Supplementary Materials for:

Functional Significance of Evolving Protein Sequence in Dihydrofolate Reductase from Bacteria to Human

Classification: *Biological Sciences*: Biochemistry

Authors: C. Tony Liu[1], Philip Hanoian[1], Jarrod B. French[1], Thomas H. Pringle[2]*, Sharon Hammes-Schiffer[3]*, Stephen J. Benkovic[1]*

Affiliations:

[1]Department of Chemistry, Pennsylvania State University, University Park, PA 16802, USA.

[2]The Sperling Foundation, Eugene, OR 97405, USA.

[3]Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, IL 61801-3364, USA.

*To Whom Correspondence should be addressed. E-mails: tom@cyber-dyne.com (T.H.P.), shs3@illinois.edu (S.H.-S.), sjb1@psu.edu (S.J.B.).

Contents **page**

1. Evolutionary analysis S3
    8.1. Comparison between human and E. coli DHFR S3
    8.2. Sequence analysis S4
    8.3. PCE analysis S13
    8.4. Sequence variability test Materials. S20

2. Kinetics and pH/rate profiles. S26

3. Thermodynamic binding of ecDHFR mutants. S30

4. Kinetic isotope effect S30

5. Crystallization S31
    7.1. Crystallization and data collection S31
    7.2. Data processing, structure determination and refinement S31

6. Empirical Valence Bond Molecular Dynamics Simulations S34

7. Isothermal titration calorimetry (ITC) S43

8. References S43

1. Evolutionary analysis

1.1. Comparison between human and *E. coli.* DHFR. E. coli DHFR has 26% identity (alignment below) as compared to human DHFR. In view of the trillions of generations that E. coli has undergone since its divergence with human, the 26% identity may represent a floor to the divergence possible with retention of function. The 26% identity is a mix of strictly invariant residues important to the fold or active site, probabilistic agreement at reduced alphabet wobble positions, and accidental agreement at unconstrained positions. This level of divergence was reached long ago (as implied by reconstructed Cambrian human ancestral DHFR) and is consistent with a steadfast core role in thymidylate biosynthesis (1).

Another measure of conservation relevant here is root-mean-square spatial comparison of *E. coli* and human folds. Using 3F8Y (PDB) for human, the DaliLite server (2) aligns the 1DDS structure of *E. coli* to a root mean-square difference of 2.0 angstroms and found 38% similarity. Structural alignment will not be in complete agreement with homological alignment due to differences in handling gaps.

```
                        10        20        30        40        50        60
                         |         |         |         |         |         |
DHFR_homSap   MVGSLNCIVAVSQNMGIGKNGDLPWPPLRNEFRYFQRMTTTSSVEGKQNLVIMGKKTWFSIPE
DHFR_escCol   ---MISLIAALAVDRVIGMENAMPWN-LPADLAWFKRNTLNKPV-------IMGRHTWESI--
Consensus            I A   #  IG #  $PW  L  #   F R T    V       IMG  TW SI

                        73        83        93       103       113       123
                         |         |         |         |         |         |
DHFR_homSap   KNRPLKGRINLVLSRELKEPPQGAHFLSRSLDDALKLTEQPELANKVDMVWIVGGSSVYKEAM
DHFR_escCol   -GRPLPGRKNIILSSQPGTDDRVTWV--KSVDEAIAA------CGDVPEIMVIGGGRVYEQFL
Consensus       RPL GR N !LS #             S D#A            V  ! !!GG  VY # $

                       136       146       156       166       176       186
                         |         |         |         |         |         |
DHFR_homSap   NHPGHLKLFVTRIMQDFESDTFFPEIDLEKYKLLPEYPGVLSDVQEEKGIKYKFEVYEKND
DHFR_escCol   --PKAQKLYLTHIDAEVEGDTHFPDYEPDDWESVFSE---FHDADAQNSHSYCFEILERR-
Consensus       P   KL% T I  # E DT FP# # #             D # #    Y FE! E
```

Alignment data :
Alignment length : 187
Residues conserved (upper-case letters) : 49 is 26.20 %
Residues not conserved (white space): 118 is 63.10 %
IV conserved positions (!) : 5 is 2.67 %
LM conserved positions ($) : 2 is 1.07 %
FY conserved positions (%) : 1 is 0.53 %
NDQEBZ conserved positions (#): 5 is 2.62 % (B = D or N, Z = E or Q)
Sequence 0001 : DHFR_homSap (187 residues).
Sequence 0002 : DHFR_escCol (159 residues).

```
 No:  Chain   Z    rmsd lali  nres  %id
 301:  1dds-A 19.8  2.0  155   160   30

DSSP  llLEEEEEEELLLLEEELlLLLLLLLLHHHHHHHHHHHHlllllllEEEEEEEHHHHHHL
Query vgSLNCIVAVSQNMGIGKnGDLPWPPLRNEFRYFKRMTTtssvegkQNLVIMGKKTWFSI   60
ident       | |      ||         |      ||| |           ||||  || ||
Sbjct --MISLIAALAVDRVIGM-ENAMPWNLPADLAWFKRNTL-------NKPVIMGRHTWESI   50
DSSP  --LEEEEEEELHHHLLLL-LLLLLLLLHHHHHHHHHHL-------LLEEEEEHHHHHHH

DSSP  LhhhLLLLLLEEEEELLLlLLLLLLLlEEELLHHHHHHHLLlllllLEEEEEELLLHHH
Query PeknRPLKGRINLVLSRElKEPPQGAhFLSRSLDDALKLTEqpelanKVDMVWIVGGSSV  120
ident     ||| || |  ||               | | |            |      ||  |
Sbjct G---RPLPGRKNIILSSQ-PGTDDRV-TWVKSVDEAIAACG------DVPEIMVIGGGRV   99
DSSP  L---LLLLLLEEEEELLL-LLLLLLL-EEELLHHHHHHHHL------LLLLEEEEELLHH

DSSP  HHHHHLLlleEEEEEEELLLLLLLEELLLLLLLLEELLlllllLLLEEE---LLEEE
Query YKEAMNHpghLKLFVTRIMQDFESDTFFPEIDLEKYKLLPeypgvlSDVQEE---KGIKY  177
ident |          ||  | |    | || ||                |            |
Sbjct YEQFLPK--aQKLYLTHIDAEVEGDTHFPDYEPDDWESVF------SEFHDAdaqNSHSY  151
DSSP  HHLLHHH--eEEEEEEELLLLLLLLLLLLLHHHEEEEE------EEEELLlllLLLLE

DSSP  EEEEEEEl
Query KFEVYEKNd  186
ident  ||  |
Sbjct CFEILERRx  160
DSSP  EEEEEEEL l
```

1.2. Sequence analysis. The 233 dihydrofolate reductase sequences provided here are hand-curated for greater accuracy, necessary when DHFR proteins are implicit in genome project assemblies but not provided as such at GenBank. As predicted by unsupervised bioinformatics algorithms, gene models can be unacceptably inaccurate (modelled without use of synergistic homological data and carrying erroneous start codons, skipped or diverged exons, retained short introns, homopolymer run frameshift errors, translations by inappropriate genetic code, and confusion with paralogs and pseudogenes). Such errors would obscure evaluation of evolutionary change in conserved features; hence the need for manual curation of each protein sequence.

The method of curation used here begins with a seed set of thoroughly studied experimental sequences in model species where there is no doubt about the completeness of the sequence, nor its accuracy. Using the four main divisions of GenBank (nr, ESTs, transcriptome projects, whole genome assemblies) and pre-calculated whole genome alignments of vertebrates at University of California, Santa Cruz, the seed set is slowly expanded in closely related species by orthologous representatives sharing homology, syntenic location and exon pattern.

The build-out of the reference sequence collection improves recursively in accuracy because of four independent tools: an ever-growing custom blast classifier (3), a phylogenetically aware sequence multi-aligner (4), a pre-computed best-blast phylogenetic overview of neighboring genes (5) and a 46-species whole genome alignment based gene predictions (6,7).

The blast classifier allows homologs extracted from raw DNA contigs and 46-way gene models to be assigned to the appropriate sequence class. The multi-aligner highlights anomalies in the sequence collection that required additional curational focus (such as incorrect exon boundaries and regions temporarily out of reading phase). The synteny browser sorts out segmental and whole genome duplications of DHFR. However synteny dissipates fairly rapidly with

phylogenetic distance, so the much deeper conservation of intron position and phase becomes critical to refining gene models and retaining orthology.

Despite these improvements, the set of sequences below will still be imperfect because of initial errors in GenBank data (sequencing lab contamination, systemic errors in read technology, mis-assembled contigs, gaps in coverage, premature truncation of contigs, high levels of polymorphism, inadvertantly studied hybrids, endoparasites and commensals, taxonomic misclassification, a single animal sequenced unrepresentative of its species for this gene, lineage sorting, horizontal gene transfer, and inevitable data handling errors). Some clades, such as tunicates and nematodes, evolve so rapidly that their sequences seem implausible; even if valid, these DHFR are not informative to comparative genomics. Re-sequencing questionable gene models was beyond the scope of the project here.

The DHFR sequence set is not intended to be exhaustive; indeed close to 3,000 could be recovered from bacterial genome projects alone. Instead, the intensity of curational effort sought to evenly sample each of the phylogenetic divergence nodes separating human from its last common ancestor with bacteria, subject to data availability which for some nodes is limited by too few extant species. This allows inference of ancestral states by parsimony. For example when a given residue is conserved over two or more consecutive divergences, we take that residue value as ancestral over the internodal time period.

The topology of the phylogenetic tree is largely agreed-upon today, though controversy persists over some internal node arrangements. How residual issues are eventually resolved is not relevant here because the analysis of featured sites here is completely robust to commonly proposed tree alternatives. Single gene trees are not reliable; we do not infer a tree from DHFR data but instead subordinate it to the generally accepted tree derived from multi-gene concatenation. Divergence nodes on the phylogenetic tree are also reliably dated for the most part by relaxed molecular clock methods and the fossil record; only approximate dates are needed here to estimate summed branch length.

We assume a given residue can be conserved for orders of magnitude longer than a neutral residue only when it is maintained by selective pressure. Neutral sites in processed pseudogenes (including those of DHFR) decay over million-year time scales; the branch lengths supporting conserved residues here sum to billions. Despite clade-specific variations in the tempo and mode of evolution, such disproportionate persistence implies that mutational changes at these conserved sites are not fixed because they are maladaptive to DHFR functionality.

Representative alignment DHFR from 233 species is shown below (the complete sequence alignment and the full genus-species abbreviations are provided at ref. (8); http://genomewiki.ucsc.edu/index.php/DHFR_dihydrofolate) in modified fasta format (ie. headers structured as small flat-file databases). The fields are 6 letter genus-species acronym; full genus, species, common name, GenBank accession number; PDB structural accession; GenBank overall taxonomy; PubMed identifiers; and comment field. When no suitable GenBank accession was available, the sequence was derived from a blast or blat (7) query to a genome project. Genomic contig accession numbers are not provided because they are unstable to assembly iterations; to validate a given sequence, it is best to re-blast against the latest GenBank data set.

The fasta header lines are simple space-delimited databases showing first gene name, then genus, species, common name, accession number if not a simple genomic blat or whole genome alignment output, PubMed accession if specifically studied in a journal article, followed by an unstructured comment field. The headers and exons are reformatted into a spreadsheet by replacing spaces and paragraph returns with tabs.

The sequences are provided in phylogenetic order relative to human. For subclades (e.g. rodents), the sequences are phylogenetically ordered relative to the most intensively sequenced species (thus mouse). It is important that the alignment tool used be capable of retaining input order. Some sequences are incomplete and others are evolving rapidly, throwing off the natural order if the tool derives a gene tree and re-orders accordingly. Here the species tree is already fixed from broader considerations and the DHFR gene tree is clamped to it.

The representative (up to the first 90 residues) sequence alignment shown below covers the region of interest in this study. Human DHFR is at the very top and *E. coli* DHFR is near the bottom of the alignment list. The numbering at the top of the alignment accommodates sequences with insertions, such as DHFR_milFar or DHFR_natPel and so does not correspond to either human or *E. coli* numbering.

```
                    10        20        30        40        50        60        70        80        90
                     |         |         |         |         |         |         |         |         |
DHFR_homSap   --------MVGSLNCIVAVSQ---NMGIGKNGDLPWPPLR--NEFRYFQRMTTTS---SVEGKQNLVIMGKKTWFSIPEK-NRPLKGRIN
DHFR_panTro   --------MVGSLNCIVAVSQ---NMGIGKNGDLPWPPLR--NEFRYFQRMTTTS---SVEGKQNLVIMGKKTWFSIPEK-NRPLKGRIN
DHFR_gorGor   --------MVGSLNCIVAVSQ---NMGIGKNGVLPWPPLR--NEFRYFQRMTTTS---SVEGKQNLVIMGKKTWFSIPEK-NRPLKGRIN
DHFR_ponAbe   --------MVGSLNCIVAVSQ---NMGIGKNGDLPWPPLR--NEFRYFQRMTTTS---SVEGKQNLVIMGKKTWFSIPEK-NRPLKDRIN
DHFR_nomLeu   --------MVGSLNCIVAVSQ---NMGIGKNGDLPWPPLR--NEFRYFQRMTTTS---SVEGKQNLVIMGKKTWFSIPEK-NRPLKGRIN
DHFR_macMul   --------MVSSLNCIVAVSQ---NMGIGKNGDLPWPPLR--NEFRYFQRMTTTS---SVEGKQNLVIMGKKTWFSIPEK-NRPLKGRIN
DHFR_papAnu   --------MARSLNCIVAVSQ---NMGIGKNGDLPWPPLR--SEFRYFQRMTTTS---SVEGKQNLVIMGKKTWFSIPEK-NRPLKGRIN
DHFR_calJac   --------MVQLLNCIVAVSQ---NMGIGKNGELPWPPLR--NEFRYFQRMTTTS---SVEDKQNLVIMGRKTWFSIPEK-NRPLKDRIN
DHFR_saiBol   --------MVQLLNCIVAVSQ---NMGIGKNGDLPWPPLR--NEFRYFQRMTTTS---SVEGKQNLVIMGRKTWFSIPEK-NRPLKDRIN
DHFR_tarSyr   --------MVRPLYCIVAVSQ---NMGIGKDGTLPWPPLR--NEFRYFQRMTTMS---SVEGKQNLVIMGRKTWFSIPEK-NRPLKDRIN
DHFR_otogar   --------MVRPLNCIVAVSQ---NMGIGKNGDLPWPPLR--NEFNYFQRMTKTS---SVEGKQNLVIMGRKTWFSIPEK-NRPLKDRIN
DHF1_micMur   --------MVRTLNCIVAVSQ---NMGIGKNGDLPWPPLR--NEFRYFQKMTTTS---SVEGKQNLVIMGRKTWFSIPEK-SRPLKDRIN
DHFR_tupBel   --------MVRPLNCIVAVSQ---NMGIGKNGDLPWPPLR--NEFKYFQRMTTTS---SVEGKQNIVIMGRKTWFSIPEK-NRPLKDRIN
DHFR_musMus   --------MVRPLNCIVAVSQ---DMGIGKNGDLPWPPLR--NEWKYFQRMTTTS---SVEGKQNLVIMGRKTWFSIPEK-NRPLKDRIN
DHFR_ratNor   --------MVRPLNCIVAVSQ---NMGIGKNGDLPWPLLR--NEFKYFQRMTTTS---SVEGKQNLVIMGRKTWFSIPEK-NRPLKDRIN
DHFR_criGri   --------MVRPLNCIVAVSQ---NMGIGKNGDFPWPMLR--NEFKYFQRMTTTS---SVEGKQNLVIMGRKTWFSIPEK-NRPLKDRIN
DHFR_perMan   --------MVRPLNCIVAVSQ---NMGIGKNGDLPWPPLR--NEFKYFQRMTTTS---SVEGKQNLVIMGRKTWFSIPEK-NRPLKDRIN
DHFR_perPol   --------MVRPLNCIVAVSQ---NMGIGKNGDLPWPLLR--NEFKYFQRMTTTS---SVEGKQNLVIMGRKTWFSIPEK-NRPLKDRIN
DHFR_dipOrd   --------MGRSLNCIVAVSQ---NMGIGKNGDLPWPPLR--NEFKYFQRMTTTS---TVEGKQNLVIMGRKTWFSIPEK-NRPLKDRIN
DHFR_speTri   --------MVRPLNCIVAVSQ---NMGIGKNGDLPWPPLR--NEFKYFQRMMTTP---SVEGKQNLVIIGRKTWFSIPEK-NRPLNDRIN
DHFR_cavPor   --------MVRPLNCIVAVSQ---NMGIGKNGDLPWPPLR--NEFGYFQRMTTTP---SVEGKQNLVIMGRKTWFSLPEK-SRPLKDRIN
DHFR_oryCun   --------MVRPLNCIVAVSQ---NMGIGKNGDLPWPPLR--NEFRYFQRMTTTS---SVEGKQNLVIMGRKTWFSIPEK-NRPLKDRIN
DHFR_ochPri   --------MVRPLNCIVAVSQ---NMGIGRNGDLPWPPLR--NEFKYFQRMTTTS---TVEGKQNLVIMGRKTWFSIPEK-NRPLKDRIN
DHFR_felCat   --------MVRPLNCIVAVSQ---NMGIGKNGDLPWPPLR--NEFKYFQRMTTTS---SVEGKQNLVIMGRKTWFSIPEK-NRPLKDRIN
DHFR_canFam   --------MVRTLNCIVAVSQ---NMGIGRNGDLPWPPLR--NEFKYFQRMTTNS---SVEGKQNLVIMGRKTWFSIPEK-NRPLKDRIN
DHFR_vulVul   --------MVRTLNCIVAVSQ---NMGIGRNGTVPWPPLR--NEFKYFQRMTTNS---SVEGKQNLVIMGRKTWFSIPEK-NRPLKDRIN
DHFR_musPut   --------MVRTLNCIAAVSQ---NMGIGKNGDLPWPPLR--NEFKFFQRMTTTS---SVEGKQNLVIMGRKTWFSIPEK-NRPLKDRIN
DHFR_ailMel   --------MVRTLNCIAAVSQ---NMGIGKNGDLPWPPLR--NEFKFFQRMTTTS---SVEGKQNLVIMGRKTWFSIPEK-NRPLKDRIN
DHFR_equCab   --------MVRPLNCIVAVSQ---NMGIGKNGDLPWPPLR--NEFKYFQRMTTAS---SVEGKQNLVIMGRKTWFSIPEK-NRPLKDRIN
DHFR_vicPac   --------MVRPLDCIIAVSQ---NMDMGKNGDLPWPPLR--NEFKYFQRMTTAS---SAEGKQNLVIMGRKTWFSIPEK-NRPLKDRIN
DHFR_susScr   --------MVRPLNCIVAVSQ---NMGIGKNGDLPWPPLR--NEYKYFQRMTTTS---SVEGKQNLVIMGRKTWFSIPEK-NRPLKDRIN
DHFR_turTru   --------MVRPLNCIVAVSQ---NMGIGKNGDLPWPPLR--NEYRYFQRMTTTS---SVEGKQNLVIMGRKTWFSIPEK-NRPLKDRIN
DHFR_oviAri   --------MVRPLNCIVAVSQ---NMGIGKNGNLPWPPLR--NEFKYFQRMTTVS---SEEGKQNLVIMGRKTWFSIPEK-NRPLKDRIN
DHFR_capHir   --------MVRPLNCIVAVSQ---NMGIGKNGNLPWPPLR--NEFKYFQRMTTVS---SEEGKQNLVIMGRKTWFSIPEK-NRPLKDRIN
DHFR_bosTau   --------MVRPLNCIVAVSQ---NMGIGKNGNLPWPPLR--NEFQYFQRMTTVS---SVEGKQNLVIMGRKTWFSIPEK-NRPLKDRIN
DHFR_myoLuc   --------MVRTLNCIAAVSQ---NMGIGKNGDLPWPPLR--NEFKYFQRMTTTS---SVEGKQNLVIMGRKTWFSIPKK-NRPLKDRIN
```

```
                        10        20        30        40        50        60        70        80        90
DHFR_pteVam   --------MVLSLNCIVAVSQ---NMGIGKNGDLPWPPLR--NEFKYFQRMTTTS---SVEGKRNLVIMGRKTWFSIPKK-NRPLKDRIN
DHFR_eriEur   --------MVRPLNCMVAVSQ---NMGIGKNGELPWPPLR--NEFKHFRTMTSTP---SVEGKQNLVIMGRKTWFSIPEK-NRPLKDRIN
DHFR_sorAra   --------MVRPLNCIVAVSQ---NMGIGKNGELPWPPLR--NEFTYFRKMTTTS---SVEGKQNLVIMGRKTWFSIPEK-NRPLKDRIN
DHFR_loxAfr   --------MVRPLNCIVAVSQ---NMGIGKNGDLP-PPLR--NEFKYFQRMTTTS---SVEGKQNLVIMGRKTWFSIPEK-NRPLKDRIN
DHFR_proCap   --------MVRPLSCIVAVSQ---NMGIGKNGDLPWPPLR--NEFKHFRTMTSTP---SVEGKQNLVIMGRKTWFSIPEK-NRPLKDRIN
DHFR_dasNov   --------MVRPLSCYDAVYQ---NMGIGKNGDMPWPPLR--NEYKYFQKMTTVS---SVEGKQNLVIMGRKTWFSIPEK-NRPLKDRIN
DHFR_monDom   --------MVRQLNCIAAVSK---NMGIGKDGDLPWPLLR--NEFKYFQKMTTTP---SVEGKQNLVIMGKKTWFSIPEK-CRPLKDRIN
DHFR_macEug   --------MAPTINCIVAVAQ---NMGMGKNGDLPWPPLR--KEFQYFQKMTTTP---SVEGKQNLLIMGKKTWFSIPEK-NRPLKDRIN
DHFR_sarHar   --------MVRPINCIAAVSQ---NMGIGKNGDLPWPPLR--NEFNYFQKMTTTS---SVEGKQNLLIMGKKTWFSIPEK-HRPLKDRIN
DHFR_triVul   --------MVRTINCIVAVSQ---NMGIGKNGDLPWPPLR--NDFKHFQKMTTIP---SVEGKQNLLITGKKTWFSIPEK-SRPLKDRIN
DHFR_ornAna   --------MGPLLNCIVAVAK---NGGIGNKGDLPWPALR--NEFRYFQKMTTTP---TVKGKQNLVIMGKKTWFSIPEK-SRPLKDRIN
DHFR_tacAcu   --------MGRPLNCIAAVAK---NGGIGNKGDLPWPPLR--NEYRYFQKMTTTP---TVEGKENLVIMGKKTWFSIPEK-SRPLKDRIN
DHFR_galGal   --------MVRSLNSIVAVCQ---NMGIGKDGNLPWPPLR--NEYKYFQRMTSTS---HVEGKQNAVIMGKKTWFSIPEK-NRPLKDRIN
DHFR_lagLag   --------MVRSLNSIVAVCQ---NMGIGKDGNLPWPPLR--NEYKYFQRMTSTS---HVEGKQNALIMGKKTWFSIPEK-NRPLKDRIN
DHFR_anaPla   --------MVRSLNSIVAVCQ---NMGIGKDGNLPWPPLR--NEYKYFQRMTSTS---HVEGKQNALIMGKKTWFSIPEK-NRPLKDRIN
DHFR_taegut   --------MPRSLNSIVAISQ---NMGIGKDGRLPWPPLR--NDYKYFQRMTSAS---PVEGKKNAVIMGRKTWFSIPEK-NRPLKDRIN
DHFR_ficHyp   --------MVRSLNSIVAVSQ---NMGIGKDGRLPWPPLR--NEYKYFQRMTSTP---RVEGKQNALIMGKKTWFSIPEK-NRPLKDRIN
DHFR_melUnd   --------MVRSLNSIVAVCQ---NMGIGKDGSLPWPPFR--NEYKYFQRMTSTS---QVEGKQNALIMGKKTWFSIPEK-NRPLKDRIN
DHFR_allMis   --------MVSSLNAIAAVSQ---NMGIGKNGTLPWPPLR--NEFKYFQRMTTTP---SVPGKQNVLIMGKKTWFSIPEK-NRPLKDRIN
DHFR_croPor   --------MVSSLNAIAAVSQ---NMGIGKNGTLPWPPLR--NEFKYFQRMTTTP---SVPGKQNVLIMGKKTWFSIPEK-NRPLKDRIN
DHFR_chrPic   --------MVRPLNCIAAVCQ---NMGIGKNGDLPWPPLR--NEFKYFQRMTTTT---TVEGKQNVLIMGKRTWFSIPEK-NRPLKDRIN
DHFR_anoCar   --------MVLSLNSIAAVCQ---NMGIGKNGQLPWPPLR--NEFKYFQKMTMTP---TQEGKQNVVIMGKRTWFSIPEK-NRPLKDRIN
DHFR_pytMol   --------MVASLHSIVAVCN---NMGIGKDGKLPWPPLR--NEFKHFQKMTMTT---KEEGKQNVVIMGKKTWFSIPEK-HRPLKNRIN
DHFR_ambMex   --------MGRILNAIVAVCP---SMGIGKDGNLPWPILR--NEFKYFQRMTMTA---TQEGKQNVVIMGRKTWLSIPEK-NRPLKGRIN
DHFR_xenTro   -------MRNPFLHAVVAVCPP--NQGIGKEGSLPWPLLR--NEFKHFQRLTMTP---TVEDKKNVVIMGRKTWFSIPEK-NRPLKERIN
DHFR_xenLae   -------MRNQFLHAVVAVCPP--NQGIGKGGSLPWPLLR--NEFKHFQRLTMTP---TVEGKKNVVIMGRKTWFSIPEK-NRPLKERIN
DHFR_latCha   ------MGAARLLNSIVAVCP---NLGIGKDGNLPWHPKRLSNEFRYFQKMTTTP---TVEGKQNVVIMGRKTWFSIPEK-NRPLKGRVN
DHFR_lepOcu   --------MPRPINCIVAVCP---NMGIGHNGNLPWHPKRLSNEFKYFQKMTMTP---TLEGQQNAVIMGRKTWFSIPER-NRPLKNRIN
DHFR_gadMor   --------MSRVLNCIVAVCP---DAGIGYKGDLPWHPTRLNNEFKHFRRLTVTP---GAEDKQNVVIMGRKTWNSIPEK-NRPLNDRIN
DHFR_tetNig   --------MARVLNAIVAVCP---DLGIGRNGDLPWHPIRLDNEFKHFRKMTSTP---SVNGKQNVVIMGRKTWFSIPEK-HRPLANRIN
DHFR_hipHip   --------MSRILNGIVAVCP---DLGIGNRGNLPWHPVRLSNEFKHFRSMTATP---SEKGKQNVVIMGRKTWFSIPEK-NRPLNNRIN
DHFR_solSen   --------MSRVLNGIVAVCP---DMGIGMTGNLPWHPVRLSNEFKHFRTMTATS---SVKGKKNVVIMGRKTWFSIPEK-NRPLNNRIN
DHFR_oreNil   --------MPRVLNAIVAVCP---DRGIGNKGNLPWHPIRLSKEFAHFRKMTATP---SVKGKQNVVIMGKKTWYSIPEK-NRPLSNRIN
DHFR_dicLab   --------MSRILNGIVAVCP---DLGIGMNGNLPWHPVRLGNEFKHFRTMTATP---SVKGKQNVVIMGRKTWFSIPEK-NRPLNNRIN
DHFR_perFla   --------MSRILNGIVAVCP---DLGIGNNGNLPWHPVRLNNEFKHFRTMTATP---SVNGKQNVVIMGRKTWFSIPEK-NRPLNNRIN
DHFR_spaAur   --------MSRIVNGIVAVCP---DLGIGNNGNLPWHPVRLNREFKHFRTMTATP---SVKGKQNVVIMGRKTWFSIPDK-FRPLNNRIN
DHFR_gasAcu   --------MSRVLNGIVAVCP---DLGIGCHGNLPWHPLRLSNEFKHFRRMTATA---SVKDKQNVVIMGRKTWFSIPEK-NRPLNNRIN
DHFR_oryLat   --------MTRTLNGIVAVCP---DLGIGKGGNLPWHPLRLSKDFALFRKMTSTP---LVAGRQNVVIMGRKTWFSIPEK-NRPLQNRIN
```

```
                        10        20        30        40        50        60        70        80        90
DHFR_anoFim   --------MSRVLNAIVAVCP---DLGIGRNGDLPWHPVRLNNEFKHFRRMTSTP---SVEGKQNVVIMGRKTWFSIPEK-NRPLNNRIN
DHFR_esoLuc   --------MSRVLNCIVAVCP---NMGIGNKGNLPWHPKRLNNEFKYFQKMTMTP---FVEGKQNVDIMGRKTWFSIPER-NRPLKNRIN
DHFR_salSal   --------MSRVPNCIVAVCP---DVGIGNNGNLPWHPKRLNNEFKYFQKMTMTS---SVEGKQNVVIMGRKTWFSIPER-NRPLKNRIN
DHFR_oncMyk   --------MSRVLNCIVAVCP---DMGIGNNGNLPWHPKRLNNEFKYFQKMTMTS---SVEGKQNAVIMGRKTWFSIPER-NRPLKNRIN
DHFR_danRer   --------MSRILNCIVAVCP---DMGIGKNGNLPWHPIRLSNELKHFQKMTMTP---SDEGKKNVVIMGRKTWFSIPAA-HRPLKNRIN
DHFR_cteIde   --------MSRILNCIVAVCP---DMGIGRKGNLPWHPIRLSNEFKHFQKMTMTP---SVEGKKNVVIMGRKTWFSIPAQ-NRPLKNRIN
DHFR_cypCar   --------MSRILNSIVAVCP---DMGIGKNGNLPWHPIRLSNEFKHFQKMTMTP---LVEGKKNVVLMGRKTWFSIPAA-NRPLKNRIN
DHFR_ictPun   --------MGRVLNCIVAVCP---DMGIGRNGNLPWHPIRLSKEFKHFQKMTMTP---TVEGKKNVVIMGRKTWFSIPAQ-NRPLKNRIN
DHFR_leuEri   --------MTRLINSIVAVCP---NMGIGNNGNFPWHPIRLSKEFKHFQRMTSTP---SVEGKCNAVIMGRKTWHSIPEK-NRPLKNRIN
DHFR_squAca   --------MPRLVNCIVAVCP---NMGIGKDGNFPWHPIRLSKELKHFQKMTATP---SVEGKRNAVIMGRKTWFSIPEK-HRPLKDRLN
DHFR_eptBur   -------MAQNPVNVIAAVLP---NMGIGWKGGLPWHSKSLVKEMKHFTRLTSAA---A-EGKQNAVVMGRKTWESIPEK-FRPLRNRIN
DHFR_cioInt   -----MPAKDIQIHSVVACCN---NGGIGFKGRLPW---RLPKEMKYFKRITTGE---VEEGRRNAIIIGRKTWESIPKS-FKPLKDRIN
DHFR_cioSav   -----MPAKELKIHSIVACCN---NRGIGNKGRLPW---RLPKEMKHFTSLTTGD---VENGKRNAVVVGRKTWESIPKS-FKPLKNRVN
DHFR_oikDio   -------MNKSGWNMILAADI---KGGIGLRNDLPW---RIPQDLKHFQMLTKGT-----PNQESVVVMGRNTWQSIPEK-FRPLKGRVN
DHFR_braFlo   -------MKTKKLSLVVAAC---NNMGIGVDGKIPW---TLRGDMKFFSRLTSGT---EEAGKQNAVVMGRKTWFSIPDR-FRPLPKRLN
DHFR_sacKow   ---------MKKISLVAAAC---NNMGIGKNGDLPW---RLRKEMSFFTKVTSET---KEDGKQNAVVMGRKTWFSIPEK-YRPLAGRYN
DHFR_balCla   ---------MQKISPVAAAC---NSMGIGKNGNLPW---RLRKEMKYFTNVTSET---VEEEKQNAVIMGRKTWFSIPEK-YRPLNNRFN
DHFR_strPur   -------MAEKKLNLIAAACTSKGKMGIGINGNLPW---RLRQEMAYFERLTKTA---QMEGMKNAVIMGRKTWDSIPEK-FRPLKDRVN
DHFR_parLiv   -------MADKRLNLIAAACTSNGKMGIGINGNLPW---RLRQEMAYFERLTKTS---QMEGMKNAVVMGRKTWFSIPEK-FRPLKDRVN
DHFR_lytVar   -------MAEKKLNLIAAACTSSGKMGIGINGNLPW---RLRQEMAYFERLTKTP---QLEGMKNAVIMGRKTWFSIPEK-FRPLKDRIN
DHFR_patPec   ------MAGQKQCNLIVAACKCKDSLGIGINGTIPW---KLRTDMKFFSTQTSTT---AENDKKNAVIMGRKTWLSIPDK-FRPLPNRVN
DHFR_droMel   ---------MLRFNLIVAVCE---NFGIGIRGDLPW---RIKSELKYFSRTTKRT---SDPTKQNAVVMGRKTYFGVPES-KRPLPDRLN
DHFR_gloMor   ---------MLKFNLIVAVSK---NFGIGLKGGLPW---ELKSELRYFSELTKRV---FDSTKRNVVIMGRKTYFGIPLN-NRPLRNRLN
DHFR_haeIrr   ---------MLKFSLIVAVCE---NFGIGIKGDLPW---LLKSELKYFSTTTKRV---HDPSKRNVVIMGRKTYFGVPES-KRPLQQRLN
DHFR_sarCra   ---------MLKFSLIVAVCE---NFGIGIKGDLPW---KIKSELKYFSSTTKRV---RDPSKRNVAIMGRKTYFGIPES-KRPLPERLN
DHFR_culQui   ---------MKKFSLIVAVCS---NGGIGIKGDLPW---RLRSELRHFARMTKRV---ADPGKRNAIIMGRKTYFGIPEG-RRPLPDRLN
DHFR_anoGam   --------MSKKFSCIVAVCE---NRGIGINGDLPW---KLKQELKYFSHTTKKV---NDADKRNAVIMGRKTYFGVPES-KRPLPERLN
DHFR_aedAlb   ---------MKKFSLIVAVCA---NGGIGIKGDLPW---RLRQELKYFSRMTKKI---QDSGKRNAIIMGRKTYFGVPES-KRPLPERLN
DHFR_aedAeg   ---------MKKFSLIVAVCA---NGGIGIKGDLPW---RLRQELKYFSRMTKKI---TDTSKRNAIIMGRKTYFGVPES-KGPLPERLN
DHFR_armSub   ---------MKKFSLIVAVCA---NGGIGIKGDLPW---RLKQELKYFSRMTKKK---QDTSKRNAIIMGRKTYFGVPES-KRPLPERLN
DHFR_danPle   -------MSKVKLNLIAAACE---NMGIGSNGSLPW---RLKKEMEYFTTMTTKV---KDASKINAVIMGRRTWDCIPNK-NRPLENRLN
DHFR_bomMor   -------MSRTQLNLIAAACE---NMGIGINGTLPW---KLKKEMAYFTTMTTSV---SDKKKVNAVIMGRRTWDCIPIK-YRPLSNRIN
DHFR_helVir   -------MSQVKLNLIAAACD---NMGIGVNGALPW---RLKKEMAYFTTMTSKV---SEPTKVNAVIMGRRTWDCIPDK-YRPLQDRVN
DHFR_triCas   --------MVIKFDLIAAACE---NMGIGKNNDLPW---RLKSELAFFSQMTTQT---SDESKKNVVLMGRKTWDSIPPK-FKPLHQRFN
DHFR_denPon   --------MALKLNLIVAASE---NWGIGKNGTLPW---KLKKEMAFFRTMTSAT---EDKSKKNVVIMGRRTWESIPAQ-FKPLPNRIN
DHFR_apiMel   ----------MNLHLIAATCE---GMGIGIKGTLPW---KLKSELAFFTYMTTNT---KNPNKRNVVLMGRRTWESIPKE-NRPLKNRIN
DHFR_bomImp   ----------MNLNLIAAACE---GMGIGVKGTLPW---RLKSEMAFFTSMTTNT---KDPNKKNVVLMGRRTWECIPKK-YKPLKNRIN
DHFR_euqCor   --------MSLNLNIIAAVCE---GMGIGVKGSLPW---KLKSEMAFFTSMTTQT---KDP-KKNVVIMGRRTWECIPTK-YRPLKDRIN
```

```
                        10        20        30        40        50        60        70        80        90
DHFR_nasVit   --------MQVKLKLIAAACE---NMGIGVNGDLPW---RLRKEMDFFTKMTSTT---KDENKKNVVLMGRRTWESIPKK-FKPLSNRIN
DHFR_copFlo   --------MQVKLKLIAAACD---NMGIGINGDLPW---RLRKEMDFFTKMTSTT---KDPNKQNVVLMGRKTWESIPKK-YKPLANRIN
DHFR_attCep   --------MPPKLELIAAACE---NMGIGVNGNLPW---RLKAEMAYFTRMTTNT---RNKNKRNVVLMGRRTWECIPEK-YRPLKDRIN
DHFR_camFlo   --------MPPKLELIAAACE---NMGIGINGGLPW---RLKTEMAFFTRMTTDT---KDKDKKNIVLMGRRTWDCIPEK-YKPLENRIN
DHFR_harSal   --------MLPKLELIAAACE---NMGIGIKGDLPW---RLKTEMEYFTRMTTET---KDKNKKNVVLMGRRTWDCLPET-YRPLRNRIN
DHFR_linHum   --------MQLKLELIAAACE---NMGIGNNGALPW---RLKTEMEYFTRMTIDT---KDKNKKNVVLMGRRTWDSIPIK-YRPLNDRIN
DHFR_pogBar   --------MSHKLELIAAACE---NMGIGINGDLPW---RLKTEMAFFTRMTTET---NHTNKKNVVLMGRRTWECIPDK-YRPLKDRIN
DHFR_solInv   --------MPPKLELIAAACE---NMGIGVNGDLPW---RLKTEMAFFTRMTTDT---KQNNKKNVVLMGRRTWECIPKK-YRPLKDRIN
DHFR_bemTab   -------MSSLRLSIIVAMTA---DNGIGFKNSLPW---SLPNELRNFAKLTKNC---KDKSKQNAVVMGRNTWESIPAQ-HRPLKDRLN
DHFR_acyPis   ----------MVYSVIAAVSK---NGGIGYKGNLPW---KIKKEMEYFNLMTTRV---NLKGVQNAVIMGRCTWQSIPDK-YRPLKGRIN
DHFR_blaGer   --------MNLKLXLIVAVSE---NMGIGLNGDLPW---RLRKELAHFSRLTKRT---IDSNKQNAVIMGRRTWESIPIK-NRPLPERLN
DHFR_pedHum   --------MVNNFNLIAAVCS---NNGIGYKGNLPW---NLRKELQYFNRMTKDV---KNPEKKNAVIMGRKTWDSLPHN-WKPLPGRYN
DHFR_onyArc   --------MGPKLYIIAAACE---NMGIGYKNDLPW---KLRNEMAYFNRMTQIV---ENPELKNAVIMGRRTWDSIPEK-FRPLRNRLN
DHFRcalCle    ----MTSLKCVKINVIAAACK--SSRGIGKDNDLPWK---LPTDMKYFREKTAST---TEI--RVGVIMGRRTWESVPPK-FRPFKNRFN
DHFR_lepSal   -----MSKSIPTLHVIAAACR--PTNGIGKNNTLPWN---LPTDLKYFKTTTSST---PDS--KVAVIMGRRTWESIPSK-PRPLKNRIN
DHFR_litVan   --------MSVRLNIIVAACE---NHGIGINGELPW---KLREEMKYFSRMTKAT---KSLEKQNAVMMGRKTWESIPAK-FRPLPGRLN
DHFR_celPug   --------MGPRLNIIVAVAE---NHGIGKGGELPW---RLREEMKHFSRMTKRL---DSPATQNVVLMGRKTWESIPLK-FRPLPGRLN
DHFR_dapPul   ----------MKLNLIVATAS---NMGIGFQGTIPW---RLKKDMALFAKLTKNT---NDTNKQNAVVMGRKTWESIPEK-NRPLSNRLN
DHFR_ixoSca   --MCPGVQESLQYFAIAAMCH---NRGIGVLNTLPW---RLKKEMAYFSRITSQ----AAEGKTNAVVMGRNTWDSIPPK-YKPLPGRVN
DHFR_ambMac   --MCPSHKSVVSCFAIAAMCR---NRGIGFKNALPW---RLKKEMAFFKRMTSE----AAEGKQNAVVMGRNTWESIPPK-FRPLNNRIN
DHFR_perSed   --------MHRYLNLVVAVCN---NLGIGIGGELPW---RLRGDMKFFSKLTSET---KDSEKTNVVIMGRKTWASIPEK-FRPLPKRTN
DHFR_milTar   ----------------VAACR---NGGIGIRGDLPW---TLRGDMRFFTKITSQT---KDPQKKNAVIMGRKTWFSIPER-FRPLSKRIN
DHFR_triSpi   ---------MAQVNIIVAICE---KYGIGKKNSLPW---HLSKEMQHFKKMTTSV---SDPNKINAVIMGRNTWYSIPEK-YRPLSGRFN
DHFR_xipInd   ------------MNLIVAACD---NMGIGHGNELPWP--KLPNESKHFLKLTAGT---RDPGKQNAVIMGRKTWEIIPVE-HRPLKHRLN
DHFR_caeEle   ---------MRKMNLIVAMDA---EGGIGKNGVLPW---RIKKDMQYFASVTKNV---SDQSKRNAVLMGRKCWESIPVT-RRPLAGRLN
DHFR_melInc   ------------MNIIAAVDE---NFGIGKNNSLPW---RLPKEYKHFINLTTTT---KNPNKINAVLMGRKCWESIPEK-YRPLKNRLN
DHFR_ascSuu   -----MVSPKLPINIIVAMDS---RGGIGKNGALPW---HIPEDLKYFQTMTTKT---IDPTKQNAIVMGRKVWESLPAK-WRPLKNRLN
DHFR_schMed   ---------MKRLNLIVAACE---NQGIGKNGKLPW---NLKNEMIYFNNITTSV---SDSNKQNVVIMGRITWESIPNK-FRPMPKRIN
DHFR_schMan   ----------MRLNVVVAVSE---NWGIGKGGGLPW---KIKKDMEFFKTVTTK----AHPGLKNAVVMGRVTWESIPES-FKPLKDRIN
DHFR_taeSol   -------MGLRRLNVIAAVAK---NGGIGKENKLPW---HIREDMAFFSRITST----AQEGKKNAVVLGRRTWLSFPPK-FRPLPNRVN
DHFR_aplCal   -------MTSTKLNIVVAVCT---NMGIGIEGRLPW---RLKQDMAFFKQLTVET---QDEQKKNMVIMGKKTWMSIPTK-FRPLQDRVN
DHFR_lotGig   -------MCTKRLNIVVAACK---NNGIGVNGSIPW---RLKKDMAMFRHITSDT---VDESKQNAVIMGRKTWMSIPDK-FRPLKNRVN
DHFR_phyAcu   ---------MPKLNIIVAACN---NNGIGIEGRLPW---RLKSDMSFFKQITLRT---QDAEKKNAVIMGKNTWFSIPSK-FRPLVGRIN
DHFR_pinMax   -------MSKCKLNLVVAACN---NRGIGIDGQLPW---RLRKDMDFFKKITTET---NDPEKRNAVIMGRKTWFSIPEK-FRPLSKRIN
DHFR_mytCal   -----MSKTKSTLNLVVAACN---SRGIGINGKLPW---RIRKDMDFFKKITMET---KNPDKKNVVIMGRKTWFSIPEK-FRPLPKRIN
DHFR_dreRos   -------MSMPRLNLVVAICN---NNGIGINGSLPW---KLRKDMDFFKTITMTT---TDPEKQNAVIMGRKTWQSIPEN-FRPLRNRIN
DHFR_alvPom   --------MPPTLNLIVAMCN---NNGIGIQGKLPW---KIKGDMAFFRKMTTET---KNPAKMNAVIMGRKTWFSIPEK-NRPLNKRMN
DHFR_helRob   --------MQPKLQIVVALCV--KNRGIGLNNSIPW---KLPGDMTFFRKLTSET---SILGSKNAILMGRKTWDSIPSN-LKPLKNRLN
```

```
                        10        20        30        40        50        60        70        80        90
DHFR_nemVec   -----MAAVGVRFSCIVAMDL---KRGIGKNNDLPW---KIKGDMKFFTEKTSEV---KTEGKQNAVIMGRKVWESIPEK-FRPLKGRLN
DHFR_acrMil   ----------------VAVAD---NGGIGKENRLPW---HLKGDMKFFSHLTSTV---ATEGKQNAVIMGRNTWESIPKK-YRPLPRRLN
DHFR_hydMag   ------MVQLRKVHCIAAIDS---EWGIGLKGKLPW---NLPREYKFFQNITTKV---ISEGKQNAVIMGKNTWFSIPQQ-HRPLKNRLN
DHFR_mneLei   --------MKLRFNIIVATDL---EGGIGKNNNLPW---RCRKEFRHFTDTTSKL---QN-GKQGAAVMGRKTWESIPEK-FRPLKNRHN
DHFR_triAdh   --------MQKNFRMIAAMTR---DRGIGYKNDLPW---KLRNEYKYFARLTTST---KDSAKKNAVLMGRNTWVSIPEK-NRPLRNRIN
DHFR_subDom   ------------MYIMVAKEI---KSGIGNKGKIPWP--RINTDMNFLETMTKKT---RNKTKQNAAIMGRLTWESIPAK-YRPLPERFN
DHFR_monBre   ----------MRVALVVAATK---AGGIGHQGQLPW---HLKADMQHFRDITTKT---EQPAAINAVIMGRKTFESIPDK-FRPLHGRLN
DHFR_canAlb   ---------MLKPNVAIIVAALKPALGIGYKGKMPW---RLRKEIRYFKDVTTRTTKPNT---RNAVIMGRKTWESIPQK-FRPLPDRLN
DHFR_canGla   ---------MSKVPVVGIVAALLPEMGIGFQGNLPW---RLAKEMKYFREVTTLTNDNSK---QNVVIMGRKTWESIPQK-FRPLPKRIN
DHFR_pneCar   ---------MNQQKSLTLIVALTTSYGIGRSNSLPW---KLKKEISYFKRVTSFVPTFDSFESMNVVLMGRKTWESIPLQ-FRPLKGRIN
DHFR_schSti   -------MSEIKPTIAIIVAALKPDLAIGFQGKMPW---RLRKEIRYFKDVTTKTSDPSK---INAVVMGRKTWDSIPAR-FRPLPDRIN
DHFR_spaPas   -----MTMTEHKPTISIVVAALKPKYGIGYQGKMPW---RLRKEIQYFKNVTSKTTKSNG---INAVIMGRKTWESIPKK-FRPLPDRIN
DHFR_lodElo   -------MTSEKPTISIIVAALKPSYGIGNKGKLPW---RLRKEMAYFKRVTLRTTHTQ-----NAVIMGRKTWESIPVK-FRPLPDRLN
DHFR_debHan   MTAAN---GKVTPPIAMIVAALLPELGIGIKGKMPW---RLRQEIKYFKNVTTTTKDPNG---INAVIMGRKTWESIPTK-FRPLPGRLN
DHFR_meyGui   ---------MRQPPIALIVAALLPDLGIGFGGALPW---RLKQEIKYFRDVTSNAPDGS----INAVIMGRRTWESIPPR-FRPLPNRIN
DHFR_milFar   MTSSGHSLGRGGIPIVMIVAALVPEMGIGLKGKMPW---RLKNEMAYFRKVTTTTTDNEA---INAVIMGRKTWQSIPEK-FRPLPKRLN
DHFR_claLus   ------MIKSLKPVTSVIVAALSPKYGIGAQGKLPW---RLKQEMKYFKDVTSAARAGS----INAVVMGRKTWESIPKK-FRPLPNRLN
DHFR_komPas   --------MSRPFKVNIIVAALIPGHGIGLKGKLPW---SLKNEMKYFRLLTTNTISPTS---KNVVIMGRNTWESIPSR-FRPLQNRLN
DHFR_ogaPar   -----------MTKVSLIVAALVPKYGIGYKGQLPW---ALKEEMRYFRRVTTQTADKNK---KNAVVMGRKTWESIPER-FRPLKGRVN
DHFR_rhiDel   ----------MTQRKLALIAAATEELGIGRKGDLPW---HLPRDLRYFRDVTTKIPKGNS--AQNVVIMGRVTWESIPPK-YRPLDNRFN
DHFR_encHel   -----------MFALIVAFAK---NKGIGKNNRLPWP--PLRTDMSWFRTLSQSVPLITSPSSSNAVVMGRKTWDSIPPR-FQPLPNRTN
DHFR_encRom   -----------MFALIVALAR---NKGIGRANRLPWSR-PLRTDMSWFRTLSQSIPLITLPSPSNAVVMGRKTWESIPPK-FRPLPNRMN
DHFR_encCun   -----------MLALVVALAS---HRGIGNANALPWPR-PLAADMAWFRTLSQSIPLISAPSASNAVVMGRRTWDSIPSR-FRPLANRIN
DHFR_encInt   -----------MIALIVALAA---NQGIGRHGKLPWKH-PLKTDMAWFRILSQSVPILSSPSKSNVVVMGRKTWHSIPSR-FRPLQNRIN
DHFR_harCan   ----------MNGFALVVAVT--KTMGIGNKGGLPWS--SIRGDMNFFKKITSET---KDAAKQNAVVMGRKTYVGIPAK-FRPLSNRVN
DHFR_polPal   ---------MVKISFVVAVSL---NNVIGKDGGIPW---RLKGDLAFFSKTTTGHPIVMGRKTFDSIPLPNRHSIVITSS-GKLYN-RDG
DHFR_dicDis   ----------MKVSIIVAVSK---NNVIGTAGDIPW---ELPKDLKHFRDTTHGFPCIIGRISLEAFLLPNRFNIIVSSK-KEFINIREK
DHFR_araTha   ----MTSKPQSTYQVVVAATK---EMGIGKDGKLP---WNLPTDLKFFKDLTLST---SDSAKKNAVVMGRKTWESIPKK-YRPLSGRLN
DHFR_popTri   ----QQSDPRKTYQVVVAATK---DMGIGKDGKLP---WKLPSDLKFFKDLTLTT---MDSGKKNAVIMGRKTWESIPLE-HRPLPGRLN
DHFR_phyPat   ----THSEKLRGFQIVVAATR---ERGIGKQGHLP---WKLPTDMKFFKTVTSVT---TSSSKKNAVIMGRHTWESIPEK-FRPLPGRLN
DHFR_selMoe   ------AATLRAFQVVVAATR---DLGIGKEGKLP---WSVPSDMAFFKRITSEA---ADARKKNAVVMGRRTWESIPAK-FRPLPGRVN
DHFR_ostTau   -------MEMRAFQVVVAATV---DGGIGKDNGLP---WKLTKDMGHFKKLTSET---RDANAVNAVVMGRKTWESIPAK-FRPLPGRLN
DHFR_micPus   ----MEASTVRPFQVVVAATK--GEMGIGYQGGLP---WNLPKDMAYFKAVTSQV---NEPGTKNAVVMGRKTWESIPAK-FRPLPGRVN
DHFR_chlVar   ----TPEQARRTIQIVVAATK---QWGIGKGGSLP---WSLPGDMKYFRELTSRT---ADPAKQNVVIMGRKTWESIPAK-FRPLAGRIN
DHFR_volCar   ------MTSHKGFQLVVAATP---SLGIGKNGTLP--GWKLPGDMAYFRELTSRT---REPSCQNVAIMGRKTWESIPNK-FRPLKGRIN
DHFR_chlRei   -----MDTSRKGFQLVIAATP---SLGIGKNGKLP--DWQLPGDMAYFKELTSRT---RDGAHTNAVIMGRKTWESIPAK-FRPLKGRLN
DHFR_phyInf   -------MTEQSIRVVVAALE--TTGGIGLRQHIP---WRLPSDMKHFRALTTAS---PDSSVQHAVIMGRKTWESLPAK-VRPMPKRYN
DHFR_albLai   --------MVRTISIVVAATE--DCWGIGWRQSIP---WKLAKDLKHFRDLTTRT---NSQNMRNAVIMGRKTWESLPPS-YQPLPNRYN
```

```
                    10        20        30        40        50        60        70        80        90
DHFR_blaHom   -------MSIIRFSIVAAMTT---KRGIGLNGGLP---WRIKQDMKFFVDLTTTT---TDSEKQNAVIIGKNTYFSFPEK-FRPLKNRIN
DHFR_aurAno   ----APTNDLPQLTCIVAAVA--GSYGIGKDGTLP---WKLAGDMKYFKKVTSA----APEGKTNACVMGRKTWLSIPPK-FRPLGGRKN
DHFR_phaTri   ---------------VVAAAA--SHRGIGYQGSLP---WRLPGDLRHFQAVTTQP---PETGGTNAVIMGRKTWDSIPDR-FRPLPGRIN
DHFR_thaPse   ---------------VVAAAA--GSRGIGHQGKLP---WRLPGDMNHFKKVTTTP---PSPGLTNAVIMGRKTWDSIPSK-FRPLDGRVN
DHFR_perMar   -------MALPQLSVIVAHTC---KWGIGKDGQLP--WKSLPEDMKRFKKITTGG---CNDNVKNVCIMGRKTWESIPER-FRPLRDRIN
DHFR_tetThe   -------MKTRHFDIVLAQTL--KKQGIGYKNSLP---WRLPNELKNFKKITTET----NKGLQNAVIMGKNTWEALPKK-QQPLKDRLN
DHFR_cryHom   -------MSEKNVSIVVAASVL--SSGIGINGQLPW---SISEDLKFFSKITNNK---CDSNKKNALIMGRKTWDSIGR---RPLKNRII
DHFR_tryCru   ----GTRLALRAFSLVVAVDE---HGGIGDGRSIP---WNVPEDMKFFRDLTTKL---PSPAKRNAVVMGRKTWDSIPPK-FRPLPGRLN
DHFR_leiTro   ----FAFPSLRAFSIVVALDK---QHGIGDGESIP---WRVPEDMAFFKDQTTLL---PTEKKRNAVVMGRKTWESVPVK-FRPLKGRLN
DHFR_criFas   ----FAFPSLRAFSIVVAADQ---QHGIGDGETIP---WTVPETLAFFKDQTTLL---PTEKKRNAVVMGRKTWE-VPLK-FRPLKGRLN
DHFR_ectSil   ----ASPSTSRSFVVIVAATA--GSLGIGKNGALP---WRLAADMAYFKRCTSTP---TTTDKINAVIMGRKTWQSIPER-FRPLAGRRN
DHFR_toxGon   --------MQKPVCLVVAMTP---KRGIGINNGLPWP--HLTTDFKHFSRVTKTT---SVGKRFNAVVMGRKTWESMPRK-FRPLVDRLN
DHFR_plaFal   ------KVESKNEGKKNEVFNNYTFRGLGNKGVLPWKCNSL--DMKYFCAVTTYV---NSKKLQNVVVMGRTNWESIPKK-FKPLSNRIN
DHFR_babBov   ---STVYEGCGGLTIYVAIAL---NRVIGHQNQIPWP--HIIHDFRFLRNGTTYI---KNPNIQNVVIFGRKTYESIPKA-SLPLKNRIN
DHFR_thePar   ---AEDYSGLPFVKLLVAITP---ENGIGISNGLPWP--HIKRDFLFMFRATTYV---KHPGAQNVIIIGRKTYDSLPEG-TFPLKNRIN
DHFR_naeGru   ------MMQSTSPPMEAVVAVLLNSNGIGLNGNLPWLQDGLTEDMKHFVSVTCI-----NDQLMNAVIMGRKTWESIPSK-FKPLSKNRH
DHFR_escCol   -----------MISLIAALAV---DRVIGMENAMPW---NLPADLAWFKRNTLN----------KPVIMGRHTWESI----GRPLPGRKN
DHFR_breLat   -----------MLSSIFAMGQ---NRVIGRDNQLPW---RLPEDLKYFRRITTG----------HAIIMGRKTYESI----GKPLPNRRN
DHFR_marPos   -----------MLSLIVAMST---NRTIGINNSLPW---HLPNDLKYFKQATMG----------KPIVMGRKTFESI----GKPLPGRRN
DHFR_salEnt   -----MNPESVRIYLVAAMGA---NRVIGNGPDIPW---KIPGEQKIFRRLTES----------KVVVMGRKTFESI----GKPLPNRHT
DHFR_klePne   -----MNPELVRIYLVAAMGA---NRVIGNGPDIPW---KIPGEQKIFRRLTEG----------KVVVMGRKTFESI----GKPLPNRRT
DHFR_halNea   -----MNQQAVRIYLVAAMGS---NRVIGNGSNIPW---RIPGEQKIFRTLTEG----------KVVVMGRKTFESI----GKPLPNRHT
DHFR_pseAla   -----------MIALIAACDR---HRLIGDHGRIPW---RIPGEQARFQKLTTG----------HVVIMGRQTYAEI----GRPLPRRQT
DHFR_macCas   ----------MIISLIAAISS---NYVIGKDKDIPW---KIPGEQVRFKDLTMG----------KSVIMGRKTFESI----GQPLPNRKT
DHFR_cloCel   -----------MISLIVAVAK---NNVIGNNGIIPW---KIKGEQKRFKELTIG----------KTIIMGRKSFEEI----GKPLPNRKT
DHFR_geoUra   ----------MVISLIAAMAE---NRVIGRNNAIPW---DIPADRKRFRALTLG----------HPVIMGRKTFESL----AGPLPGRKN
DHFR_oxaFor   ---------MMTISHVVAMAE---NRVIGKDGQMPW---HIPGEQKIFRELTVG----------KALILGRKTHESI----GRVLPDRIT
DHFR_nocSpp   -----MTPGGKRVVLVAAVAR---NGVIGDGPDIPW---QLPGEQRLFKGLTWG----------HILVMGRATYDSI----GRPLPGRTT
DHFR_halPau   ---------MTRLVCIAAVAE---NGVIGRDGDMPW---HYSADLRHFKETTMG----------HPVVMGRTTYESIAGQLDGPLPGRTN
DHFR_natPel   MEAAAALEPARELVGIVAVAE---NGVIGRDGDMPW---HVPADLQHFKETTMG----------HPVIMGRVTYEGILETLGEPLPGRTT
DHFR_halXan   AESALAAETDRELVGIVAVAD---NGVIGKDGDMPW---HIPADLQHFKETTMD----------CPVIMGRVTYEGILEALGEPLPGRTT
DHFR_natMag   ADTDPVVEPDHELVGIVAVAD---NGVIGKDGDMPW---HIPEDLQHFKETTMA----------HPVIMGRVTYESIVDALGEPLPGRTT
DHFR_halLac   SALEAAEATDRELVGIVAVAD---NGVIGADGEMPW---HLPADLAHFKETTMD----------HPVIMGRVTYEGILETLGDPLPGRTT
DHFR_halWal   HDVDIDIDIDIDIVLIAAVAA---NDIIGRDGEMPW---HIPADLQQFKRRTMG----------HPVILGRRTYEAIIGALGEPFPGRTS
DHFR_halBor   RDTTDANAESIEYVLVAAVAE---NGVIGRDGRMPW---HFSEDMAHFKQTTMG----------HPVILGRKTYENIVDAIGEPFPGRMS
DHFR_halMed   --MSEHTDDSVRFVLVAAVAE---NRVIGRDGDMPW---HLPEDLKHFKATTMG----------HPVVMGRTTYESIARQIDGPLPGRRN
DHFR_natPha   ----------MKLVLIAAVAE---NGVIGTDGEMPW---HYPEDLKRFKETTMG----------HPVIMGRTTYESITGQLGGPLPGRTN
Consensus                    ivA      n gIG  g lPw    l  #   F   T          k na!!MGRkTwes!p k  rPL  R n
```

1.3. PCE analysis. The criteria for identifying PCEs from sequence alignment are briefly summarized in the main text. The systematic way to identify PCEs is through a difference alignment as illustrated below. As mentioned in the main text, phylogenetically coherent events (PCEs) are defined as changes at a long-conserved amino acid position at which both the newly 'altered residue' and the unaltered 'ancestral residue' remain invariant over subsequent geological time in all (studied) speciation lineages. Such events have only become identifiable in the large-scale genomic era (9) because of the large number of species required to establish pre- and post-invariance with adequate confidence (i.e. with summed branch lengthyears supporting the event orders of magnitude longer than neutral drift decay) A strongly supported PCE seems to require positive (possibly differently driven) Darwinian selection to be operative at the PCE site in both descendent lineages.

Identification of PCEs involves ordering the sequences according to initial phylogenetic bias (as shown below, taking human first leads to the ordering of primates, rodents, laurasiatheres; taking opossum first would lead to another order), using the alignment tool, Mulalin (4), which retains input order. Unsupported idiosyncratic features from low-coverage species are interpreted as distracting sequence error and corrected in the alignment (through retained in the fasta set). It was not a given that DHFR would have any PCEs. Thus the PCEs that it has are worthy of special experimental attention. While we have sufficient taxon sampling density here to be confident in PCE identification, DHFR remains un-sequenced in many thousands of species. Thus by claiming certain events as PCEs, we are in effect predicting that, as additional DHFR are sequenced they will continue to strengthen support for our PCE classification. However limitations on extant species (and sequencing of paleo DNA) mean that relatively little additional branch length is available for some nodes (e. g. coelocanth).

Below the DHFR sequences (same sequences as in previous pages) are re-oriented to human DHFR (homSap), with dots representing same residue at the same site in other species. The numbering system is consistent with the aligned sequences in previous pages. For example, in the region of interest, DHFR_gorGor has the same amino acid sequence as DHFR_homSap except at position 33 (V in DHFR_gorGor instead of D in DHFR_homSap). Scanning through the phylogenetically ordered alignment of 233 species, a PCE can be picked out visually as a <u>column of dots</u> over a <u>residual column of a fixed letter</u>. PCEs not relative to human can also be found as <u>columns of a fixed letter</u> over a <u>residual column of a different fixed letter</u>. Gaps can be treated as an amino acid for this purpose: <u>columns of dots</u> or <u>a fixed letter</u> over <u>columns of gaps</u> (and vice versa) are taken as PCEs. It is important that the columns of dots or fixed letters stay constant over a long period (i.e. in many species) both before and after the divergence event. In the alignment below, dots represent differences relative to human, dashes represent gaps.

The three PCEs studied here are highlighted in blue (PWPPLRNEF region in human) and yellow (PEKN).

<u>PWPP:</u>
As shown below, the evolutionary sequence of events in vertebrates and earlier deuterostomes shows an interesting deletional/insertional history at the PWPPLRNEF position.

1) PWPPLR--NEF turned into PWHPKRLSNEF as illustrated by a column of dots at the end of PWP (position 37) shifting into a column of H residue around DHFR_latCha. Also, there are columns of dots at positions 41 and 42, which turned into L and S/N respectively.

2) PWHPKRLSNEF turned into PWRLP KEMKYFKR around DHFR_cioInt. This is illustrated by the appearance of a gap, which persisted from DHFR_cioInt all the way to the end of the alignment (DHFR_natPha).

We interpret these changes as somehow advantageous to the altered clade because each has been fixed for hundreds of millions of years of summed branch length, inconsistent with functionally deleterious, or even neutral, changes. Note the deletions are occurring at the end of exon 1, not in the second exon which begins NEF.

<u>L28 in E. coli to F32 in human:</u>

This PCE is shown as a column of dots (at position 45 in the alignment below), which turned into a column of mostly L or M. The wobbling between L and M as well as the rarity of L→ F mutation are explained in the next section.

<u>PEKN:</u>

The P residue at position 78 in the alignment below stays as a column of dots (except a handful of rare cases, which are not significant) all the way back to DHFR_escCol. After that the sequence length stays constant from DHFR_escCol to DHFR_natPha. Positions 79 and 80 vary throughout most analyzed species but both turn into two gaps at DHFR_escCol too. Again, the chance for G or P residue to change is discussed in the next section.

The PCE analysis performed here is restricted to deuterostomes. It should also be noted that there might be other PCEs and we did not look for all possible PCEs. We mainly focused on PCEs around the enzyme active site to increase the chance of producing experimentally detectable changes in enzyme activity due to PCE-guided mutagenesis studies.

```
                         10        20        30        40        50        60        70        80        90
                          |         |         |         |         |         |         |         |         |
DHFR_homSap   --------MVGSLNCIVAVSQ---NMGIGKNGDLPWPPLR--NEFRYFQRMTTTS---SVEGKQNLVIMGKKTWFSIPEK-NRPLKGRIN
DHFR_panTro   ......................................................................................
DHFR_gorGor   ................................V.....................................................
DHFR_ponAbe   ......................................................................................D...
DHFR_nomLeu   ......................................................................................
DHFR_macMul   .........S............................................................................
DHFR_papAnu   .........AR.............................S............................................
DHFR_calJac   ..........QL....................E.............................D........R...............D...
DHFR_saiBol   ..........QL.....................................................R...............D...
DHFR_tarSyr   ..........RP.Y................D.T.....................M.................R...............D...
DHFR_otogar   ..........RP.................................N......K..................R...............D...
DHF1_micMur   ..........RT.....................................K.....................R..........S....D...
DHFR_tupBel   ..........RP.....................................K.....................I....R...............D...
DHFR_musMus   ..........RP............D...................WK.....................R...............D...
DHFR_ratNor   ..........RP.......................L.......K.........................R...............D...
DHFR_criGri   ..........RP......................F...M.......K.........................R...............D...
DHFR_perMan   ..........RP.................................K.........................R...............D...
DHFR_perPol   ..........RP.......................L.......K.........................R...............D...
DHFR_dipOrd   .........GR.....................................K............T..........R...............D...
DHFR_speTri   ..........RP.....................................K.....M..P.............I.R.............ND...
DHFR_cavPor   ..........RP.....................................G........P.............R.....L....S....D...
DHFR_oryCun   ..........RP.....................................................................R...............D...
DHFR_ochPri   ..........RP..................R...............K............T...........R...............D...
DHFR_felCat   ..........RP.....................................K.........................R...............D...
DHFR_canFam   ..........RT..................R...............K.......N.................R...............D...
DHFR_vulVul   ..........RT..................R..TV...........K.......N.................R...............D...
DHFR_musPut   ..........RT....A...........................KF.........................R...............D...
DHFR_ailMel   ..........RT....A...........................KF.........................R...............D...
DHFR_equCab   ..........RP.................................K.......A.................R...............D...
DHFR_vicPac   ..........RP.D..I.........DM.................K.......A.....A...........R...............D...
DHFR_susScr   ..........RP...............................YK.........................R...............D...
DHFR_turTru   ..........RP...............................Y..........................R...............D...
DHFR_oviAri   ..........RP.....................N...........K.......V.....E...........R...............D...
DHFR_capHir   ..........RP.....................N...........K.......V.....E...........R...............D...
DHFR_bosTau   ..........RP.....................N...........Q.......V.................R...............D...
DHFR_myoLuc   ..........RT....A...........................K.........................R.......K.......D...
DHFR_pteVam   ..........L.................................K..................R......R.......K.......D...
DHFR_eriEur   ..........RP...M................E...........KH.RT..S.P...............R...............D...
DHFR_sorAra   ..........RP.....................E...........T..RK.....................R...............D...
DHFR_loxAfr   ..........RP.....................-...........K.........................R...............D...
DHFR_proCap   ..........RP.S..............................KH.RT..S.P...............R...............D...
DHFR_dasNov   ..........RP.S.YD..Y.............M..........YK...K...V................R...............D...
DHFR_monDom   ..........RQ....A...K.........D......L.......K...K....P.........................C....D...
DHFR_macEug   .........APTI......A.......M..............K..Q...K....P...........L...................D...
DHFR_sarHar   ..........RPI...A...............................N...K................L..............H....D...
DHFR_triVul   ..........RTI...............................D.KH..K...IP...........L.T.............S....D...
```

```
                        10        20        30        40        50        60        70        80        90
DHFR_ornAna   .........GPL.......AK....G...NK......A...........K....P...T.K...................S....D...
DHFR_tacAcu   .........GRP....A..AK....G...NK............Y....K....P...T....E...................S....D...
DHFR_galGal   ..........R...S....C..........D.N...........YK......S.....H......A...................D...
DHFR_lagLag   ..........R...S....C..........D.N...........YK......S.....H......AL..................D...
DHFR_anaPla   ..........R...S....C..........D.N...........YK......S.....H......AL..................D...
DHFR_taegut   .........PR...S...I...........D.R..........DYK......SA....P....K.A....R..............D...
DHFR_ficHyp   ..........R...S...............D.R...........YK......S.P...R......AL..................D...
DHFR_melUnd   ..........R...S....C..........D.S.....F.....YK......S.....Q......AL..................D...
DHFR_allMis   ..........S...A.A.............T............K........P.....P....VL..................D...
DHFR_croPor   ..........S...A.A.............T............K........P.....P....VL..................D...
DHFR_chrPic   ..........RP....A..C..........................K........T...T......VL....R.............D...
DHFR_anoCar   ..........L...S.A..C............Q...........K...K..M.P...TQ.....V.....R.............D...
DHFR_pytMol   ..........A..HS....CN.........D.K...........KH..K..M.T...KE.....V...............H....N...
DHFR_ambMex   .........GRI..A....CP...S.....D.N....I.......K......M.A...TQ.....V....R...L.............
DHFR_xenTro   .......MRNPF.HAV...CPP...Q....E.S....L.......KH...L.M.P...T..D.K.V....R...............E...
DHFR_xenLae   .......MRNQF.HAV...CPP...Q....G.S....L.......KH...L.M.P...T....K.V....R...............E...
DHFR_latCha   ......MGAARL..S....CP....L....D.N..H.K.LS.......K....P...T......V....R................V.
DHFR_lepOcu   .........PRPI......CP........H..N..H.K.LN-..K...K..M.P...TL..Q..A....R........R......N...
DHFR_gadMor   .........SRV.......CP...DA...YK.....H.T.LN...KH.R.L.V.P...GA.D...V....R...N.........ND...
DHFR_tetNig   .........ARV..A....CP...DL...R......H.I.LD...KH.RK..S.P.....N....V....R..........H...AN...
DHFR_hipHip   .........SRI..G....CP...DL...NR.N..H.V.LS...KH.RS..A.P....EK....V....R...............NN...
DHFR_solSen   .........SRV..G....CP...D....MT.N..H.V.LS...KH.RT..A.......K..K.V....R...............NN...
DHFR_oreNil   .........PRV..A....CP...DR...NK.N..H.I.LSK..AH.RK..A.P.....K....V........Y..........SN...
DHFR_dicLab   .........SRI..G....CP...DL...M..N..H.V.LG...KH.RT..A.P.....K....V....R...............NN...
DHFR_perFla   .........SRI..G....CP...DL...N..N..H.V.LN...KH.RT..A.P.....N....V....R...............NN...
DHFR_spaAur   .........SRIV.G....CP...DL...N..N..H.V.LNR..KH.RT..A.P.....K....V....R.......D..F...NN...
DHFR_gasAcu   .........SRV..G....CP...DL...CH.N..H...LS...KH.R...A.A.....KD...V....R...............NN...
DHFR_oryLat   .........TRT..G....CP...DL....G.N..H...LSKD.AL.RK..S.P...L.A.R..V....R...............QN...
DHFR_anoFim   .........SRV..A....CP...DL...R......H.V.LN...KH.R...S.P..........V....R...............NN...
DHFR_esoLuc   .........SRV.......CP........NK.N..H.K.LN...K...K..M.P...F......VD...R........R......N...
DHFR_salSal   .........SRVP......CP...DV...N..N..H.K.LN...K...K..M............V....R........R......N...
DHFR_oncMyk   .........SRV.......CP...D....N..N..H.K.LN...K...K..M............A....R........R......N...
DHFR_danRer   .........SRI.......CP...D......N..H.I.LS..LKH..K..M.P....D...K.V....R.......AA.H....N...
DHFR_cteIde   .........SRI.......CP...D....RK.N..H.I.LS...KH..K..M.P........K.V....R.......AQ......N...
DHFR_cypCar   .........SRI..S....CP...D......N..H.I.LS...KH..K..M.P...L....K.V.L..R.......AA......N...
DHFR_ictPun   .........GRV.......CP...D....R..N..H.I.LSK..KH..K..M.P...T....K.V....R.......AQ......N...
DHFR_leuEri   .........TRLI.S....CP........N..NF..H.I.LSK..KH.....S.P........C.A....R...H...........N...
DHFR_squAca   .........PRLV......CP.........D.NF..H.I.LSK.LKH..K..A.P........R.A....R...........H....D.L.
DHFR_eptBur   .......MAQNPV.V.A..LP........WK.G...HSKSLVK.MKH.T.L.SAA...A-.....A.V..R...E......F...RN...
DHFR_cioInt   .....MPAKDIQIHSV..CCN....G...FK.R...----.LPK.MK..K.I..GE...VE..RR.AI.I.R...E...KS.FK...D...
DHFR_cioSav   .....MPAKELKIHS...CCN....R...NK.R...----.LPK.MKH.TSL..GD...VEN..R.A.VV.R...E...KS.FK...N.V.
DHFR_oikDio   .......MNKSGW.M.L.ADI...KG...LRN.....----.IPQDLKH..ML.KGT...--PNQESV.V..RN..Q......F......V.
DHFR_braFlo   .......MKTKK.SLV..AC-..N.....VD.KI...---TLRGDMKF.S.L.SGT...EEA....A.V..R.......DR.F...PK.L.
DHFR_sacKow   ........-MKKISLVA.AC-..N.............---.LRK.MSF.TKV.SET...KED....A.V..R..........Y...A..Y.
DHFR_balCla   ........-MQKISPVA.AC-..NS.......N.....---.LRK.MK..TNV.SET...VE.E...A....R..........Y...NN.F.
DHFR_strPur   .......MAEKK..L.A.ACTSKGK....I..N.....---.LRQ.MA..E.L.K.A...QM..MK.A....R...D......F....D.V.
DHFR_parLiv   .......MADKR..L.A.ACTSNGK....I..N.....---.LRQ.MA..E.L.K.....QM..MK.A.V..R..........F....D.V.
DHFR_lytVar   .......MAEKK..L.A.ACTSSGK....I..N.....---.LRQ.MA..E.L.K.P...QL..MK.A....R..........F....D...
```

```
                        10        20        30        40        50        60        70        80        90
DHFR_patPec   ......MAGQKQC.L...ACKCKDSL...I..TI..---KLRTDMKF.STQ.S.T...AEND.K.A....R...L...D..F...PN.V.
DHFR_droMel   ........-MLRF.L....CE....F...IR.......---.IKS.LK..S.T.KRT....DPT...A.V..R..Y.GV..S.K...PD.L.
DHFR_gloMor   ........-MLKF.L.....K....F...LK.G.....---ELKS.L...SEL.KRV...FDST.R.V....R..Y.G..LN.....RN.L.
DHFR_haeIrr   ........-MLKFSL....CE....F...IK.......---LLKS.LK..STT.KRV...HDPS.R.V....R..Y.GV..S.K...QQ.L.
DHFR_sarCra   ........-MLKFSL....CE....F...IK.......---KIKS.LK..SST.KRV...RDPS.R.VA...R..Y.G...S.K...PE.L.
DHFR_culQui   ........-MKKFSL....CS....G...IK.......---.LRS.L.H.A...KRV...ADP..R.AI...R..Y.G...G.R...PD.L.
DHFR_anoGam   ........SKKFS.....CE....R...I.......---KLKQ.LK..SHT.KKV...NDAD.R.A....R..Y.GV..S.K...PE.L.
DHFR_aedAlb   ........-MKKFSL....CA....G...IK.......---.LRQ.LK..S...KKI...QDS..R.AI...R..Y.GV..S.K...PE.L.
DHFR_aedAeg   ........-MKKFSL....CA....G...IK.......---.LRQ.LK..S...KKI...TDTS.R.AI...R..Y.GV..S.KG..PE.L.
DHFR_armSub   ........-MKKFSL....CA....G...IK.......---.LKQ.LK..S...KKK...QDTS.R.AI...R..Y.GV..S.K...PE.L.
DHFR_danPle   .......MSKVK..L.A.ACE........S..S.....---.LKK.ME..TT...KV...KDAS.I.A....RR..DC..N......EN.L.
DHFR_bomMor   .......MSRTQ..L.A.ACE........I..T.....---KLKK.MA..TT...SV....DKK.V.A....RR..DC..I..Y...SN...
DHFR_helVir   .......MSQVK..L.A.ACD........V..A.....---.LKK.MA..TT..SKV....EPT.V.A....RR..DC..D..Y...QD.V.
DHFR_triCas   ..........IKFDL.A.ACE.........N.......---.LKS.LAF.SQ...QT....D.S.K.V.L..R...D...P..FK..HQ.F.
DHFR_denPon   .........ALK..L...A.E....W......T.....---KLKK.MAF.RT..SAT...EDKS.K.V....RR..E...AQ.FK..PN...
DHFR_apiMel   ........--MN.HL.A.TCE...G....IK.T.....---KLKS.LAF.TY...NT...KNPN.R.V.L..RR..E...KE......N...
DHFR_bomImp   ........--MN..L.A.ACE...G....VK.T.....---.LKS.MAF.TS...NT...KDPN.K.V.L..RR..EC..K..YK...N...
DHFR_eugCor   .........SLN..I.A..CE...G....VK.S.....---KLKS.MAF.TS...QT...KDP-.K.V....RR..EC..T..Y....D...
DHFR_nasVit   .........QVK.KL.A.ACE........V.......---.LRK.MDF.TK..S.T...KD.N.K.V.L..RR..E...K..FK..SN...
DHFR_copFlo   .........QVK.KL.A.ACD........I.......---.LRK.MDF.TK..S.T...KDPN...V.L..R...E...K..YK..AN...
DHFR_attCep   .........PPK.EL.A.ACE........V..N.....---.LKA.MA..T....NT...RNKN.R.V.L..RR..EC.....Y....D...
DHFR_camFlo   .........PPK.EL.A.ACE........I..G.....---.LKT.MAF.T....DT...KDKD.K.I.L..RR..DC.....YK..EN...
DHFR_harSal   .........LPK.EL.A.ACE........IK.......---.LKT.ME..T....ET...KDKN.K.V.L..RR..DCL..T.Y...RN...
DHFR_linHum   .........QLK.EL.A.ACE........N..A.....---.LKT.ME..T...IDT...KDKN.K.V.L..RR..D...I..Y...ND...
DHFR_pogBar   .........SHK.EL.A.ACE........I.......---.LKT.MAF.T....ET...NHTN.K.V.L..RR..EC..D..Y....D...
DHFR_solInv   .........PPK.EL.A.ACE........V.......---.LKT.MAF.T....DT...KQNN.K.V.L..RR..EC..K..Y....D...
DHFR_bemTab   .......MSSLR.SI...MTA...DN...FKNS.....---SLP..L.N.AKL.KNC...KDKS...A.V..RN..E...AQ.H....D.L.
DHFR_acyPis   ........--MVYSV.A...K....G...YK.N.....---KIKK.ME..NL...RV...NLK.V..A....RC..Q...D..Y.......
DHFR_blaGer   .........NLK.XL.....E........L.......---.LRK.LAH.S.L.KRT...IDSN...A....RR..E...I......PE.L.
DHFR_pedHum   ..........NNF.L.A..CS....N...YK.N.....---NLRK.LQ..N...KDV...KNPE.K.A....R...D.L.HN.WK..P..Y.
DHFR_onyArc   .........GPK.YI.A.ACE........YKN.......---KLR..MA..N...QIV...ENPELK.A....RR..D......F...RN.L.
DHFRcalCle    ....MTSLKCVKI.V.A.ACK..SSR....DN....K---LPTDMK..REK.AST...TEI--RVG....RR..E.V.P..F..F.N.F.
DHFR_lepSal   .....MSKSIPT.HV.A.ACR..PTN.....NT..N---LPTDLK..KTT.SST...PDS--KVA....RR..E...S..P....N...
DHFR_litVan   .........SVR..I...ACE....H...I..E.....---KLRE.MK..S...KAT...KSLE...A.M..R...E...A..F...P..L.
DHFR_celPug   .........GPR..I....AE....H....G.E.....---.LRE.MKH.S...KRL...DSPAT..V.L..R...E...L..F...P..L.
DHFR_dapPul   ........--MK..L...TAS........FQ.TI.....---.LKKDMAL.AKL.KNT...NDTN...A.V..R...E..........SN.L.
DHFR_ixoSca   ..MCPGVQESLQYFA.A.MCH....R...VLNT.....---.LKK.MA..S.I.SQ-...AA...T.A.V..RN..D...P..YK..P..V.
DHFR_ambMac   ..MCPSHKS.V.CFA.A.MCR....R...FKNA.....---.LKK.MAF.K...SE-...AA.....A.V..RN..E...P..F...NN...
DHFR_perSed   .........HRY..LV...CN....L...IG.E.....---.LRGDMKF.SKL.SET...KDSE.T.V....R...A......F...PK.T.
DHFR_milTar   ..........--------..ACR....G...IR.......---TLRGDM.F.TKI.SQT...KDPQ.K.A....R........R.F...SK...
DHFR_triSpi   ........-MAQV.I...ICE...KY....KNS.....---HLSK.MQH.KK...SV....DPN.I.A....RN..Y......Y...S..F.
DHFR_xipInd   ..........----M.L...ACD........HGNE.....--KLP..SKH.LKL.AGT...RDP....A....R...EI..VE.H....H.L.
DHFR_caeEle   ........-MRKM.L...MDA...EG......V.....---.IKKDMQ..ASV.KNV....DQS.R.A.L..R.C.E...VT.R...A..L.
DHFR_melInc   ..........----M.I.A..DE....F.....NS.....---.LPK.YKH.INL...T...KNPN.I.A.L..R.C.E......Y....N.L.
DHFR_ascSuu   .....MVSPKLPI.I...MDS...RG......A.....---HIPEDLK...T...KT...IDPT...AIV..R.V.E.L.A..W....N.L.
DHFR_schMed   ........-MKR..L...ACE....Q......K.....---NLK..MI..NNI..SV....DSN...V....RI..E...N..F..MPK...
DHFR_schMan   ........--MR..VV....E....W....G.G.....---KIKKDMEF.KTV..K-...AHP.LK.A.V..RV..E....S.FK...D...
DHFR_taeSol   .......MGLRR..V.A..AK....G....ENK.....---HIREDMAF.S.I.S.-...AQ...K.A.VL.RR..L.F.P..F...PN.V.
```

```
                    10        20        30        40        50        60        70        80        90
DHFR_aplCal   .......MTSTK..IV...CT........IE.R..----.LKQDMAF.KQL.VET...QD.Q.K.M........M...T..F...QD.V.
DHFR_lotGig   .......MCTKR..IV..ACK....N...V..SI..----.LKKDMAM.RHI.SDT...VD.S...A....R...M...D..F....N.V.
DHFR_phyAcu   .........-MPK..I...ACN....N...IE.R..----.LKSDMSF.KQI.LRT...QDAE.K.A.....N......S..F...V....
DHFR_pinMax   .......MSKCK..LV..ACN....R...ID.Q..----.LRKDMDF.KKI..ET...NDPE.R.A....R...........F...SK...
DHFR_mytCal   .....MSKTKST..LV..ACN...SR...I..K..----.IRKDMDF.KKI.MET...KNPD.K.V....R...........F...PK...
DHFR_dreRos   .......MSMPR..LV..ICN....N...I..S..---KLRKDMDF.KTI.M.T...TDPE...A....R...Q....N.F...RN...
DHFR_alvPom   .........PPT..L...MCN....N...IQ.K..---KIKGDMAF.RK...ET...KNPA.M.A....R..............NK.M.
DHFR_helRob   .........QPK.QIV..LCV..K.R...L.NSI..---KLPGDMTF.RKL.SET....IL.SK.AIL..R...D...SN.LK...N.L.
DHFR_nemVec   .....MAAVGVRFS....MDL...KR.....N...---KIKGDMKF.TEK.SEV...KT.....A....R.V.E......F......L.
DHFR_acrMil   ...................AD....G....ENR..---HLKGDMKF.SHL.S.V...AT.....A....RN..E...K..Y...PR.L.
DHFR_hydMag   .....MVQLRKVH..A.IDS...EW...LK.K..---NLPR.YKF..NI..KV...IS.....A.....N......QQ.H....N.L.
DHFR_mneLei   .........KLRF.I...TDL...EG.....NN..----.CRK...H.TDT.SKL...QN-...GAAV..R...E......F....N.H.
DHFR_triAdh   .........QKNFRM.A.MTR...DR...YKN...---KLR..YK..A.L..ST...KDSA.K.A.L..RN..V..........RN...
DHFR_subDom   .........----MYIM..KEI...KS...NK.KI...--.INTDMNFLET..KKT...RNKT...AA...RL..E...A..Y...PE.F.
DHFR_monBre   ........--MRVALV..ATK...AG...HQ.Q..---HLKADMQH.RDI..KT...EQPAAI.A....R..FE...D..F...H..L.
DHFR_canAlb   ........-MLKP.VAII.AALKPAL...YK.KM..----.LRK.I...KDV..RTTKPNT---R.A....R...E...Q..F...PD.L.
DHFR_canGla   ........-MSKVPVVGI.AALLPE....FQ.N..----.LAK.MK..REV..LTNDN.K----.V....R...E...Q..F...PK...
DHFR_pneCar   ........-MNQQKSLTLIVALTTSY...RSNS..---KLKK.IS..K.V.SFVPTFDSFESM.V.L..R...E...LQ.F........
DHFR_schSti   .......MSEIKPTIAII.AALKPDLA..FQ.KM..----.LRK.I...KDV..KTSDP.K---I.A.V..R...D...AR.F...PD...
DHFR_spaPas   .....MTMTEHKPTISIV.AALKPKY...YQ.KM..----.LRK.IQ..KNV.SKTTKSNG---I.A....R...E...K..F...PD...
DHFR_lodElo   .......MTSEKPTISII.AALKPSY...NK.K..----.LRK.MA..K.V.LRTTHTQ-----.A....R...E...V..F...PD.L.
DHFR_debHan   MTAAN...GKVTPPIAMI.AALLPEL...IK.KM..----.LRQ.IK..KNV...TKDPNG---I.A....R...E...T..F...P..L.
DHFR_meyGui   ........-MRQPPIALI.AALLPDL...FG.A..----.LKQ.IK..RDV.SNAPDG.----I.A....RR..E...PR.F...PN...
DHFR_milFar   MTSSGHSLGR.GIPIVMI.AALVPE....LK.KM..----.LK..MA..RKV...TTDNEA---I.A....R...Q......F...PK.L.
DHFR_claLus   ......MIKSLKPVTS.I.AALSPKY...AQ.K..----.LKQ.MK..KDV.SAARAG.----I.A.V..R...E...K..F...PN.L.
DHFR_komPas   .........SRPFKVNII.AALIPGH...LK.K..---SLK..MK..RLL..NTISPTS---K.V....RN..E...SR.F...QN.L.
DHFR_ogaPar   ..........---MTKVSLI.AALVPKY...YK.Q..---ALKE.M...R.V..QTADKNK---K.A.V..R...E....R.F......V.
DHFR_rhiDel   ........--MTQRKLALIAAATEEL...RK.....---HLPRDL...RDV..KIPKGNS--A..V....RV..E...P..Y...DN.F.
DHFR_encHel   .........---MFAL...FAK....K.....NR....--PLRTDMSW.RTLSQSVPLITSPSSS.A.V..R...D...PR.FQ..PN.T.
DHFR_encRom   .........---MFAL...LAR....K...RANR...SR-PLRTDMSW.RTLSQSIPLITLPSPS.A.V..R...E...P..F...PN.M.
DHFR_encCun   .........---M.ALV..LAS...HR...NANA....R-PLAADMAW.RTLSQSIPLI.APSAS.A.V..RR..D...SR.F...AN...
DHFR_encInt   .........---MIAL...LAA....Q...RH.K...KH-PLKTDMAW.RILSQSVPIL.SPS.S.V.V..R...H...SR.F...QN...
DHFR_harCan   ........--MNGFAL.VAVT..KT....NK.G...S--SIRGDMNF.KKI.SET...KDAA...A.V..R..YVG..A..F...SN.V.
DHFR_polPal   ........-MVKISFV....L....NV...D.GI..----.LKGDLAF.SKT..GHPIVMGRKTFDSIPLPNRHSIV.TSS.GKLYN-.DG
DHFR_dicDis   .........--MKVSI.....K....NV..TA..I..---ELPKDLKH.RDT.HGFPCIIGRISLEAFLLPNRFNIIVSS..KEFINI.EK
DHFR_araTha   ....MTSKPQSTYQVV..ATK...E.....D.K..---WNLPTDLKF.KDL.LST....DSA.K.A.V..R...E...K..Y...S..L.
DHFR_popTri   ....QQSDPRKTYQVV..ATK...D.....D.K..---WKLPSDLKF.KDL.L.T...MDS..K.A....R...E...LE.H...P..L.
DHFR_phyPat   ....THSEKLRGFQIV..ATR...ER....Q.H..---WKLPTDMKF.KTV.SVT...TSSS.K.A....RH..E......F...P..L.
DHFR_selMoe   ......AATLRAFQVV..ATR...DL....E.K..---WSVPSDMAF.K.I.SEA...ADAR.K.A.V..RR..E...A..F...P..V.
DHFR_ostTau   .......MEMRAFQVV..ATV...DG....DNG..---WKLTKDMGH.KKL.SET...RDANAV.A.V..R...E...A..F...P..L.
DHFR_micPus   ....MEAST.RPFQVV..ATK..GE....YQ.G..---WNLPKDMA..KAV.SQV...NEP.TK.A.V..R...E...A..F...P..V.
DHFR_chlVar   ....TPEQARRTIQIV..ATK...QW....G.S..---WSLPGDMK..REL.SRT...ADPA...V....R...E...A..F...A....
DHFR_volCar   ......MTSHKGFQLV..ATP...SL......T..--GWKLPGDMA..REL.SRT...REPSC..VA...R...E...N..F........
DHFR_chlRei   .....MDTSRKGFQLVI.ATP...SL......K..--DWQLPGDMA..KEL.SRT...RDGAHT.A....R...E...A..F......L.
DHFR_phyInf   .......MTEQ.IRVV..ALE..TTG...LRQHI.---W.LPSDMKH.RAL..A....PDSSV.HA....R...E.L.A..V..MPK.Y.
DHFR_albLai   ..........RTISIV..ATE..DCW...WRQSI.---WKLAKDLKH.RDL..RT...NSQNMR.A....R...E.L.PS.YQ..PN.Y.
DHFR_blaHom   .......MSIIRFSIVA.MTT...KR...L..G..---W.IKQDMKF.VDL...T...TDSE...A..I..N.Y..F....F....N...
DHFR_aurAno   ....APTNDLPQ.T....AVA..GSY....D.T..---WKLAGDMK..KKV.SA-...AP...T.ACV..R...L...P..F...G..K.
```

```
                         10        20        30        40        50        60        70        80        90
DHFR_phaTri   .......-------V..AAA..SHR...YQ.S..---W.LPGDL.H..AV..QP...PET.GT.A....R...D...DR.F...P....
DHFR_thaPse   .......-------V..AAA..GSR...HQ.K..---W.LPGDMNH.KKV...P...PSP.LT.A....R...D...S..F...D..V.
DHFR_perMar   .......MALPQ.SV...HTC...KW....D.Q..--WKSLPEDMKR.KKI..GG...CNDNVK.VC...R...E....R.F...RD...
DHFR_tetThe   .......MKTRHFDIVL.QTL..KKQ...YKNS..---W.LP..LKN.KKI..ET...-NK.L..A.....N..EAL.K..QQ...D.L.
DHFR_cryHom   .......MSEKNVSIV..A.VL..SS...I..Q..---SISEDLKF.SKI.NNK...CDSN.K.AL...R...D..GR-.-....N..I
DHFR_tryCru   ....GTRLALRAFSLV...DE...HG...DGRSI.---WNVPEDMKF.RDL..KL...PSPA.R.A.V..R...D...P..F...P..L.
DHFR_leiTro   ....FAFPSLRAFSIV..LDK...QH...DGESI.---W.VPEDMAF.KDQ..LL...PT.K.R.A.V..R...E.V.V..F......L.
DHFR_criFas   ....FAFPSLRAFSIV..AD....QH...DGETI.---WTVPETLAF.KDQ..LL...PT.K.R.A.V..R...E-V.L..F......L.
DHFR_ectSil   ....ASPSTSR.FVV...ATA..GSL......A..---W.LAADMA..K.C.S.P...TTTD.I.A....R...Q....R.F...A..R.
DHFR_toxGon   .........QKPVCLV..MTP...KR...I.NG.....--HLTTD.KH.S.V.K.T.....GKRF.A.V..R...E.M.R..F...VD.L.
DHFR_plaFal   ......KVESKNEGKKNE.FNNYTFR.L.NK.V...KCNSL.-DMK..CAV..YV...NSKKL..V.V..RTN.E...K..FK..SN...
DHFR_babBov   ...STVYEGC.G.TIY..IAL....RV..HQNQI.....--HIIHD..FLRNG..YI...KNPNI..V..F.R..YE...KA.SL...N...
DHFR_thePar   ...AEDYSGLPFVKLL..ITP...EN...ISNG.....--HIKRD.LFMF.A..YV...KHP.A..VI.I.R..YD.L..G.TF...N...
DHFR_naeGru   ......MMQST.PPMEAV.AVLLNSN...L..N...LQDGLTEDMKH.VSV.CI-....-NDQLM.A....R...E...S..FK..SKNRH
DHFR_escCol   .........---MISL.A.LAV...DRV..MENAM..---NLPADLAW.K.N.LN-...------KP....RH..E..---.G...P..K.
DHFR_breLat   .........---M.SS.F.MG.....RV..RDNQ.....-.LPEDLK..R.I..G-...------HAI...R..YE..---.GK..PN.R.
DHFR_marPos   .........---M.SL...M.T....RT..I.NS...---HLP.DLK..KQA.MG-...------KPIV..R..FE..---.GK..P..R.
DHFR_salEnt   .....MNPESVRIYLVA.MGA....RV..NGP.I...---KIPG.QKI.R.L.ES-...------KV.V..R..FE..---.GK..PN.HT
DHFR_klePne   .....MNPELVRIYLVA.MGA....RV..NGP.I...---KIPG.QKI.R.L.EG-...------KV.V..R..FE..---.GK..PN.RT
DHFR_halNea   .....MNQQAVRIYLVA.MGS....RV..NGSNI....-.IPG.QKI.RTL.EG-...------KV.V..R..FE..---.GK..PN.HT
DHFR_pseAla   ........---MIAL.A.CDR...HRL..DH.RI....-.IPG.QAR..KL..G-...------HV....RQ.YAE.---.G...PR.QT
DHFR_macCas   ........--MIISL.A.I.S....YV...DK.I....-KIPG.QVR.KDL.MG-...------KS....R..FE..---.GQ..PN.KT
DHFR_cloCel   ........---MISL....AK....NV..N..II...-KIKG.QKR.KEL.IG-...------KTI...R.SFEE.---.GK..PN.KT
DHFR_geoUra   ........--MVISL.A.MAE....RV..R.NAI...-DIPADRKR.RAL.LG-...------HP....R..FE.L---.AG..P..K.
DHFR_oxaFor   ........-MMTISHV..MAE....RV...D.QM...-HIPG.QKI.REL.VG-...------KAL.L.R..HE..---.G.V.PD..T
DHFR_nocSpp   .....MTPGGKRVVLVA..AR....GV..DGP.I....-QLPG.Q.L.KGL.WG-...------HILV..RA.YD..---.G...P..TT
DHFR_halPau   ........-MTR.V..A..AE....GV..RD..M...-HYSADL.H.KET.MG-...------HP.V..RT.YE..AGQLDG..P..T.
DHFR_natPel   MEAAAALEPARE.VG....AE....GV..RD..M...-HVPADLQH.KET.MG-...------HP....RV.YEG.L.TLGE..P..TT
DHFR_halXan   AESALAAETDRE.VG....AD....GV...D..M...-HIPADLQH.KET.MD-...------CP....RV.YEG.L.ALGE..P..TT
DHFR_natMag   ADTDPVVEPDHE.VG....AD....GV...D..M...-HIPEDLQH.KET.MA-...------HP....RV.YE..VDALGE..P..TT
DHFR_halLac   SALEAAEATDRE.VG....AD....GV..AD.EM...-HLPADLAH.KET.MD-...------HP....RV.YEG.L.TLGD..P..TT
DHFR_halWal   HDVDIDIDIDIDIVL.A..AA....DI..RD.EM...-HIPADLQQ.K.R.MG-...------HP..L.RR.YEA.IGALGE.FP..TS
DHFR_halBor   RDTTDANAESIEYVLVA..AE....GV..RD.RM...-HFSEDMAH.KQT.MG-...------HP..L.R..YEN.VDAIGE.FP..MS
DHFR_halMed   ..MSEHTDDSVRFVLVA..AE....RV..RD..M...-HLPEDLKH.KAT.MG-...------HP.V..RT.YE..ARQIDG..P..R.
DHFR_natPha   ........--MK.VL.A..AE....GV..TD.EM...-HYPEDLKR.KET.MG-...------HP....RT.YE..TGQLGG..P..T.
Consensus                  ivA      n gIG  g lPw    l  #   F   T          k na!!MGRkTwes!p k  rPL  R n
```

1.4. Sequence variability test. An internal control was performed to show the evolution of certain conserved motifs in DHFR. Being within DHFR, this is perhaps a better guideline for conservation than discussing the neutral rate of evolution in junk DNA or pseudogenes.

We pulled out the 7 positions between the ultra-conserved GIG and PW regions (shown below; positions 29 to 35 in all of the above alignments) into separate spreadsheet columns (In the table below, they are in the first 7 columns right after the species names). There are no deletions or insertions between them, always 7 intervening residues so no alignment ambiguity. Next, each column is sorted alphabetically (the last 7 columns in the table below) in turn and scored with the summary report function of the spreadsheet: how many different amino acids were acceptable in at least one species? and, was the 225 count spread evenly among these? The analysis reports are given immediate after the table below. This process discards phylogenetic information to look at how rapidly change can get fixed at residues that are simply placeholders. Side chain properties such as bulk or charge don't matter at all (note these residues could still be contributing backbone hydrogen bonds, etc.). The G and P residues in this region represent the opposite extremes. G and P are completely invariant up to sequencing error and mutation. *So we are seeing that our featured PCEs (G51PEKN and N23PP in ecDHFR) are very special in their conservation: their change is clade-coherent, it does not wobble randomly or within a reduced alphabet with preferences. The occurrence of L to F mutation is also rare.*

The first block of sequence shows as is, the second block each residue column has been sorted alphabetically (which loses species association).

		Residues in the analysis								**Alphabetically rearrangement of the residues in the analysis**						
1	DHFR_homSap	G	K	N	G	D	L	P		G	A	A	E	A	F	P
2	DHFR_panTro	G	K	N	G	D	L	P		G	A	A	E	A	F	P
3	DHFR_gorGor	G	K	N	G	V	L	P		G	C	A	G	A	F	P
4	DHFR_ponAbe	G	K	N	G	D	L	P		G	D	D	G	A	I	P
5	DHFR_nomLeu	G	K	N	G	D	L	P		G	D	D	G	A	I	P
6	DHFR_macMul	G	K	N	G	D	L	P		G	D	D	G	A	I	P
7	DHFR_papAnu	G	K	N	G	D	L	P		G	D	D	G	A	I	P
8	DHFR_calJac	G	K	N	G	E	L	P		G	D	D	G	A	I	P
9	DHFR_saiBol	G	K	N	G	D	L	P		G	F	D	G	A	I	P
10	DHFR_tarSyr	G	K	D	G	T	L	P		G	F	D	G	D	I	P
11	DHFR_otogar	G	K	N	G	D	L	P		G	F	D	G	D	I	P
12	DHF1_micMur	G	K	N	G	D	L	P		G	F	D	G	D	I	P
13	DHFR_tupBel	G	K	N	G	D	L	P		G	F	D	G	D	I	P
14	DHFR_musMus	G	K	N	G	D	L	P		G	F	D	G	D	I	P
15	DHFR_ratNor	G	K	N	G	D	L	P		G	F	D	G	D	I	P
16	DHFR_criGri	G	K	N	G	D	F	P		G	H	D	G	D	I	P
17	DHFR_perMan	G	K	N	G	D	L	P		G	H	D	G	D	I	P
18	DHFR_perPol	G	K	N	G	D	L	P		G	H	D	G	D	I	P
19	DHFR_dipOrd	G	K	N	G	D	L	P		G	H	D	G	D	I	P
20	DHFR_speTri	G	K	N	G	D	L	P		G	H	D	G	D	I	P
21	DHFR_cavPor	G	K	N	G	D	L	P		G	I	D	G	D	I	P
22	DHFR_oryCun	G	K	N	G	D	L	P		G	I	D	G	D	I	P
23	DHFR_ochPri	G	R	N	G	D	L	P		G	I	D	G	D	I	P
24	DHFR_felCat	G	K	N	G	D	L	P		G	I	D	G	D	I	P
25	DHFR_canFam	G	R	N	G	D	L	P		G	I	D	G	D	I	P

26	DHFR_vulVul	G	R	N	G	T	V	P		G	I	D	G	D	L	P
27	DHFR_musPut	G	K	N	G	D	L	P		G	I	D	G	D	L	P
28	DHFR_ailMel	G	K	N	G	D	L	P		G	I	D	G	D	L	P
29	DHFR_equCab	G	K	N	G	D	L	P		G	I	D	G	D	L	P
30	DHFR_vicPac	G	K	N	G	D	L	P		G	I	D	G	D	L	P
31	DHFR_susScr	G	K	N	G	D	L	P		G	I	D	G	D	L	P
32	DHFR_turTru	G	K	N	G	D	L	P		G	I	D	G	D	L	P
33	DHFR_oviAri	G	K	N	G	N	L	P		G	I	D	G	D	L	P
34	DHFR_capHir	G	K	N	G	N	L	P		G	I	D	G	D	L	P
35	DHFR_bosTau	G	K	N	G	N	L	P		G	I	D	G	D	L	P
36	DHFR_myoLuc	G	K	N	G	D	L	P		G	I	D	G	D	L	P
37	DHFR_pteVam	G	K	N	G	D	L	P		G	I	E	G	D	L	P
38	DHFR_eriEur	G	K	N	G	E	L	P		G	I	E	G	D	L	P
39	DHFR_sorAra	G	K	N	G	E	L	P		G	I	E	G	D	L	P
40	DHFR_loxAfr	G	K	N	G	D	L	P		G	I	E	G	D	L	P
41	DHFR_proCap	G	K	N	G	D	L	P		G	I	E	G	D	L	P
42	DHFR_dasNov	G	K	N	G	D	M	P		G	I	E	G	D	L	P
43	DHFR_monDom	G	K	D	G	D	L	P		G	I	E	G	D	L	P
44	DHFR_macEug	G	K	N	G	D	L	P		G	I	G	G	D	L	P
45	DHFR_sarHar	G	K	N	G	D	L	P		G	I	G	G	D	L	P
46	DHFR_triVul	G	K	N	G	D	L	P		G	I	G	G	D	L	P
47	DHFR_ornAna	G	N	K	G	D	L	P		G	I	G	G	D	L	P
48	DHFR_tacAcu	G	N	K	G	D	L	P		G	I	G	G	D	L	P
49	DHFR_galGal	G	K	D	G	N	L	P		G	I	G	G	D	L	P
50	DHFR_lagLag	G	K	D	G	N	L	P		G	I	G	G	D	L	P
51	DHFR_anaPla	G	K	D	G	N	L	P		G	I	G	G	D	L	P
52	DHFR_taegut	G	K	D	G	R	L	P		G	I	G	G	D	L	P
53	DHFR_ficHyp	G	K	D	G	R	L	P		G	K	G	G	D	L	P
54	DHFR_melUnd	G	K	D	G	S	L	P		G	K	G	G	D	L	P
55	DHFR_allMis	G	K	N	G	T	L	P		G	K	G	G	D	L	P
56	DHFR_croPor	G	K	N	G	T	L	P		G	K	G	G	D	L	P
57	DHFR_chrPic	G	K	N	G	D	L	P		G	K	G	G	D	L	P
58	DHFR_anoCar	G	K	N	G	Q	L	P		G	K	G	G	D	L	P
59	DHFR_pytMol	G	K	D	G	K	L	P		G	K	H	G	D	L	P
60	DHFR_ambMex	G	K	D	G	N	L	P		G	K	H	G	D	L	P
61	DHFR_xenTro	G	K	E	G	S	L	P		G	K	H	G	D	L	P
62	DHFR_xenLae	G	K	G	G	S	L	P		G	K	K	G	D	L	P
63	DHFR_latCha	G	K	D	G	N	L	P		G	K	K	G	D	L	P
64	DHFR_lepOcu	G	H	N	G	N	L	P		G	K	K	G	D	L	P
65	DHFR_gadMor	G	Y	K	G	D	L	P		G	K	K	G	D	L	P
66	DHFR_tetNig	G	R	N	G	D	L	P		G	K	K	G	D	L	P
67	DHFR_hipHip	G	N	R	G	N	L	P		G	K	K	G	D	L	P
68	DHFR_solSen	G	M	T	G	N	L	P		G	K	K	G	D	L	P
69	DHFR_oreNil	G	N	K	G	N	L	P		G	K	K	G	D	L	P
70	DHFR_dicLab	G	M	N	G	N	L	P		G	K	K	G	D	L	P
71	DHFR_perFla	G	N	N	G	N	L	P		G	K	K	G	D	L	P
72	DHFR_spaAur	G	N	N	G	N	L	P		G	K	K	G	D	L	P
73	DHFR_gasAcu	G	C	H	G	N	L	P		G	K	K	G	D	L	P
74	DHFR_oryLat	G	K	G	G	N	L	P		G	K	K	G	D	L	P
75	DHFR_anoFim	G	R	N	G	D	L	P		G	K	K	G	D	L	P
76	DHFR_esoLuc	G	N	K	G	N	L	P		G	K	K	G	D	L	P
77	DHFR_salSal	G	N	N	G	N	L	P		G	K	K	G	D	L	P

78	DHFR_oncMyk	G	N	N	G	N	L	P		G	K	K	G	D	L	P
79	DHFR_danRer	G	K	N	G	N	L	P		G	K	K	G	D	L	P
80	DHFR_cteIde	G	R	K	G	N	L	P		G	K	K	G	D	L	P
81	DHFR_cypCar	G	K	N	G	N	L	P		G	K	K	G	D	L	P
82	DHFR_ictPun	G	R	N	G	N	L	P		G	K	K	G	D	L	P
83	DHFR_leuEri	G	N	N	G	N	F	P		G	K	K	G	D	L	P
84	DHFR_squAca	G	K	D	G	N	F	P		G	K	K	G	D	L	P
85	DHFR_eptBur	G	W	K	G	G	L	P		G	K	K	G	D	L	P
86	DHFR_cioInt	G	F	K	G	R	L	P		G	K	K	G	E	L	P
87	DHFR_cioSav	G	N	K	G	R	L	P		G	K	K	G	E	L	P
88	DHFR_oikDio	G	L	R	N	D	L	P		G	K	K	G	E	L	P
89	DHFR_braFlo	G	V	D	G	K	I	P		G	K	K	G	E	L	P
90	DHFR_sacKow	G	K	N	G	D	L	P		G	K	K	G	E	L	P
91	DHFR_balCla	G	K	N	G	N	L	P		G	K	K	G	E	L	P
92	DHFR_strPur	G	I	N	G	N	L	P		G	K	K	G	E	L	P
93	DHFR_parLiv	G	I	N	G	N	L	P		G	K	K	G	E	L	P
94	DHFR_lytVar	G	I	N	G	N	L	P		G	K	K	G	E	L	P
95	DHFR_patPec	G	I	N	G	T	I	P		G	K	K	G	E	L	P
96	DHFR_droMel	G	I	R	G	D	L	P		G	K	K	G	G	L	P
97	DHFR_gloMor	G	L	K	G	G	L	P		G	K	K	G	G	L	P
98	DHFR_haeIrr	G	I	K	G	D	L	P		G	K	K	G	G	L	P
99	DHFR_sarCra	G	I	K	G	D	L	P		G	K	K	G	G	L	P
100	DHFR_culQui	G	I	K	G	D	L	P		G	K	K	G	G	L	P
101	DHFR_anoGam	G	I	N	G	D	L	P		G	K	L	G	G	L	P
102	DHFR_aedAlb	G	I	K	G	D	L	P		G	K	N	G	G	L	P
103	DHFR_aedAeg	G	I	K	G	D	L	P		G	K	N	G	G	L	P
104	DHFR_armSub	G	I	K	G	D	L	P		G	K	N	G	G	L	P
105	DHFR_danPle	G	S	N	G	S	L	P		G	K	N	G	G	L	P
106	DHFR_bomMor	G	I	N	G	T	L	P		G	K	N	G	G	L	P
107	DHFR_helVir	G	V	N	G	A	L	P		G	K	N	G	H	L	P
108	DHFR_triCas	G	K	N	N	D	L	P		G	K	N	G	H	L	P
109	DHFR_denPon	G	K	N	G	T	L	P		G	K	N	G	I	L	P
110	DHFR_apiMel	G	I	K	G	T	L	P		G	K	N	G	K	L	P
111	DHFR_bomImp	G	V	K	G	T	L	P		G	K	N	G	K	L	P
112	DHFR_eugCor	G	V	K	G	S	L	P		G	K	N	G	K	L	P
113	DHFR_nasVit	G	V	N	G	D	L	P		G	K	N	G	K	L	P
114	DHFR_copFlo	G	I	N	G	D	L	P		G	K	N	G	K	L	P
115	DHFR_attCep	G	V	N	G	N	L	P		G	K	N	G	K	L	P
116	DHFR_camFlo	G	I	N	G	G	L	P		G	K	N	G	K	L	P
117	DHFR_harSal	G	I	K	G	D	L	P		G	K	N	G	K	L	P
118	DHFR_linHum	G	N	N	G	A	L	P		G	K	N	G	K	L	P
119	DHFR_pogBar	G	I	N	G	D	L	P		G	K	N	G	K	L	P
120	DHFR_solInv	G	V	N	G	D	L	P		G	K	N	G	K	L	P
121	DHFR_bemTab	G	F	K	N	S	L	P		G	K	N	G	K	L	P
122	DHFR_acyPis	G	Y	K	G	N	L	P		G	K	N	G	K	L	P
123	DHFR_blaGer	G	L	N	G	D	L	P		G	K	N	G	K	L	P
124	DHFR_pedHum	G	Y	K	G	N	L	P		G	K	N	G	K	L	P
125	DHFR_onyArc	G	Y	K	N	D	L	P		G	K	N	G	K	L	P
126	DHFRcalCle	G	K	D	N	D	L	P		G	K	N	G	K	L	P
127	DHFR_lepSal	G	K	N	N	T	L	P		G	K	N	G	K	L	P
128	DHFR_litVan	G	I	N	G	E	L	P		G	K	N	G	K	L	P
129	DHFR_celPug	G	K	G	G	E	L	P		G	K	N	G	K	L	P

130	DHFR_dapPul	G	F	Q	G	T	I	P		G	K	N	G	K	L	P
131	DHFR_ixoSca	G	V	L	N	T	L	P		G	K	N	G	K	L	P
132	DHFR_ambMac	G	F	K	N	A	L	P		G	K	N	G	N	L	P
133	DHFR_perSed	G	I	G	G	E	L	P		G	K	N	G	N	L	P
134	DHFR_milTar	G	I	R	G	D	L	P		G	K	N	G	N	L	P
135	DHFR_triSpi	G	K	K	N	S	L	P		G	K	N	G	N	L	P
136	DHFR_xipInd	G	H	G	N	E	L	P		G	K	N	G	N	L	P
137	DHFR_caeEle	G	K	N	G	V	L	P		G	K	N	G	N	L	P
138	DHFR_melInc	G	K	N	N	S	L	P		G	K	N	G	N	L	P
139	DHFR_ascSuu	G	K	N	G	A	L	P		G	K	N	G	N	L	P
140	DHFR_schMed	G	K	N	G	K	L	P		G	K	N	G	N	L	P
141	DHFR_schMan	G	K	G	G	G	L	P		G	K	N	G	N	L	P
142	DHFR_taeSol	G	K	E	N	K	L	P		G	K	N	G	N	L	P
143	DHFR_aplCal	G	I	E	G	R	L	P		G	K	N	G	N	L	P
144	DHFR_lotGig	G	V	N	G	S	I	P		G	K	N	G	N	L	P
145	DHFR_phyAcu	G	I	E	G	R	L	P		G	K	N	G	N	L	P
146	DHFR_pinMax	G	I	D	G	Q	L	P		G	K	N	G	N	L	P
147	DHFR_mytCal	G	I	N	G	K	L	P		G	K	N	G	N	L	P
148	DHFR_dreRos	G	I	N	G	S	L	P		G	K	N	G	N	L	P
149	DHFR_alvPom	G	I	Q	G	K	L	P		G	L	N	G	N	L	P
150	DHFR_helRob	G	L	N	N	S	I	P		G	L	N	G	N	L	P
151	DHFR_nemVec	G	K	N	N	D	L	P		G	L	N	G	N	L	P
152	DHFR_acrMil	G	K	E	N	R	L	P		G	L	N	G	N	L	P
153	DHFR_hydMag	G	L	K	G	K	L	P		G	L	N	G	N	L	P
154	DHFR_mneLei	G	K	N	N	N	L	P		G	L	N	G	N	L	P
155	DHFR_triAdh	G	Y	K	N	D	L	P		G	L	N	G	N	L	P
156	DHFR_subDom	G	N	K	G	K	I	P		G	L	N	G	N	L	P
157	DHFR_monBre	G	H	Q	G	Q	L	P		G	L	N	G	N	L	P
158	DHFR_canAlb	G	Y	K	G	K	M	P		G	L	N	G	N	L	P
159	DHFR_canGla	G	F	Q	G	N	L	P		G	M	N	G	N	L	P
160	DHFR_pneCar	G	R	S	N	S	L	P		G	M	N	G	N	L	P
161	DHFR_schSti	G	F	Q	G	K	M	P		G	M	N	G	N	L	P
162	DHFR_spaPas	G	Y	Q	G	K	M	P		G	N	N	G	N	L	P
163	DHFR_lodElo	G	N	K	G	K	L	P		G	N	N	G	N	L	P
164	DHFR_debHan	G	I	K	G	K	M	P		G	N	N	G	N	L	P
165	DHFR_meyGui	G	F	G	G	A	L	P		G	N	N	G	N	L	P
166	DHFR_milFar	G	L	K	G	K	M	P		G	N	N	G	N	L	P
167	DHFR_claLus	G	A	Q	G	K	L	P		G	N	N	G	N	L	P
168	DHFR_komPas	G	L	K	G	K	L	P		G	N	N	G	N	L	P
169	DHFR_ogaPar	G	Y	K	G	Q	L	P		G	N	N	G	Q	L	P
170	DHFR_rhiDel	G	R	K	G	D	L	P		G	N	N	G	Q	L	P
171	DHFR_encHel	G	K	N	N	R	L	P		G	N	N	G	Q	L	P
172	DHFR_encRom	G	R	A	N	R	L	P		G	N	N	G	Q	L	P
173	DHFR_encCun	G	N	A	N	A	L	P		G	N	N	G	Q	L	P
174	DHFR_encInt	G	R	H	G	K	L	P		G	N	N	G	Q	L	P
175	DHFR_harCan	G	N	K	G	G	L	P		G	N	N	G	Q	L	P
176	DHFR_polPal	G	K	D	G	G	I	P		G	N	N	G	Q	L	P
177	DHFR_dicDis	G	T	A	G	D	I	P		G	N	N	G	Q	L	P
178	DHFR_araTha	G	K	D	G	K	L	P		G	N	N	G	R	L	P
179	DHFR_popTri	G	K	D	G	K	L	P		G	N	N	G	R	L	P
180	DHFR_phyPat	G	K	Q	G	H	L	P		G	N	N	G	R	L	P
181	DHFR_selMoe	G	K	E	G	K	L	P		G	N	N	G	R	L	P

182	DHFR_ostTau	G	K	D	N	G	L	P		G	N	N	G	R	L	P
183	DHFR_micPus	G	Y	Q	G	G	L	P		G	R	N	G	R	L	P
184	DHFR_chlVar	G	K	G	G	S	L	P		G	R	N	G	R	L	P
185	DHFR_volCar	G	K	N	G	T	L	P		G	R	N	G	R	L	P
186	DHFR_chlRei	G	K	N	G	K	L	P		G	R	N	G	R	L	P
187	DHFR_phyInf	G	L	R	Q	H	I	P		G	R	N	G	R	L	P
188	DHFR_albLai	G	W	R	Q	S	I	P		G	R	N	K	R	L	P
189	DHFR_blaHom	G	L	N	G	G	L	P		G	R	N	N	S	L	P
190	DHFR_aurAno	G	K	D	G	T	L	P		G	R	N	N	S	L	P
191	DHFR_phaTri	G	Y	Q	G	S	L	P		G	R	N	N	S	L	P
192	DHFR_thaPse	G	H	Q	G	K	L	P		G	R	N	N	S	L	P
193	DHFR_perMar	G	K	D	G	Q	L	P		G	R	N	N	S	L	P
194	DHFR_tetThe	G	Y	K	N	S	L	P		G	R	N	N	S	L	P
195	DHFR_cryHom	G	I	N	G	Q	L	P		G	R	N	N	S	L	P
196	DHFR_tryCru	G	D	G	R	S	I	P		G	R	N	N	S	L	P
197	DHFR_leiTro	G	D	G	E	S	I	P		G	R	N	N	S	L	P
198	DHFR_criFas	G	D	G	E	T	I	P		G	R	N	N	S	L	P
199	DHFR_ectSil	G	K	N	G	A	L	P		G	R	N	N	S	L	P
200	DHFR_toxGon	G	I	N	N	G	L	P		G	R	N	N	S	L	P
201	DHFR_plaFal	G	N	K	G	V	L	P		G	S	N	N	S	L	P
202	DHFR_babBov	G	H	Q	N	Q	I	P		G	T	N	N	S	L	P
203	DHFR_thePar	G	I	S	N	G	L	P		G	T	N	N	S	L	P
204	DHFR_naeGru	G	L	N	G	N	L	P		G	V	N	N	S	L	P
205	DHFR_escCol	G	M	E	N	A	M	P		G	V	Q	N	S	L	P
206	DHFR_breLat	G	R	D	N	Q	L	P		G	V	Q	N	S	L	P
207	DHFR_marPos	G	I	N	N	S	L	P		G	V	Q	N	S	L	P
208	DHFR_salEnt	G	N	G	P	D	I	P		G	V	Q	N	T	M	P
209	DHFR_klePne	G	N	G	P	D	I	P		G	V	Q	N	T	M	P
210	DHFR_halNea	G	N	G	S	N	I	P		G	V	Q	N	T	M	P
211	DHFR_pseAla	G	D	H	G	R	I	P		G	V	Q	N	T	M	P
212	DHFR_macCas	G	K	D	K	D	I	P		G	V	Q	N	T	M	P
213	DHFR_cloCel	G	N	N	G	I	I	P		G	W	Q	N	T	M	P
214	DHFR_geoUra	G	R	N	N	A	I	P		G	W	Q	N	T	M	P
215	DHFR_oxaFor	G	K	D	G	Q	M	P		G	Y	Q	N	T	M	P
216	DHFR_nocSpp	G	D	G	P	D	I	P		G	Y	Q	N	T	M	P
217	DHFR_halPau	G	R	D	G	D	M	P		G	Y	R	N	T	M	P
218	DHFR_natPel	G	R	D	G	D	M	P		G	Y	R	N	T	M	P
219	DHFR_halXan	G	K	D	G	D	M	P		G	Y	R	P	T	M	P
220	DHFR_natMag	G	K	D	G	D	M	P		G	Y	R	P	T	M	P
221	DHFR_halLac	G	A	D	G	E	M	P		G	Y	R	P	T	M	P
222	DHFR_halWal	G	R	D	G	E	M	P		G	Y	R	Q	T	M	P
223	DHFR_halBor	G	R	D	G	R	M	P		G	Y	S	Q	V	M	P
224	DHFR_halMed	G	R	D	G	D	M	P		G	Y	S	R	V	M	P
225	DHFR_natPha	G	T	D	G	E	M	P		G	Y	T	S	V	V	P

<u>Analysis Summary:</u>

G: completely invariant up to sequencing error and mutation

K: 16 different amnio acids occur there (see distribution table below); most common ones do not share side chain attributes. Our 225 sequences are overweighted to vertebrates; K is common especially in tetrapods where it may have acquired some importance. I is the second most common amino acid (32 times) and N is the third most common one (21 times).

K	96
I	32
N	21
R	18
Y	11
L	10
V	9
F	7
D	5
H	5
M	3
A	2
T	2
W	2
C	1
S	1

N: 12 different amno acids but predominantly a reduced alphabet of N (103 cases), K (39 cases), and D (33 cases).

N	103
K	39
D	33
G	15
Q	12
E	7
R	6
A	3
H	3
S	2
L	1
T	1

G: 8 amino acids but overwhelmingly G with a few N. Looking at phylogenetic coherence, there is none: the N's are just sprinkled in. Classical reduced alphabet situation with preference for G (185 cases), acceptability of N (30 cases).

G	185
N	30
P	3
E	2
Q	2
K	1
R	1
S	1

D: 13 amino acids of which 10 are above sequence quality level. There is a preference for D and N. K, S, and T are ok. GREDAQ not show-stoppers as substitutions. There is little phylogenetic conservation within mammals even, despite the "inertia" that keeps a residue fixed over short time intervals even when no selective pressure supports it.

D	76
N	37
K	22
S	19
T	15
G	11
R	11
E	10
A	9
Q	9
V	3
H	2
I	1

L: classical first column of genetic code: L preferred (182 occurrences), I (22 occurrences) and M (17 occurrences) are not as good but ok. F (3 cases) and V (1 case) are really marginal.

L	182
I	22
M	17
F	3
V	1

P: completely invariant up to sequencing error and mutation.

2. Kinetics and pH/rate profiles.

Both the pre-steady-state and steady-state kinetic experiments were performed using an Applied Photophysics stopped-flow spectrophotometer at 25 ^{o}C. The reactions were carried out in MTEN buffer (composed of 50 mM MES, 25 mM Tris, 25 mM ethanolamine, and 100 mM NaCl) following the published procedures . One of the syringes in the stopped-flow analyzer was loaded with 20 μM enzyme, 250 μM NADPH, 2 mM DTT, and 50 mM MTEM buffer (according to [MES]). The other syringe contained 200 μM DHF, 2mM DTT, and 50mM MTEM buffer. After combining DHFR and NADPH as described above, the mixtures were incubated on ice for 5 minutes prior to the onset of the chemical reaction. The other syringe contained 200 μM DHF, 2mM DTT, and 50 mM MTEM buffer. Upon mixing, the final concentrations of the individual species in the reaction chamber were halved (10 μM enzyme, 125 μM NADPH, 100 μM DHF, 2 mM DTT, and 50 mM MTEM buffer). For the pre-steady-state kinetics, the progress of the DHFR-catalyzed hydride transfer reaction was monitored by the loss of fluorescence resonance energy transfer from the enzyme to NADPH under single turnover conditions. The reaction mixture was excited at 290 nm and the emission was measured using a 400 nm cut-off output filter. The measure of absorbance vs. time trace (of burst phase) was fit to standard single exponential decay to obtain the hydride transfer rate (k_{hyd}). To construct the pH/rate profiles, at

least 5 separate kinetic runs were performed at each pH condition and the averaged k_{hyd} values were used for the analysis. Steady-state kinetics experiments were performed following similar experimental conditions as described above with the exception that the reaction progress was monitored at 340 nm. Kinetic isotope effect (KIE) experiments were conducted according to the concentrations and conditions listed above. Parallel experiments were performed using NADPH or NADPD.

The binding affinity of DHF to the binary E:NADPH complex was examined under pre-steady-state conditions. For the binding experiments, the final concentrations of DHFR, DTT, MTEM and NADPH in the stopped-flow reaction chamber were 5 μM, 2 mM, 50 mM, and 100 μM, respectively. The [DHF] varied from 100 nM to 100 μM, while the k_{hyd} values stayed constant. The dissociation constants of E:NADPH:DHF into E:NADPH and DHF were estimated through iterations of mathematical fitting as described in section 3.

The pH/rate profiles for both the pre-steady-state hydride transfer step (k_{hyd}) and the steady-state turnover process (k_{cat}) were constructed for each mutant, and the data can be fitted into eq. 1, which is derived from the mechanistic scheme (10,11) illustrated in Fig. S1. The k_{obs} values in the pH/rate profiles (Fig. S2) are averages of at least 5 separate kinetic runs. The reaction mechanism (Fig. S1) involves one ionization event in the observable rate constants. The kinetic pK_a values measured for the hydride transfer reaction are comparable between mutants, and with the wild-type enzyme (10). When the observed rate constants are similar, the kinetic pK_a value determined from the steady-state kinetics is different from the value obtained from pre-steady-state kinetics. This is consistent with earlier report (10) showing that in higher pH domains, the steady-state rate constant is contaminated with the hydride transfer process, which becomes rate-limiting. The scheme in Fig. S1 assumes that upon deprotonation of the ES or EP complexes (at higher pH), the forward reactions associated with the deprotonated complexes are either significantly slower than those observed for the active species (at low pH) or zero (within experimental errors). This assumption is supported by the pH/rate profile fits. It should be noted that the kinetic data reported in the human DHFR study (12) can also be analyzed with a slight modification to the scheme in Fig. S1 by increasing the contribution from the lower parallel pathway.

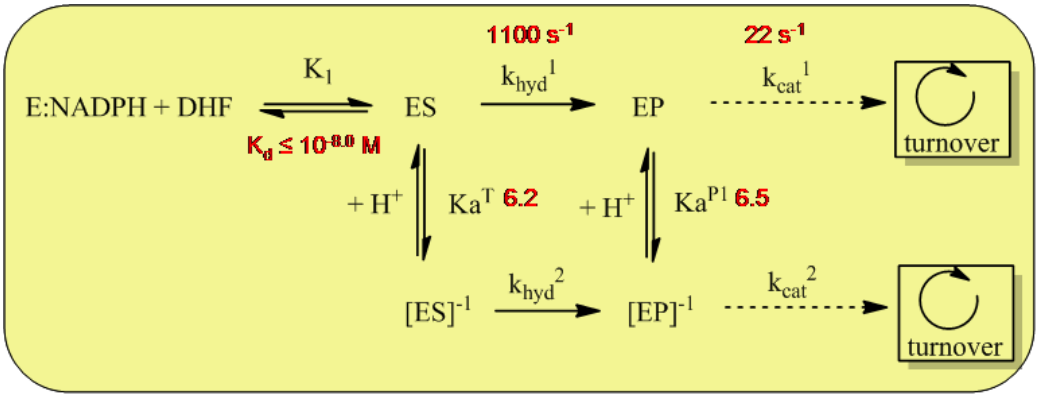

Figure S1. Proposed mechanistic scheme with representative kinetic data (at 298K) for the N23PP/G51PEKN ecDHFR mutant shown in red. The scheme involves a simplified reaction pathway with K_1, k_{hyd}^1, and k_{cat}^1 values. Ka^T and Ka^{P1} are the acid dissociation constants for the ternary E:NADPH:DHF (ES) complex and the product E:$NADP^+$:THF (EP) complex, respectively.

$$\log(k_{obs}) = \log\left(k_1 \frac{[H^+]}{K_a + [H^+]} + k_2 \frac{K_a}{K_a + [H^+]} \right) \qquad \textbf{eq. (1)}$$

In eq. (1), k_{obs} is the apparent observed rate constant (either k_{hyd} or k_{cat}). K_a is the kinetic acid dissociation constant determined from the pH/rate profiles. k_1 and k_2 are the unimolecular rate constants (of either k_{hyd} or k_{cat}) for the active and the 'inactive' complexes, respectively. Eq. (1) is derived from considering mass balance on all kinetically relevant species. It is important to point out that eq. (1) is a universal expression that can be applied to both pre-steady-state and steady-state analysis. Since the pre-steady-state and steady-state kinetics involved different observables, they were monitored separately. For the pre-steady-state analysis, the k_1, k_2, and K_a terms in eq. (1) represent the k_{hyd}^1, k_{hyd}^2, and K_a^T terms in Fig. S1. Since the observable was the conversion of ES to EP, the subsequent steps after EP are irrelevant to the observed rate constants. Similarly, for the steady-state turnover experiments the unimolecular disappearance of EP (likely the release of THF from the product complex as mentioned in the main text) was monitored, meaning that the the k_1, k_2, and K_a terms in eq. (1) represent the k_{cat}^1, k_{cat}^2, and K_a^{p1} terms in Fig. S1. In this case, the steps prior to EP in the scheme were not captured spectrophotometrically, and they are not part of the observed rate constants. In all cases, only the more reactive species (k_{hyd}^1 and k_{cat}^1) were used for data analysis since we are interested in how the various mutations would affect the optimized kinetic rates. However it should be noted that all the pH/rate profiles were fit to eq. (1) without setting the k_2 value to zero. The presence of another plateau at higher pH is suggestive of a two reactive DHFR ternary complexes that are separated by one ionizable group. Our lab has began to probe the nature of the second plateau, and the results from that study should be made available in due time.

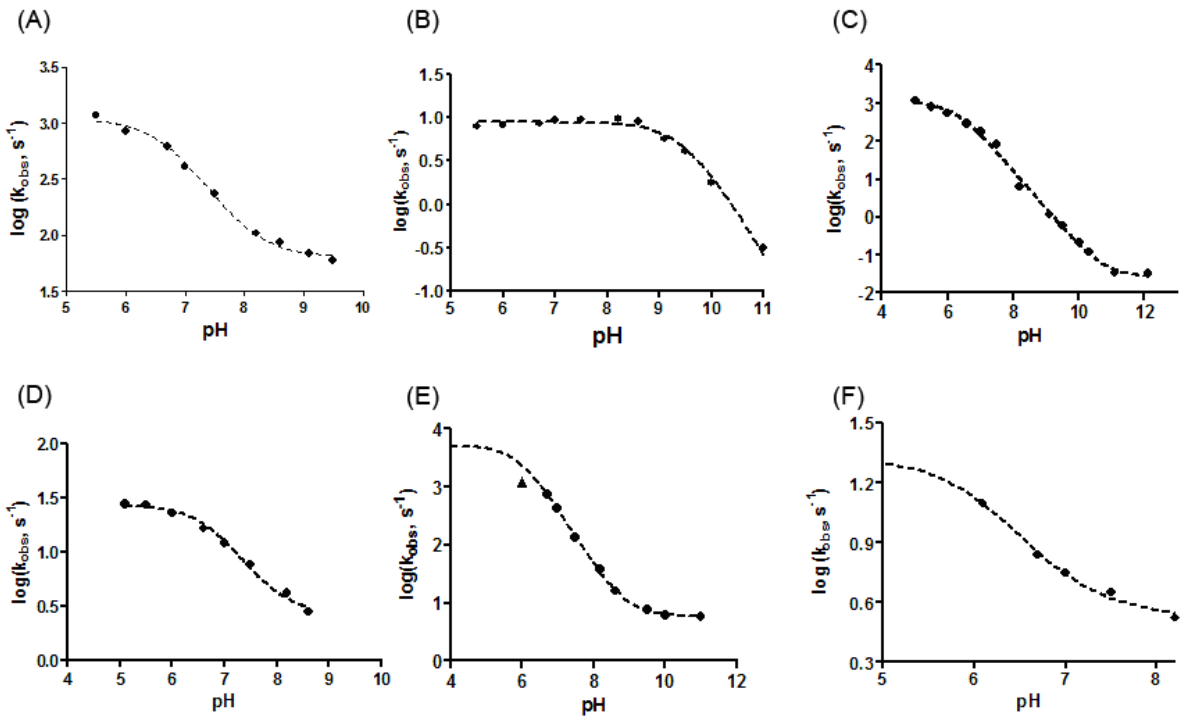

Figure S2. pH/rate profiles for the various ecDHFR variants studied: (A) Plot of averaged pre-steady-state k_{obs} values vs. pH for the G51PEKN ecDHFR-catalyzed hydride transfer reaction with 10 μM of enzyme, 125 μM of NADPH, and 100 μM of DHF in 50 mM of aqueous MTEM buffer at 25 ^{o}C. The data were fit into eq. (1) to yield maximum k_{hyd} value of (1100 ± 80) s^{-1}, pK_a value of 6.77 ± 0.07, and k_2 = (64 ± 4) s^{-1}. (B) Plot of averaged steady-state k_{obs} values vs. pH for the G51PEKN ecDHFR-catalyzed hydride transfer reaction with 10 μM of enzyme, 125 μM of NADPH, and 100 μM of DHF in 50 mM of aqueous MTEM buffer at 25 ^{o}C. The data were fit into eq. (1) to yield maximum k_{cat} value of (8.9 ± 0.4) s^{-1} and pK_a value of 9.49 ± 0.05. (C) Plot of averaged pre-steady-state k_{obs} values vs. pH for the N23PP/G51PEKN ecDHFR-catalyzed hydride transfer reaction with 10 μM of enzyme, 125 μM of NADPH, and 100 μM of DHF in 50 mM of aqueous MTEM buffer at 25 ^{o}C. The data were fit into eq. (1) to yield maximum k_{hyd} value of (1100 ± 100) s^{-1}, pK_a value of 6.20 ± 0.06, and k_2 = (0.027 ± 0.003) s^{-1}. (D) Plot of averaged steady-state k_{obs} values vs. pH for the N23PP/G51PEKN ecDHFR-catalyzed hydride transfer reaction with 10 μM of enzyme, 125 μM of NADPH, and 100 μM of DHF in 50 mM of aqueous MTEM buffer at 25 ^{o}C. The data were fit into eq. (1) to yield maximum k_{cat} value of (26.9 ± 1.6) s^{-1}, pK_a value of 6.85 ± 0.09, and k_2 = (2.7 ± 0.3) s^{-1}. (E) Plot of averaged pre-steady-state k_{obs} values vs. pH for the N23PP/L28F/G51PEKN ecDHFR-catalyzed hydride transfer reaction with 10 μM of enzyme, 125 μM of NADPH, and 100 μM of DHF in 50 mM of aqueous MTEM buffer at 25 ^{o}C. The data were fit into eq. (1) to yield maximum k_{hyd} value of (5100 ± 1200) s^{-1}, pK_a value of 5.9 ± 0.1, and k_2 = (5.9 ± 0.1) s^{-1}. The datum point (▲) at pH 6 was not included in the fit because it is outside of the stopped-flow detection capability. (E) Plot of averaged steady-state k_{obs} values vs. pH for the N23PP/L28F/G51PEKN ecDHFR-catalyzed hydride transfer reaction with 10 μM of enzyme, 125 μM of NADPH, and 100 μM of DHF in 50

mM of aqueous MTEM buffer at 25 ^{o}C. The data were fit into eq. (1) to yield maximum k_{cat} value of (17.6 ± 5.1) s^{-1}, pK_a value of 6.1 ± 0.2, and k_2 = (3.5 ± 0.2) s^{-1}.

3. Thermodynamic binding of ecDHFR mutants. The affinity of DHF to the binary E:NADPH complex was probed using pre-steady-state kinetic analysis for all ecDHFR mutants studied here. Figure S3 shows a representative pre-steady-state kinetics graph for the N23PP/L28F/G51PEKN ecDHFR mutant. 5 μM of the enzyme was mixed with excess (100 μM) NADPH, and the mixture was incubated on ice for at least five minutes prior to the introduction of DHF. The observed hydride transfer rate constant remained the same between 5 μM to 0.1 μM of [DHF]. The hydride transfer rate in Fig. S3 is also comparable with the value in Fig. S2, where [DHF] is 100 μM. This suggests that the dissociation constant, K_d, for E:NADPH:DHF into E:NADPH + DHF is at least 10^{-7} M. As seem in Fig. S3, fitting the kinetic data (open circles) to a standard one-site binding expression (solid line) results in a sharp break immediately after the left-most datum point due to the strength of binding. This also means that meaningful binding constant cannot be extracted simply from the fitted expression. Instead, manual guesses for the hydride transfer rate constant at 5 x 10^{-8} M of DHF were made in stepwise fashion (decreasing value from the plateau value of ~180 s^{-1}) to approximate the dissociation constant. An estimated dissociation constant was made when the inserted guess produced significant deviation (Fig. S3, dotted line) from the fitted curve (Fig. S3, solid line). All together, the dissociation constant, K_d, for E:NADPH:DHF into E:NADPH + DHF was estimated to be ~10^{-7} – 10^{-8} M, which are similar to the wild-type value (10).

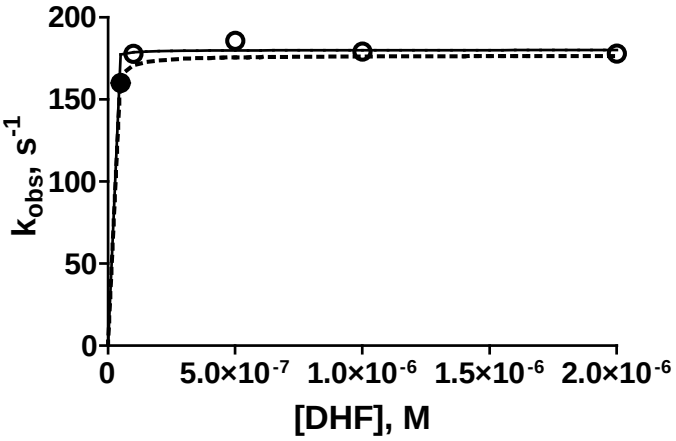

Figure S3. Plot of averaged pre-steady-state k_{obs} (open circles) vs. [DHF] for the N23PP/L28F/G51PEKN ecDHFR mutant-catalyzed hydride transfer reaction with 5 μM of enzyme and 100 μM of NADPH in 50 mM of aqueous MTEM buffer at pH 7.3 and 25 ^{o}C. The data were fit (solid line) into a standard one-site binding model but the strength of the binding prevents accurate determination of the binding constant. A guess (●) of k_{obs} 160 s^{-1} at [DHF] = 5 x 10^{-8} M starts to generate noticeable deviation of the fitted line (dotted line). The dotted line yields a K_d value of ~$10^{-8.4}$ M.

4. Kinetic Isotope Effect. The KIE data for the ecDHFR mutants are summarized in Table S1. In low pH conditions where the hydride transfer reaction is well separated (much faster) from the turnover process, normal primary KIE was found for the pre-steady-state kinetics while unity KIE was observed for the steady-state kinetics. Under basic conditions (pH 11), the steady-state rates showed KIE values of between 2.07 – 1.98. This is because at higher pH the hydride transfer rate becomes more rate-limiting (10).

Table S1. Kinetic isotope effect data for the pre-steady-state and the steady-state kinetics obtained for the ecDHFR mutants. The KIE values given below are determined from the averages of at least 4 kinetic runs with either NADPH or NADPD as the substrate. The reactions were performed with 10 μM of enzyme, 125 μM of NADPH, and 100 μM of DHF in 50 mM of aqueous MTEM buffer at 25 ^{o}C.

	$\frac{k_{hyd}(NADPH)}{k_{hyd}(NADPD)}$	$\frac{k_{hyd}(NADPH)}{k_{hyd}(NADPD)}$	$\frac{k_{cat}(NADPH)}{k_{cat}(NADPD)}$
G51PEKN	2.7 ± 0.4 (pH 6.0)	2.8 ± 0.4 (pH 8.6)	1.14 ± 0.1 (pH 5.5)
N23PP/G51PEKN	2.7 ± 0.3 (pH 5.5)	2.7 ± 0.4 (pH 11)	1.14 ± 0.2 (pH 5.5)
N23PP/L28F/G51PEKN	2.1 ± 0.2 (pH 6.7)	2.7 ± 0.4 (pH 10)	0.93 ± 0.1 (pH 5.5)

5. Crystallization

5.1. Crystallization and data collection. Crystallization was performed by the hanging-drop vapor diffusion method at 20°C. Drops were set up using approximately 25 mg/mL protein in 10 mM Tris, pH 7.5 containing 1 mM methotrexate and 1 mM NADPH. Crystals formed after 3 – 4 days in 100 mM calcium acetate, 36% Peg 400 and 100 mM Hepes, pH 7.0. Crystals were harvested, briefly soaked in a solution of 100 mM calcium acetate, 36% Peg 400 and 100 mM Hepes containing 1 mM methotrexate and 1 mM NADPH (freshly made), and flash frozen in liquid nitrogen. Data were collected at 100 K at the A1 beamline of the Cornell High Energy Synchrotron Source (CHESS). Data were collected over 180° with a 1° oscillation range and extended to approximately 1.8 Å. Data collection statistics are provided in Table S2.
.

5.2. Data processing, structure determination and refinement. The data were indexed, integrated and scaled using HKL2000 (13). The crystals contained two molecules per asymmetric unit and had an approximate solvent content of 54%. Molecular replacement was employed for phasing, using MOLREP (14) with the structure of *E. coli* DHFR (PDB code 1RH3) as the search model. The resulting structure was refined using alternating cycles of refinement using REFMAC5 (15) and manual model building with Coot (16). The addition of water molecules took place only after the refinement converged and was followed by an additional round of refinement. The ligands were placed into difference density using the models available from the PDB (MTX and NADPH) and were included in the model for a final round of refinement. Data refinement statistics are provided in Table S3.
.

Table S2. Data Collection Statistics.

	N23PP/G51PEKN DHFR
resolution (Å)	50.0 – 1.85
wavelength (Å)	0.987
beam line	CHESS A1
space group	$P2_1$
a (Å)	52.25
b (Å)	62.77
c (Å)	62.44
β (°)	106.8
no. of reflections	83,707
unique reflections	31,876
average I/σ	12.1 (2.2)
redundancy	2.6 (2.6)
completeness (%)	95.2 (80.5)
R_{sym}[a] (%)	7.8 (32.7)

Numbers in parentheses correspond to the highest resolution shell

[a] $R_{\mathrm{sym}} = \Sigma\Sigma_{\mathrm{i}} \left| I_i - \langle I \rangle \right| / \Sigma \langle I \rangle$, where $\langle I \rangle$ is the mean intensity of the N reflections with intensities I_i and common indices h, k, l

Table S3. Data Refinement Statistics.

	N23PP/G51PEKN DHFR
resolution (Å)	50.0 – 1.85
no. of protein atoms	2544
no. of ligand atoms	165
no. of water atoms	169
no. of reflections in working set	30,241
no. of reflections in test set	1,631 (5.1 %)
R factor[a] (%)	20.3
R_{free}[b] (%)	25.4
rmsd bonds (Å)	0.018
rmsd angles (^{o})	1.8
mean *B* factor (Å^2)	22.3
Ramachandran plot	
most favored (%)	98.1
additionally allowed (%)	1.9
generously allowed (%)	0.0
disallowed (%)	0.0

[a] *R* factor $= \Sigma_{hkl} \left\| F_{obs} \right| - k \left| F_{calc} \right\| / \sum_{hkl} \left| F_{obs} \right|$, where F_{obs} and F_{calc} are observed and calculated structure factors respectively.

[b] For R_{free}, the sum is extended over a subset of reflections (5.1 %) excluded from all stages of refinement.

6. Empirical Valence Bond Molecular Dynamics Simulations. Empirical valence bond (EVB) molecular dynamics (MD) simulations were performed for four systems: wild-type ecDHFR, N23PP ecDHFR, N23PP/G51PEKN ecDHFR, and wild-type hsDHFR. The initial configuration for the WT ecDHFR simulations was the crystal structure of WT ecDHFR in the closed state with bound $NADP^+$ and folate (PDB: 3QL3) (11). For consistency, the Asp37Asn point mutation contained in this structure was reverted to Asp37 using the utility Profix (17). The initial configuration for the N23PP simulations was the crystal structure of N23PP/S148A DHFR with bound $NADP^+$ and folate (PDB: 3QL0) (11). The S148A mutation was removed using the utility Profix (17). Since the MD simulations for the N23PP/G51PEKN mutant were performed prior to obtaining the crystal structure published herein, the N23PP/G51PEKN mutant was modeled by local minimization of the RMSD of the backbone C_α atoms in the preceding and proceeding four residues about position 51 in the initial structure of the N23PP ecDHFR simulations with the corresponding residues in the hsDHFR crystal structure (PDB: 2W3M) (18). This modeled structure agrees well with the crystal structure of ecDHFR N23PP/G51PEKN (PDB: 4GH8) with a C_α RMSD of 0.84 Å. The validity of this model is further supported by the similarity of the N23PP/G51PEKN ecDHFR crystal structure and the WT hsDHFR crystal structure, as shown in the main text. The initial structure for the hsDHFR simulations was the crystal structure of hsDHFR complexed with NADPH and folate (PDB: 2W3M). Protons were added using AMBER leap. In each system, His45 was protonated at the epsilon position, and all remaining histidines were doubly protonated. Each system was embedded in a truncated octahedral periodic box with 8199 water molecules and 11 sodium counter-ions.

The empirical valence bond method and the mapping potential approach, as well as our application to hydride transfer in dihydrofolate reductase, have been described in detail previously (e.g., Refs. 19,20,21). The EVB potential consists of a 2×2 matrix, where the diagonal elements, V_{11} and V_{22}, are the potential energies of the reactant and product diabatic states, respectively. The diabatic states are represented by the AMBER99SB force field for the protein (22,23), the TIP3P water model (24), NADPH and $NADP^+$ parameters from Ref. (25), and DHF-H^+ and THF charges calculated using the restrained electrostatic potential (RESP) method (26) with parameters from the generalized AMBER force field (GAFF) (27). To improve the determination of the ESP for each species and to localize charge differences upon reduction to the pterin ring, charges were calculated using three fragments: both the reduced and oxidized forms of the pterin, capping the C9-N10 bond with a proton, and the para-aminobenzoylglutamate (pABG) moiety, capping the C9-C6 bond with a proton. To obtain an integral charge for each species (DHF-H^+ or THF), the charges on the C9 protons were averaged between the relevant pterin fragment and the pABG fragment, and the remaining charge was assigned to the C9 carbon. Gaussian 03 (28) was used for the electronic structure calculations required for the RESP method. The atom types and partial charges used for these species are provided in Table S4.

Using classical MD in the pure reactant state, first the solvent and ions were equilibrated for 500 ps at constant NPT using the Berendsen (29) thermostat and barostat with harmonic restraints on protein and ligand atoms of 100 kcal/mol$Å^2$. (Note that the Berendsen thermostat was used only for the initial equilibration and not for data collection.) Next the solvent and ions were energy minimized, followed by minimization of the full system using the conjugate gradient algorithm. The full system was then annealed from 50 K to 300 K in increments of 50 K, holding the temperature constant at each temperature for 100 ps at constant NPT. The full system was then equilibrated for 1 ns at constant NPT and 5-20 ns at constant NVT. A 10 Å real

space non-bonded cut-off with Particle Mesh Ewald (PME) (30) for long-range electrostatics was used in all calculations. All bonds involving hydrogen atoms were constrained to their equilibrium bond lengths during these simulations using SHAKE (31). System preparation and equilibration were performed using the AmberTools program and the AMBER 11 program (32), respectively.

Following this extensive classical MD equilibration, the coordinates and topology of each system were transferred to a modified version of DLPROTEIN (33). All of the simulations with DLPROTEIN were performed at constant NVT using the Nosé-Hoover thermostat (34,35). The charge on the hydride was incorporated into the donor carbon charge for the reactant state and the acceptor carbon charge for the product state. The donor-hydride and acceptor-hydride constrained harmonic bonds were replaced by a Morse potential with a dissociation energy (D_e) of 103 kcal/mol, an equilibrium bond length (R_{eq}) of 1.09 Å, and α of 1.817 $Å^{-1}$, corresponding to the frequency of the CT-HC harmonic bond in the AMBER99SB force field. All bonds involving hydrogen atoms and not involving the hydride remained constrained in these simulations. The van der Waals parameters for the hydride were treated consistently with the AMBER force field, except that the non-bonded interactions of the hydride with the donor and the acceptor were excluded in the product and reactant states, respectively. Each system was re-equilibrated for 100 ps using the EVB mapping potential with $\lambda = 0.95$ (95% reactant state). Following this equilibration, subsequent windows were generated from the configuration following 10 ps of equilibration in the previous window, reducing λ in increments of 0.05 until reaching $\lambda = 0.05$ (95% product state). Each window was propagated for 600 ps, with the first 100 ps taken as equilibration. The diabatic energies V_{11} and V_{22} were sampled every 1 fs, and configurations were saved every 100 fs. This procedure was performed three times for each ecDHFR system studied in order to generate three independent data sets. However, in one of the independent data sets generated for the N23PP mutant, a conformational change was observed in the βF-βG loop where it partially unfolds, leading to interatomic distance changes of several Angstroms. This data set was therefore discarded for data analysis purposes. Two independent data sets were generated for WT hsDHFR.

The free energy profiles were generated from a series of 19 trajectories with different mapping potentials (i.e., windows) and combined using the weighted histogram analysis method (WHAM) (36). Three independent sets of trajectories were propagated for wild-type ec DHFR and N23PP/G51PEKN ecDHFR, and two independent sets of trajectories were propagated for N23PP ecDHFR and WT hsDHFR. Independent data sets were combined to obtain a total of 28.5 ns for wild-type ecDHFR and N23PP/G51PEKN ecDHRR and a total of 19.0 ns for N23PP ecDHFR and WT hsDHFR. A bin size of 1 kcal/mol was used, and bins with less than 50 configurations sampled in each window were discarded. Although the quantitative free energy values depend on these parameters, the free energy differences between systems ($\Delta\Delta G^{\neq}$ and $\Delta\Delta G°$) are robust with respect to these details. The parameters V_{12} and Δ in the EVB potential correspond to the coupling between the two diabatic states and a constant energy shift of the second state relative to the first state. These parameters were fit to reproduce the experimental free energy of activation ($\Delta G^{\neq} = 13.4$ kcal/mol) and free energy of reaction ($\Delta G° = -4.4$ kcal/mol) for wild-type ecDHFR on the ground state EVB surface, resulting in $V_{12} = 44.15$ kcal/mol and $\Delta = -60.86$ kcal/mol. These parameters were then kept fixed for the calculations of the free energy profiles for the ecDHFR mutants and for WT hsDHFR. The free energy barriers and free energies of reaction for independent data sets are shown in Table S5. Due to the use of the AMBER99SB force field for the diabatic states rather than the GROMOS force field used

previously in our group, the parameters of the EVB potential have changed relative to our previous EVB MD simulations of ecDHFR (20,37,38).

Donor-acceptor distances and average inter-C_α distances were calculated using the data combined from all trajectories by thermally averaging each distance along the collective reaction coordinate with a bin size of 2 kcal/mol in the energy reaction coordinate. The inter-C_α distance changes from RS to TS were computed as the difference between the transition state value and the reactant state value for each pair of C_α atoms with adjacent averaging over +/- 10 kcal/mol. Root-mean-square fluctuations (RMSFs; Fig. S4, S5) in the RS (TS) were calculated by first generating the thermally averaged structure for the RS (TS) and then calculating the RMSF of each C_α atom with respect to this thermally averaged structure for all configurations corresponding to an energy gap reaction coordinate within 10 kcal/mol of the value associated with the RS (TS). The configurations were weighted according to the probabilities determined from the WHAM used to generate the free energy profiles. Table S6 compares the differences in the RMSF data between wild-type hsDHFR and the N23PP/G51PEKN ecDHFR variant.

Table S4. Atom types and partial charges used for protonated dihydrofolate (DHF-H+) and tetrahydrofolate (THF). Note Generalized Amber Force Field (GAFF) atom types were used for these ligands. H6 is the transferring hydride and therefore is not present in DHF-H+.

Atom Name	DHF-H+ atom type	DHF-H+ charge	THF atom type	THF charge
N5	nh	-0.238406	nh	-0.744490
HN5	hn	0.378463	hn	0.398993
C4A	cd	-0.150270	cd	0.050147
C4	c	0.434047	c	0.473314
O4	o	-0.509820	o	-0.585297
N3	n	-0.390588	n	-0.468894
HN3	hn	0.348003	hn	0.352579
C2	cd	0.774744	cd	0.722468
N2	nh	-0.936436	nh	-0.922606
H21	hn	0.452596	hn	0.402068
H22	hn	0.452596	hn	0.402068
N1	nc	-0.652360	nc	-0.693288
C8A	cc	0.457140	cc	0.429205
N8	nh	-0.469164	nh	-0.550984
HN8	hn	0.357414	hn	0.355896
C7	c3	0.084249	c3	0.017010
H71	h1	0.097248	h1	0.065980
H72	h1	0.097248	h1	0.065980
C6	c2	0.298617	c3	0.250024
H6			h1	0.001647
C9	c3	0.066499	c3	0.013756

H91	h1	0.102817	h1	0.060939
H92	h1	0.102817	h1	0.060939
N10	nh	-0.685880	nh	-0.685880
H10	hn	0.372838	hn	0.372838
C14	ca	0.260186	ca	0.260186
C15	ca	-0.198155	ca	-0.198155
C16	ca	-0.152153	ca	-0.152153
H16	ha	0.185699	ha	0.185699
H15	ha	0.104079	ha	0.104079
C13	ca	-0.198155	ca	-0.198155
H13	ha	0.114899	ha	0.114899
C12	ca	-0.152153	ca	-0.152153
H12	ha	0.152660	ha	0.152660
C11	ca	-0.090484	ca	-0.090484
C	c	0.692208	c	0.692208
O	o	-0.652769	o	-0.652769
N	n	-0.540544	n	-0.540544
HN	hn	0.287739	hn	0.287739
CA	c3	0.073583	c3	0.073583
HA	h1	0.036915	h1	0.036915
CT	c	0.799292	c	0.799292
O1	o	-0.809254	o	-0.809254
O2	o	-0.809254	o	-0.809254
CB	c3	-0.011474	c3	-0.011474
HB1	hc	0.016041	hc	0.016041
HB2	hc	0.016041	hc	0.016041
CG	c3	-0.056192	c3	-0.056192
HG1	hc	-0.020784	hc	-0.020784
HG2	hc	-0.020784	hc	-0.020784
CD	c	0.820027	c	0.820027
OE2	o	-0.845813	o	-0.845813
OE1	o	-0.845813	o	-0.845813

Table S5. Comparison of hydride transfer free energy barriers ($\Delta G^{\neq}$) and free energies of reaction ($\Delta G°$) from independent data sets. Free energy barriers were calculated using WHAM with a bin size of 1 kcal/mol, V_{12} = 44.15 kcal/mol, and Δ = -60.86 kcal/mol. The variation in $\Delta G^{\neq}$ among independent data sets is ~1 kcal/mol. Note that the free energies of reaction exhibit more variation among data sets because of difficulties sampling the product state.

	$\Delta G^{\neq}$ (kcal/mol)				$\Delta G°$ (kcal/mol)			
System	all sets	set 1	set 2	set 3	all sets	set 1	set 2	set 3
WT ecDHFR	13.4	13.2	14.0	13.1	-4.4	-5.5	-2.5	-5.1

N23PP	14.6	15.0	14.3		-4.3	-4.0	-5.1	
N23PP/G51PEKN	13.2	12.8	13.1	13.8	-7.2	-7.7	-7.4	-6.3
WT hsDHFR	13.1	13.2	13.0		-5.8	-5.8	-5.7	

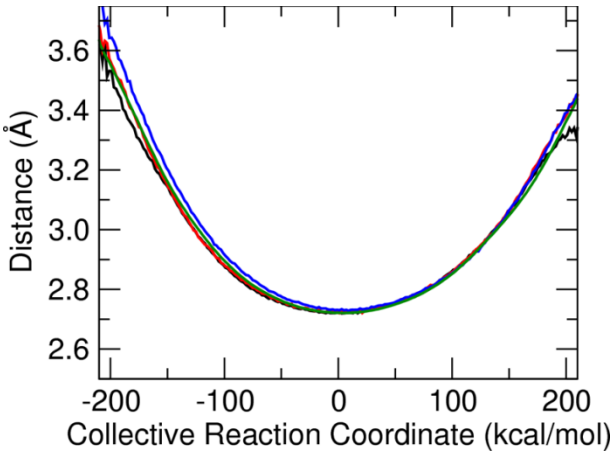

Figure S4. Thermally-averaged donor-acceptor distances along the collective reaction coordinate for wild-type ecDHFR (black), N23PP ecDHFR (red) N23PP/G51PEKN ecDHFR (blue), and wild-type hsDHFR (green).

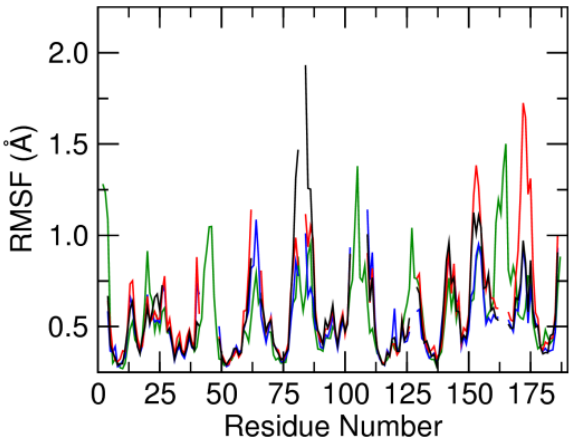

Figure S5. Root-mean-square fluctuations (RMSFs) of C_α atoms in the transition state for wild-type ecDHFR (black), N23PP ecDHFR (red), N23PP/G51PEKN ecDHFR (blue), and wild-type hsDHFR (green). These RMSFs were calculated relative to the thermally averaged structure of each system in the transition state. Residue numbering corresponds to wild-type hsDHFR.

Table S6. Comparison of RMSF (Å) between N23PP/G51PEKN ecDHFR mutant and wild-type hsDHFR. The residue numbering corresponds to wild-type hsDHFR. The data are sorted by the degree of absolute differences (from high to low) between hsDHFR and N23PP/G51PEKN ecDHFR mutant.

hsDHFR residue number	N23PP/G51PEKN ecDHFR RMSF (Å)	WT hsDHFR RMSF (Å)	Difference (hsDHFR - ecDHFR mutant) (Å)
162	0.514	1.143	0.629
161	0.546	1.034	0.488

125	0.418	0.834	0.416
168	0.454	0.861	0.407
153	1.123	0.741	-0.382
40	0.782	0.408	-0.374
124	0.394	0.768	0.374
172	0.917	0.555	-0.362
109	0.973	0.658	-0.315
126	0.48	0.783	0.303
167	0.474	0.766	0.292
22	0.523	0.811	0.288
20	0.657	0.937	0.28
21	0.499	0.778	0.279
111	0.781	0.523	-0.258
166	0.496	0.752	0.256
85	0.629	0.852	0.223
19	0.478	0.692	0.214
169	0.573	0.779	0.206
66	0.498	0.701	0.203
173	0.788	0.596	-0.192
110	0.658	0.467	-0.191
84	0.867	0.681	-0.186
174	0.52	0.702	0.182
13	0.669	0.488	-0.181
142	0.798	0.622	-0.176
81	0.656	0.826	0.17
186	0.885	0.72	-0.165
157	0.509	0.672	0.163
129	0.585	0.739	0.154
67	0.444	0.59	0.146
86	0.73	0.874	0.144
175	0.65	0.781	0.131
49	0.457	0.327	-0.13
117	0.483	0.353	-0.13
171	0.67	0.542	-0.128
14	0.645	0.523	-0.122
131	0.429	0.55	0.121
158	0.527	0.644	0.117
18	0.375	0.491	0.116
4	0.597	0.711	0.114
59	0.455	0.564	0.109
24	0.517	0.625	0.108

87	0.662	0.556	-0.106
147	0.374	0.478	0.104
144	0.784	0.681	-0.103
152	0.794	0.692	-0.102
141	0.669	0.572	-0.097
178	0.469	0.375	-0.094
102	0.922	0.83	-0.092
80	0.733	0.646	-0.087
7	0.413	0.328	-0.085
140	0.534	0.449	-0.085
89	0.485	0.401	-0.084
26	0.66	0.577	-0.083
62	0.716	0.633	-0.083
12	0.457	0.375	-0.082
65	0.781	0.703	-0.078
77	0.37	0.444	0.074
112	0.536	0.462	-0.074
184	0.373	0.447	0.074
72	0.396	0.325	-0.071
23	0.464	0.533	0.069
154	1.052	1.12	0.068
6	0.388	0.321	-0.067
78	0.518	0.584	0.066
130	0.603	0.666	0.063
5	0.4	0.461	0.061
150	0.478	0.537	0.059
58	0.358	0.416	0.058
156	0.796	0.741	-0.055
185	0.473	0.528	0.055
79	0.631	0.685	0.054
143	0.641	0.587	-0.054
180	0.342	0.394	0.052
181	0.363	0.415	0.052
70	0.554	0.503	-0.051
100	0.474	0.524	0.05
64	0.841	0.89	0.049
56	0.36	0.408	0.048
69	0.47	0.516	0.046
30	0.389	0.434	0.045
57	0.336	0.381	0.045
27	0.642	0.598	-0.044

33	0.417	0.461	0.044
55	0.354	0.398	0.044
28	0.425	0.467	0.042
159	0.543	0.585	0.042
63	0.715	0.756	0.041
73	0.367	0.326	-0.041
31	0.305	0.342	0.037
50	0.344	0.308	-0.036
61	0.524	0.488	-0.036
97	0.374	0.41	0.036
98	0.472	0.508	0.036
71	0.398	0.363	-0.035
101	0.541	0.576	0.035
160	0.586	0.621	0.035
34	0.353	0.386	0.033
92	0.432	0.465	0.033
94	0.497	0.464	-0.033
145	0.59	0.557	-0.033
17	0.337	0.369	0.032
54	0.318	0.35	0.032
96	0.443	0.475	0.032
25	0.509	0.54	0.031
95	0.57	0.54	-0.03
138	0.315	0.345	0.03
93	0.437	0.466	0.029
38	0.412	0.439	0.027
51	0.318	0.291	-0.027
60	0.523	0.55	0.027
99	0.538	0.565	0.027
170	0.557	0.584	0.027
177	0.472	0.445	-0.027
122	0.359	0.385	0.026
8	0.284	0.309	0.025
41	0.548	0.523	-0.025
53	0.31	0.335	0.025
35	0.322	0.346	0.024
68	0.419	0.443	0.024
149	0.403	0.427	0.024
88	0.505	0.482	-0.023
116	0.307	0.329	0.022
120	0.404	0.426	0.022

132	0.41	0.432	0.022
151	0.665	0.687	0.022
183	0.377	0.399	0.022
146	0.64	0.66	0.02
123	0.452	0.47	0.018
10	0.298	0.282	-0.016
179	0.377	0.393	0.016
37	0.506	0.491	-0.015
74	0.313	0.298	-0.015
9	0.292	0.279	-0.013
11	0.313	0.3	-0.013
90	0.412	0.4	-0.012
36	0.413	0.402	-0.011
155	0.912	0.901	-0.011
32	0.381	0.391	0.01
139	0.373	0.363	-0.01
76	0.359	0.35	-0.009
121	0.366	0.375	0.009
176	0.531	0.522	-0.009
39	0.368	0.376	0.008
115	0.311	0.319	0.008
114	0.342	0.349	0.007
15	0.445	0.451	0.006
16	0.413	0.419	0.006
52	0.287	0.293	0.006
113	0.374	0.38	0.006
119	0.43	0.436	0.006
75	0.325	0.32	-0.005
91	0.383	0.378	-0.005
118	0.36	0.365	0.005
136	0.314	0.319	0.005
182	0.366	0.37	0.004
137	0.277	0.28	0.003
133	0.357	0.355	-0.002
135	0.308	0.31	0.002
29	0.454	0.453	-0.001
134	0.321	0.321	0
148	0.449	0.449	0

7. Isothermal titration calorimetry (ITC). ITC experiments were done using MicroCal Auto-iTC200 (GE) while the raw data were analyzed by OneSites model using Origin 7. In a typical experiment, 400 μL of solution containing 20 μM of protein in 5 mM sodium phosphate buffer at pH 7 was loaded into the reaction chamber thermostatted at 25°C. The injection syringe was loaded with 200 μL of solution containing 200 μM of NADPH or $NADP^{+}$ or TMP in 5mM of sodium phosphate buffer at pH 7. The reaction protocol involved 25 injections (1.5μL aliquots, over 3 seconds) with 180 seconds spacing time, reference power of 5 μCal/sec, and high feedback mode. Duplicate runs were done and the values were averaged.

8.References and Notes:

1. Carreras C, Santi D (1995) The catalytic mechanism and structure of thymidylate synthase. *Annu Rev Biochem* 64:721-762.

2. Holm L, Rosenström P (2010) Dali server: conservation mapping in 3D. *Nucl Acids Res* 38(web server issue):W545-549.

3. Henikoff S, Endow SA, Greene EA (1996) Connecting protein family resources using the proWeb network. *TIBS* 21(11): 444-445.

4. Corpet F (1988) Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res* 16(22):10881-10890.

5. Von Mering C, et al. (2007) STRING 7-recent developments in the integration and prediction of protein interactions. *Nucleic Acid Res* 35(database issue):D358-D362.

6. Kent WJ, et al. (2002) The human genome browser at UCSC. *Genome Res* 12(6):996-1006.

7. Kent WJ (2002) BLAT - the BLAST-like alignment tool. *Genome Res* 12(4):656-64.

8. Pringle TH: for genus species abbreviations and the full set of full length sequences are provided at http://genomewiki.ucsc.edu/index.php/DHFR_dihydrofolate.

9. Genome 10K Community of Scientists (2009) Genome 10K: a proposal to obtain whole-genome sequence for 10000 vertebrae species. *J Hered* 100(6):659-674.

10. Fierke CA, Johnson KA, Benkovic SJ (1987) Construction and evoluation of the kinetic scheme associated with dihydrofolate reductase from *Escherichia coli*. *Biochemistry* 26(13):4085-4092.

11. Bhabha G. et al. (2011) A dynamic knockout reveals that conformational fluctuations influence the chemical step of enzyme catalysis. *Science* 332(6026):234-238.

12. Appleman JR, et al. (1990) Unusual transient- and steady-state kinetic behavior is predicted by the kinetic scheme operational for recombinant human dihydrofolate reductase. *J Biol Chem* 265(5):2740-2748.

13. Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* 276(A):307-326.

14. Vagin A, Teplyakov A (1997) MOLREP: an automated program for molecular replacement. *J Appl Cryst* 30(6):1022-1025.

15. Murshudov GN et al. (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr Sect D* 67(4):355-367.

16. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr Sect D* 60(12):2126-2132.

17. Xiang JZ, Honig B (2002) JACKAL: A Protein Structure Modeling Package. Columbia University and Howard Hughes Medical Institute, New York.

18. Leung AKW, et al. *RCSB PDB* doi:10.2210/pdb2w3a/pdb.

19. Warshel, A (1991) in *Computer Modeling of Chemical Reactions in Enzymes and Solutions*, (John Wiley & Sons, Inc., New York).

20. Agarwal PK, Billeter SR, Hammes-Schiffer S (2002) Nuclear quantum effects and enzyme dynamics in dihydrofolate reductase catalysis. *J Phys Chem B* 106(12):3283-3293.

21. Åqvist J, Warshel A (1993) Simulation of Enzymes Using Valence Bond Force Fields and Other Hybrid Quantum/Classical Approaches. *Chem Rev* 93(7):2523-2544.

22. Hornak V, et al. (2006) Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins: Struct Funct Bioinf* 65(3):712-725.

23. Cornell WD, et al. (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J Am Chem Soc* 117(5):5179-5197.

24. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of Simple Potential Functions for Simulating Liquid Water. *J Chem Phys* 79(2):926-935.

25. Holmberg N, Ryde U, Bülow L (1999) Redesign of the coenzyme specificity in L-Lactate dehydrogenase from *Bacillus stearothermophilus* using site-directed mutagenesis and media engineering. *Protein Engineering* 12(10):851-856.

26. Bayly CI, Cieplak P, Cornell WD, Kollman P (1993) A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: The RESP model. *J Chem Phys* 97(40):10269-10280.

27. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) Development and testing of a general Amber force field. *J Comput Chem* 25(9):1157-1174.

28. Frisch MJ et al. (2003) *Gaussian 03*, revision E.01 Gaussian, Inc., Pittsburgh, PA.

29. Berendsen HJC, Postma JPM, Van Gunsteren WF, DiNola A, Haak JR (1984) Molecular Dynamics with coupling to an external bath. *J Chem Phys* 81(8):3684-3690.

30. Darden T, York D, Pedersen L (1993) Particle mesh Ewald: An N.Log(N) method for Ewald sums in large systems. *J Chem Phys* 98(12):10089-10092.

31. Ryckaert JP, Ciccotti G, Berendsen HJC (1977) Numerical integration of the cartesian equations of motion of a system with cosntraints: Molecular dynamics of n-alkanes. *J Comput Phys* 23:327.

32. Case DA, et al. (2010) *AMBER 11*. University of California, San Francisco.

33. Melchionna S, Cozzini S (2001) *DLPROTEIN Version 2.1 Molecular Dynamics Software Package for Macromolecules*, INFM UDr SISSA, Triste, Italy.

34. Nosé SA (1984) molecular dynamics method for simulations in the canonical ensemble. *Mol Phys* 52(2):255-268.

35. Hoover WG (1985) Canonical dynamics: Equilibrum phase-space distriutions. *Phys Rev A* 31(3):1695-1697.

36. Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA (1992) The weighted histogram analysis method for free-energy calculations on biomolecules. I. the method. *J Comput Chem* 13(8):1011-1021.

37. Watney JB, Agarwal PK, Hammes-Schiffer S (2003) Effect of Mutation on Enzyme Motion in Dihydrofolate Reductase. *J Am Chem Soc* 125(13):3745-3750.

38. Wong KF, Watney JB, Hammes-Schiffer S (2004) Analysis of Electrostatics and Correlated Motions for Hydride Transfer in Dihydrofolate Reductase. *J Phys Chem B* 108(32):12231-12241.