Supplementary material for:
Periodic distributions of hydrophobic amino acids allows to define fundamental building blocks to align distantly related proteins
by J. Baussand, C. Deremble, A. Carbone

		Non overlap			Overlap			
	Sequence	rss	hb	thb	rss	hb	thb	Concerned residues
%identity	11.8	20.0	20.5	21.1	15.1	12.9	13.5	
Polarity	0.51	0.67	0.60	0.62	0.56	0.49	0.51	N, Q, S, T, K, R, H, D, E
Apolarity	0.53	0.72	0.82	0.83	0.60	0.59	0.61	A, C, G, P, I, L, M, F, W, Y, V
Hydrophobicity	0.75	1.02	1.25	1.27	0.85	0.78	0.81	I, L, M, F, W, Y, V
Acidity	0.60	0.77	0.73	0.75	0.64	0.59	0.61	D, E, N, Q
Basicity	0.74	0.99	0.88	0.91	0.81	0.72	0.74	K, R, H
H-Bonding	0.48	0.62	0.55	0.56	0.53	0.48	0.49	C, W, N, Q, S, T, Y, K, R, H, D, E
Ionizability	0.52	0.68	0.62	0.64	0.57	0.53	0.54	D, E, H, C, Y, K, R
Aromaticity	0.96	1.23	1.48	1.48	1.04	0.99	1.02	F, W, Y, H
Aliphaticity	0.76	0.75	0.83	0.85	0.63	0.62	0.64	G, A, V, L, I
Volume 60-100 Å^3	0.58	0.70	0.64	0.66	0.60	0.37	0.56	A,G,S
Volume 100-120 Å^3	0.53	0.53	0.50	0.50	0.50	0.24	0.43	N,D,B,C,P,T
Volume 120-150 Å^3	0.49	0.65	0.66	0.68	0.55	0.43	0.53	Q,E,Z,V
Volume 150-185 Å^3	0.52	0.69	0.75	0.77	0.58	0.46	0.58	H,I,L,K,M,R
Volume 185-230 Å^3	1.20	1.55	1.87	1.86	1.30	0.98	1.28	F,Y,W

Supplementary Table 1 . Conservation rate of physico-chemical properties in a pattern (that is sequence, rss, hb and thb) calculated as $f^{pat}_{obs_{SS}}/f^{seq}_{exp_{SS}}$, where $f^{pat}_{obs_{SS}}$ is the observed frequency of pairs of aa satisfying the same physico-chemical property S in the pattern and $f^{seq}_{exp_{SS}}$ is the expected frequency of pairs of aa satisfying the same physico-chemical property S in the sequence. (Pairs with gaps are included in the calculation.) Results are presented for the 2 hypothesis *non overlap* (where both residues of the pairs considered belong to the pattern) and *overlap* (where at least one residue of the pair belongs to the pattern).

AAA		1VJS.-	(3-393)	1F3H.A	(5-140)	1JKW.-	(11-287)
1E94.E	(1-443)	1GCY.A	(1-357)	**blmb**		**cys**	
1E32.A	(191-458)	1AVA.A	(1-346)	1SML.A	(2-267)	1AIM.-	(1-212)
1D2N.A	(505-750)	1EHA.A	(91-490)	1K07.A	(36-311)	1THE.A	(1-253)
aadh		1BF2.-	(163-637)	1ZNB.A	(20-249)	1DEU.A	(1-242)
1LEH.A	(1-364)	1G5A.A	(77-554)	1QH5.A	(1-260)	**cyt3**	
1HWX.A	(1-501)	1GJW.A	(68-572)	**bv**		3CYR.-	(1-107)
1BGV.A	(1-449)	**alpha-amylase_NC**		2TBV.A	(102-273)	2CY3.-	(1-118)
aat		1BF2.-	(163-750)	1SMV.A	(65-260)	1AQE.-	(2-111)
3TAT.A	(5-408)	1EHA.A	(91-557)	1STM.A	(17-157)	**cytb**	
1B8G.A	(3-430)	1G94.A	(1-448)	1BMV.1	(1001-1185)	1SOX.A	(3-93)
1BJW.A	(1-382)	1BAG.-	(1-425)	1CWP.A	(42-190)	1CYO.-	(1-88)
1BW0.A	(4-415)	1BVZ.A	(121-585)	2STV.-	(12-195)	1CXY.A	(6-86)
ABC_tran		1UOK.-	(1-558)	**cah**		1KBI.A	(1-92)
1F3O.A	(2-232)	7TAA.-	(1-476)	2CAB.-	(5-260)	**dCMP_cyt_deam**	
1G29.1	(1-240)	1QHP.A	(1-495)	1KOQ.A	(5-226)	1JTK.A	(1-131)
1G6H.A	(4-257)	1VJS.-	(3-482)	1ZNC.A	(5-259)	1CTT.-	(48-171)
1E69.A	(1-1164)	1GCY.A	(1-418)	**cbm12**		1CTT.-	(189-294)
Acetyltransf		1AVA.A	(1-403)	1AIW.-	(1-62)	**DEATH**	
1BO4.A	(25-160)	1G5A.A	(77-628)	1E15.A	(449-498)	1E41.A	(89-192)
1QST.A	(49-208)	1GJW.A	(68-636)	1ED7.A	(655-698)	1ICH.A	(327-413)
1B87.A	(1-181)	**ANK**		**CBM_20)**		1NGR.-	(334-418)
1CJW.A	(30-195)	1AWC.B	(5-127)	1QHP.A	(577-686)	1DDF.-	(201-327)
ACPS		1IKN.D	(143-272)	1KUM.-	(509-616)	1D2Z.A	(28-129)
1QR0.A	(105-211)	1YCS.B	(327-455)	1CQY.A	(418-516)	1D2Z.B	(23-172)
1F7L.A	(1-118)	**AP_endonuc_2**		**cbp**		**dhfr**	
1QR0.A	(1-104)	1K77.A	(2-260)	1AJ4.-	(2-161)	8DFR.-	(1-186)
ALBUMIN		1I60.A	(2-277)	1BR1.B	(3-150)	4DFR.A	(1-159)
1BJ5.-	(3-197)	1QTW.A	(1-285)	2SCP.A	(1-174)	3DFR.-	(1-162)
1BJ5.-	(198-389)	**Asp_Glu_race_D**		2SAS.-	(1-185)	**DHOdehase**	
1BJ5.-	(390-584)	1JFL.A	(1-115)	**CBS**		1H7W.A	(533-844)
aldosered		1B74.A	(1-105)	1ZFJ.A	(95-158)	1EP3.A	(1-311)
1AFS.A	(1-319)	1JFL.A	(116-226)	1B3O.B	(178-231)	1D3G.A	(30-396)
1A80.-	(2-278)	1B74.A	(106-213)	1ZFJ.A	(160-220)	2DOR.A	(1-311)
1QRQ.A	(36-360)	**az**		**CH**		**DMRL_synthase**	
Ald_Xan_dh_2		1CUO.A	(1-129)	1BHD.A	(147-254)	1C2Y.A	(1-155)
1HLR.A	(194-907)	1PMY.-	(1-123)	1AOA.-	(121-251)	1C41.A	(5-199)
1FO4.A	(537-1332)	1KDJ.-	(1-102)	1AOA.-	(260-375)	1DI0.A	(11-155)
1N62.B	(6-809)	1QHQ.A	(2-140)	**chromo**		**dsrm**	
alpha-amylase_C		1ID2.A	(1-106)	1G6Z.A	(-2-69)	1QU6.A	(91-179)
1VJS.-	(394-482)	1GY1.A	(2-155)	1AP0.-	(8-80)	1QU6.A	(1-90)
7TAA.-	(382-476)	**Bac_DNA_binding**		1DZ1.A	(102-171)	1STU.-	(1-68)
1GJW.A	(573-636)	1WTU.A	(1-99)	**CPSase_L_chain**		**DUF170**	
1SMA.A	(506-588)	1IHF.A	(2-97)	1A9X.A	(556-676)	1F0Z.A	(1-66)
1UOK.-	(480-558)	1IHF.B	(1-94)	1DV1.A	(2-114)	1FM0.D	(1-81)
1G5A.A	(555-628)	**Binary_toxA**		1GSO.A	(2-103)	1JSB.A	(4-73)
1BVZ.A	(503-585)	1G24.A	(41-250)	**CPS**		**ech**	
1CYG.-	(403-491)	1QS1.A	(265-461)	1A9X.A	(128-402)	1NZY.A	(1-269)
1BAG.-	(348-425)	1QS1.A	(60-264)	1A9X.A	(677-935)	2DUB.A	(32-290)
1G94.A	(355-448)	**biotin_lipoyl**		1BNC.A	(115-330)	1DCI.A	(53-327)
1AVA.A	(347-403)	1BDO.-	(77-156)	**Cu_amine_oxid**		**egf**	
alpha-amylase		1PMR.-	(1-80)	1A2V.A	(18-672)	1URK.-	(6-46)
1G94.A	(1-354)	1QJO.A	(1-80)	1SPU.A	(91-724)	1ESL.-	(121-157)
1BAG.-	(1-347)	1FYC.-	(1-106)	1KSI.A	(6-647)	1HRE.-	(175-241)
1BVZ.A	(121-502)	**BIR**		**cyclin**		1EPI.-	(1-53)
1UOK.-	(1-479)	1G73.C	(256-344)	1VIN.-	(181-432)	4TGF.-	(1-50)
7TAA.-	(1-381)	1C9Q.A	(1-117)	1BU2.A	(22-250)	1DAN.L	(47-86)
1QHP.A	(1-407)	1F3H.A	(5-140)	1JKW.-	(11-287)	1RFN.B	(86-142)

Supplementary Table 2. Proteins coming from HOMSTRAD database used to construct matrices IHBM and OHBM. For each family, the PDB name and the region considered are reported.

EGF_Lam	
1KLO.-	(66-121)
1KLO.-	(11-65)
1KLO.-	(122-172)
Epimerase	
1BXK.A	(1-350)
1UDC.-	(1-338)
1BWS.A	(3-316)
1DB3.A	(1-357)
fabp	
1EIO.A	(1-127)
1GGL.A	(1-134)
1MDC.-	(1-131)
FAD_binding_4	
1I19.A	(57-273)
1F0X.A	(9-273)
1VAO.A	(6-273)
FAD-oxidase_C	
1DII.A	(243-521)
1VAO.A	(274-560)
1F0X.A	(274-567)
1I19.A	(274-613)
FAD-oxidase_NC	
1VAO.A	(6-560)
1F0X.A	(9-567)
1I19.A	(57-613)
fer2	
1FRR.A	(1-95)
1AYF.A	(6-108)
1B9R.A	(1-105)
fer4	
1DUR.A	(1-55)
7FD1.A	(1-106)
1BLU.-	(1-80)
1XER.-	(1-103)
1HFE.L	(2-86)
1FXD.-	(1-58)
1K0T.A	(1-80)
flav	
2FCR.-	(1-173)
1AKR.-	(2-148)
5NLL.-	(1-138)
fn3	
1FNF.-	(1327-1415)
1FNF.-	(1142-1235)
1FNF.-	(1416-1509)
1CFB.-	(610-709)
1CTO.-	(1-109)
1EBP.A	(119-220)
3HHR.B	(131-234)
1BOY.-	(107-213)
1BOY.-	(3-106)
1CFB.-	(710-814)
1EBP.A	(22-118)
GAF	
1MC0.A	(402-555)
1MC0.A	(233-372)
1F5M.A	(4-179)
GATase	
1QDL.B	(1-195)
1GPM.A	(3-207)
1A9X.B	(1653-1880)
GEL	
1D0N.A	(629-755)
1D0N.A	(263-383)
1SVR.-	(158-251)
2VIL.-	(1-126)
ghf18	
2EBN.-	(5-289)
1CNV.-	(1-283)
1NAR.-	(1-289)
ghf1	
1E73.M	(3-333)
1QOX.A	(2-450)
1QVB.A	(1-481)
ghf33	
2SLI.-	(277-759)
1EUR.-	(47-407)
3SIL.-	(4-382)
ghf5	
1EGZ.A	(1-291)
1BQC.A	(1-302)
1EDG.-	(1-380)
1CEO.-	(1-340)
1ECE.A	(1-358)
glob	
1LH1.-	(1-153)
1HLB.-	(1-157)
2LHB.-	(1-149)
2HBG.-	(1-147)
3SDH.A	(2-146)
1MBA.-	(1-146)
1ITH.A	(1-141)
1ECD.-	(1-136)
1A6M.-	(1-151)
1CG5.B	(1-141)
gluts	
17GS.A	(0-209)
1GUL.A	(4-220)
3GTU.B	(1-224)
Glyco_hydro_18_D2	
1D2K.A	(293-354)
1EDQ.A	(444-516)
1E9L.A	(267-337)
1E15.A	(292-379)
Glyco_hydro_18	
1D2K.A	(36-427)
1EDQ.A	(133-563)
1E15.A	(3-446)
1E9L.A	(22-393)
Glyco_hydro_2	
1DP0.A	(220-333)
1DP0.A	(626-730)
1BHG.A	(226-328)
grs	
1GER.A	(3-450)
1GER.A	(3-450)
1LVL.-	(1-458)
1NPX.-	(1-447)
Haloperoxidase	
1B6G.-	(1-310)
1CQW.A	(15-309)
1EHY.A	(2-294)
1CR6.A	(228-544)
1BRT.-	(1-277)
1C4X.A	(3-283)
HATPase_c	
1EI1.A	(2-220)
1B63.A	(-2-202)
1H7S.A	(29-217)
1AH6.-	(2-214)
helicase_C	
1D9X.A	(415-595)
1FUK.A	(233-394)
1HEI.A	(326-480)
hexapep	
1LXA.-	(1-180)
1QRE.A	(9-175)
3TDT.-	(102-236)
HGTP_anticodon	
1EVL.A	(533-642)
1HC7.A	(281-395)
1ATI.A	(395-505)
1H4V.B	(325-421)
1KMM.A	(326-424)
1QE0.A	(326-420)
histone	
1AOI.C	(11-106)
1AOI.A	(59-135)
1AOI.D	(26-100)
1AOI.B	(20-102)
HLH	
1MDY.A	(1-166)
1AN4.A	(196-260)
1HLO.A	(3-82)
1AM9.A	(319-398)
1A0A.A	(0-62)
HMA	
2HQI.-	(1-72)
2AW0.-	(1-72)
1CPZ.A	(1-68)
HMG_box	
1HRZ.A	(3-75)
1CKT.A	(7-77)
1HMA.-	(2-74)
2LEF.A	(1-86)
hom	
1FTT.-	(0-67)
1B8I.B	(205-259)
1LFB.-	(13-89)
hormone_rec	
1LBD.-	(225-462)
2LBD.-	(182-419)
1PRG.A	(207-476)
1PRG.A	(207-476)
3ERT.A	(306-551)
1A28.A	(682-932)
igcon	
1FC1.A	(238-341)
3HFL.H	(119-223)
1CQK.A	(4-104)
igI	
1VCA.A	(1-95)
1TIT.-	(1-89)
1TLK.-	(33-135)
1NCT.-	(-6-91)
1WIT.-	(1-93)
2NCM.-	(1-99)
1ZXQ.-	(1-93)
1ITB.B	(104-205)
1ITB.B	(206-315)
igps	
1PII.-	(1-255)
1PII.-	(256-452)
1NSJ.-	(1-205)
igV	
1TVD.B	(1-116)
1B88.A	(1-114)
3CD4.-	(1-97)
1NEU.-	(1-119)
il8	
1B3A.A	(2-68)
2EOT.-	(1-74)
1ROD.A	(1-72)
1SDF.-	(1-67)
inositol_P	
1QGX.A	(2-355)
2HHM.A	(5-276)
1INP.-	(1-400)
intb	
2FGF.-	(19-144)
1I1B.-	(3-153)
1IRA.X	(7-151)
int	
1IDO.-	(132-315)
1ATZ.A	(925-1108)
1AUQ.-	(498-705)
1AOX.A	(139-339)
kinase	
1CKI.A	(1-304)
1A06.-	(10-316)
1PHK.-	(15-291)
1CDK.A	(8-350)
1TKI.A	(18-338)
1BLX.A	(5-309)
1LR4.A	(7-333)
1B6C.B	(175-500)
kunitz	
5PTI.-	(1-58)
1TAP.-	(1-60)
1BF0.-	(1-60)

Supplementary Table 2. (Continued)

1BFJ.-	(1-111)
2PLD.A	(1-105)
1ZFP.E	(56-153)
sh3	
1GRI.A	(157-217)
1YCS.B	(457-519)
2HSP.-	(1-71)
1ARK.-	(1-60)
1PHT.-	(3-85)
1BB9.-	(12-94)
Sm	
1D3B.B	(7-87)
1B34.B	(26-118)
1B34.A	(2-81)
1D3B.A	(4-75)
Stap_Strp_toxin	
1ESF.A	(1-233)
1AN8.-	(3-208)
3TSS.-	(5-194)
subt	
1GT9.1	(1-357)
1GA6.A	(2-370)
1EA7.A	(1-310)
1IC6.A	(1-279)
Sulfotransfer	
1EFH.A	(4-286)
1FMJ.A	(8-349)
1NST.A	(579-879)
TNF	
2TNF.A	(9-157)
1D4V.B	(119-281)
1ALY.-	(116-261)
Toprim	
1D6M.A	(1-135)
1ECL.-	(3-140)
1DD9.A	(259-341)
toxin_2	
1C56.A	(1-40)
1CMR.-	(1-31)
1PNH.-	(1-31)
1CHL.-	(1-36)
TPR	
1A17.-	(19-177)
1ELW.A	(2-118)
1ELR.A	(222-349)
1E96.B	(2-186)
1FCH.A	(280-480)
1IHG.A	(220-364)
tRNA_bind	
1B7Y.B	(39-151)
1GD7.A	(1-109)
1FL0.A	(150-255)
tRNA-synt_2b	
1ADJ.A	(2-325)
1QF6.A	(241-531)
1ATI.A	(1-394)
uce	
1J7D.A	(6-145)
1AYZ.A	(2-154)
1C4Z.D	(4-147)
WW	
1PIN.A	(6-39)
1E0L.A	(1-37)
1EG4.A	(47-84)
Yers_vir_YopE	
1HY5.A	(1100-1218)
1G4W.R	(171-290)
1HE1.A	(95-229)

Supplementary Table 2. (Continued)

193l.-.-	153l.-.-	1dhk.A.-	1bag.-.-	1pot.-.-	1sbp.-.-
1aba.-.-	1gp1.A.-	1dvr.A.-	1dts.-.-	1ptv.A.-	1ytn.-.-
1acf.-.-	1pne.-.-	1dyn.B.-	1irs.A.-	1qpa.A.-	2cyp.-.-
1afi.-.1	1aps.-.1	1dyn.B.-	1mai.-.-	1ris.-.-	1spb.P.-
1agj.A.-	1elg.-.-	1eaf.-.-	3cla.-.-	1rmg.-.-	1bhe.-.-
1agq.D.-	1tgj.-.-	1ece.A.-	1edg.-.-	1ryt.-.-	1afr.F.-
1aiz.B.-	1rcy.-.-	1ecm.B.-	1csm.B.-	1ryt.-.-	1xik.A.-
1ak1.-.-	2dri.-.-	1ecp.A.-	1ula.-.-	1ryt.-.-	1xsm.-.-
1aoy.-.1	2dtr.-.-	1elg.-.-	1hav.A.-	1sbp.-.-	2abh.-.-
1apm.E.-	1erk.-.-	1elg.-.-	2alp.-.-	1spb.P.-	1nue.A.-
1apm.E.-	1irk.-.-	1erk.-.-	1irk.-.-	1ste.-.-	2tss.A.-
1aps.-.1	1spb.P.-	1esl.-.-	1lit.-.-	1ste.-.-	3ull.A.-
1ash.-.-	1bin.A.-	1eur.-.-	2sim.-.-	1taf.A.-	1bfm.A.1
1ash.-.-	1bvd.-.-	1fec.A.-	1nhq.-.-	1taf.A.-	1taf.B.-
1ash.-.-	1cpc.A.-	1fmb.-.-	1sme.B.-	1tdj.-.-	1psd.A.-
1ax4.A.-	1cl1.A.-	1fna.-.-	1msp.A.-	1tdj.-.-	2tys.B.-
1bbh.B.-	1nbb.B.-	1fui.A.-	1bhs.-.-	1tii.D.-	3ull.A.-
1bbp.D.-	1hbq.-.-	1gsa.-.-	1bnc.A.-	1ulo.-.-	2ayh.-.-
1bcp.L.-	1prt.B.-	1gsa.-.-	1iow.-.-	1urn.A.-	1spb.P.-
1bdi.A.-	2dri.-.-	1gtq.A.-	1gtp.A.-	1vin.-.-	1ad6.-.-
1bdm.B.-	1bhs.-.-	1hce.-.-	1i1b.-.-	1vlt.A.-	1nbb.B.-
1bdm.B.-	6ldh.-.-	1hce.-.-	4fgf.-.-	1wba.-.-	1i1b.-.-
1bfm.A.1	1taf.B.-	1hfc.-.-	1iag.-.-	1wba.-.-	4fgf.-.-
1bin.A.-	1bvd.-.-	1hrd.A.-	1leh.A.-	1wkt.-.1	1amm.-.-
1bin.A.-	2hbg.-.-	1i1b.-.-	4fgf.-.-	1xik.A.-	1afr.F.-
1bnk.A.-	1fmt.A.-	1idk.-.-	1air.-.-	1xik.A.-	1xsm.-.-
1brz.-.1	1gps.-.1	1iow.-.-	1bnc.A.-	1xsm.-.-	1afr.F.-
1btn.-.-	1dyn.B.-	1irs.A.-	1mai.-.-	2alp.-.-	1hav.A.-
1btn.-.-	1irs.A.-	1kpc.D.-	1hxq.B.-	2blt.B.-	3pte.-.-
1btn.-.-	1mai.-.-	1lea.-.-	1ruo.B.-	2dri.-.-	1rnl.-.-
1bvd.-.-	2hbg.-.-	1lmb.3.-	1pou.-.1	2fb4.H.-	1fna.-.-
1cew.I.-	1mol.A.-	1lpe.-.-	1nbb.B.-	2fb4.H.-	1tup.A.-
1cew.I.-	1oun.A.-	1lpe.-.-	1vlt.A.-	2gmf.B.-	1rcb.-.-
1cnv.-.-	1nar.-.-	1lti.D.-	1bcp.L.-	2hfh.-.1	1hst.A.-
1cpc.A.-	1col.A.-	1lti.D.-	1prt.B.-	2hhm.A.-	1spi.D.-
1cpc.A.-	1cpc.B.-	1lti.D.-	1tii.D.-	2lef.A.1	1hma.-.1
1cpc.A.-	2hbg.-.-	1ndh.-.-	1fnb.-.-	2pia.-.-	1fnb.-.-
1ctj.-.-	1cxc.-.-	1ndh.-.-	2pia.-.-	2pii.-.-	1aps.-.1
1ctj.-.-	2mta.C.-	1nul.A.-	1hgx.A.-	3nll.-.-	1qrd.B.-
1dat.-.-	1afr.F.-	1oun.A.-	1std.-.-	3ull.A.-	1bcp.L.-
1dat.-.-	1ryt.-.-	1pgs.-.-	1phm.-.-	6ldh.-.-	3nll.-.-
1den.-.1	1tcp.-.1	1plq.-.-	2pol.A.-		

Supplementary Table 3. Domingues dataset

<12%	
1ajs.A	2dkb.-
1bbt.3	1bbt.2
1aym.3	1bbt.2
1bbt.2	1bbt.1
2sga.-	1svp.A
1hav.A	1svp.A
1idy.-	1aoy.-
1hst.A	1tc3.C
1neq.-	1au7.A
1neq.-	1a04.A
1eny.-	1dhr.-
1grx.-	2trc.P

12-15%	
1ycs.B	1ihv.A
1pht.-	1ihv.A
1ihv.A	1bb9.-
1ajs.A	2gsa.A
1ajs.A	1ax4.A
2dkb.-	1ax4.A
1aym.3	1aym.1
1aym.3	1bbt.1
1bbt.2	1aym.1
1hav.A	2sga.-
1agj.A	1svp.A
5ptp.-	1svp.A
1idy.-	1tc3.C
1hst.A	1aoy.-
1hst.A	1jhg.A
1tc3.C	1aoy.-
1tc3.C	1jhg.A
1r69.-	1neq.-
1r69.-	1a04.A
1au7.A	1a04.A
1prt.F	1sap.-
1prt.F	1lt5.D
1ubi.-	1gua.B
1ubi.-	1awd.-
1gua.B	1awd.-
1uky.-	1bif.-
1tlk.-	1hnf.-
2myr.-	1edg.-
2myr.-	1ceo.-
1gow.A	1ceo.-
1thx.-	2trc.P
1grx.-	1erv.-
1nhp.-	2tmd.A

1fcd.A	2tmd.A
yb00_yeast	s50366
put3_yeast	yb00_yeast
yhx8_yeast	cat8_yeast
yhx8_yeast	yb00_yeast
yhx8_yeast	s50366
cat8_yeast	s50366

15-17%	
1abo.A	1ihv.A
2gsa.A	1ax4.A
1bbt.3	1aym.1
1bbt.3	1bbt.1
1oxa.-	2hpd.A
1agj.A	2sga.-
1idy.-	1jhg.A
1lvl.-	1fcd.A
1bf2.-	1vjs.-
1smd.-	1bf2.-
1smd.-	1vjs.-
1sbp.-	1mrp.-
1sbp.-	1atg.-
1mrp.-	1atg.-
1tvx.A	1prt.F
1dvr.A	1gky.-
1dvr.A	1bif.-
1gky.-	1bif.-
1hnf.-	1vca.A
1wit.-	1vca.A
1tlk.-	1vca.A
1fds.-	1dhr.-
1gow.A	1edg.-
1thx.-	1grx.-
3grs.-	2tmd.A
4enl.-	2mnr.-
put3_yeast	yhx8_yeast
put3_yeast	cat8_yeast
put3_yeas	s50366

17-20%	
1abo.A	1pht.-
1abo.A	1bb9.-
1ycs.B	1pht.-
1phd.-	2hpd.A
1hav.A	1agj.A
1hav.A	5ptp.-
1aoy.-	1jhg.A
1idy.-	1hst.A

3grs.-	1nhp.-
3grs.-	1fcd.A
1nhp.-	1fcd.A
1pam.A	1jdc.A
1smd.-	1jdc.-
1jdc.-	1bf2.-
1ped.A	1qor.A
1r69.-	1au7.A
1sbp.-	1pot.-
1mrp.-	1pot.-
1pot.-	1atg.-
1tvx.A	1lt5.D
1sap.-	1lt5.D
1ubi.-	1alo.-
1wit.-	2ncm.-
1wit.-	1hnf.-
1tlk.-	2ncm.-
2ncm.-	1hnf.-
2hsd.A	1dhr.-
1edg.-	1ceo.-
2pia.-	1fnc.-
2pia.-	1fdr.-
2pia.-	1ndh.-
1fnc.-	1fdr.-
1fnc.-	1ndh.-
1erv.-	2trc.P
3grs.-	1fcd.A
1nhp.-	1fcd.A
4enl.-	1muc.A
daf1_caeel	1csn.-
dmk_human	kpro_maize
kpro_maize	1csn.-
1mnm.C	1au7.A

20-22%	
1aym.1	1bbt.1
1cpt.-	2hpd.A
1agj.A	5ptp.-
5ptp.-	2sga.-
1pam.A	1smd.-
1pam.A	1bf2.-
1jdc.-	1vjs.-
1sbp.-	1wod.-
1wod.-	1pot.-
1gua.B	1alo.-
1alo.-	1awd.-
1wit.-	1tlk.-
1fds.-	1eny.-

cat8_yeast	yb00_yeast
kcc2_yeast	daf1_caeel
kcc2_yeast	1csn.-
dmk_human	daf1_caeel
1ftz.6	1au7.A

22-25%	
hmgl_trybr	hmgb_chite
1abo.A	1ycs.B
1ycs.B	1bb9.-
2dkb.-	2gsa.A
1bbt.3	1aym.3
1oxa.-	1phd.-
1ftz.-	1akh.A
1mnm.C	1akh.A
glms_bascu	pur1_vigac
glms_bascu	pur1_rat
glms_bascu	1ecf.A
1gow.-	1myr.-
cr2_human	cfhd_human
1hip.-	2hip.A
1ldg.-	1hyh.A
1hyh.A	1hlp.B
1lvl.-	1nhp.-
1pam.A	1vjs.-
1qor.A	2ohx.A
1gsu.-	2gsq.-
1gsu.-	1gsd.-
1wod.-	1mrp.-
1tvx.A	1sap.-
1uky.-	1dvr.A
1uky.-	1gky.-
2ncm.-	1vca.A
cah1_human	cah2_chlre
cah4_rat	cah2_chlre
cah6_human	cah2_chlre
2hsd.A	1fds.-
1fdr.-	1ndh.-
1thx.-	1erv.-
3grs.-	1nhp.-
cy2_rhoge	c550_bacsu
c550_bacsu	1etp.-
c550_bacsu	1cch.-
arp3_b_tau	arpx_c_ele
kcc2_yeast	dmk_human
kcc2_yeast	kpro_maize
dmk_human	1csn.-
kpro_maize	daf1_caeel

Supplementary Table 4. BAliBASE dataset

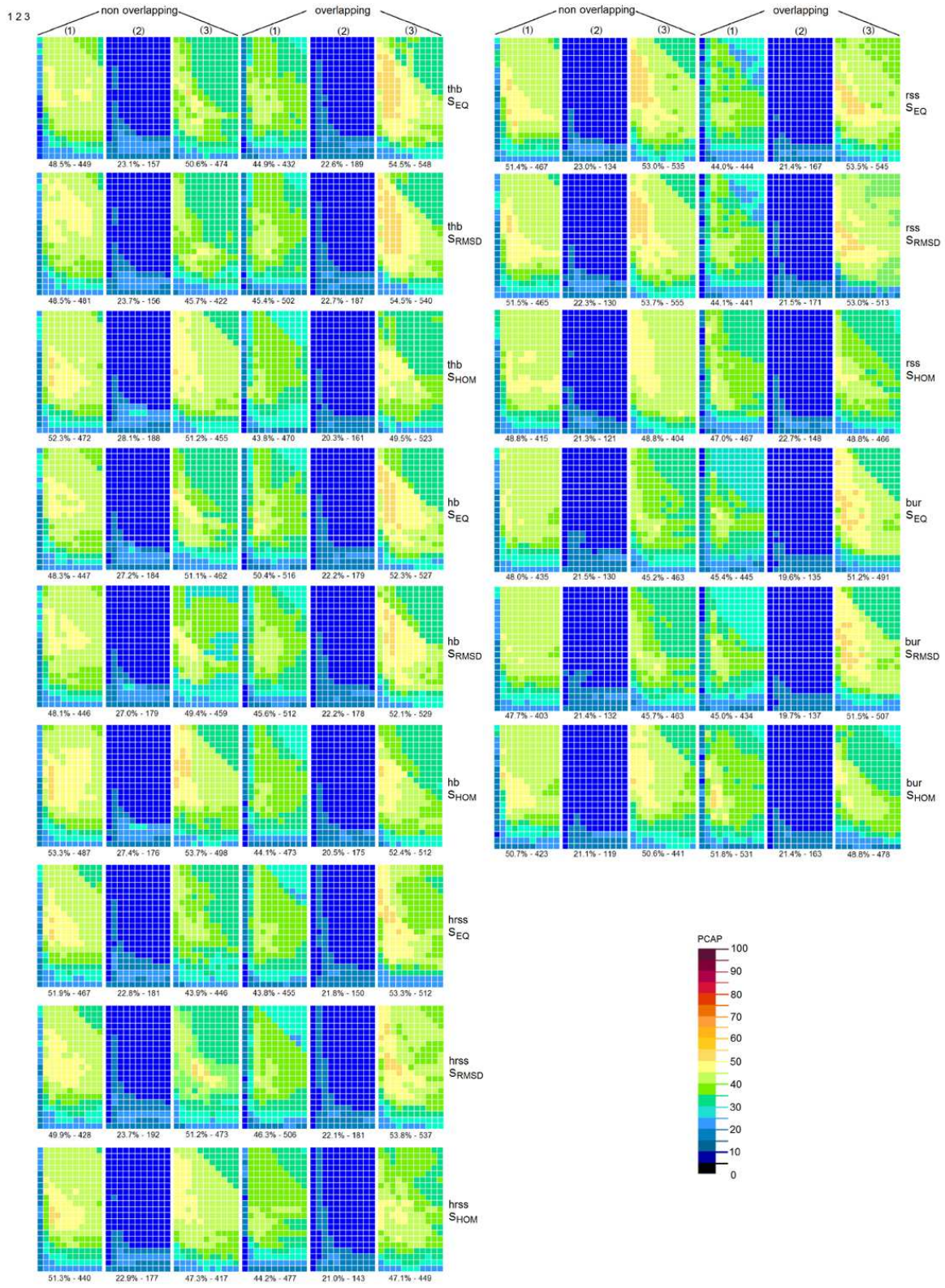

Figure 1: Landscapes associated to the 90 pairs of matrices used for the selection of the best fitting pair of substitution matrices. All combination of hypothesis discussed in Results are reported. Best PCAP and corresponding CAP are indicated under each landscape.

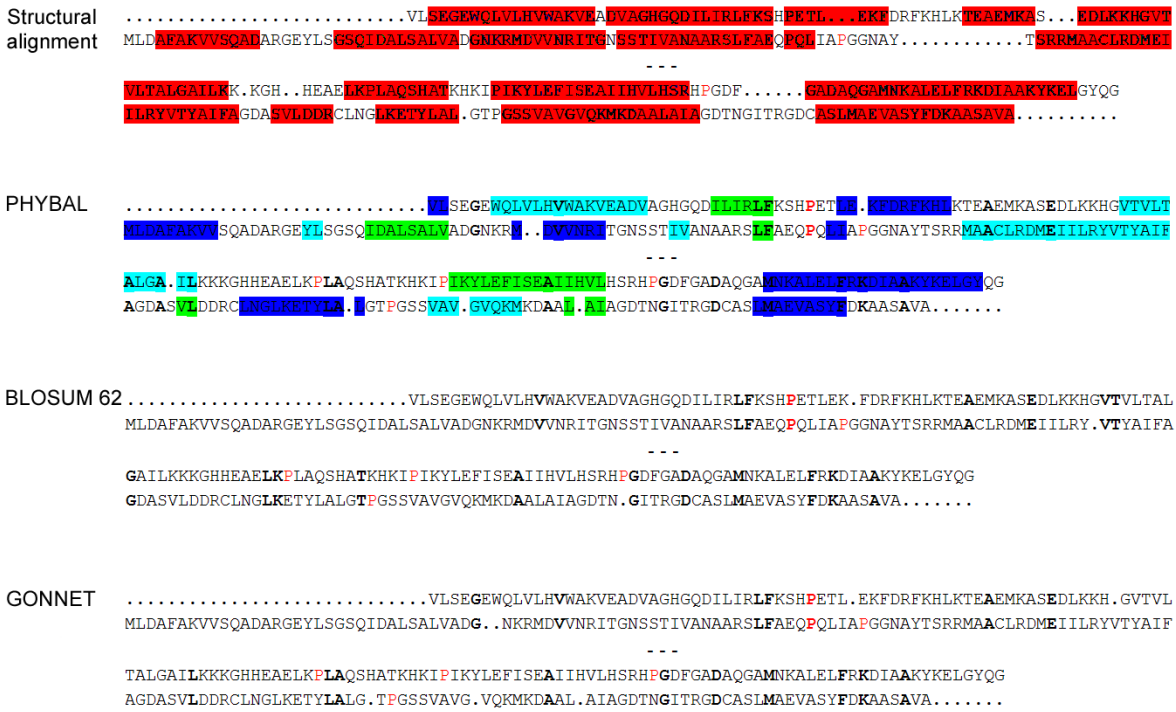

Figure 2: Sequence alignments realized on protein pair P5 with PHYBAL, PHYBAL-2D run with Blosum 62 and Gonnet matrices, and structural alignment. From Figure 5 in the main text, Blosum matrices perform the best with an alignment which is however an artifact of the algorithm, rather than an appropriate treatment of P5. In fact, large gap penalties induce no insertion of small gaps in the alignment but favor a large insertion in the N-ter region instead; since P5 structural alignment does not present any insertion in a large chunk of the N-ter of the protein, residue substitution values in Blosum help to properly detect the starting point of the chunk and reach a PCAP of 41.4%.

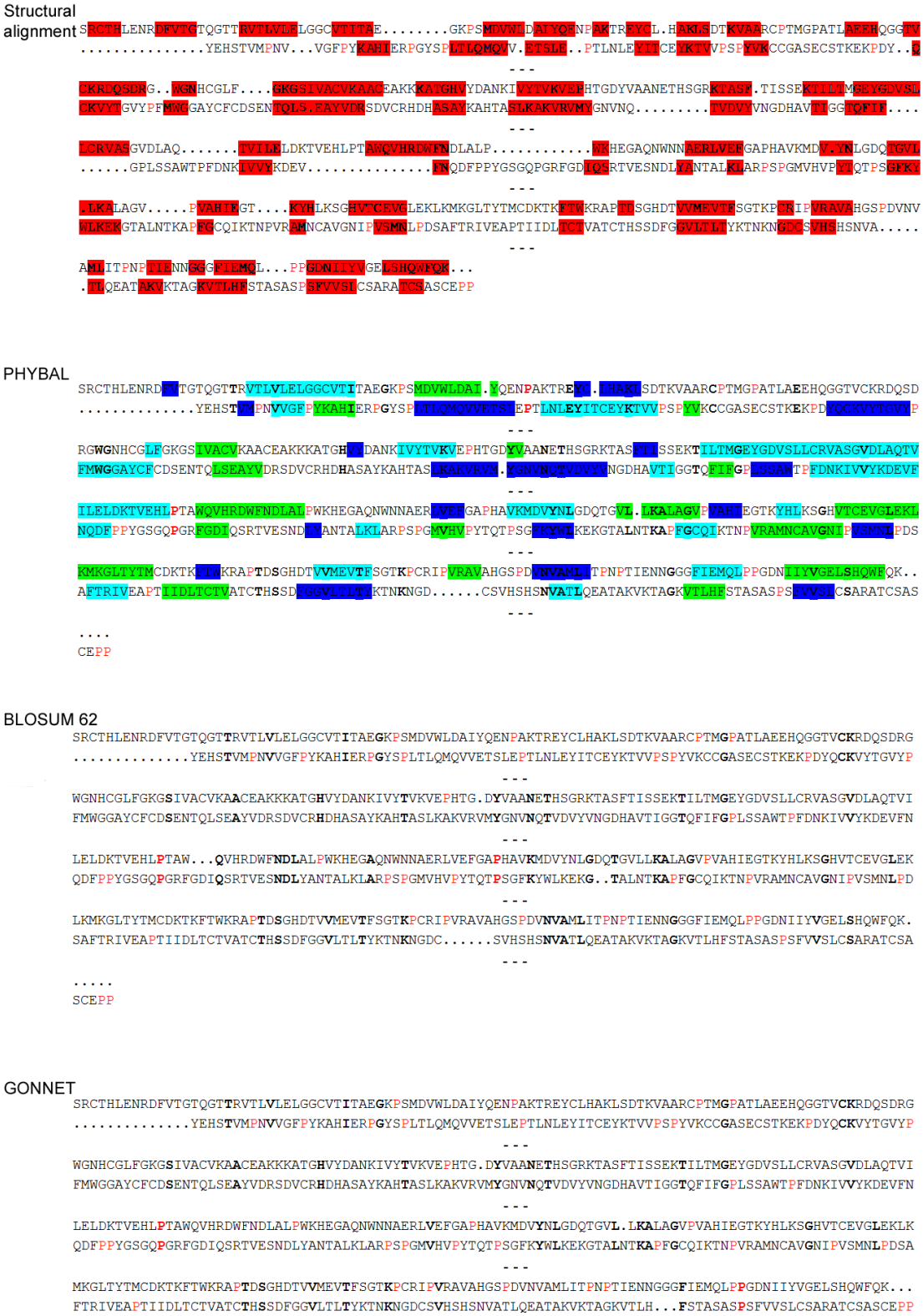

Figure 3: Sequence alignments made on protein pair P7 with PHYBAL, PHYBAL-2D run with Blosum 62 and Gonnet matrices, and structural alignment. Divergence of P7 is high both in sequence and structure as illustrated by SAD-coefficient, RMSD and % identity in Figure 4-C. Even automatic structural alignment is controversial: a long chunk with no insertions within the C-ter region is detected by MATRAS but not by SSM. These characteristics explain the low PCAP observed for all alignment methods applied.

Portability: the program runs on Linux environment with full ANSI conforming C compiler.

Usage: `phybal 'list of input file names' ... '-option' ['option arguments'] ...`

– *INPUT*: list of files containing sequences or parameters supported format XML, FASTA, MSF, FOLDFIT

-i-nogap	remove gaps in input sequences
-i-noblock	remove blocks in input sequences
-i-nocolor	clear amino acid colors in input sequences
-i-tree 'filename'	specify an input file name for guide tree (for future development)

– *OUTPUT*:

-o 'filename'	specify the output file name
-o-xml	output format = XML
-o-xml-3l	output format = XML with 3-letter amino acid code
-o-fam	output format = fasta
-o-msf	output format = MSF
-o-ffa	output format = FoldFit Alignment
-o-par	specify the output file name for parameters
-o-mat	specify the output file name for the distance matrix (for future development)

– *ALIGNMENT PARAMETERS*: substitution matrices, hydrophobic amino acids, number of hydrophobic neighbors defining a block and other parameters can be changed or specified in an xml file given as first parameter in the command line. The format of the xml file must be the same as for the xml file output given by the -o-par option.

-gappenalties 'gop' 'gep' 'bgop' 'bgep'	specify gap penalties for gop, gep, bgop and bgep (default: 10 1 15 2)
-th-nonoverlap	specify to use the IHBM matrix for pairs of residues both belonging to hb (default : overlap)
-globloc	specify to use the global-local alignment algorithm (default : global)

– *ANALYSIS*:

-noblock	do not analyse blocks
-nofillblocks	do not fill block (scattering method)
-noalign	do not align sequences
-noscoreconserved	do score conserved residues
-noblockcolor	do not color blocks
-nocolor	do not color amino acids
-noblockbreakercolor	do not color block breakers
-alignpairwise	align all sequences pairwise to the first sequence
-constructblock	construct blocks from predefined patterns in the input xml file

– *HELP*:

-h	display this help
–help	display this help

Supplementary Table 5. Available options for PHYBAL.