

CoMoDo: Identifying Dynamic Protein Domains based on Covariances of Motion

Silke A. Wieninger^{1,2} and G. Matthias Ullmann^{1,*}

¹Structural Biology / Bioinformatics, University of Bayreuth,
Universitätsstr. 30, BGI, 95447 Bayreuth, Germany

²Current address: Unité de Bioinformatique Structurale, Institut Pasteur,
25-28 rue du docteur Roux, 75015 Paris, France

* Email: Matthias.Ullmann@uni-bayreuth.de

Supporting Information

Table S1. Cutoff radii and force constant ratio of the reduced parameter set^a

$r_{\text{cut}} [\text{\AA}]$	7		8			9			10		11
$\log \frac{k_{\text{ncov}}}{k_{\text{cov}}}$	-0.25	0.0	-1.0	-0.75	-0.5	-1.5	-1.25	-1.0	-1.5	-1.25	-1.75

^aThese GNM parameter pairs lead to an accordance of at least 90% in the classification as 1-domain or multidomain protein by manual methods and DomainTester. The nine parameter pairs printed in bold lead to an accordance in domain number of multidomain proteins of at least 65% analyzing all eighty proteins assigned as multidomain by manual methods.

Table S2. According assignment as 1-domain or multidomain protein by DomainTester and manual methods^a

accordance [%]	min	max	average	Bfact
full set	84.4	92.6	90.0	90.2
reduced set	91.0	92.6	92.2	91.8

^amin, max: accordance of the parameter pair from the corresponding set giving the minimal or maximal value. average: accordance averaged over all parameter pairs of the set. Bfact: GNM parameter pairs are chosen separately for each protein by comparison to experimental B-factors. Parameter pairs of the reduced set are given in Table S1 of the Supporting Information. Data set: 122 proteins of the pDomains dataset with available experimental B-factors.

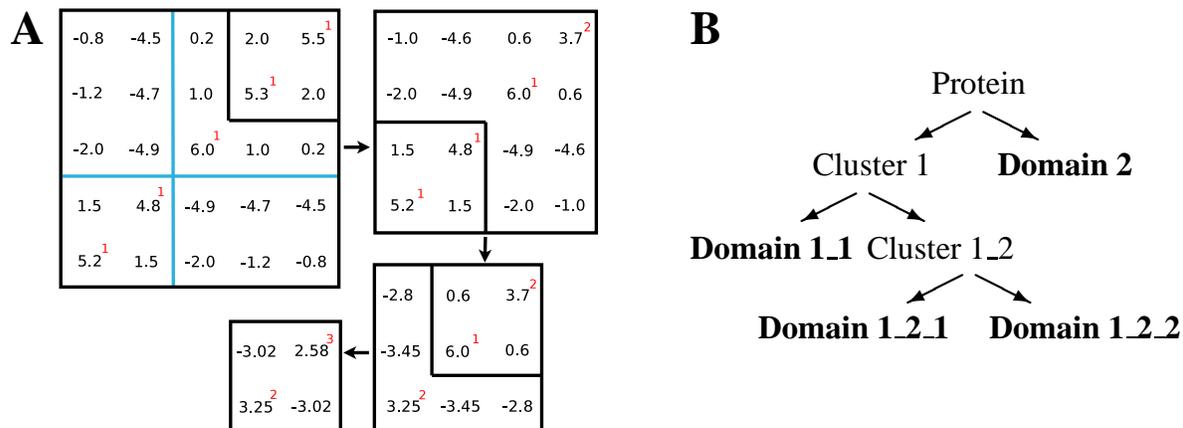


Figure S1. Identification of dynamic protein domains using CoMoDo. A) DomainClusterer example calculation of inter- and intracluster covariances upon merging of nodes. The red digits indicate the number of nodes within the cluster. The two clusters in the black box are merged next, because they have the highest intercluster covariance. The blue boxes show the final classification of the nodes into two clusters. Averaging over all covariances within one blue box gives the final intracluster covariance value, while averaging over all covariances between the two blue boxes results in the final intercluster covariance. B) Overall CoMoDo clustering scheme and domain numbering. The domain names indicate the different hierarchical levels. Domains which only differ in the last digit of the name can be combined into the precursor cluster.

Table S3. Number of domains assigned to the proteins of the pDomains dataset. Manual predictions are compared to CoMoDo predictions using different GNM parameter pairs. For the reduced and the full parameter set, the pair used in the calculation is chosen based on comparison between experimental and theoretical B-factors. 5.7 gives the number of dynamic domains calculated using a cutoff radius of 7Å and a force constant ratio $\frac{k_{\text{ncov}}}{k_{\text{cov}}}$ of 0.5. $\log K$ gives the logarithm of the force constant ratio, r the cutoff radius, corr the linear correlation coefficient between the B-factors and $\#\text{dom}$ the assigned number of dynamic domains. Proteins with bold name are assigned as multidomain both manually and by CoMoDo for all GNM parameter pairs.

PDB	chain	manual	reduced parameter set				full parameter set				5.7
		#dom	$\log K$	r	corr	#dom	$\log K$	r	corr	#dom	#dom
1a8y		3	-1.25	10	0.52	3	-1.25	10	0.52	3	3
1aba		1	0	7	0.61	1	0	6	0.64	1	1
1alc		1	-1	9	0.51	1	-1	9	0.51	1	1
1aog	A	3	-1.25	10	0.71	3	-1.5	11	0.71	3	3
1aps		1	–	–	–	–	–	–	–	–	1
1au7	A	2	0	7	-0.05	2	0	10	0.11	2	2
1aua		2	-1	8	0.64	1	-2	7	0.72	2	2
1b24	A	2	-1.5	10	0.7	2	-2	11	0.7	2	2
1bbw	A	2	-1.75	11	0.52	2	-1	11	0.53	2	2
1bc5	A	2	-0.25	7	0.59	3	-2	7	0.66	4	3
1bds		1	-1.25	10	0.81	1	-1	11	0.85	1	1
1bhg	A	3	-1.25	10	0.34	2	-1.25	10	0.34	2	3
1bi3	A	3	0	7	0.71	3	0	7	0.71	3	3
1bkb		2	-0.25	7	0.72	2	-2	11	0.73	2	2

Table S3 – continued

PDB	chain	manual	reduced parameter set				full parameter set				5_7
		#dom	log K	r	$corr$	#dom	log K	r	$corr$	#dom	#dom
1bpm		2	-1.5	9	0.65	2	-1.5	9	0.65	2	2
1btc		1	-1.75	11	0.64	2	-1.25	11	0.66	1	2
1byh		1	-0.5	8	0.54	1	-0.5	8	0.54	1	1
1c0m	A	2	-1.25	10	0.51	2	-1.25	11	0.52	2	2
1c5a		1	0	7	0.36	1	0	9	0.41	1	1
1cad		1	-0.5	8	0.72	1	0	10	0.73	1	1
1caj		1	-1.25	10	0.83	1	-1.25	10	0.83	1	1
1cgi	I	1	-1.5	9	0.7	1	-2	6	0.78	1	1
1chc		1	–	–	–	–	–	–	–	–	1
1chu	A	3	0	7	0.85	2	0	7	0.85	2	2
1ciy		3	-1	9	0.73	3	-1.25	11	0.74	3	3
1cne		2	–	–	–	–	–	–	–	–	2
1cqx	A	3	-0.25	7	0.54	2	-1	10	0.55	2	2
1crx	A	2	-0.5	8	0.08	2	0	11	0.15	2	3
1cs6	A	4	-1.75	11	0.53	3	-2	8	0.63	4	4
1csg	A	1	–	–	–	–	–	–	–	–	1
1ctn		3	0	7	0.33	3	0	11	0.4	2	3
1cun	A	2	-0.5	8	0.62	2	-1	11	0.63	2	2
1cwv	A	5	0	7	0.22	4	0	11	0.28	4	4
1d0g	T	3	-0.5	8	0.47	2	0	9	0.48	2	2
1dce	A	3	-1.5	10	0.55	6	0	10	0.55	5	6
1dfq	A	2	-1.75	11	0.62	2	-1.5	11	0.63	2	2

Table S3 – continued

PDB	chain	manual	reduced parameter set				full parameter set				5_7
		#dom	log K	r	$corr$	#dom	log K	r	$corr$	#dom	#dom
1dg3	A	2	0	7	0.71	4	0	11	0.79	3	4
1djz	A	3	-1	9	0.7	2	-0.5	11	0.73	2	3
1dnk	A	1	-1.75	11	0.83	1	-1.5	11	0.83	1	1
1duy	A	2	-1.25	10	0.63	2	-1	11	0.66	2	2
1dvp	A	2	0	7	0.02	3	0	11	0.13	2	3
1e1l	A	2	-1.25	10	0.62	2	-1.25	10	0.62	2	2
1e2v	A	2	-1.5	9	0.81	1	-1.25	7	0.83	2	1
1e39	A	3	-1.75	11	0.7	4	-1.25	11	0.71	2	3
1eaf		1	-1.5	10	0.4	1	-1.75	7	0.44	3	1
1ega	A	2	-1.5	10	0.29	2	-2	6	0.35	3	2
1eif		2	-0.25	7	0.75	2	0	6	0.79	2	2
1eqf	A	2	-1.75	11	0.59	2	-1.75	11	0.59	2	2
1fba	A	1	-1.75	11	0.38	1	-1.5	6	0.46	4	1
1fbl		2	-1.75	11	0.49	3	-1.5	11	0.49	2	2
1fc2	C	1	-1	8	0.53	1	-1	8	0.53	1	1
1fdn		1	-0.25	7	0.69	1	-0.5	6	0.71	1	1
1ffh		2	-1.75	11	0.55	2	-1.5	11	0.56	2	2
1ffu	A	2	-1.25	9	0.72	1	-0.5	6	0.72	1	1
1fha		1	0	7	0.43	2	0	7	0.43	2	2
1fmt	A	2	-1	9	0.71	2	-1	11	0.75	2	2
1fnm	A	5	-0.25	7	0.33	5	-0.75	6	0.46	6	5
1fxi	A	1	-0.5	8	0.4	1	0	8	0.4	1	1

Table S3 – continued

PDB	chain	manual	reduced parameter set				full parameter set				5_7
		#dom	log K	r	$corr$	#dom	log K	r	$corr$	#dom	#dom
1gdc		1	–	–	–	–	–	–	–	–	1
1gdd		2	-1.25	10	0.46	2	0	11	0.54	1	2
1gps		1	–	–	–	–	–	–	–	–	1
1grj		2	0	7	0.37	2	0	6	0.53	2	2
1hcc		1	–	–	–	–	–	–	–	–	1
1hiv	A	1	-1.25	10	0.12	1	0	6	0.17	1	1
1hjp		3	0	7	0.68	2	0	7	0.68	2	2
1hum	A	1	-1.75	11	0.87	1	-0.5	11	0.9	1	1
1ira	Y	3	0	7	0.45	3	-1.5	7	0.47	3	3
1jaw		2	-1.5	9	0.67	2	-1.5	9	0.67	2	2
1kit		3	-1.75	11	0.28	3	-2	11	0.29	4	3
1ksi	A	3	-1.5	9	0.6	3	-2	9	0.62	5	3
1lck	A	2	0	7	0.41	2	0	10	0.48	2	2
1mda	A	1	-1.75	11	0.22	1	-2	6	0.3	1	1
1mla		2	-1.25	9	0.69	2	-1.25	9	0.69	2	2
1mrr	A	1	0	7	0.39	1	0	10	0.4	1	1
1mup		1	0	7	0.69	1	0	7	0.69	1	1
1myt		1	-1.5	9	0.7	1	-2	9	0.72	1	1
1ovb		1	-0.25	7	0.65	1	-0.75	7	0.66	1	1
1pdz		2	-1.75	11	0.6	3	-2	11	0.6	3	3
1pky	A	3	-1.75	11	0.72	3	-1.75	11	0.72	3	3
1prs		2	–	–	–	–	–	–	–	–	2

Table S3 – continued

PDB	chain	manual	reduced parameter set				full parameter set				5_7
		#dom	log K	r	$corr$	#dom	log K	r	$corr$	#dom	#dom
1prt	B	2	-1.25	10	0.47	2	-1	11	0.49	2	2
1qba		4	-1.25	10	0.72	4	-1	10	0.72	4	4
1qd1	A	2	-0.5	8	0.55	2	0	8	0.56	2	2
1qdn	A	2	-1.75	11	0.71	1	-2	11	0.72	1	1
1qlf	A	2	-1.75	11	0.68	2	-1.25	11	0.7	2	2
1qmd	A	2	-1	9	0.38	2	-0.75	11	0.41	2	2
1qmh	A	2	-0.25	7	0.46	3	-0.25	7	0.46	3	3
1qni	A	2	-1.25	10	0.32	3	-0.25	10	0.33	2	3
1qnt	A	2	-1.75	11	0.47	2	-1.25	11	0.49	1	2
1rcb		1	0	7	0.62	1	0	6	0.64	1	1
1rve	A	1	0	7	0.51	1	0	7	0.51	1	1
1sky	E	3	-1	8	0.75	2	-1.5	8	0.78	3	2
1tbr	S	2	0	7	-0.23	2	0	11	-0.06	2	2
1tnr	A	1	-1	9	0.83	1	-1	9	0.83	1	1
1ttb	A	1	-1.75	11	0.6	1	-2	11	0.64	1	1
1tui	A	3	-1.75	11	0.43	2	0	6	0.48	3	3
1ula		1	0	7	0.57	1	-1.5	7	0.59	1	1
1urk		2	–	–	–	–	–	–	–	–	2
1ush		2	-1.25	9	0.79	2	-1.25	9	0.79	2	2
1utg		1	-1.75	11	0.61	1	-2	11	0.62	1	1
1vdc		2	-1.25	10	0.64	2	-1	10	0.64	2	2
1vfa	B	1	-1.75	11	0.57	1	-1.75	11	0.57	1	1

Table S3 – continued

PDB	chain	manual	reduced parameter set				full parameter set				5_7
		#dom	log K	r	$corr$	#dom	log K	r	$corr$	#dom	#dom
1vol	A	2	-0.25	7	0.68	2	-1	7	0.76	2	2
1vvc		2	-0.5	8	0.61	2	0	8	0.62	1	1
1wgt	A	4	-1.25	10	0.65	3	-0.75	10	0.67	3	4
1whe		2	–	–	–	–	–	–	–	–	1
1wrp	R	1	-0.25	7	0.73	2	-1.75	7	0.75	2	2
1xim	A	1	0	7	0.2	4	0	6	0.29	4	4
1ytf	D	2	-1.75	11	0.72	2	-2	11	0.73	2	2
2bs2	B	2	-1.5	9	0.63	2	-2	10	0.64	2	2
2cbl	A	3	-1.75	11	0.59	3	-2	11	0.6	3	2
2cgp	A	2	-1.25	10	0.66	2	-1	10	0.66	2	2
2ech		1	–	–	–	–	–	–	–	–	1
2fcr		1	–	–	–	–	–	–	–	–	1
2gb0	A	2	-1.75	11	0.76	2	-2	11	0.77	2	2
2gli	A	5	0	7	0.26	2	0	11	0.36	2	2
2mad	L	1	-1.75	11	0.67	1	-2	11	0.67	1	1
2msb	A	1	-1	9	0.65	1	-0.25	6	0.68	1	1
2pcd	M	1	-0.25	7	0.44	1	-1.25	6	0.53	1	1
2pf2		1	-1.25	10	0.36	1	0	11	0.45	1	1
2por		1	-0.5	8	0.26	1	0	10	0.32	1	1
2shp	A	3	-1.25	10	0.69	3	-1.25	10	0.69	3	4
2xsc	A	1	-1.75	11	0.26	1	-1.5	11	0.26	1	1
3chy		1	-0.5	8	0.79	1	-0.5	8	0.79	1	1

Table S3 – continued

PDB	chain	manual	reduced parameter set				full parameter set				5_7
		#dom	log K	r	$corr$	#dom	log K	r	$corr$	#dom	#dom
3hdh	A	2	-0.25	7	0.6	2	-2	6	0.68	4	2
3tec	I	1	-1.5	9	0.15	1	-2	6	0.39	1	1
3tf4	A	2	-1	8	0.51	4	0	11	0.53	2	3
4cp4		1	-1.75	11	0.78	2	-1.75	11	0.78	2	1
4htc	I	1	-1.25	10	-0.18	1	0	11	-0.15	1	1
4icb		1	-1	8	0.69	1	-0.75	7	0.72	1	1
4pti		1	-0.25	7	0.76	1	-1.25	7	0.81	1	1
5eau		2	-0.75	8	0.7	4	-0.75	8	0.7	4	4
5nn9		1	-1.75	11	0.55	2	-1.75	10	0.55	3	2
5p21		1	-0.75	8	0.52	1	-0.75	8	0.52	1	1
6ebx	A	1	-0.75	8	0.78	1	-0.75	8	0.78	1	1