# Protein structure networks provide insight into active site flexibility in novel esterase/lipases from the carnivorous plant *Drosera capensis*

Vy T. Duong,<sup>†,‡</sup> Megha H. Unhelkar,<sup>†</sup> John E. Kelly,<sup>†</sup> Suhn H. Kim,<sup>†</sup> Carter T.

Butts, \*\*,  $\P$ , §,  $\parallel$  and Rachel W. Martin\*, †, ‡

Department of Chemistry, University of California, Irvine, Irvine, CA 92697, Department of Molecular Biology and Biochemistry, University of California, Irvine, Irvine CA 92697, Department of Sociology, UC Irvine, Irvine, CA, 92697 USA, Department of Electrical

Engineering and Computer Science, UC Irvine, Irvine, CA, 92697 USA, and Department of Statistics, UC Irvine, Irvine, CA, 92697 USA

<sup>\*</sup>To whom correspondence should be addressed

<sup>&</sup>lt;sup>†</sup>Department of Chemistry, University of California, Irvine, Irvine, CA 92697

 $<sup>^{\</sup>ddagger}\text{D}\text{e}\text{partment}$  of Molecular Biology and Biochemistry, University of California, Irvine, Irvine CA 92697

<sup>&</sup>lt;sup>¶</sup>Department of Sociology, UC Irvine, Irvine, CA, 92697 USA

<sup>&</sup>lt;sup>§</sup>Department of Electrical Engineering and Computer Science, UC Irvine, Irvine, CA, 92697 USA

Department of Statistics, UC Irvine, Irvine, CA, 92697 USA

## Sequence Alignments

Sequence alignments for the esterase/lipases from D. capensis are shown along with annotation reference sequences from other plants. Cluster 1 (Figure S1) contains enzymes with the traditional GDSL motif, including GDL1 CARPA from Carica papaya. Cluster 2 (Figure S2) contains only sequences from *D. capensis*, while Cluster 3 (Figure S3) contains two reference sequences from Arabidopsis thaliana. Cluster 4 is split into two figures for legibility (Figures S4 and S5). The alignment figures are annotated to highlight chemical properties of the amino acid residues as well as important sequence features. The amino acid attributes are color-coded as follows: cysteines are yellow, positively charged residues are blue, negatively charged residues are red, hydrophobic residues are green, and all others are black. Highly conserved residues are indicated with a dot above the sequence position. The catalytic triad residues are marked with colored arrows. Signal  $P 4.1^1$  is used to predict the signal peptide cleavage site, which is specified by underlining the residues on either end of the cleavage point. The signal peptide itself is highlighted in light orange. Strikethrough text indicates sequence regions that are absent in the active enzyme, in this case the N-terminal signal peptide that is expressed but removed during maturation. Functional blocks I-IV are highlighted with colored boxes. Annotations were performed by homology to the annotations reference sequences from C. papaya and A. thaliana found in the UniProt database and identified by their UniProt IDs.



Figure S 1: Sequence alignment for Cluster 1 esterase/lipases, annotated by homology to the reference sequence GDL1\_CARPA. The four functional blocks that are critical for enzyme function are highlighted using outlined colored boxes. The N-terminal signal peptide is highlighted in light orange. Colored arrows indicate the catalytic triad residues. Conserved residues are marked using colored dots: acidic (red), basic (blue), hydrophobic (green), and hydrophilic (black) residues.

DCAP_0448	MHWQV HPK	-
DCAP_8086		ь
DCAP_0434	MRGAAG ADRPPCQFPA IYNEGDSNSD TGGISAAFEP IPCAYGATEF GKAARRNSDG RLIDLIGIY TLSTLKGKS	Ν
DCAP_4098		ь
DCAP_5529	MESCRRIANL-VPLLLLIF-ITQCTAKRAA SPTEPCKFPA IFNFGDSNSD TGGFSAAFGQ AAPPAGETFF GRPAGRYSDG RLIDFIAQG LCVPYLSAY	L
DCAP_5165		ь
	BIOCK I BIOCK II	
DO3D 0440		
DCAP_0448		
DCAP_8086	QSIGDURHG ANIATLASTV LLPHTSLEVS GLSPFSLAIQ LNQMEFRAR VIESKSN-HG	
DCAP_0434	SSIGARESHG ANFARKOSTI REQUEITEL GICSFSCOMS ROLLTINGER GOLLSTRAK SP-SHETRIS EINPRENKSN RINSH A-LLIVMLQQ	
DCAP_4098	NSLGARFSHG ANFATGGSTI RROMETIFEY GISPFSLOVQ LWHDUGFRSR SSDLH-NQVT DP-THRS ILP	
DCAP_5529	DAVGSMESHG ANFATAASTI REQUNTINGG GISPISIAVQ WIQETERGAR TQIRKSQGIP HSKLHIPDMIPIQITDI ILGIWVETD KMILFSLKSQ	
DCAP_5165	DSVGTNFSHG ANFATAGSTI RPQNTTLRQS GFSPVSLLDQ FVEFSDFYHK APA	
DCAD 0449		
DCAP_0440		_
DCAP_8086	STELESS HIERATIFF IEGENDERSN -LARIGISG WUTLEVING FATINELIA LGARTEWIN AFRIGETAR IVE-LENDDS DVDUGGT	5
DCAP_0434	TRITATORY DUTATION	G
DCAP_4098	EDEFENSESAATTE DIGONDISVA -FRALSVEQU TAADMILGU FSTAVQHILE QGARLEWIHS TERGELEVA VMITTERAEG FLDQIGEL	G
DCAP_JJ29	GVIELLEVE BIFSBALTIE DIGUNDING FERNSIDEV KSEVPDENDE KIIKNIIG GGKEWENIN MEDNICELEV MEDSILLAAS QVDAGGAS	r D
DCAP_J10J	BIOCONDITAGE IN STREET DE CONDITAGE INSTREET DE CONTRELLA CONTRELL	P
		_
DCAP 0448	RNYNK MIKEALTETE ATLADASATY UDVHSVMLEL FERFTSHGLK HETKACCEEG GEAVNENDKU VCENT-KVIN GTETTAAACK DESNYVEND	G
DCAP 8086	YNAAVVNYNK MIKEALTETE AALADASLIY VDVHSVMLEL FEHPTEHGLK HGTKACCGEG GGAYNENPKV YCGNT-KVIS GIKTTASACK DESNYVSWD	G
DCAP 0434	ONDMATEINE OLKDAVVKLE TOLPEAATTY VOLYAAKYGL ISTTKSPGFV DPLKICCGYH -EKDGNKGLTYN GTAVFGSSCA KEEVHVSWD	G
DCAP 4098	ONDMATEINE OLKDAVIKLE TOLPEAATTY VOLYAAKYEL ISTTKSOGEV DELKICCGYH -EKDGNV WCGOKGLTYN GTAVFGSSCA KPERYYSWD	G
DCAP 5529	FNEVAOFFNO ALKEAVAOLK KELPLAALTY VDVYSLKYDL ISNAROHGFR EPLRACCGHG -GKYNFNSHV GCGGKIKVKG KELTVGKSCK DPSVALWAD	G
DCAP 5165	FNEVAKYFNS KLKEAVSOLR DELPLAALTY VDVNSVMYDL YEHATKYGFO HPLKACCGHG -GKFNYNOHH GCGSPLGOSCK EPSKALWWD	G
	BIOCK	v
DCAP 0448	THATEAANKI VALSIMNGSL LILLSRFTSS VAYITS	
DCAP 8086	THATEAANKN VALSIINGSL FDPPFPIQKL C-LLHHIG A SIGNAL GRAVAGE Site V active Ser residue	
DCAP_0434	VHYSEAANKW FADHILNRSL SRFLSPML	
DCAP_4098	VHYSEAANKW FTDHILNGSL SDPSIPITHA CYRN signal sequence 🕴 active Asp residue	
DCAP_5529	VHYTEAANKW VFDRIVDGSF SYPPVPLKMA CRRVE	
DCAP_5165	VHYTEAGNKW VFDKIVNGSY SDPPIPIHMA CHRNP • Conserved residue + active His residue	
	Block IV	

Figure S 2: Sequence alignment and annotation for Cluster 2. The four block regions are determined by sequence conservation and outlined with colored boxes. Three D. capensis esterase/lipases contain the N-terminal signal sequence (highlighted in light orange) and three lack it. The catalytic triad is indicated using colored arrows. Colored dots denote conserved residues.

GLIP6 ARATH					MSSSSSMD	-LLMCLLL	LI		SPVVL	AKSSSTVPAI
GDL77 ARATH					MTIS-	TVIAFMSMFL	VFV		-MSGP	IVVEGRAF
DCAP_1840									MMGEG	TDVNKRLGAS
DCAP_1460				<mark>MR</mark> -	-MARVDDHPHR-	RLFMVCATFV			-LSMKLFING	-VSGDPQVPCY
DCAP_1380	METVGPLGQD	YDWAEEDDDD	DMNGMEHACM	GYISLLHHAT	LIRRTALPFE	IQQSAAYF	HIVKLKTNTS	FVTERKSNGA	GVPAPHDVAD	AAGSLQVPSM
DCAP_0405					<mark>MQN</mark>	HSLKWMALYL	SIV	AASLCSCA	<u>A</u>	<u>-V</u> VTDPHFPAM
DCAP_4465								-MQTSFCSGA	A	VVTDPYFPAM
	• •••• •		• •	•		•			• •	•
GLIP6_ARATH	FTFGDSIFDA	GNNHYNKNCT	AQADFPPYGS	SFFHRPT	GRFTNGRTVA	DFISEFVGLP	LQ-KPFLELQ	IQILNGTSNF	SNGINFASAG	SGLLLDTNKF
GDL77_ARATH	FVFGDSLVDS	GNNNYL-VTT	ARADSPPYGI	DFPTRRPT	GRFSNGLNIP	DLISEAIGNE	EPPLPYLSPE	LRGRSL	LNGANFASAG	IGILNDTGFQ
DCAP_1840	FIFGDSLVDA	GNNNFL-PTL	SRANVTPNGI	DFKASGGTPT	GRYSNGRTMS	DIIGEELGQS	NYAVPFLAPN	STGKAI	LHGVNYASGG	GGIINSTGSI
DCAP_1460	FIFGDSLVDN	GNNNNI-ASL	ARSNYLPYGI	DFPQGPT	GRFSNGKTTV	DVITQLLGFD	DY-IP-PHAT	ASGEQI	LKGVNFASAA	AGIREETGQQ
DCAP_1380	FVFGDSIIDV	GNNNFL-NSV	AKSNFWPYGC	DFSRGPT	GRFSNGKTVV	DFIGELLGIS	NI-PAFADPA	TAGTKV	VTGINYASAA	AGILDETGRH
DCAP_0405	FVFGDSVIDS	GNNNYL-NSA	AKANYFPYGI	EFEQGPT	GRFCNGRTFM	DYLAEMLGLP	RI-PAFANPL	ETGHGI	LHGVNYASAS	AGILEETGIA
DCAP_4465	FVFGDSVIDS	GNNNYL-NSV	AKANYFPYGI	DFEQGPT	GRFCNGRTFI	DYLAEMLGLR	RI-PAFANPF	ETGHGI	LHGINYASAS	AGILEETGLA
	BIOCK I				BIOCK II					
CIIDS ADAMU	MCST		OOFOTIVEO	N	TTERSTOPS		TE NVELDED		DAVIDIANT DO	UNIVERTIDATIVE
GLIPO ARATH	MGV	TPIQTQL	QUFUTLVEQ-	SDI ICKD	OTODIVEO	LUITTUCCND	TE-NIFLPER	APTLSP	DAYVNAMLDQ	VNKTIDUNC
GDL//_ARAIN	FIN	THE CHEVOL	DIFQQIQQRV	SKLIGKP===	-QIQKLVSQA	TROITICGND	PUNNIFLEFT	SAR-SROFIL	PDIVRELISE	INNOLUDINO
DCAP_1840	FIASVVIHWQ	DIGENCOV	NNYKNEVCOT	UNIL COED	-TRNIIMHES	IF SITIGSND	FLNNILLPF5	SVGSRITUSP	TAP V DDM15Q	VCOOLOGI VN
DCAP_1400	LGG	RISENGQV	INFRATOR	VNILGDED	ENMOOILIAKC	MUENCECOND	VINNVILDEL	VEC CEEVED	UDEANLIINU	VADOTIALUS
DCAP_1380	WGD	RISLSQQV	ONFESTIGUE	RNTMATANPT DCDMER	FNMSQHLARS	MVFMSFGSND	VINNVILDE	VTC DUTCOT	VDFANLLINH	IARQILALHS
DCAP_0405	LGD	RESIREQV	ONFRIMINOI	DCOLUER	EMKDYLDNU	LAMMAN COND	VINNVIIDEID	VDC CNLVVD	ENENDITIE	VNDOTUTI UC
DCAF_4405	19D	Krölikigv	QNEBIILIQL	NGQ11B1	-BRINDI DE NV	Block III	I DININ I DIDE I D	TES-SMEINE	ENTADIDIDE	INNOTVILLID
						DIOCK III				
GLIP6 ARATH	LGARRIAFES	LGPVGCVPAR	AMLENAPTNK	CEGKMNVMAK	MYNKRLEDTV	NTTPTKYPGA	TAVEGAVYGT	THREOTYPAR	YGESDVSNAC	CGN-GTLGGL
GDL77 ARATH	LGVGRVLVTG	AGPLGCAPAE	LARSGTSNGR	CSAELORAAS	LYDPOLLOMT	NELNKKIGRN	VETAANTNOM	OEDFLSTPRR	YGEVTSKVAC	CGO-GPYNGM
DCAP 1840	LDARKESVGS	VGPTGCTPYO	KIINOLNADO	CADLENKMAL	AYNSKLKDLL	TOLNKNLPGA	TEVYANVYDI	VMELTSNYKA	YGEVTATOAC	CNGGGTFAGT
DCAP 1460	FGARKMALTG	TGOIGCSPNE	LANNSPDGKT	CDGKINSANO	MENGKLKSLV	DOLNSOFSDA	HETYTNAYDT	FODMLDNPAA	YGERVTNAGC	CGV-GRNNGO
DCAP 1380	IGLRKFFLAG	IAPLGCIPNO	RASGLAPAGR	CVDSVNOMLG	SFNEGLRSLV	SLLNSKYPGS	IFVYGNTYGA	VGDILNNPTR	NGFSVVDRGC	CGL-GRNOGO
DCAP 0405				ASI	TSRPWLEKLA	DOLNSNHPDA	KFVYGDTYAA	SMDMIANGST	YGFNVADEGC	CGI-GRNRGO
DCAP 4465	LGLRKFFMAG	IGPLGCIPNQ	RASGAGPPGK	CVSAVNDMVL	MFNLRLDKLA	DOLNSNHPDA	KFVYGDTYAA	SMDMIANGST	YGFKVVDEGC	CGI-GRNRGQ
-										
	•		↓ ↓							
GLIP6_ARATH	MQCGREGYKI	CNNPNEFLFW	DFYHPTEHTY	RLMSKALWNG	N-KNHIRPFN	LMALATNKIT	F			
GDL77_ARATH	$\operatorname{GL}{\mathbf{C}}\operatorname{TVL}{-}\operatorname{SNL}$	CPNRELYVFW	DAFHPTEKAN	RMIVRHILTG	T-TKYMNPMN	LSSALAL	<ul> <li><u>XX</u> signa</li> </ul>	I cleavage sit	e 🖌 active	Ser residue
DCAP_1840	VPCGPT-SSL	CQDRHKHVFW	DPYHPSEAAN	LILAKQLMDG	D-TKYVSPIN	LRQLRDL	-			
DCAP_1460	ITCLPM-QQP	CPNRNEYLFW	DAYHPTEAAN	IVVGTRSYRA	QSASDAYPYD	IQHLAQL	- signa	l sequence	🖌 active	Asp residue
DCAP_1380	ITCLPY-AVP	CANRTQYVFW	DAFHVTEAVN	SLLARRAFFG	P-PTDCYPIN	IQQLALL	-			
DCAP_0405	VSCLPL-LPP	CANRDEYVFW	DAFHPTQAAN	KTLAAEAYKI	I-LSKF		- • cons	erved residue	e 🚽 active	His residue
DCAP_4465	VSCLPL-LPP	CANRDEYVFW	DAFHPTQAAN	KILAAEAYKI	I-LSKF		-			
		В	lock IV							

Figure S 3: Sequence alignment and annotation for Cluster 3. Reference sequences are GLIP6\_ARATH and GDL77\_ARATH. All but three Cluster 3 esterase/lipases contain a N-terminal signal peptide (highlighted in light orange). Functional block regions are outlined using colored boxes. Colored dots indicate conserved residues.

							1.1			
						•				
DCAP_6218				MAVLY	NSRGSNSFVI	PENETVPALI	VFGDSIVDPG	NTDRLDTICR	ADHSPYGRDF	EGGKATGRFS
DCAP_6260	MIYSGKGRMH	QKMPSPT	KIS	AISIIIHVLV	LHRGTHSQSL	PQNVTVPALF	FFGDSVVDPG	NSDWILSVCR	ADHPPYGRDF	EGGVATGRES
EXLS_ARATH		DI VICINAL OK	COCONOSMA-	-IC-LLSVLF-	LTETITAVKL	PPRLIIPAVI	AFGDSIVDTG	MNNNVKTVVK	CDFLFIGINF	QSGVATGRFC
DCAP_1/61	MQ	KLISSMN-QK	SLAAKLVVFQ	-FIATLHM	LELLOPAULI	DNCEEUDAVI	VEGDSIVDIG	NNNILQTIGK	CNEDDVCKDE	PTGLPTGRIS
DCAP_0217		-MMERININ-KIN	MELOTIT	LEVOLUEDI	LELISEAVIL	DDNUTUDATE	AFGDSIVDPG	MINDYDULCK	UNIT PPIGRDF	FGGRAIGRES
DCAP_J401 DCAP_0159		MUVECEEN	T TEVENCIET	-bragentore	- FUNAECSKI	PERVIVEALE	VECDSTUDSC	MININUTTTTCP	SNEVEVCEDE	FECKPECPET
DCAP 2088						PNNATVPCIE	VEGDSVVDTG	NUNYNSTLCK	SNEPPYCEDE	PCCOATCRES
DCAP 2089					MMT.	PTNVTVPGIF	VEGDSTVDSG	NNNYISTICK	SNEPPYGRDE	PGGOATGRES
2000						11111110	Block I		0111110101	Block II
	• •	• •	•	•• •	•					•
DCAP_6218	NAKIPTDYLA	ESLGIKDLLP	AYLDPTIETE	DLLTGVSFAF	ACCGYDPLTP	QFFVSIHFHF	DFYSSSTCIA	LVVVHGWILS	VTERVPSLQD	QLEYFKEYKK
DCAP_6260	NAKIPSDYTA	QEFGVKELLP	AYLDISLRTE	DLLTGVSFAF	TCSGYDPMTS	L-IF			QVPTLTD	QLQHFREWQL
EXL3_ARATH	DGRVPADLLA	EELGIKSIVP	AYLDPNLKSK	DLLTGVSFAS	GGSGYDPITP	K-LV			AVISLED	QLSYFEEYIE
DCAP_1761	NGKVPADLIV	EKLGIKEYLP	PYLNQSLEFQ	DLVTGVNFAT	GATGYDPVSA	Q-LA			TVKSLDD	QLELFKDYKT
DCAP_6217	NGKIPTDLFA	ELLGIKELLP	AYLDPTLTTQ	DLLTGVSFGS	GVAGYDPVSA	A-LL			ATLSLDA	QLNLFKEYQS
DCAP_5461	NGYIPTDILA	QELGIKEMVP	SYLDEGLSPN	DMLTGVSFAC	GCSGWDPSTS	RARY			TARTLAE	QYGYF <b>EE</b> YIV
DCAP_0158	NGKVPSDLFA	EAFGVKELVP	PYLDPSLTMD	DLLTGVNFAS	ACSGYLPATA	LHLS			PSLSLED	QLDLFKEYIS
DCAP_2088	DGIVPSDIVA	QAFGVKKFVP	AYLDPSLSTD	EMLTGVNFAS	ACSGYLPLTA	TYKC			ISLSLEN	QLDLFKQYIV
DCAP_2089	DGIVPSDIVG	NSERKLS-	NE		SSNHLV	LFQY			FSLSLEN	QLDLFKQYVV
DO3 D (010	•						•		• •	•
DCAP_6218	KLIAAVGEER	TSFIVSKSIY	VVVAGSNDET	FTINLFKL	REMNMSSIT	DIMISQASII	LQALIDLG	ARRIGVLQLL	DOCOLDANDE	
DCAP_6260	KVKTAVGGKK	ASIVISESVI	LIVIGNNDET	FNIIGS-LFR	SLQINVSSIC	DILLITFASTF	LQELINLG	ARRIGATGLP	PUGCLPAMRT	SAEGHSRPCN
EALS_ARATH	RVENIVGEAR	ACKLICKCAY	LIFACENDIA	NTITI-R-A	RPEIDVDSIT	TLMSDSASEF	VTRLIGIG	VRRVAVEGAP	PIGCVPSQRT	EACCLOBECH
DCAP_1701	KHIAAIGNSK	MARLISKONY	LUCACENDUT	DEVENE DED	KPQ1DVD511	DILVSPASNE	U VELVENC	ADVICUESUD	RECEIPSON	LECCLEDECV
DCAP_0217	NMEL EVODEN	VESTISONEY	LVCTCVNDEL	MNEDDDCCEP	PPNUDI A SVT	DITTERACING	IKAIVSIC	ARRIGUESVE	DVCUTDERE	TACCUEDADU
DCAR_3401	KITARVCEOK	TRATICONT	LIATCONDEN	VEVDTO		DILUNYTOCE	ITPTWI	ARRIGITGIT	FIGHTFFSRT	INGGVENNEV
DCAP_0130	KDIXEVGEQK	1331130011	HIAT GONDEN	1F1D1Q	-MNGN11AFB	DINVINITIOUP	NNKOPI VDTC		DOCCUPANE	NYCCLIPETV
DCAP 2089	KUKAAVGEER	TTRISOSTE	TICTOSNDEL	VVVFTO	-KSGNMSAVT	DSWWWASCE	LEVERLYDIG	APPIVLCCAP	POGCLPAART	NYCCLLEFIV
	It vitility Oblity	111111000	Block III	111010	ROOMIDITI	Dovininooi	DIVINIDIDIO	marci v 166mi	1000011111111	NIGOLINIIV
		•								1 1
DCAP 6218		TLY	LRGKNPDSRF	VYLDLYNPVL	RLVONPTOSG	FQISNVG	CCATGTSETS	IFCNSLFD	LFSTCKNESE	YLFWDAYHPT
DCAP 6260	KOYNONAMLF	NSKLESLMGS	LGKNLTGAKL	VYLDLYGPLL	QLVKNPENFV	LPESPVVTKA	RKGDTNREMV	-FCALRLISR	ETLRCTECSS	YKASLSISKM
EXL3 ARATH	DNYNEAAKLF	NSKLSPKLDS	LRKTLPGIKP	IYINIYDPLF	DIIQNPANYG	FEVSNKG	CCGTGAIEVA	VLCNKIT	S-SVCPDVST	HVFWDSYHPT
DCAP 1761	DEYNKAALLF	NSKLNTEIES	LNRNLSGVAM	FFLDVYAPLL	DLINNPSQAG	FEVVDKG	CCGTGNIEVS	FLCNRLE	NLLTCKDATK	YIFWDSFHPT
DCAP 6217	PQFDQLALLF	NSKLQETVVD	LNKNLTGAKL	AYIDLYQPLA	HLINNGS <mark>E</mark> YG	FQVVNRG	CCGTGLFEAS	ILCNPFD	ITCKNDSQ	YIFWDAFHPT
DCAP 5461	TEFNNAATLF	NFKLQTLIDS	LNRNFPGAKF	GYL <mark>DIYSKLM</mark>	YVIENAADFG	CKVNDRG	CCGTGLVEMG	VACNGLV	DIFSCKNNSE	YVFWDAAHPT
DCAP 0158					VCIFFLVG	FEVVNRG	CCGTGLFEEG	FLCNVFS	IPFSCKHTSK	YVFWDGSHPT
DCAP 2088	ESFNQDSLDF	NLKLQAMLKS	LKNTLQGSRF	AYFDLYYTVL	DLIQKPHEYG	FEVVGKG	CCGTGFFEEG	PLCNIFS	SLISCPNASK	YVFWDASHPT
DCAP 2089	ESFNQDTLNF	NLKLQTMLKS	LQNTLQGSRF	IYFDFYYTVL	DLVQKPHDYG	FEVVDKG	CCGTGLFEEG	PLCNIVS	TLISCPNASK	YVFWDASHPT
-										Block IV
DCAP_6218	DRANKIIIEE	LFGKKMVSMA	CINFGCPEGL	SVELAGRIGN	RTLASITRAF	VQSQRGDIV				
DCAP_6260	TMIY						vv cianal	ologyago cita	1	0
EXL3_ARATH	EKTYKVLVSL	LINKFVNQFV					AA signal	cieavage sile	🕴 🛉 active	Ser residue
DCAP_1761	EATYRIIVDK	AVKDNINFFY							1	
DCAP_6217	QRTYQILVNG	LVNTTINDLY	N				signal	sequence	active.	Asp residue
DCAP_5461	EQINQYLVTT	LTAENLHKFF						nued readdure	1	Liter an etal.
DCAP_0158	QAAYKHVLDR	ILNVT'MPHFF					<ul> <li>consei</li> </ul>	iveu residue	active	HIS residue
DCAP_2088	QAAYNIVLAK	NFNQTMSKFF								
DCAP_2089	QAAYNIIVAR	NINQTMSQFF								

Figure S 4: Sequence alignment and annotation of Cluster 4a (first set), annotated by homology to EXL3\_ARATH. Cluster 4 is separated into two parts (4a and 4b) for clarity. Block regions I-IV are shown in colored boxes with active site residues marked by colored arrows. Colored dots indicate conserved residues. When present, the N-terminal signal peptide is highlighted in light orange.

										•
DCAP 5138	MEQ-KQDKSV	/ MIKMEMKFEN	VLSSFVI-IL	VLVLAAENIN	AQPLVPALII	FGDSSVDTGN	NDYIHTLFKA	NYPPYGRDFI	DKKATGRFCN	GKLATDITAD
APG2_ARATH		MDR	CTSSFLLLTL	VSTLSILQIS	-FAQLVPAIMT	FGDSVVDVGN	NNYLPTLFRA	DYPPYGRDFA	NHKATGRFCN	GKLATDITAE
DCAP 1365		MAHVSTAF	FAATYKILLL	SLLLLVFLVT	-CEAKVPAIIV	FGDSSVDPGN	NNQVPTMARS	NFAPYGRDLP	GGQPTGRFCN	GKLVPDFISE
DCAP_5587		MVI	FHHGFIILIF	LAQLITLQHV	LGSKVPAIIV	FGDSSVDTGN	NNVIATVLKS	NFRPYGRDFD	GGRPTGRFSN	GRVPADFISE
DCAP 2187			MARNPA	LWNTSKVTSS	SPSQVPAVFV	FGDSTVDPGN	NNYIGTIFTS	NYAPYGRDLP	NHIPTGRFSN	GRLATDFIAS
DCAP_4076	MEQFSTNSTI	L VIVFLIVSLI	ICIPIQTRGG	IGFGRHVKNG	SDPMVPAILV	FGDSTMDPGN	NDYIATTFRS	NFAPYGRDFA	NHEATGRFTN	GRLVTDFVAS
						Block I			Block II	
	•	•• •	• •••		•	•• •	•		• ••	••
DCAP_5138	TLGFTSYPPS	S YLSPQAAGKN	LAIGANFASA	GSGYDDLTAY	LSHAIPLSQQ	LEYYKEYQGK	LSVLVGSSNA	NSTLTGALYV	IGAGNSDFVQ	NYFLNPVL
APG2_ARATH	TLGFTKYPP/	A YLSPEASGKN	LLIGANFASA	ASGYDDKAAL	LNHAIPLYQQ	VEYFKEYKSK	LIKIAGSKKA	DSIIKGAICL	LSAGSSDFVQ	NYYVNPLL
DCAP_1365	AIGLKQLVPA	A YLDPGYTIAD	FATGVSFASA	GTGYDTTTSN	VLSVIPLWRE	VEYYKEFQQR	LRDYLGEEKA	NVIISEALYM	TSIGTNDFLE	NYYTLPQR
DCAP_5587	FFRLKKTVP	A YLDPMYDISN	FSTGVCFASA	GTGYDNSTAA	VLSVIPLWKE	VEYYKEYQAR	LREYLGEREA	NHILSEAVYL	ISIGTNDFLE	NYYLPT-GAR
DCAP_2187	DLGVKEYVPI	P YLDPKLMANE	LITGVSFASA	GSGYDPYTAQ	LGGVITMQKQ	LEYFREYKAK	IEKLVGKEKS	RHIIENAAYI	VSAGANDFAF	NYYLPPLMSR
DCAP_4076	YLGIKKYVP	A YLDPSLSDQE	LLTGVSFASG	GNGYDPLTPQ	LSGVISMQRQ	LEYFKEYKSR	IEKLVGEEKA	DHIVGSAGYV	ISAGTNDFVV	NYYSTALPIR
								В	lock III	
DCAD 5139						CDUQNNC		VNV VIN		DDI VI WVENT
APC2 ARATH	VKUVTUDAV	SELIDNESTE	TROUVANGAR	KIGVTSLPPT	GCLPAARTLE	GEHEKCC	VERLINTDAON	FNKKLN	PINUGROKOA	SDLKIVVEDI
DCAP 1365	PTOFTVEOV	DVIVETAEDE	LEOLYGLGAR	KESEGGLEPEM	GCLPLERATN	VOGOOGC	NEEVNSVAWD	FNAKLV	EMUDRINUOI.	PGLKMALGNP
DCAP 5587	SLOFTIEOY	) NELVGIARNE	TVOTYNLGAR	KMSLGGLPPM	GCLPLERTTN	TIGGNAC	TSKYNNVAKO	FNSKLE	VLVNOLNGDL	PDAKVLESNP
DCAP 2187	R-NLSVEOY	) PELLOIAODE	VOELWKEGAR	KIGVVGLAPL	GCLPMVITIN	TENTTOPROC	LKSPGLVVSS	YNOOLOMKLN	DIONOLAVSD	ORTKLEYLDI
DCAP 4076	RNTYSVGEY	) KELLONVOEF	TOGLWKEGAR	KLAVVGLPPM	GCLPEVITTN	SESTFOREC	TESUSSISE	YNOOFOSOLD	VLOROFAAEV	OTTKLAYIDI
		E	- <u>c</u>						- <u></u>	2
				-	•	_↓ ↓_				
DCAP_5138	YDAVYDLIQ	R PQDFGFAESR	KGCCGTGVIE	TTIFLCNPLS	IGTCRNATEY	VFWDAVHPSE	AANELLASSL	LIQGI <mark>D</mark> LIS-	-	
APG2_ARATH	YSPLY <mark>D</mark> LVQN	N PSKSGFTEAT	KGCCGTGTVE	TTSLLCNPKS	FGTCSNATQY	VFWDSVHPSE	AANEILATAL	IGQGFSLLG-	-	
DCAP_1365	FYILQELVLY	PQDHGFEVTN	RGCCGTGQYE	MG-YVCNQ-S	PVTCPDASKY	VFWDSFHPTE	RTNQLVADHL	VKNYLVELLH	-	
DCAP_5587	YGILLRMIRE	R PSLYGFEETS	RGCCATGRFE	MS-YLCNEFS	PFTCTDATKF	IFWDSFHPTE	KANFIIAKHV	FERSLGPKFL	-	
DCAP_2187	NTPTLSYIQ	D PTRFGFEEVA	RGCCGTGYLE	LS-FLCNPTT	I-SCPDPSKY	VFWDSIHPSQ	RTCRLVVDTF	RPVLDEMKAS	Т	
DCAP_4076	YTPIANMVQ	E PSKYGFEEVN	RGCCGTGYVE	II-FLCNPIS	N-TCTDDSKY	VFF <b>D</b> AIH <b>PTE</b>	KAYNIIFQYI	RPVIDSLKLI	G	
	XX	signal cleavad	ne site	active Corr	opiduo	Block IV				
		olgital oleavag		active Ser residue						
		signal sequence		active Asp residue C-terminal domain of DCAP 4076 not sl			not shown			
	•	conserved residue		active His re	esidue					

Figure S 5: Sequence alignment and annotation of Cluster 4b (second set), annotated by homology to APG2\_ARATH. Cluster 4 is separated into two parts (4a and 4b) for clarity. Block regions I-IV are shown in colored boxes with active site residues marked by colored arrows. Colored dots indicate conserved residues. When present, the N-terminal signal peptide is highlighted in light orange. DCAP\_4076 has an additional C-terminal domain (shown in Figure S8).

### Preliminary Structural Models and In silico Maturation

Preliminary models for the esterase/lipases were produced using the online Robetta implementation<sup>2</sup> of Rosetta<sup>3</sup>. The Rosetta structures contain the full sequences, including the N-terminal signal peptides that are cleaved during maturation. We performed *in silico* maturation, which we have previously described for cysteine proteases,<sup>4</sup> for each protein. The initial Rosetta structure for each enzyme includes the signal peptide and lacks posttranslational modifications. During *in silico* maturation, the signal sequence is removed and the structure is equilibrated for 500 ps in explicit TIP3P solvent using NAMD.<sup>5</sup> Figures of predicted structures were generated using Chimera.<sup>6</sup> Figure S6A shows the workflow of the overall enzyme discovery process. Panels (B) and (C) show an example of a Cluster 2 esterase/lipase, DCAP\_8086, before (B) and after (C) the *in silico* maturation process. Further comparison of a Cluster 3 esterase/lipase (DCAP\_1460) to Cluster 4 enzymes and a cutin synthase from *Solanum lycopersicum* (tomato), G1DEX3\_SOLLC, is shown in Figure S7. Functional sequence blocks DCAP\_1460 and G1DEX3\_SOLLC are highlighted by color (Figure S7). DCAP\_4076, has an additional C-terminal domain. A PSI-BLAST search for the sequence of this domain indicated that it is related to the negative regulator of systemic acquired resistance proteins previously discovered in other plants,<sup>7</sup> with approximately 36% sequence identity to the SNI1 proteins from *Arabidopsis thaliana* (Uniprot ID: SNI1\_ARATH) and *Glycine max* (Uniprot ID: Q0ZFU8\_SOYBN). The *Arabidopsis* protein negatively regulates DNA recombination and gene expression during short-term stress responses. It has been suggested that SNI1\_ARATH provides a scaffold for other proteins involved in regulation of transcription to bind;<sup>8</sup> it is possible that this domain is playing a similar role here. DCAP\_4076 lacks the N-terminal secretion signal common to many of the esterase/lipases, suggesting an intracellular function (Figure S8).

The template structures used by Rosetta to calculate the predicted structures for a representative esterase / lipase, DCAP\_0434, are tabulated in Supplementary Table S1.



Figure S 6: (A) Flow chart illustrating the overall strategy for identifying enzymatic targets from genomic DNA. The workflow is indicated with solid arrows, while dotted arrows represent steps where information from a later stage of the pipeline enables refinement of earlier stages in an iterative manner. After genome sequencing, assembly, and gene discovery, target proteins are identified based on putative enzymatic activity. Functional sequence features are identified by analogy to annotation reference sequences found in the UniProt database. Structures are predicted using the Rosetta software, and equilibrated in explicit solvent after removal of sequence regions not present in the mature enzyme. Structures are compared using network analytic methods, enabling strategic selection of enzymes for experimental characterization in a future study. (B) DCAP\_8086 before and (C) after *in silico* maturation. The light orange helix in part A is the N-terminal signal sequence, which is cleaved upon maturation. Important residues are color-coded as follows: dark cyan (catalytically active serine), red (active site aspartic acid), purple (active site histidine).



Figure S 7: Comparison of DCAP\_1460 (Cluster 3) to *D. capensis* esterase/lipases from each of the other clusters. These pairwise alignments of structural models provide an indication of the type and magnitude of structural differences between clusters: in general, the overall fold and secondary structural elements is conserved, although considerable variation can be observed in their relative positions and the conformations of loops and termini. Alignment was performed using the matchmaker feature of Chimera with default settings.<sup>6</sup> Functional block regions I-IV are colored accordingly while the catalytic triad (Ser-His-Asp) residues are colored dark cyan, red, and purple. Active site residues are located in block I and IV, binding residues in block II-III. A. Comparison of DCAP\_1460 to esterase/lipase DCAP\_6260 (Cluster 4a). B. Comparison of DCAP\_1460 to DCAP\_5587 (Cluster 4b). C. Comparison of DCAP\_1460 to DCAP\_2088[@Cluster 4a). D. Comparison of DCAP\_1460 to model esterase/lipase, G1DEX3\_SOLLC, from *Solanum lycopersicum* (tomato).



Figure S 8: A. Sequence alignment of the C-terminal domain of DCAP\_4076 with the SNI1 proteins from *Arabidopsis thaliana* (Uniprot ID: SNI1\_ARATH) and *Glycine max* (Uniprot ID: Q0ZFU8\_SOYBN). B. Ribbon structure of DCAP\_4076, with the catalytic domain in light blue and the C-terminal domain in dark blue. C. Structural model of DCAP\_4076 showing the surface representation. The active site D (red) and H (magenta) residues are visible at the top of the model.

All initial and equilibrated structures available for download as PDB files are tabulated in Supplementary Tables 1 and 2, respectively.

Protein	Organism	Sequence Elements included	File Name
GDL1 CARPA	Carica papaua	signal, active region	GDL1 CARPA m1.pdb
DCAP 3343	D. capensis	signal, active region	DCAP $3343 \text{ m1.pdb}$
DCAP 6947	D. capensis	signal, active region	DCAP 6947 m1.pdb
DCAP 0448	D. capensis	signal, active region	DCAP 0448 m1.pdb
DCAP 8086	D. capensis	signal, active region	DCAP 8086 m1.pdb
DCAP 0434	D. capensis	active region	DCAP 0434 m1.pdb
DCAP 4098	D. capensis	active region	DCAP 4098 m1.pdb
DCAP 5529	D. capensis	signal, active region	DCAP 5529 m1.pdb
DCAP 5165	D. capensis	active region	DCAP 5165 m1.pdb
GLIP6 ARATH	A. thaliana	signal, active region	GLIP6 ARATH m1.pdb
GDL77 ARATH	A. thaliana	signal, active region	GDL77 ARATH m1.pdb
$DCA\overline{P}$ 1840	D. capensis	active region	DCAP 1840 $m1.pdb$
DCAP 1460	D. capensis	signal, active region	DCAP 1460 m1.pdb
DCAP 1380	D. capensis	active region	DCAP 1380 m1.pdb
DCAP 0405	D. capensis	signal, active region	DCAP 0405 m1.pdb
DCAP 4465	D. capensis	active region	DCAP 4465 m1.pdb
DCAP 6218	D. capensis	active region	DCAP 6218 m1.pdb
DCAP 6260	D. capensis	active region	DCAP 6260 m1.pdb
EXL3 ARATH	A. thaliana	signal, active region	EXL3 ARATH m1.pdb
DCAP 1761	D. capensis	active region	DCAP 1761 $m1.pdb$
DCAP 6217	D. capensis	signal, active region	DCAP 6217 m1.pdb
DCAP 5461	D. capensis	signal, active region	DCAP 5461 m1.pdb
DCAP 0158	D. capensis	signal, active region	DCAP 0158 m1.pdb
DCAP 2088	D. capensis	active region	DCAP 2088 m1.pdb
DCAP 2089	D. capensis	active region	DCAP 2089 m1.pdb
$DCAP^{-}5138$	D. capensis	active region	DCAP 5138 m1.pdb
APG2 ARATH	A. thaliana	signal, active region	APG2 ARATH m1.pdb
DCAP 1365	D. capensis	signal, active region	DCAP 1365 m1.pdb
$DCAP_{5587}$	D. capensis	signal, active region	DCAP_5587_m1.pdb
DCAP 2187	D. capensis	active region	DCAP 2187 m1.pdb
DCAP 4076	D. capensis	active region	DCAP 4076 m1.pdb

Supplementary Table 1: Rosetta structures for esterase / lipases (PDB files available for download)

Supplementary	Table 2:	Mature	$\operatorname{structures}$	for esterase	/ lipases	(PDB	files	available	e for
			dowr	nload)					

Protein	Organism	Sequence Elements included	File Name
GDL1_CARPA	Carica papaya	active region	GDL1_CARPA_mature_m1.pdb
$DCAP_{3343}$	D. capensis	active region	$DCAP_{3343}_{mature_m1.pdb}$
$DCAP_{6947}$	D. capensis	active region	$DCAP_{6947}mature_{m1.pdb}$
$DCAP_{0448}$	D. capensis	active region	$DCAP_0448_mature_m1.pdb$
$DCAP_{8086}$	D. capensis	active region	$DCAP_{8086}mature_{m1.pdb}$
$DCAP_{0434}$	D. capensis	active region	$DCAP_0434_mature_m1.pdb$
$DCAP_{4098}$	D. capensis	active region	$DCAP_{4098} mature_{m1.pdb}$
$DCAP\_5529$	D. capensis	active region	$DCAP_{5529}mature_m1.pdb$
$DCAP\_5165$	D. capensis	active region	$DCAP_{5165}mature_m1.pdb$
GLIP6_ARATH	A. thaliana	active region	$GLIP6\_ARATH\_mature\_m1.pdb$
GDL77_ARATH	A. thaliana	active region	GDL77_ARATH_mature_m1.pdb
$DCAP_{1840}$	D. capensis	active region	$DCAP_{1840} mature_{m1.pdb}$
DCAP_1460	D. capensis	active region	$DCAP_{1460} mature_{m1.pdb}$
DCAP_1380	D. capensis	active region	$DCAP_{1380} mature_{m1.pdb}$
$DCAP_{0405}$	D. capensis	active region	$DCAP_0405_mature_m1.pdb$
$DCAP_{4465}$	D. capensis	active region	$DCAP_{4465}mature_{m1.pdb}$
$DCAP_{6218}$	D. capensis	active region	$DCAP_6218_mature_m1.pdb$
$DCAP_{6260}$	D. capensis	active region	$DCAP_{6260}mature_{m1.pdb}$
EXL3_ARATH	A. thaliana	active region	EXL3_ARATHmaturem1.pdb
DCAP_1761	D. capensis	active region	DCAP_1761_mature_m1.pdb
$DCAP_{6217}$	D. capensis	active region	$DCAP_{6217} mature_{m1.pdb}$
$DCAP_5461$	D. capensis	active region	$DCAP_5461_mature_m1.pdb$
$DCAP_{0158}$	D. capensis	active region	$DCAP_{0158}mature_{m1.pdb}$
$DCAP_{2088}$	D. capensis	active region	$DCAP_{2088} mature_{m1.pdb}$
$DCAP_{2089}$	D. capensis	active region	$DCAP_{2089}mature_{m1.pdb}$
$DCAP\_5138$	D. capensis	active region	$DCAP_{5138}mature_{m1.pdb}$
APG2_ARATH	A. thaliana	active region	APG2_ARATHmaturem1.pdb
$DCAP_{1365}$	D. capensis	active region	$DCAP_{1365}mature_{m1.pdb}$
$DCAP\_5587$	D. capensis	active region	$DCAP_{5587}mature_m1.pdb$
$DCAP_{2187}$	D. capensis	active region	$DCAP_{2187}mature_{m1.pdb}$
DCAP 4076	D. capensis	active region	$DCAP_{4076}mature_{m1.pdb}$

Supplementary Table 3: Templates used for structure prediction of DCAP\_0434, designated by PDBID.

1 1	J 1 1	_ , 0 ,	
PDBID	protein	organism	citation
3KVN (A)	EstA	Pseudomonas aeruginosa	9
3KVN (X)	$\operatorname{EstA}$	Pseudomonas aeruginosa	9
1 ESC(A)	Streptomyces scabies esterase	$Streptomyces\ scabiei$	10
3RJT(A)	lipolytic protein	Alicyclobacillus acidocaldarius	11
3MIL(A)	isoamyl acetate- hydrolyzing esterase	Saccharomyces cerevisiae	12
4JJ6 (A)	Axe2 variant H194A	Geobacillus stearothermophilus	13
40AP (A)	Axe2 variant W190I	$Geobacillus\ stearothermophilus$	14
3W7V (A)	Axe2	$Geobacillus\ stearothermophilus$	15
4JKO (A)	Axe2 variant S15A	$Geobacillus\ stearothermophilus$	16
4HYQ (A)	phospholipase A1	Streptomyces albidoflavus NA297	17
4ZR8 (A)	uroporphyrinogen decarboxylase	Acinetobacter baumannii	18
4WSH (B)	probable uroporphyrinogen decarboxylase	Pseudomonas aeruginosa	19
4R7G(A)	Phosphoribosylformylglycinamidine synthase	Salmonella enterica	20

### Active Site Network Constraint Measures

To assess the extent to which each active site was structurally constrained, four base constraint measures and one derived measure (the first principal component of these measures following standardization) were computed as described in the main text. Figure S9 shows the values of each studied protein on five constraint measures; proteins are ordered vertically by rank on the omnibus site constraint measure.



Figure S 9: Constraint measures for the active site networks. Vertical axes indicate values on each of the four base constraint measures and the omnibus derived measure, as described in the main text.

#### References

- Petersen, T.; Brunak, S.; von Heijne, G.; Henrik Nielsen, H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods* 2011, *8*, 785–786.
- (2) Raman, S.; Vernon, R.; Thompson, J.; Tyka, M.; Sadreyev, R.; Pei, J.; Kim, D.; Kellogg, E.; DiMaio, F.; Lange, O. et al. Structure prediction for CASP8 with all-atom refinement using Rosetta. *Proteins* **2009**, *77*, 89–99.
- (3) Kim, D.; Chivian, D.; Baker, D. Protein Structure Prediction and Analysis Using the Robetta Server. Nucleic Acids Research 2004, 32, W526–31.
- (4) Butts, C. T.; Zhang, X.; Kelly, J. E.; Roskamp, K. W.; Unhelkar, M. H.; Freites, J. A.; Tahir, S.; Martin, R. W. Sequence comparison, molecular modeling, and network analysis predict structural diversity in cysteine proteases from the Cape sundew, *Drosera capensis*. *Computational and Structural Biotechnology Journal* **2016**, *14*, 271–282.
- (5) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable molecular dynamics with NAMD. Journal of Computational Chemistry 2005, 26, 1781–1802.
- (6) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of Computational Chemistry* **2004**, *25*, 1605–1612.
- (7) Yan, S.; Wang, W.; Marqués, J.; Mohan, R.; Saleh, A.; Durrant, W. E.; Song, J.;
   Dong, X. Salicylic acid activates DNA damage responses to potentiate plant immunity.
   Molecular Cell 2013, 52, 602–610.
- (8) Mosher, R. A.; Durrant, W. E.; Wang, D.; Song, J.; Dong, X. A comprehensive structure-function analysis of *Arabidopsis* SNI1 defines essential regions and transcriptional repressor activity. *The Plant Cell* **2006**, *18*, 1750–1765.

- (9) van den Berg, B. Crystal structure of a full-length autotransporter. Journal of Molecular Biology 2010, 396, 627–633.
- (10) Wei, Y.; Schottel, J.; Derewenda, U.; Swenson, L.; Patkar, S.; Derewenda, Z. A novel variant of the catalytic triad in the Streptomyces scabies esterase. *Nature Structural Biology* **1995**, *2*, 218–223.
- (11) Chang, C.; Chhor, G.; Bearden, J.; Joachimiak, A. Crystal structure of lipolytic protein G-D-S-L family from *Alicyclobacillus acidocaldarius* subsp. acidocaldarius DSM 446. *To be published* **2011**,
- (12) Ma, J.; Lu, Q.; Yuan, Y.; Ge, H.; Li, K.; Zhao, W.; Gao, Y.; Niu, L.; Teng, M. Crystal structure of isoamyl acetate-hydrolyzing esterase from *Saccharomyces cerevisiae* reveals a novel active site architecture and the basis of substrate specificity. *Proteins* **2011**, 79, 662–668.
- (13) Lansky, S.; Alalouf, O.; Solomon, V.; Alhassid, A.; Belrahli, H.; Govada, L.; Chayan, N.;
  Shoham, Y.; Shoham, G. Crystal structure of a catalytic mutant of Axe2 (Axe2-H194A),
  an acetylxylan esterase from *Geobacillus stearothermophilus*. To be published 2014,
- (14) Lansky, S.; Alalouf, O.; Solomon, V.; Alhassid, A.; Belrahli, H.; Govada, L.; Chayan, N.; Shoham, Y.; Shoham, G. An Axe2 mutant (W190I), an acetyl-xylooligosaccharide esterase from *Geobacillus stearothermophilus*. To be published **2014**,
- (15) Lansky, S.; Alalouf, O.; Solomon, V.; Alhassid, A.; Belrahli, H.; Govada, L.; Chayan, N.; Shoham, Y.; Shoham, G. Crystal Structure of Axe2, an Acetylxylan Esterase from *Geobacillus stearothermophilus*. To be published **2014**,
- (16) Lansky, S.; Alalouf, O.; Solomon, V.; Alhassid, A.; Belrahli, H.; Govada, L.; Chayan, N.; Shoham, Y.; Shoham, G. Crystal Structure of Axe2, an Acetylxylan Esterase from *Geobacillus stearothermophilus. Acta Crystallographica, Section D: Biological Crystallography* **2014**, 70, 261–278.

- (17) Murayama, K.; Kano, K.; Matsumoto, Y.; Sugimori, D. Crystal structure of phospholipase A1 from Streptomyces albidoflavus NA297. Journal of Structural Biology 2013, 182, 192–196.
- (18) Abendroth, J.; Fairman, J.; Lorimer, D.; Edwards, T. Structure of uroporphyrinogen decarboxylase from Acinetobacter baumannii. *To be published* 2015,
- (19) Abendroth, J.; Dranow, D.; Lorimer, D.; Edwards, T. Crystal structure of probable uroporphyrinogen decarboxylase (UPD) (URO-D) from *Pseudomonas aeruginosa*. To be published **2014**,
- (20) Tanwar, A.; Sindhikara, D.; Hirata, F.; Anand, R. Determination of the formylglycinamide ribonucleotide amidotransferase ammonia pathway by combining 3D-RISM theory with experiment. ACS Chemical Biology 2015, 10, 698–704.