SUPPORTING INFORMATION

Title: Dihydroorotase from the hyperthermophile, Aquifiex aeolicus, is activated by stoichiometric association with aspartate transcarbamoylase and forms a one pot reactor for pyrimidine biosynthesis

Authors: Pengfei Zhang, Philip Martin, Cristina Purcarea, Asmita Vaishnav, Joseph S. Brunzellel, Roshini Fernando, Hedeel I. Guy-Evans, David R. Evans, and Brian F. P. Edwards

Table S1: Representative structures for carbamoyl phosphate synthetase, aspartate transcarbamoylase, and dihydroorotase.

Enzyme	**Class**	**Species**	**PDB**	**Limit**	**Ligands** [1]	**No. Id.** [2]	**Citation** [3]
CPS		Escherichia coli	1JDB	2.10	None		(*1*)
ATC	A	Aquifex aeolicus (DAC)	3D6N	2.3	None (T-state	291	Paper
ATC	C	Bacillus subtilis	2AT2	3.00 (Cα)	None (T-state)	118	(*2*)
ATC	C	Pyrococcus abyssi	1ML4	1.80	PALA (R-state)	99	(*3*)
ATC	B	Escherichia coli	6AT1	2.60	None (T-state)	89	(*4*)
ATC	B	Escherichia coli	1D09	2.10	PALA (R-state)	90	(*5*)
ATC	B	Escherichia coli (C-trimer)	3CSU	1.88	None (T-state)	91	(*6*)
ATC	B	Escherichia coli (C-trimer)	1EKX	1.95	PALA (R-state)	91	(*6*)
ATC	C	Sulfolobus acidocaldarius	1PG5	2.60	None (T-state)	88	(*7*)
ATC	B	Moritella profunda	2BE7	2.85	None (T-state)	81	(*8*)
DHO	I	Aquifex aeolicus (DAC)	3D6N	2.3	FLC	422	Paper
DHO	I	Aquifex aeolicus	1XRF	1.65	None	422	(*9*)

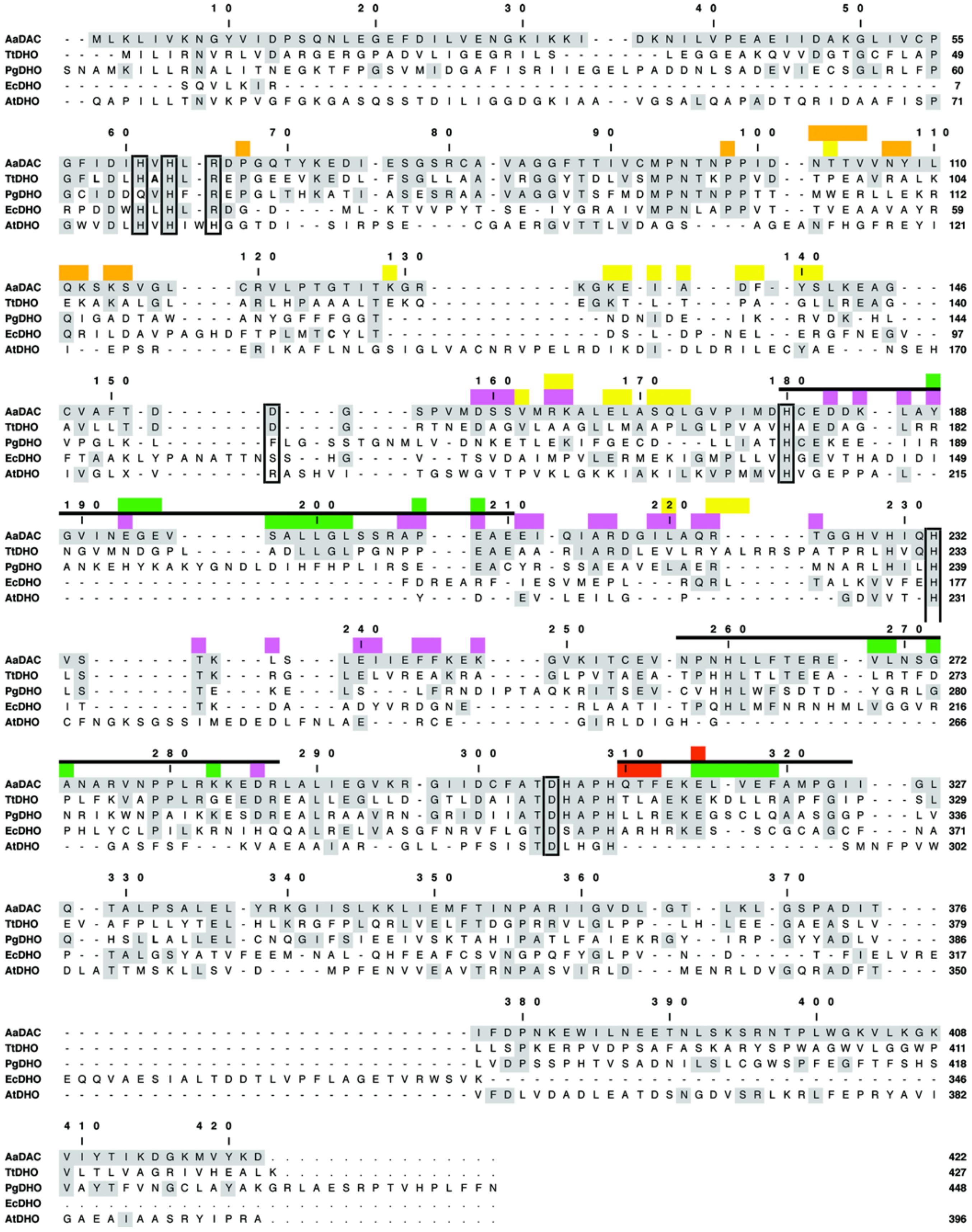
DHO	I	Aquifex aeolicus	1XRT	1.61	None	422	(*9*)
DHO	I	Thermus thermophilus	2Z00	2.42	None	158	RSGI
DHO	I	Porphyromonas gingivalis	2GWN	1.85	CAC	128	MCSG
DHO	II	Escherichia coli	1J79A 1J79B	1.70	ORO NCD	68	(*10*)
DHO	II	Escherichia coli (T109S)	2E25	2.70	FOT	68	(*11*)
DHO	II	Escherichia coli	1XGE	1.90	NCD DOR	68	(*12*)
DHO	II	Escherichia coli	2EG6 (h)	1.70	None	68	(*13*)
DHO	II	Escherichia coli	2EG7 (h)	2.00	HDDP	68	(*13*)
DHO	II	Escherichia coli	2EG8 (h)	2.20	FOT	68	(*13*)
DHO	I	Agrobacterium tumefaciens	2OGJ	2.62	None	27	NSGXRC

---

[1] Ligand codes: CAC = cacodylate ion; DOR = (4S)-2,3-dioxohexahydropyrimidine-4-carboxylic acid; FOT = 5-fluoro-2,6-dioxo-1,2,3,6-tetrahydropyrimidine-4-carboxylic acid (5-fluoroorotate); HDDP = 2-oxo-1,2,3,6-tetrahydropyrimidine-4,6-dicarboxylic acid; NCD = lysine NZ-carboxylic acid; ORO = orotic acid

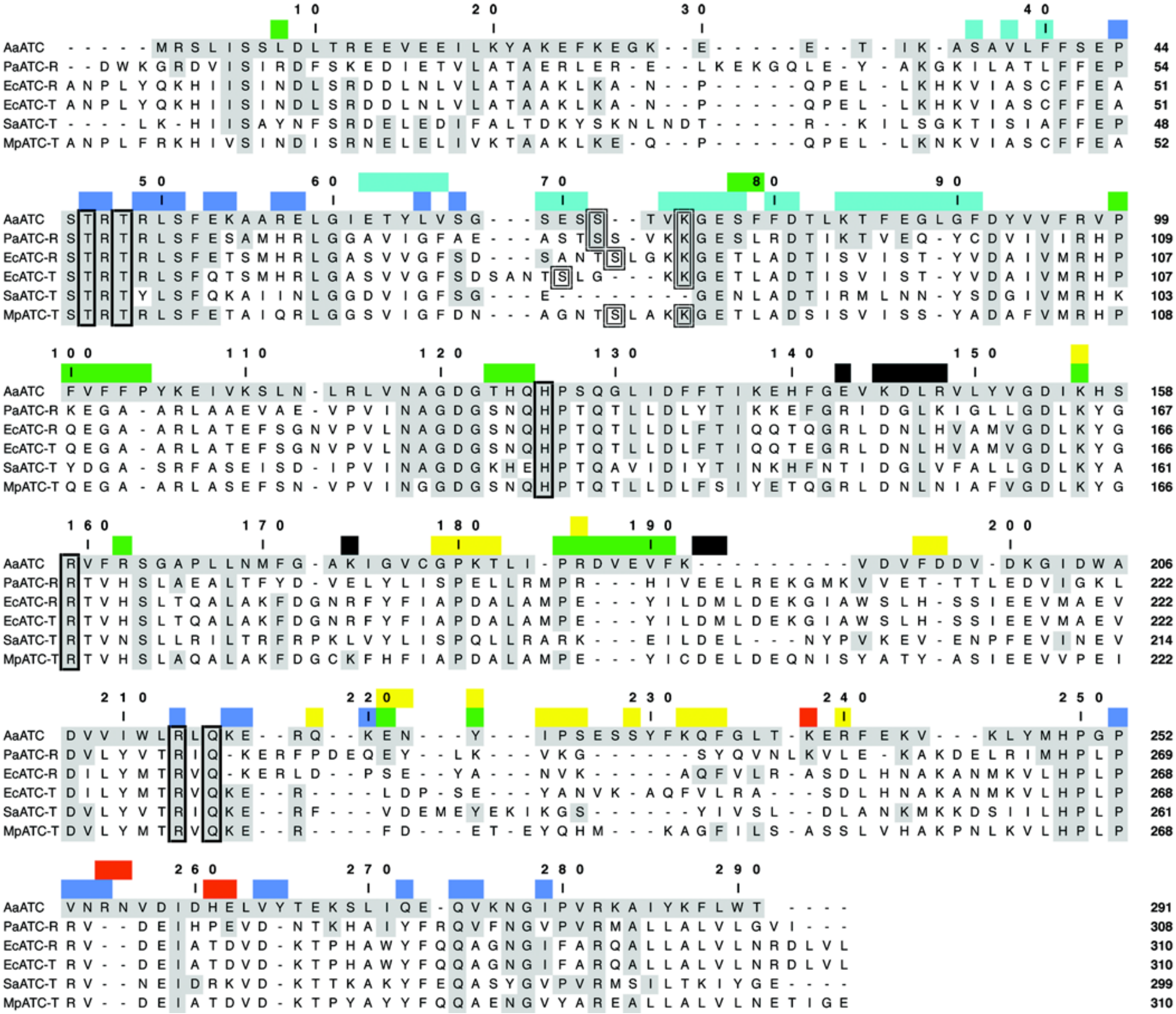
[2] Number of identical residues from a Blast search of the Protein Data Bank using the sequence of the corresponding protein from *Aquifex aeolicus*.

[3] MCSG = Midwest Center for Structural Genomics, paper “to be published”; NSGXRC = New York Structural GenomiX Research Consortium, paper “to be published”; RSGI = RIKEN Structural Genomics/Proteomics Initiative.



```
                  10        20        30           40        50
                   |         |         |            |         |
AaDAC   --MLKLIVKNGYVIDPSQNLEGEFDILVENGKIKKI---DKNILVPEAEIIDAKGLIVCP 55
TtDHO   ----MILIRNVRLVDARGERGPADVLIGEGRILS-------LEGGEAKQVVDGTGCFLAP 49
PgDHO   SNAMKILLRNALITNEGKTFPGSVMIDGAFISRIIEGELPADDNLSADEVIECSGLRLFP 60
EcDHO   --------SQVLKIR--------------------------------------------- 7
AtDHO   --QAPILLTNVKPVGFGKGASQSSTDILIGGDGKIAA--VGSALQAPADTQRIDAAFISP 71

           60         70         80         90        100         110
            |          |          |          |         |           |
AaDAC   GFIDIHVHL-RDPGQTYKEDI-ESGSRCA-VAGGFTTIVCMPNTNPPID--NTTVVNYIL 110
TtDHO   GFLDLHAHL-REPGEEVKEDL-FSGLLAA-VRGGYTDLVSMPNTKPPVD--TPEAVRALK 104
PgDHO   GCIDDQVHF-REPGLTHKATI-ASESRAA-VAGGVTSFMDMPNTNPPTT--MWERLLEKR 112
EcDHO   RPDDWHLHL-RDG-D----ML-KTVVPYT-SE-IYGRAIVMPNLAPPVT--TVEAAVAYR 59
AtDHO   GWVDLHVHIWHGGTDI--SIRPSE----CGAERGVTTLVDAGS----AGEANFHGFREYI 121

                    120       130                        140
                     |         |                          |
AaDAC   QKSKSVGL----CRVLPTGTITKGR----------KGKE-I-A---DF--YSLKEAG--- 146
TtDHO   EKAKALGL----ARLHPAAALTEKQ----------EGKT-L-T---PA--GLLREAG--- 140
PgDHO   QIGADTAW----ANYGFFFGGT---------------NDNIDE---IK--RVDK-HL--- 144
EcDHO   QRILDAVPAGHDFTPLMTCYLT---------------DS-L-DP-NEL--ERGFNEGV-- 97
AtDHO   I--EPSR-----ERIKAFLNLGSIGLVACNRVPELRDIKDI-DLDRILECYAE---NSEH 170

          150                       160       170       180
           |                         |         |         |
AaDAC   CVAFT-D-------D----G----SPVMDSSVMRKALELASQLGVPIMDHCEDDK--LAY 188
TtDHO   AVLLT-D-------D----G----RTNEDAGVLAAGLLMAAPLGLPVAVHAEDAG--LRR 182
PgDHO   VPGLK-L-------FLG-SSTGNMLV-DNKETLEKIFGECD---LLIATHCEKEE--IIR 189
EcDHO   FTAAKLYPANATTNSS--HG----V--TSVDAIMPVLERMEKIGMPLLVHGEVTHADIDI 149
AtDHO   IVGLX-V-------RASHVI----TGSWGVTPVKLGKKIAKILKVPMMVHVGEPPA-L-- 215

        190             200          210        220             230
         |               |            |          |               |
AaDAC   GVINEGEV------SALLGLSSRAP---EAEEI-QIARDGILAQR------TGGHVHIQH 232
TtDHO   NGVMNDGPL-----ADLLGLPGNPP---EAEAA-RIARDLEVLRYALRRSPATPRLHVQH 233
PgDHO   ANKEHYKAKYGNDLDIHFHPLIRSE---EACYR-SSAEAVELAER------MNARLHILH 239
EcDHO   -----------------------FDREARF-IESVMEPL---RQRL-----TALKVVFEH 177
AtDHO   ------------------------Y---D--EV-LEILG---P----------GDVVT-H 231

                           240           250        260         270
                            |             |          |           |
AaDAC   VS-------TK---LS---LEIIEFFKEK----GVKITCEV-NPNHLLFTERE--VLNSG 272
TtDHO   LS-------TK---RG---LELVREAKRA----GLPVTAEA-TPHHLTLTEEA--LRTFD 273
PgDHO   LS-------TE---KE---LS---LFRNDIPTAQKRITSEV-CVHHLWFSDTD--YGRLG 280
EcDHO   IT-------TK---DA---ADYVRDGNE-------RLAATI-TPQHLMFNRNHMLVGGVR 216
AtDHO   CFNGKSGSSIMEDEDLFNLAE---RCE--------GIRLDIGH-G--------------- 266

              280       290        300       310        320
               |         |          |         |          |
AaDAC   ANARVNPPLRKKEDRLALIEGVKR-GIIDCFATDHAPHQTFEKEL-VEFAMPGII---GL 327
TtDHO   PLFKVAPPLRGEEDREALLEGLLD-GTLDAIATDHAPHTLAEKEKDLLRAPFGIP---SL 329
PgDHO   NRIKWNPAIKKESDREALRAAVRN-GRIDIIATDHAPHLLREKEGSCLQAASGGP---LV 336
EcDHO   PHLYCLPILKRNIHQQALRELVASGFNRVFLGTDSAPHARHRKES--SCGCAGCF---NA 371
AtDHO   ---GASFSF---KVAEAAIAR--GLL-PFSISTDLHGH---------------SMNFPVW 302

            330       340       350       360           370
             |         |         |         |             |
AaDAC   Q--TALPSALEL-YRKGIISLKKLIEMFTINPARIIGVDL-GT--LKL-GSPADIT---- 376
TtDHO   EV-AFPLLYTEL-HLKRGFPLQRLVELFTDGPRRVLGLPP--LH-LEE-GAEASLV---- 379
PgDHO   Q--HSLLALLEL-CNQGIFSIEEIVSKTAHIPATLFAIEKRGY--IRP-GYYADLV---- 386
EcDHO   P--TALGSYATVFEEM-NAL-QHFEAFCSVNGPQFYGLPV--N--D-----T-FIELVRE 317
AtDHO   DLATTMSKLLSV-D----MPFENVVEAVTRNPASVIRLD---MENRLDVGQRADFT---- 350

                                      380       390       400
                                       |         |         |
AaDAC   ----------------------------IFDPNKEWILNEETNLSKSRNTPLWGKVLKGK 408
TtDHO   ----------------------------LLSPKERPVDPSAFASKARYSPWAGWVLGGWP 411
PgDHO   ----------------------------LVDPSSPHTVSADNILSLCGWSPFEGFTFSHS 418
EcDHO   EQQVAESIALTDDTLVPFLAGETVRWSVK------------------------------- 346
AtDHO   ----------------------------VFDLVDADLEATDSNGDVSRLKRLFEPRYAVI 382

        410       420
         |         |
AaDAC   VIYTIKDGKMVYKD................                               422
TtDHO   VLTLVAGRIVHEALK...............                               427
PgDHO   VAYTFVNGCLAYAKGRLAESRPTVHPLFFN                               448
EcDHO   ..............................
AtDHO   GAEAIAASRYIPRA................                               396
```

Figure S1. Structure-based sequence alignment for AaeDHO. The sequences of dihydroorotases from *T. thermophilus* (2z00A.pdb), *P. gingivalis* (2gwnA.pdb), *E. coli* (1xgeB.pdb), and *A. tumefaciens* (2ogjA.pdb) are aligned with the sequence of AaeDHO based upon the structural equivalences identified by SSM (*14*). The rmsd values for AaeATC superposed onto TthDHO, PgiDHO, EcoDHO and AtuDHO are 1.1 Å, 1.7 Å, 2,4 Å, and 3.3 Å, respectively. The ruler above each block gives the position in the sequence of AaeDHO; the sequence number for the other aligned structures are given at the end of each row. Identities with AaeDHO have a grey background. AaeDHO residues assigned to a specific interface by PISA are marked with the appropriate color in the cell above them – or in the two cells above if the residue participates in two interfaces. The primary residues in the active site are boxed. AaeDHO residues assigned to loops-A, -B, and –C, which start respectively at H180, N257, and Q310, are flagged with horizontal bars. The carboxylated lysine (KCX102) in EcoDHO is marked with an 'X'. Only the first two letters of the species prefixes are used in this figure.



```
                      10        20          30                 40
AaATC   -----MRSLISSLDLTREEVEEILKYAKEFKEGK--E-----E--T-IK-ASAVLFFSEP 44
PaATC-R --DWKGRDVISIRDFSKEDIETVLATAERLER-E--LKEKGQLE-Y-AKGKILATLFFEP 54
EcATC-R ANPLYQKHIISINDLSRDDLNLVLATAAKLKA-N--P-----QPEL-LKHKVIASCFFEA 51
EcATC-T ANPLYQKHIISINDLSRDDLNLVLATAAKLKA-N--P-----QPEL-LKHKVIASCFFEA 51
SaATC-T ----LK-HIISAYNFSRDELEDIFALTDKYSKNLNDT-----R--KILSGKTISIAFFEP 48
MpATC-T ANPLFRKHIVSINDISRNELELIVKTAAKLKE-Q--P-----QPEL-LKNKVIASCFFEA 52

             50        60           70          80        90
AaATC   STRTRLSFEKAARELGIETYLVSG---SESS--TVKGESFFDTLKTFEGLGFDYVVFRVP 99
PaATC-R STRTRLSFESAMHRLGGAVIGFAE---ASTSS-VKKGESLRDTIKTVEQ-YCDVIVIRHP 109
EcATC-R STRTRLSFETSMHRLGASVVGFSD---SANTSLGKKGETLADTISVIST-YVDAIVMRHP 107
EcATC-T STRTRLSFQTSMHRLGASVVGFSDSANTSLG---KKGETLADTISVIST-YVDAIVMRHP 107
SaATC-T STRTYLSFQKAIINLGGDVIGFSG---E--------GENLADTIRMLNN-YSDGIVMRHK 103
MpATC-T STRTRLSFETAIQRLGGSVIGFDN---AGNTSLAKKGETLADSISVISS-YADAFVMRHP 108

        100       110        120       130       140       150
AaATC   FVFFPYKEIVKSLN-LRLVNAGDGTHQHPSQGLIDFFTIKEHFGEVKDLRVLYVGDIKHS 158
PaATC-R KEGA-ARLAAEVAE-VPVINAGDGSNQHPTQTLLDLYTIKKEFGRIDGLKIGLLGDLKYG 167
EcATC-R QEGA-ARLATEFSGNVPVLNAGDGSNQHPTQTLLDLFTIQQTQGRLDNLHVAMVGDLKYG 166
EcATC-T QEGA-ARLATEFSGNVPVLNAGDGSNQHPTQTLLDLFTIQQTEGRLDNLHVAMVGDLKYG 166
SaATC-T YDGA-SRFASEISD-IPVINAGDGKHEHPTQAVIDIYTINKHFNTIDGLVFALLGDLKYA 161
MpATC-T QEGA-ARLASEFSN-VPVINGGDGSNQHPTQTLLDLFSIYETQGRLDNLNIAFVGDLKYG 166

         160       170        180        190                200
AaATC   RVFRSGAPLLNMFG-AKIGVCGPKTLI-PRDVEVFK---------VDVFDDV-DKGIDWA 206
PaATC-R RTVHSLAEALTFYD-VELYLISPELLRMPR---HIVEELREKGMKVVET-TTLEDVIGKL 222
EcATC-R RTVHSLTQALAKFDGNRFYFIAPDALAMPE---YILDMLDEKGIAWSLH-SSIEEVMAEV 222
EcATC-T RTVHSLTQALAKFDGNRFYFIAPDALAMPE---YILDMLDEKGIAWSLH-SSIEEVMAEV 222
SaATC-T RTVNSLLRILTRFRPKLVYLISPQLLRARK---EILDEL---NYPVKEV-ENPFEVINEV 214
MpATC-T RTVHSLAQALAKFDGCKFHFIAPDALAMPE---YICDELDEQNISYATY-ASIEEVVPEI 222

           210           220             230        240          250
AaATC   DVVIWLRLQKE--RQ--KEN---Y---IPSESSYFKQFGLT-KERFEKV---KLYMHPGP 252
PaATC-R DVLYVTRIQ-KERFPDEQEY--LK---VKG------SYQVNLKVLE-KAKDELRIMHPLP 269
EcATC-R DILYMTRVQ-KERLD--PSE--YA---NVK-----AQFVLR-ASDLHNAKANMKVLHPLP 268
EcATC-T DILYMTRVQKE--R----LDP-SE---YANVK-AQFVLRA---SDLHNAKANMKVLHPLP 268
SaATC-T DVLYVTRIQKE--RF---VDEMEYEKIKGS------YIVSL--DLANKMKKDSIILHPLP 261
MpATC-T DVLYMTRVQKE--R----FD---ET-EYQHM---KAGFILS-ASSLVHAKPNLKVLHPLP 268

               260       270       280       290
AaATC   VNRNVDIDHELVYTEKSLIQE-QVKNGIPVRKAIYKFLWT ----                291
PaATC-R RV--DEIHPEVD-NTKHAIYFRQVFNGVPVRMALLALVLGVI---                308
EcATC-R RV--DEIATDVD-KTPHAWYFQQAGNGIFARQALLALVLNRDLVL                310
EcATC-T RV--DEIATDVD-KTPHAWYFQQAGNGIFARQALLALVLNRDLVL                310
SaATC-T RV--NEIDRKVD-KTTKAKYFEQASYGVPVRMSILTKIYGE----                299
MpATC-T RV--DEIATDVD-KTPYAYYFQQAENGVYAREALLALVLNETIGE                310
```

Figure S2. Structure-based sequence alignment for AaeATC. The sequences of aspartate transcarbamoylase from *T. abyssi* (R-state; 1ml4A.pdb), *E. coli* (R-state: 1d09A.pdb; T-state: 6at1A.pdb), *S. acidophilus* (T-state; 1pg5A.pdb), and *M. profunda*. (T-state; 2be7A.pdb) are aligned with the sequence of AaeATC based upon the structural equivalences identified by SSM (*14*). The rmsd values relative to AaeATC are 1.9 Å, 1.9 Å, 2.2 Å, 2.1 Å, and 2.3 Å, respectively. The ruler above each block gives the position in the sequence of AaeATC; sequence numbers for the other aligned structures are given at the end of each row. Identities with AaeATC have a grey background. AaeATC residues assigned to a specific interface by PISA are marked with the appropriate color in the cell above them – or in the two cells above if the residue participates in two interfaces. The primary residues in the active site are boxed with single lines; the residues participating in the active site of an adjacent catalytic chain are boxed with double lines. Only the first two letters of the species prefixes are used in this figure.

REFERENCES

1. Thoden, J. B., Holden, H. M., Wesenberg, G., Raushel, F. M., and Rayment, I. (1997) Structure of carbamoyl phosphate synthetase: a journey of 96 A from substrate to product, *Biochemistry 36*, 6305-6316.
2. Stevens, R. C., Reinisch, K. M., and Lipscomb, W. N. (1991) Molecular structure of Bacillus subtilis aspartate transcarbamoylase at 3.0 A resolution, *Proc Natl Acad Sci U S A 88*, 6087-6091.
3. Van Boxstael, S., Cunin, R., Khan, S., and Maes, D. (2003) Aspartate transcarbamylase from the hyperthermophilic archaeon Pyrococcus abyssi: thermostability and 1.8A resolution crystal structure of the catalytic subunit complexed with the bisubstrate analogue N-phosphonacetyl-L-aspartate, *J Mol Biol 326*, 203-216.
4. Stevens, R. C., Gouaux, J. E., and Lipscomb, W. N. (1990) Structural consequences of effector binding to the T state of aspartate carbamoyltransferase: crystal structures of the unligated and ATP- and CTP-complexed enzymes at 2.6-A resolution, *Biochemistry 29*, 7691-7701.
5. Jin, L., Stec, B., Lipscomb, W. N., and Kantrowitz, E. R. (1999) Insights into the mechanisms of catalysis and heterotropic regulation of Escherichia coli aspartate transcarbamoylase based upon a structure of the enzyme complexed with the bisubstrate analogue N-phosphonacetyl-L-aspartate at 2.1 A, *Proteins 37*, 729-742.
6. Endrizzi, J. A., Beernink, P. T., Alber, T., and Schachman, H. K. (2000) Binding of bisubstrate analog promotes large structural changes in the unregulated catalytic trimer of aspartate transcarbamoylase: implications for allosteric regulation induced cell migration, *Proc Natl Acad Sci U S A 97*, 5077-5082.
7. De Vos, D., Van Petegem, F., Remaut, H., Legrain, C., Glansdorff, N., and Van Beeumen, J. J. (2004) Crystal structure of T state aspartate carbamoyltransferase of the hyperthermophilic archaeon Sulfolobus acidocaldarius, *J Mol Biol 339*, 887-900.

8. De Vos, D., Hulpiau, P., Vergauwen, B., Savvides, S. N., and Van Beeumen, J. (2005) Expression, purification, crystallization and preliminary X-ray crystallographic studies of a cold-adapted aspartate carbamoyltransferase from Moritella profunda, *Acta Crystallograph Sect F Struct Biol Cryst Commun 61*, 279-281.

9. Martin, P. D., Purcarea, C., Zhang, P., Vaishnav, A., Sadecki, S., Guy-Evans, H. I., Evans, D. R., and Edwards, B. F. (2005) The crystal structure of a novel, latent dihydroorotase from Aquifex aeolicus at 1.7A resolution, *J Mol Biol 348*, 535-547.

10. Thoden, J. B., Phillips, G. N., Jr., Neal, T. M., Raushel, F. M., and Holden, H. M. (2001) Molecular Structure of Dihydroorotase: A Paradigm for Catalysis through the Use of a Binuclear Metal Center, *Biochemistry 40*, 6989-6997.

11. Lee, M., Maher, M. J., and Guss, J. M. (2007) Structure of the T109S mutant of Escherichia coli dihydroorotase complexed with the inhibitor 5-fluoroorotate: catalytic activity is reflected by the crystal form, *Acta Crystallograph Sect F Struct Biol Cryst Commun 63*, 154-161.

12. Lee, M., Chan, C. W., Mitchell Guss, J., Christopherson, R. I., and Maher, M. J. (2005) Dihydroorotase from Escherichia coli: loop movement and cooperativity between subunits, *J Mol Biol 348*, 523-533.

13. Lee, M., Chan, C. W., Graham, S. C., Christopherson, R. I., Guss, J. M., and Maher, M. J. (2007) Structures of Ligand-free and Inhibitor Complexes of Dihydroorotase from Escherichia coli: Implications for Loop Movement in Inhibitor Design, *J Mol Biol.*

14. Krissinel, E., and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions, *Acta Crystallogr D Biol Crystallogr 60*, 2256-2268.