Supporting Information

The Assignment of Pterin Deaminase Activity to an Enzyme of Unknown Function Guided by Homology Modeling and Docking[†]

Hao Fan[a,b,c,§], Daniel S. Hitchcock[‡,§], Ronald D. Seidel III[ζ], Brandan Hillerich[ζ], Henry Lin[b], Steven C. Almo[ζ], Andrej Sali[a,b,c,]*, Brian K. Shoichet[b,]*, and Frank M. Raushel[‡,Ψ,]*

[a]Department of Bioengineering and Therapeutic Sciences, [b]Department of Pharmaceutical Chemistry, and [c]California Institute for Quantitative Biosciences, University of California, San Francisco.[‡]Department of Biochemistry & Biophysics, [Ψ]Department of Chemistry, Texas A&M University, College Station, Texas 77843. [ζ]Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, New York 10461.

Compounds tested in Physical Library Screening.

Table S1. Virtual screening of the HEI database against the template structure 1K70.

Table S2. Functional homologues of Arad3529.

Scheme S1. A variety of pterins found to be deaminated by Arad3529.

Figure S1. Sequence alignment between Arad3529 and cytosine deaminase (PDB id: 1K70). The active site residues in 1K70 that are within 5 Å of the co-crystallized inhibitor, and the correspondingly aligned residues in Arad3529 are colored. The helices and strands in the catalytic domain of 1K70 structure were marked by red HHH and blue SSS, respectively.

Figure S2. Operson Context for Arad3529. Arrows indicate individual operons. Arad3529 is located on the same operon with TauC, GlcD, and MdaB. The proteins from cog1402, LysR, TauA, TauB and TauC are found on different operons but are conserved across several genomes with Arad3529 homologs. Sequences which likely code for metabolic proteins are drawn in color, whereas transporters and transcription regulators are in black.

Compounds tested in Physical Library Screening:

Docking predictions: melamine, cytosine, thioguanine, guanine, adenine, 7-methylguanine, cytidine, 2'-deoxycytidine, pterin, biopterin, and neopterin.

Substrates of known amidohydrolases: guanosine, adenosine, 8-mercaptoguanine, 2,6-diaminopurine, xanthine, 8-oxoguanine, urate, 9-methylguanine, 6-chloropurine, 8-oxoadenine, isopentenyladenine, *N*6-methyladenine, 1-methyladenine, *cis*-zeatin, *trans*-zeatin, benzyladenine, 2-amino-6-benzylthiopurine, *N*-ethenopurine, *N*6-acyladenine, kinetin, *N*-dimethyladenine, *N*-butyladenine, isoguanine, 2-amino-6-methoxypurine, 6-methylthiopurine, 2-chloroadenine, 6-methoxyadenine, 6-mercaptopurine, 7-methyladenine, 2-dimethylaminoadenine, 3-iminisoindolinone, acycloguanosine, 2'-deoxyguanosine, xanthosine, 2'-deoxyadenosine, 3'-deoxyadenosine, 5'-deoxyadenosine, 2',5'-dideoxyadenosine, 5'-amino-5'-deoxyadenosine, 5'-chloro-5'-deoxyadenosine, 5'-deoxy-5'-methylthioadenosine, *S*-adenosylhomocysteine, *N*6-methyl-2'-deoxyadenosine, isoguanine, 2'-deoxyisoguanine, cytosine-β-D-arabinofuranose, *N*-methylcytidine, 5-methyl-2'-deoxycytidine, 5-hydroxymethylcytidine, 2,4-diaminopyrimidine, 4,6-diaminopyrimidine, 4-amino-2,6-dihydroxypyrimidine, 2-amino-4,6-dihydroxypyrimidine, 5-methlcytosine, 5-hydroxymethylcytosine, 5-carboxycytosine, 5-aminocytosine, 5-fluorocytosine, 5-chlorocytosine, 4-hydroxy-2,5,6-triaminopyrimidine, *N*-methylcytosine, 3-methylcytosine, 2,6-diaminopyrimidine, 2,4,6-triaminopyrimidine, uracil, 5-aminouracil, 5-formyluracil, 5-fluorouracil, 3-oxauracil, 5-hydroxymethyluracil, 6-hydroxyaminouracil, thiamine, 2-aminopyrimidine, 4-aminopyrimidine, toxopyrimidine, sulfamonomethoxine, 4-thiouracil, 2-thiouracil, thymine, 5-azacytosine, ammeline, ammelide, cyanuric acid, 2-amino-4-hydroxy-6-chlorotriazine, cyromazin, *N*-ethylammeline, desethylatrazine, 6-azacytosine, 2-aminopyridine, 2,6-diaminopyridine, 4,6-diamino-2-hydroxpyridine, 4-aminopyridine, 4-dimethylaminopyridine, creatinine, 5-amino-4-imidazolecarboxyamide, 5-hydantoin acetic acid, 3-amino-5-mercapto-1,2,4-triazole, 3-amino-5-carboxy-1,2,4-triazole, dihydrouracil, dihydrothymine, and dihydroorotate.

Substituted pterins tested further: 2,4-diamino-6-hydroxymethylpteridine, isoxanthopterin, pterin-6-carboxylate, sepiapterin, folate, formylpterin, pterin-7-carboxylate, hydroxymethylpterin, xanthopterin, 7,8-dihydrohydroxymethylpterin, and 7,8-dihydroneopterin.

Table S1. Virtual screening of the HEI database against the template structure 1K70

Name	KEGG ID	Rank
cytosine	C00380	10
creatinine	C00791	32
5-methylcytosine	C02376	48
adenine	C00147	50
N6-methyladenine	C08434	59
2'-deoxycytosine	C02026	73
guanine	C00242	77
toxopyrimidine	C01279	126

We noted that 9 molecules received better docking ranks than cytosine. These molecules are all low molecular weight compounds and contain multiple ionizable groups, such as Biuret. Thus, these compounds are likely false positives, and were not selected for experimental validation with either the cytosine deaminase or Arad3529.

Table S2. Functional homologues of Arad3529

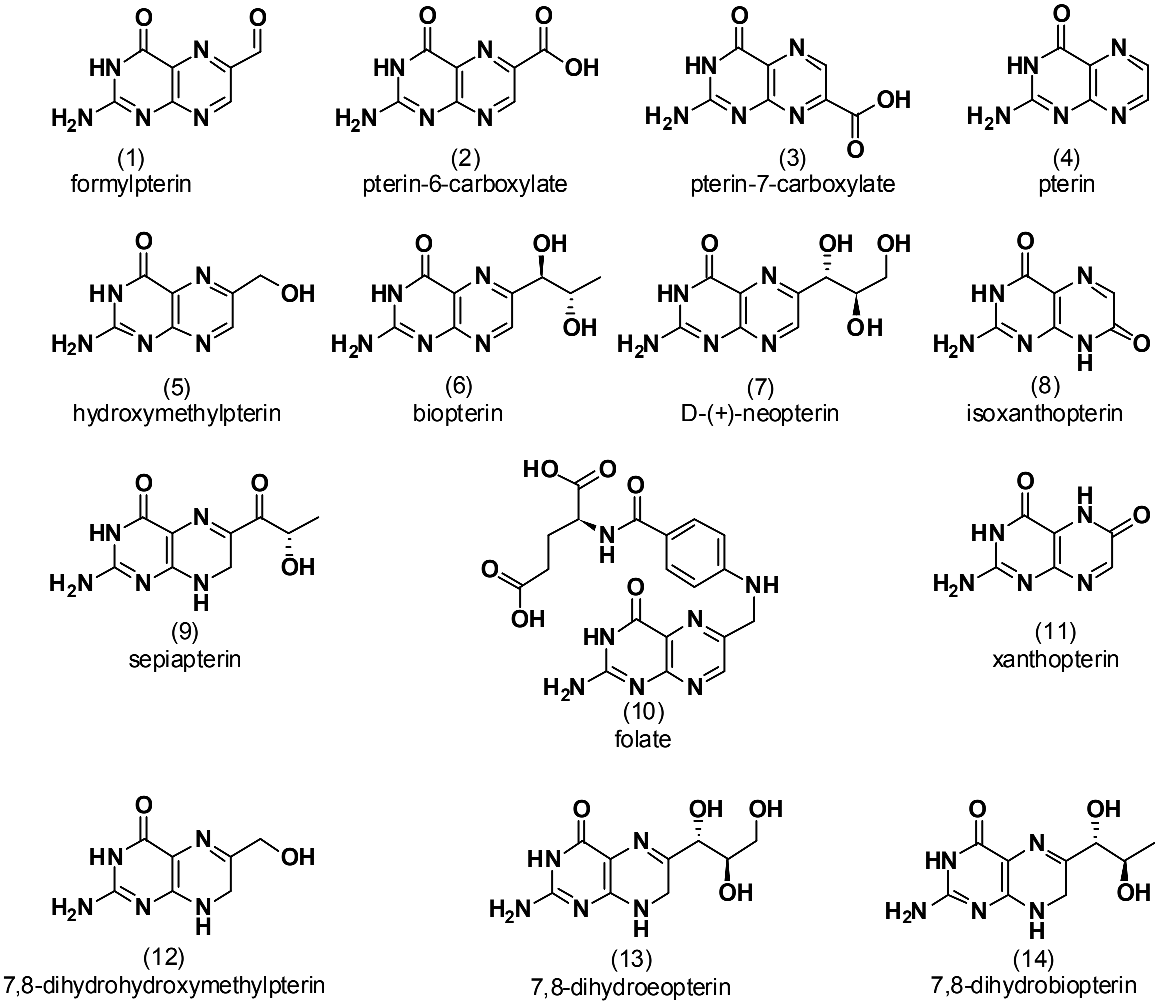
Organism	Gi Number	Locus Tag
Agrobacterium radiobacter K84	222086854	Arad_3529
Agrobacterium tumefaciens 5A	358004348	AT5A_20871
Agrobacterium vitis S4	222149611	Avi_3553
Agrobacterium vitis S4	222106464	Avi_5410
alpha proteobacterium BAL199	163793175	BAL199_25339
Arthrospira maxima CS-328	209525534	AmaxDRAFT_2897
Arthrospira platensis NIES-39	291567086	NIES39_C04920
Arthrospira platensis str. Paraca	284051729	AplaP_010100009690
Arthrospira sp. PCC 8005	376001956	ARTHRO_1140019
Burkholderia multivorans CGD2M	221198870	BURMUCGD2M_6445
Burkholderia xenovorans LB400	91780827	Bxe_C0802
Chlorella variabilis	307110611	CHLNCDRAFT_34176
Comamonas testosteroni ATCC 11996	371450227	CTATCC11996_23416
Crocosphaera watsonii WH 0003	357261488	CWATWH0003_4509
Crocosphaera watsonii WH 8501	67924588	CwatDRAFT_1740
Cyanobium sp. PCC 7001	254432712	CPCC7001_2605
Cyanothece sp. ATCC 51142	172038506	cce_3593
Cyanothece sp. ATCC 51472	354554150	Cy51472DRAFT_2251
Cyanothece sp. CCY0110	126659334	CY0110_06099
Cyanothece sp. PCC 7424	218437242	PCC7424_0235
Cyanothece sp. PCC 7822	307151773	Cyan7822_1898
Cyanothece sp. PCC 8801	218246775	PCC8801_1952
Cyanothece sp. PCC 8802	257059817	Cyan8802_1979
Dinoroseobacter shibae DFL 12	159043930	Dshi_1381
Erwinia billingiae Eb661	300717286	EbC_27110
Gluconacetobacter oboediens 174Bp2	349687131	Gobo1_010100008018
Gluconacetobacter sp. SXCC-1	330992835	SXCC_02737
Hoeflea phototrophica DFL-43	163760226	HPDFL43_08189
Hydrogenophaga sp. PBC	388566070	Q5W_0868
Jannaschia sp. CCS1	89053965	Jann_1474
Ketogulonicigenium vulgare Y25	310816506	EIO_2064
Klebsiella oxytoca 10-5246	376395439	HMPREF9690_02815
Klebsiella pneumoniae 342	206580486	KPK_2297
Klebsiella pneumoniae KCTC 2242	386035174	KPN2242_13140
Klebsiella pneumoniae subsp. pneumoniae HS11286	378979178	KPHS_30190
Klebsiella pneumoniae subsp. pneumoniae MGH 78578	152970585	KPN_02036
Klebsiella pneumoniae subsp. pneumoniae NTUH-K2044	238895079	KP1_3111
Klebsiella sp. 1_1_55	290509179	HMPREF0485_00950
Klebsiella sp. 4_1_44FAA	365137953	HMPREF1024_00689

Klebsiella sp. MS 92-3	330015330	HMPREF9538_05868
Klebsiella variicola At-22	288935109	Kvar_2241
Labrenzia aggregata IAM 12614	118592963	SIAM614_02036
Labrenzia alexandrii DFL-11	254501071	ADFL11_1107
Lyngbya sp. PCC 8106	119485483	L8106_10086
Mesorhizobium alhagi CCNWXJ12-2	359791086	MAXJ12_16641
Mesorhizobium amorphae CCNWGS0123	357026772	MEA186_18497
Mesorhizobium australicum WSM2073	354569803	
Mesorhizobium ciceri biovar biserrulae WSM1271	319783022	Mesci_3325
Mesorhizobium loti MAFF303099	14022090	mll1290
Mesorhizobium loti MAFF303099	161621451	mll1290
Mesorhizobium opportunistum WSM2075	337268293	Mesop_3816
Methylobacterium chloromethanicum CM4	218530001	Mchl_2042
Methylobacterium extorquens AM1	240138303	MexAM1_META1p1649
Methylobacterium extorquens DM4	254560845	METDI2397
Methylobacterium extorquens DSM 13060	373564592	MetexDRAFT_4416
Methylobacterium extorquens PA1	163851150	Mext_1723
Methylobacterium populi BJ001	188580356	Mpop_1091
Microcoleus chthonoplastes PCC 7420	254410796	MC7420_274
Microcoleus vaginatus FGP-2	334120965	MicvaDRAFT_2986
Microcystis aeruginosa NIES-843	166363277	MAE_05360
Microcystis aeruginosa PCC 7806	159026135	IPF_4829
Microcystis aeruginosa PCC 7941	389764552	MICAD_400012
Microcystis aeruginosa PCC 9432	389677661	MICCA_3050004
Microcystis aeruginosa PCC 9443	389731332	MICAC_5760002
Microcystis aeruginosa PCC 9701	389881376	MICAK_3820002
Microcystis aeruginosa PCC 9717	389716015	MICAB_6620002
Microcystis aeruginosa PCC 9806	389790346	MICAE_2140020
Microcystis aeruginosa PCC 9807	389802797	MICAF_3630004
Microcystis aeruginosa PCC 9808	389824008	MICAG_3490012
Microcystis aeruginosa PCC 9809	389834001	MICAH_1010004
Microcystis sp. T1-4	390441413	MICAI_4010001
Micromonas pusilla CCMP1545	303279224	MICPUCDRAFT_58167
Moorea producta 3L	332707052	LYNGBM3L_26730
Ochrobactrum anthropi ATCC 49188	153011192	Oant_3872
Octadecabacter antarcticus 307	254436998	OA307_1868
omamonas testosteroni KF-1	221070037	
Opitutaceae bacterium TAV1	390121479	ObacTA_020100014809
Opitutaceae bacterium TAV5	373851669	
Ostreococcus lucimarinus CCE9901	145349453	OSTLU_42867
Ostreococcus tauri	308806984	Ot08g02330
Pantoea sp. aB	304397339	PanABDRAFT_2477
Pantoea sp. At-9b	317046647	Pat9b_0413

Paulinella chromatophora	194476571	PCC_0086
Polymorphum gilvum SL003B-26A1	328542414	SL003B_0794
Prochlorococcus marinus str. MIT 9211	159903893	P9211_13521
Prochlorococcus marinus str. MIT 9301	126696828	P9301_14901
Prochlorococcus marinus str. MIT 9303	124023689	P9303_19891
Prochlorococcus marinus str. MIT 9312	78779784	PMT9312_1400
Prochlorococcus marinus str. MIT 9313	33862599	PMT0326
Prochlorococcus marinus str. NATL1A	124026429	NATL1_17241
Prochlorococcus marinus str. NATL2A	72382709	PMN2A_0870
Prochlorococcus marinus subsp. marinus str. CCMP1375	33240827	Pro1378
Pseudomonas sp. ADP	32455877	pADP-1_p093
Rahnella aquatilis CIP 78.65 = ATCC 33071	383189811	Rahaq2_1935
Rahnella sp. Y9602	322832607	Rahaq_1889
Rhizobium etli CFN 42	86281764	RHE_CH02043
Rhizobium etli CFN 42	162329640	RHE_CH02043
Rhizobium etli CIAT 652	190894434	RHECIAT_PC0000096
Rhizobium etli CNPAF512	327188280	RHECNPAF_9300154
Rhizobium leguminosarum bv. trifolii	1439551	
Rhizobium leguminosarum bv. trifolii WSM1325	241113027	Rleg_4666
Rhizobium leguminosarum bv. trifolii WSM2304	209549272	Rleg2_1675
Rhizobium leguminosarum bv. trifolii WSM597	392850220	Rleg9DRAFT_1550
Rhizobium leguminosarum bv. trifolii WU95	392515184	Rleg8DRAFT_0280
Rhizobium leguminosarum bv. viciae 3841	116255810	pRL110609
Rhizobium leguminosarum bv. viciae 3841	116248977	pRL120308
Rhizobium sp. PDO1-076	375053111	PDO_3140
Roseobacter denitrificans OCh 114	110678617	RD1_1293
Ruegeria sp. TM1040	99078030	TM1040_3052
Serratia odorifera 4Rx13	270262662	SOD_c02830
Serratia sp. M24T3	383816172	SPM24T3_17475
Starkeya novella DSM 506	298292809	Snov_2842
Synechococcus elongatus PCC 6301	56750963	syc0954_d
Synechococcus elongatus PCC 7942	81299379	Synpcc7942_0568
Synechococcus sp. BL107	116072321	BL107_11056
Synechococcus sp. CB0101	318040418	SCB01_010100001875
Synechococcus sp. CB0205	317970568	SCB02_010100013621
Synechococcus sp. CC9311	113953404	sync_0740
Synechococcus sp. CC9605	78212398	Syncc9605_0854
Synechococcus sp. CC9902	78185105	Syncc9902_1538
Synechococcus sp. JA-2-3B'a(2-13)	86608541	CYB_1063
Synechococcus sp. JA-3-3Ab	86606230	CYA_1567
Synechococcus sp. PCC 7335	254421902	S7335_2052
Synechococcus sp. RS9916	116073349	

Synechococcus sp. RS9917	87125519	RS9917_02061
Synechococcus sp. WH 5701	87302978	WH5701_07386
Synechococcus sp. WH 5701	87303856	WH5701_16173
Synechococcus sp. WH 7803	148240087	SynWH7803_1751
Synechococcus sp. WH 7805	88809157	WH7805_05676
Synechococcus sp. WH 8016	352096259	Syn8016DRAFT_2486
Synechococcus sp. WH 8102	33866172	
Synechococcus sp. WH 8109	260436186	SH8109_1371
Synechocystis sp. PCC 6803	16330119	slr1237
Trichodesmium erythraeum IMS101	113477958	Tery_4570
Variovorax paradoxus EPS	319796039	Varpa_5413
Variovorax paradoxus EPS	319796483	Varpa_5860
Variovorax paradoxus S110	239817717	Vapar_4756
Variovorax paradoxus S110	239818099	Vapar_5141
Verminephrobacter eiseniae EF01-2	121610081	Veis_3139

Scheme S1.



(1) formylpterin

(2) pterin-6-carboxylate

(3) pterin-7-carboxylate

(4) pterin

(5) hydroxymethylpterin

(6) biopterin

(7) D-(+)-neopterin

(8) isoxanthopterin

(9) sepiapterin

(10) folate

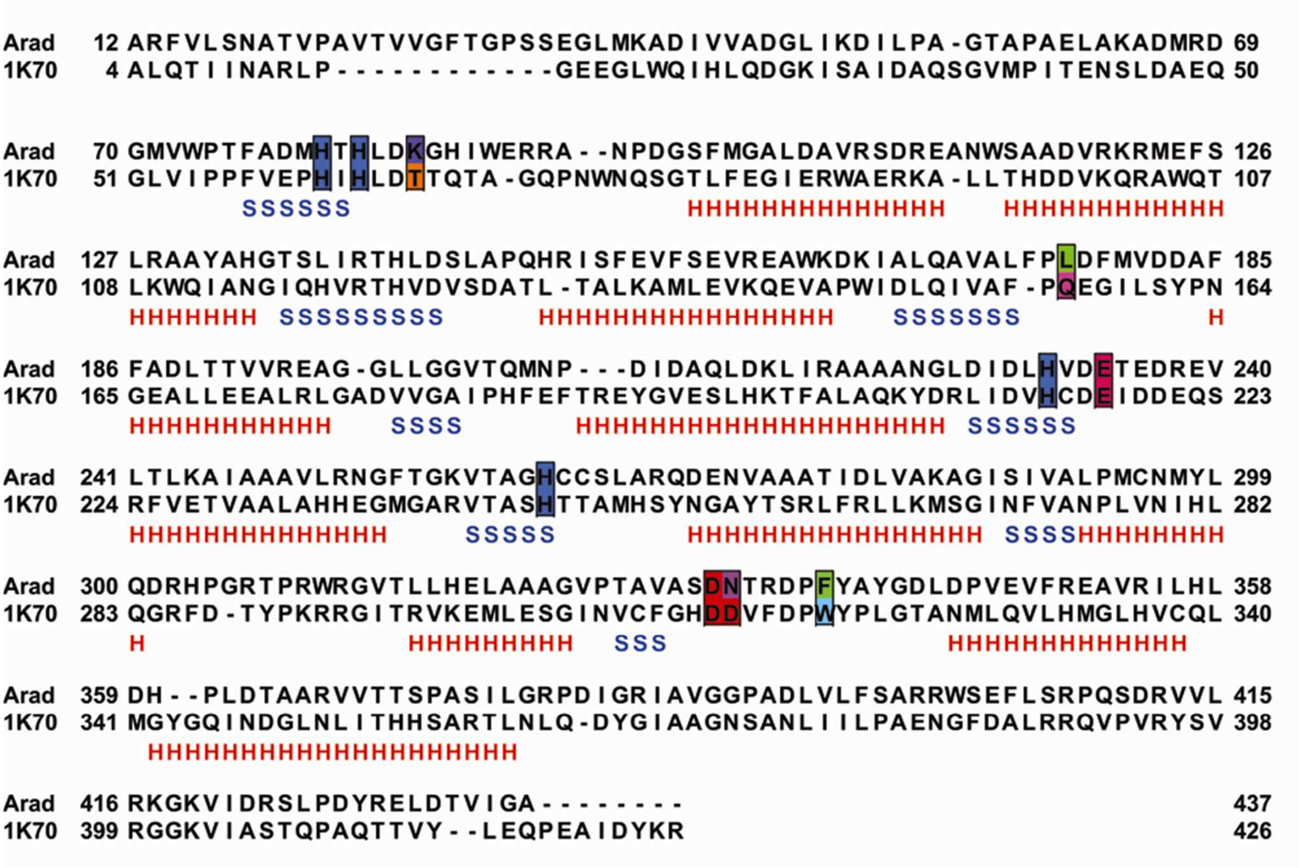
(11) xanthopterin

(12) 7,8-dihydrohydroxymethylpterin

(13) 7,8-dihydroeopterin

(14) 7,8-dihydrobiopterin

Figure S1.



```
Arad   12 ARFVLSNATVPAVTVVGFTGPSSEGLMKADIVVADGLIKDILPA-GTAPAELAKADMRD 69
1K70    4 ALQTIINARLP-------------GEEGLWQIHLQDGKISAIDAQSGVMPITENSLDAEQ 50

Arad   70 GMVWPTFADMHTHLDKGHIWERRA--NPDGSFMGALDAVRSDREANWSAADVRKRMEFS 126
1K70   51 GLVIPPFVEPHIHLDTTQTA-GQPNWNQSGTLFEGIERWAERKA-LLTHDDVKQRAWQT 107
                SSSSSS                  HHHHHHHHHHHHHH   HHHHHHHHHHHH

Arad  127 LRAAYAHGTSLIRTHLDSLAPQHRISFEVFSEVREAWKDKIALQAVALFPLDFMVDDAF 185
1K70  108 LKWQIANGIQHVRTHVDVSDATL-TALKAMLEVKQEVAPWIDLQIVAF-PQEGILSYPN 164
          HHHHHHH SSSSSSSSS     HHHHHHHHHHHHHHHH   SSSSSSS          H

Arad  186 FADLTTVVREAG-GLLGGVTQMNP---DIDAQLDKLIRAAAANGLDIDLHVDETEDREV 240
1K70  165 GEALLEEALRLGADVVGAIPHFEFTREYGVESLHKTFALAQKYDRLIDVHCDEIDDEQS 223
          HHHHHHHHHHH   SSSS      HHHHHHHHHHHHHHHHHHHH SSSSSS

Arad  241 LTLKAIAAAVLRNGFTGKVTAGHCCSLARQDENVAAATIDLVAKAGISIVALPMCNMYL 299
1K70  224 RFVETVAALAHHEGMGARVTASHTTAMHSYNGAYTSRLFRLLKMSGINFVANPLVNIHL 282
          HHHHHHHHHHHHHH    SSSSS       HHHHHHHHHHHHHHHH SSSSHHHHHHHH

Arad  300 QDRHPGRTPRWRGVTLLHELAAAGVPTAVASDNTRDPFYAYGDLDPVEVFREAVRILHL 358
1K70  283 QGRFD-TYPKRRGITRVKEMLESGINVCFGHDDVFDPWYPLGTANMLQVLHMGLHVCQL 340
          H              HHHHHHHHH  SSS               HHHHHHHHHHHHH

Arad  359 DH--PLDTAARVVTTSPASILGRPDIGRIAVGGPADLVLFSARRWSEFLSRPQSDRVVL 415
1K70  341 MGYGQINDGLNLITHHSARTLNLQ-DYGIAAGNSANLIILPAENGFDALRRQVPVRYSV 398
           HHHHHHHHHHHHHHHHHHHH

Arad  416 RKGKVIDRSLPDYRELDTVIGA--------                              437
1K70  399 RGGKVIASTQPAQTTVY--LEQPEAIDYKR                              426
```

Figure S2.