

2HNE A	217	IKL	KVC	anvqddi	RRCRLARAAI	g	P	D	IA	M	AV	DAN	QRW	253	
3RR1 A	144	FKL	NGC	eemgiidtsravdaav	ARVAEIRSAF	g	N	T	VE	F	GL	DFH	GRV	189	
2DW6 A	181	VKM	ATG	gapieedr	MRIEAVLEEI	g	K	D	AQ	L	AV	DAN	GRF	218	
3SSZ A	199	YKM	RFG	wgpkdgmmdgmrenl	KRVEAVREVI	g	Y	D	VD	L	ML	ECY	MGW	242	
2GSH A	189	GKM	pt	HWG	phdgdagirkda	AMVADMREKC	g	P	D	FW	L	ML	DCW	MSQ	232	
4E4U A	163	VKF	DPAG	pytaysghqlslevldrce	LFCCRVRREAV	g	S	K	AD	L	LF	GTH	GQM	212	
4E5T A	170	VKF	DPAG	aytiydghqpsledlerse	AFCKQIRAAV	g	T	K	AD	L	LF	GTH	GQF	219	
4E4F A	168	IRV	QCG	vpgmettygmakgkglayepatkgsleeqlwstekylldftp	KLFEAVRDKF	g	F	N	EH	L	LH	DMH	HRL	239	
4FI4 A	169	IRL	QSG	vpglpstygvsgdkmfyepadgnlptenvwstskylkhap	KLFEAAREAL	g	D	D	VH	L	LH	DVH	HRL	238	
4F4R A	169	VRV	QSG	vpgiettygvaktppgeryepadsslpaehvwstekyllnhap	KLFAAVREF	g	D	D	LH	V	LH	DVH	HRL	239	
4HCH A	188	IKL	HTW	mppvswapdvkmdl	KACAARREAV	g	P	D	IR	L	MI	DAF	HWY	231	
4GFI A	149	LKI	KTG	tadde	ARLRARAAA	g	P	E	AR	I	II	DAN	EGW	182	
4GIR A	173	IRC	QLG	fyggvpeniqtaqnptqgsyydqdyientv	EMFKNLREKY	g	K	Q	FH	I	LH	DVH	ERL	233	
4G8T A	222	FKL	KG	vldgfeea	EAVTALAKRF	g	P	D	AR	I	TL	DPN	GAW	258	
4GYP C	215	FKL	KG	vlpgeqei	DTVRALKRF	g	P	D	AR	I	TV	DPN	GAW	251	
3UCC A	191	LKG	VIK	kkyg	KDATNVGDEG	g	F	A	PNilense	ealelvkeaidkagytekivig	M	DV	AAS	EFY	251
1E9I A	192	LAK	VLK	ak	GMNTAVGDEG	g	Y	A	PNlgsnae	ealaviaeavkaagyelgkditla	M	DC	AAS	EFY	252
4EWJ A	189	LKK	ILK	kar	GLVTAVGDEG	g	F	A	PKfegted	dgvetiieaieaagyagengimig	F	DC	ASS	EFY	dke	253
1ELS A	193	LKS	LTK	kryg	ASAGNVGDEG	g	V	A	PNiqtae	ealdlivdaikaaghdgkvkig	L	DC	ASS	EFF	253
4G7F A	190	LKS	IIK	kkyg	QDAVNVGDEG	g	F	A	PPitdine	plpilmeaieqaghkgrfaic	M	DS	AAS	ETY	250
2PTW A	193	LRG	IIK	kkyg	QDAVNVGDEG	g	F	A	PPikdine	plpilmeaieeaghrkfaic	M	DC	AAS	ETY	253
3OTR A	199	LKN	VIK	knkyg	LDATNVGDEG	g	F	A	PNvatae	ealnllveaikaagyegkikia	F	DA	AAS	EFY	kq	261
3QTP A	195	LKV	VIK	kkyg	QDATNVGDEG	g	F	A	PNvsgar	ealdllveaikaagytkieia	M	DC	AAS	EFY	255
3UJ2 A	213	LAG	LLK	sk	GLATSVGDEG	g	F	A	PDlasde	eaieyileavklagyepgrdfvla	M	DA	ASS	E-W	272
1WUE A	180	VKL	KIR	pgydv	EPVALIRQHF	g	P	N	LP	L	MV	DAN	SAY	213	
2ZC8 A	160	IKL	KIK	pgwdy	EVLKAVREAF	g	P	E	AT	L	TA	DAN	SAY	193	
1WUF A	180	VKL	KIA	pnkdi	QFVEAVRKS	g	P	K	LS	L	MA	DAN	SAY	213	
1ROM A	167	IKL	KIK	pgwdv	QPVRATREAF	g	P	D	IR	L	TV	DAN	SAY	200	
2MUC A	166	FKL	KIG	anpveqdl	KHVVTIKREL	g	D	S	AS	V	RV	DVN	QYW	203	
1JPM A	159	LKI	KVG	kddiatdi	ARIQEIRKRV	g	S	A	VK	L	RL	DAN	QGW	196	
1MRA A	163	VKT	KIG	ypaldqdl	AVVRSIRQAV	g	D	D	FG	I	MV	DYN	QSL	200	
1EC7 A	204	FKL	KG	vlageeee	ESIVALAQRF	g	P	Q	AR	I	TL	DPN	GAW	240	
1BQG A	210	FKL	KG	vlrgedei	AAVTALSERF	g	P	D	AR	I	TL	DPN	GAW	246	
4HN8 A	223	FKL	KG	vrxgaeex	EAIRAIKARF	g	P	D	AR	V	TL	DPN	GAW	259	
1YEY A	220	IKL	KVC	anvqddi	RRCRLARAAI	g	P	D	IA	X	AV	DAN	QRW	256	
position																			

Fig. S1. Misaligned, correlated sequences within an alignment of enolases. The nine sequences between the two lines are misaligned; the insert residues shown in red structurally correspond to the first 10 columns shown. Note that these misaligned sequences share two distinguishing features: (i) they contain a 27-30 residue insertions that the other sequences lack and they conserve a glycine (G) residue in the seventh column instead of the consensus arginine (R) residue. GISMO relies on such features to identify and realign clusters of correlated sequences.