	CD	3.8364
GLY 76	CA	MET 101	CA	3.189

K48*

GLY 75	CA	GLU 151	CD	3.4615
GLY 75	C	GLU 151	CD	3.7598
GLY 76	CA	LYS 148	CE	3.2525
GLY 76	C	LYS 148	CE	3.7417

S3: Details of the PCA analysis

To account for the highly dispersed nature of the data, PCA analysis was performed by using the mean centered correlation matrix (X) of the original data (D) corresponding to Table 2 (except H_{bonds} per 100Å^2). Eigenvalues (λ) and eigenvectors (V) were obtained after diagonalization of the X matrix. Eigenvectors were ranked in descending order according to their respective eigenvalues. The percentage of variance explained (or accounted for) by each eigenvalue (λ_j) is obtained by dividing the corresponding eigenvalue (λ_j) by the sum of all the eigenvalues.

The first principal plane (PC1 & PC2) accounts for 78.9% of the total information provided. Adding the third component/axis (PC3) will result in 90.6% of the information being covered (Table S3).

Table S3: Percent of the variance explained by each eigenvalue. Only the first 7 PCs are shown.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Eigenvalue	5.3	4.2	1.4	0.8	0.2	0.1	0.0
% variance explained	44.1	34.8	11.8	6.4	2.1	0.7	0.2
total variance covered by this and all preceding PCs	44.1	78.9	90.6	97.1	99.1	99.8	100.0

Eigenvectors provide the principal component coefficients also known as loadings. These coefficients represent coordinates of each descriptor in the principal axes system of the PC components. Each coefficient (c_{ij}) is obtained as $c_{ij} = \sqrt{\lambda_j} V_{ij}$, where j is the eigenvalue index, i numerates the variables, and V_{ij} is the element of the eigenvector matrix V. The corresponding coefficients for the 12 variables considered here are shown in Table S4.

For facilitate the analysis of the relationship between the principal components and the variables of interest, it is convenient to plot these principal component coefficients on the correlation circle (Figure 4a,c, main text). The quality of the representation of each variable i by a given principal component j (Table S5) can be assessed by the squared cosine of the angle (β) formed by the principal axis PCj and the vector connecting the corresponding point with the origin. The corresponding expression is: $\cos^2 \beta_{ij} = c_{ij}^2$. The sum of $\cos^2\beta$ values over the two PC components gives an assessment of the quality of the representation of a variable on the corresponding plane.

Table S4: Correlation between the first four principal components and the original variables

variables	PC1	PC2	PC3	PC4
E_{inter}	0.50	-0.83	0.10	0.21
E_{vdw}	0.32	-0.72	-0.42	-0.38
E_{elec}	0.50	-0.80	0.17	0.28
BSA	-0.89	0.31	-0.25	0.17
E_{rest}	0.52	0.85	-0.07	0.03
E_{unamb}	0.07	0.02	0.99	0.03
E_{amb}	0.51	0.85	-0.10	0.03
H_{bonds}	-0.72	0.58	0.27	-0.14
H_{ydro}	-0.87	-0.45	-0.04	0.18
E_{bind}	0.91	-0.07	0.10	-0.37
H_{score}	0.74	0.30	-0.24	0.53
ASA_{hp}	0.84	0.38	0.05	-0.01

Table S5: The corresponding squared cosines given for the first four principal components

variables	PC1	PC2	PC3	PC4
E_{inter}	0.25	0.69	0.01	0.05
E_{vdw}	0.10	0.52	0.17	0.14
E_{elec}	0.25	0.65	0.03	0.08
BSA	0.80	0.09	0.06	0.03
E_{rest}	0.27	0.72	0.00	0.00
E_{unamb}	0.01	0.00	0.99	0.00
E_{amb}	0.26	0.72	0.01	0.00
H_{bonds}	0.52	0.34	0.07	0.02
H_{ydro}	0.75	0.20	0.00	0.03
E_{bind}	0.83	0.01	0.01	0.14
H_{score}	0.55	0.09	0.06	0.28
ASA_{hp}	0.70	0.15	0.00	0.00

The representation of the original data in the principal axes system is known as scores (Table S6). Their respective components (Y_{ij}) are obtained according to Y=D.V, D being the original data matrix (Table 2) and *V* the matrix containing column eigenvectors. The quality of the representation of the data on a principal axis is given by its squared cosine (Table S7). It is defined by $\cos\alpha_{ij} = OH_j / Oi$. Here *Oi* is the length of the vector connecting the origin and the i^{th} data point, and OH_j is the projection of this vector onto the j^{th} axis. The squared cosine of the angle α can then be expressed as $\cos^2 \alpha_{ij} = Y_{ij}^2 / \sum_{j=1}^{p} Y_{ij}^2$ where p is the total number of principal components.

Table S6: **Representation of the coordinates of the data in the new principal axes system**.

Only the first four PCs are shown

linkage	PC1	PC2	PC3	PC4
K48	-4.60	0.51	-0.31	-0.90
K63	2.42	-0.58	0.36	0.18
K6	-0.96	-1.74	0.26	0.44
K33	3.25	-1.98	-0.26	-0.80
K29	1.08	2.37	-1.33	-1.16
K27	-0.81	-0.61	2.51	-0.25
K11	-1.16	-1.96	-1.65	1.27
HT	0.79	4.00	0.43	1.21

Table S7: **Squared cosine of the scores.**

Only the first four PCs are shown

linkage	PC1	PC2	PC3	PC4
K48	0.94	0.01	0.00	0.04
K63	0.86	0.05	0.02	0.01
K6	0.17	0.55	0.01	0.04
K33	0.69	0.26	0.00	0.04
K29	0.12	0.56	0.18	0.14
K27	0.08	0.05	0.81	0.01
K11	0.14	0.39	0.28	0.16
HT	0.03	0.87	0.01	0.08

Table S8: Summary of HADDOCK parameters for the best structure of the best cluster for each of the eight linkages.

Chain inkage	E_{inter} (kcal/mol)	E^{a}_{vdw} (kcal/mol)	E^{a}_{elec} (kcal/mol)	BSA (Å^2)	E_{rest} (kcal/mol)	E^{b}_{unamb} (kcal/mol)	E^{c}_{amb} (kcal/mol)	H_{bonds}	H^{d}_{ydro}	E^{e}_{bind}	H_{score} (a.u.)	ASA^{f}_{hp} (Å^2)	H_{bond} per 100 Å
K48	-346.4	-7.2	-339.2	1362	4.8	4.0	0.8	9	14	-11.3	-135.3	63.1	0.66
K63	-143.8	1.0	-144.8	896	26.3	2.2	24.1	4	0	-7.0	-73.4	263.8	0.45
K6	-191.5	-5.7	-185.8	1072	7.0	3.1	3.9	6	11	-9.1	-81.2	53.7	0.56
K33	-92.2	4.7	-96.9	820	19.5	2.9	16.6	2	0	-5.4	-72.8	253.2	0.24
K29	-279.6	-3.0	-276.5	1020	53.6	2.3	51.4	5	0	-6.9	-90.9	239.9	0.49
K27	-261.0	-8.5	-252.5	1139	12.2	4.0	8.2	7	5	-9.2	-101.3	181.2	0.61
K11	-182.3	-12.5	-169.8	1405	3.2	2.3	0.9	5	14	-11.5	-84.1	140.3	0.36
HT	-243.9	-24.4	-219.5	1200	61.7	2.2	59.4	5	0	-8.3	-36.3	270.9	0.42
K48*	-144.0	7.4	-151.4	315	1.8	1.8	N/A	1	0	-4.8	-54.7	273.8	0.32

Table S9: **Percent of the variance explained by each eigenvalue for the best structure of the best cluster.** Only the first 7 PCs are shown.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Eigenvalue	6.7	3.0	1.6	0.4	0.3	0.1	0.0
% variance explained	55.7	24.6	13.5	3.5	2.1	0.6	0.0
total variance covered by this and all preceding PCs	55.7	80.3	93.8	97.3	99.4	100.0	100.0

Table S10: Correlation between the first four principal components and the original variables

variables	PC1	PC2	PC3	PC4
E_{inter}	0.67	-0.65	-0.36	-0.06
E_{vdw}	0.22	-0.76	0.53	0.29
E_{elec}	0.67	-0.59	-0.43	-0.10
BSA	-0.78	0.40	-0.44	0.08
E_{rest}	0.64	0.74	0.15	0.06
E_{unamb}	-0.71	-0.23	0.45	-0.48
E_{amb}	0.66	0.73	0.13	0.07
H_{bonds}	-0.90	0.32	0.20	-0.08
H_{ydro}	-0.89	-0.21	-0.36	0.14
E_{bind}	0.89	-0.15	0.40	-0.06
H_{score}	0.78	0.25	-0.50	-0.22
ASA_{hp}	0.88	0.25	0.11	-0.05

Table S11: The corresponding squared cosines given for the first four principal components

variables	PC1	PC2	PC3	PC4
E_{inter}	0.45	0.42	0.13	0.00
E_{vdw}	0.05	0.58	0.29	0.08
E_{elec}	0.45	0.35	0.19	0.01
BSA	0.60	0.16	0.19	0.01
E_{rest}	0.42	0.55	0.02	0.00
E_{unamb}	0.50	0.05	0.20	0.23
E_{amb}	0.43	0.54	0.02	0.01
H_{bonds}	0.81	0.10	0.04	0.01
H_{ydro}	0.79	0.05	0.13	0.02
E_{bind}	0.79	0.02	0.16	0.00
H_{score}	0.62	0.06	0.25	0.05
ASA_{hp}	0.77	0.06	0.01	0.00

Table S12: **Representation of the coordinates of the data in the new principal axes system**. Only the first four PCs are shown

linkage	PC1	PC2	PC3	PC4
K48	-4.87	0.37	1.08	0.16
K63	2.52	-1.00	0.22	0.38
K6	-1.22	-1.32	-0.62	-0.14
K33	3.18	-2.48	0.48	-0.38
K29	1.52	1.70	1.58	1.05
K27	-1.63	-0.06	0.95	-1.04
K11	-1.58	-0.59	-2.53	0.64
HT	2.07	3.37	-1.16	-0.68

Table S13: **Squared cosine of the scores.**

Only the first four PCs are shown

linkage	PC1	PC2	PC3	PC4
K48	0.95	0.01	0.05	0.00
K63	0.80	0.13	0.01	0.02
K6	0.29	0.35	0.08	0.00
K33	0.60	0.36	0.01	0.01
K29	0.26	0.33	0.28	0.12
K27	0.54	0.00	0.19	0.22
K11	0.25	0.03	0.64	0.04
HT	0.25	0.65	0.08	0.03

Figure S1: Principal component analysis for the best structure of the best cluster for each generated chains

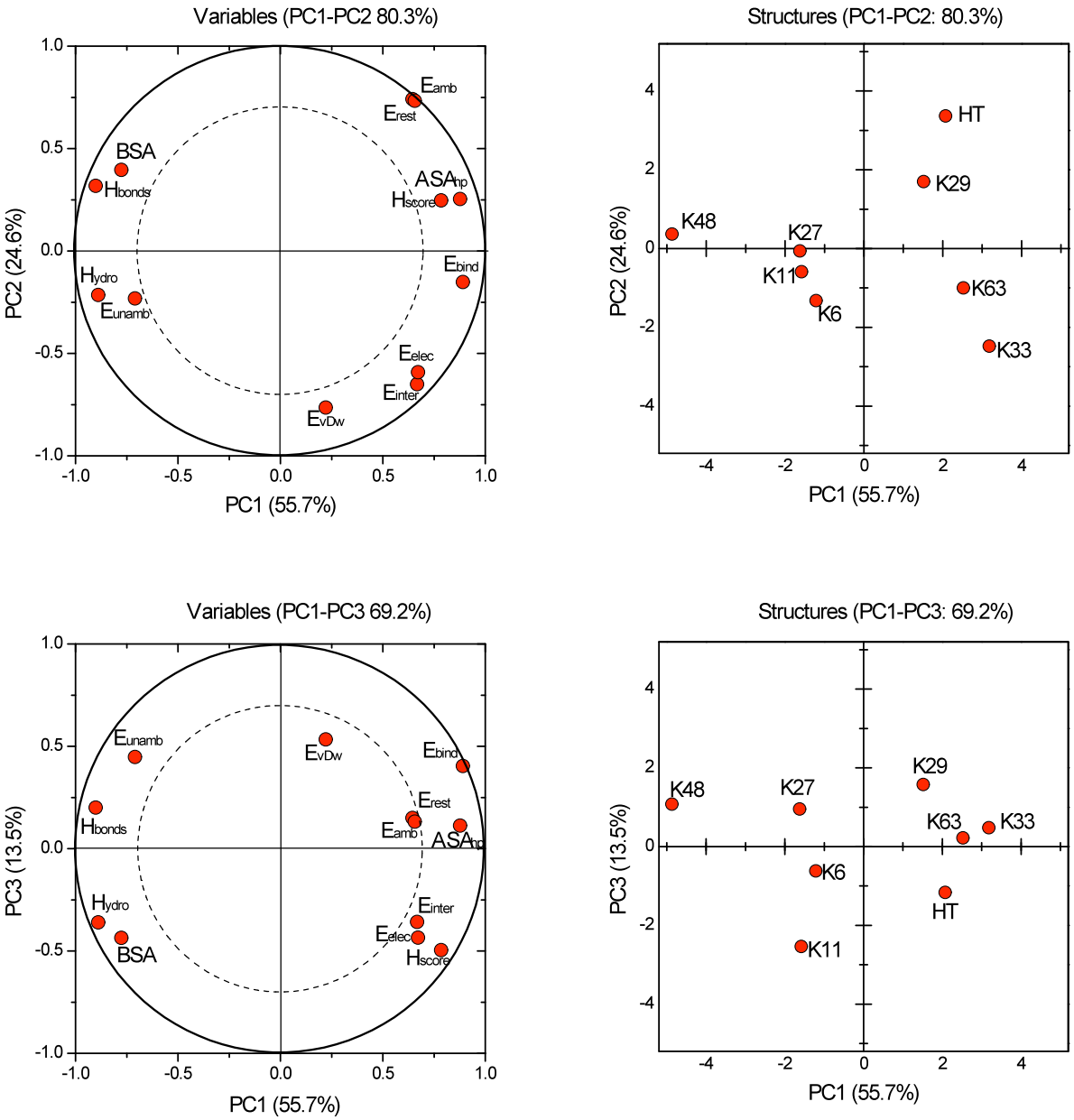
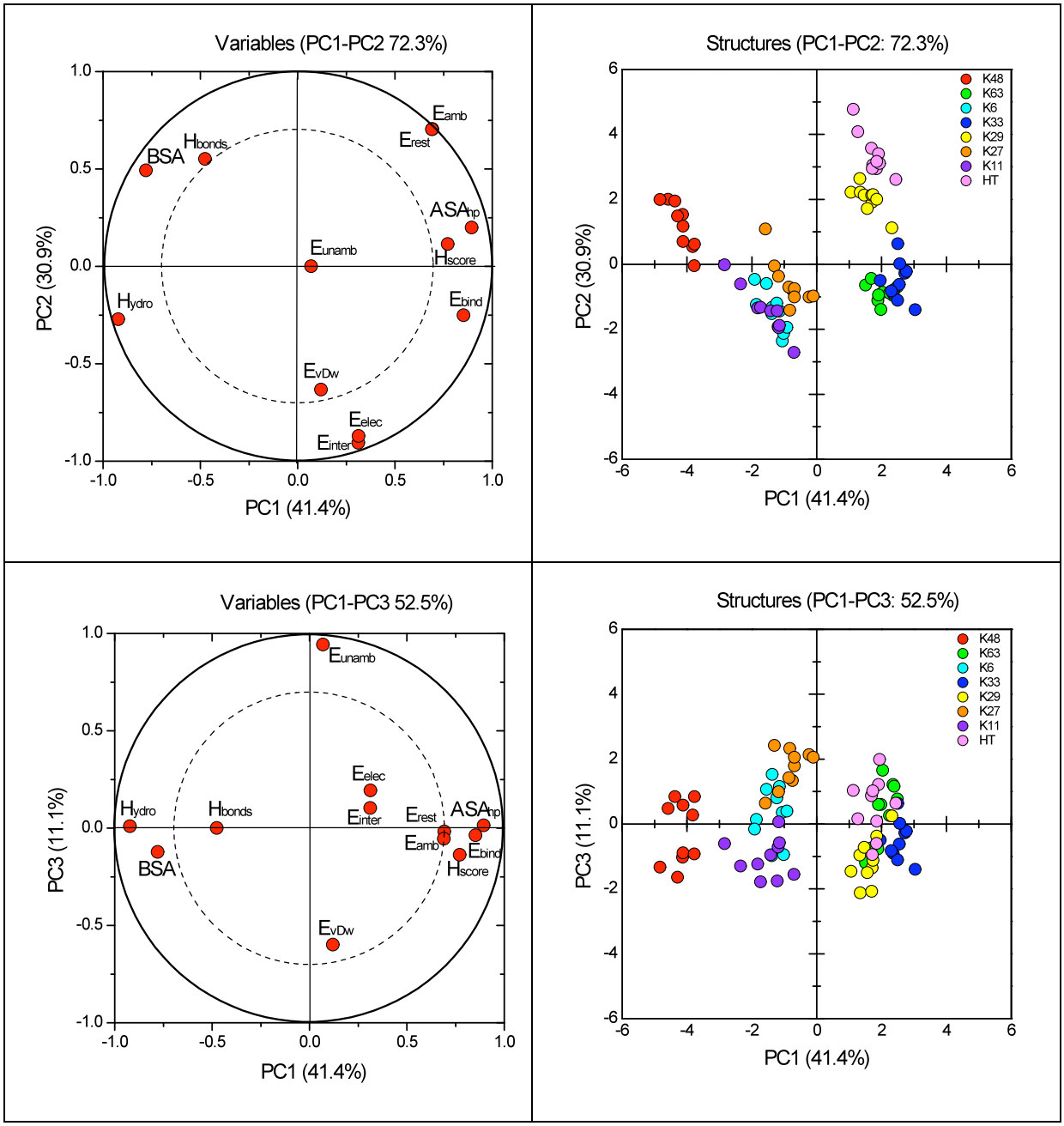


Figure S2: Principal component analysis for the 10 best structures for each generated chains



References:

1. Saha, R. P., Bahadur, R. P., Pal, A., Mandal, S. & Chakrabarti, P. (2006). ProFace: a server for the analysis of the physicochemical features of protein-protein interfaces. *BMC Struct Biol* **6**, 11.