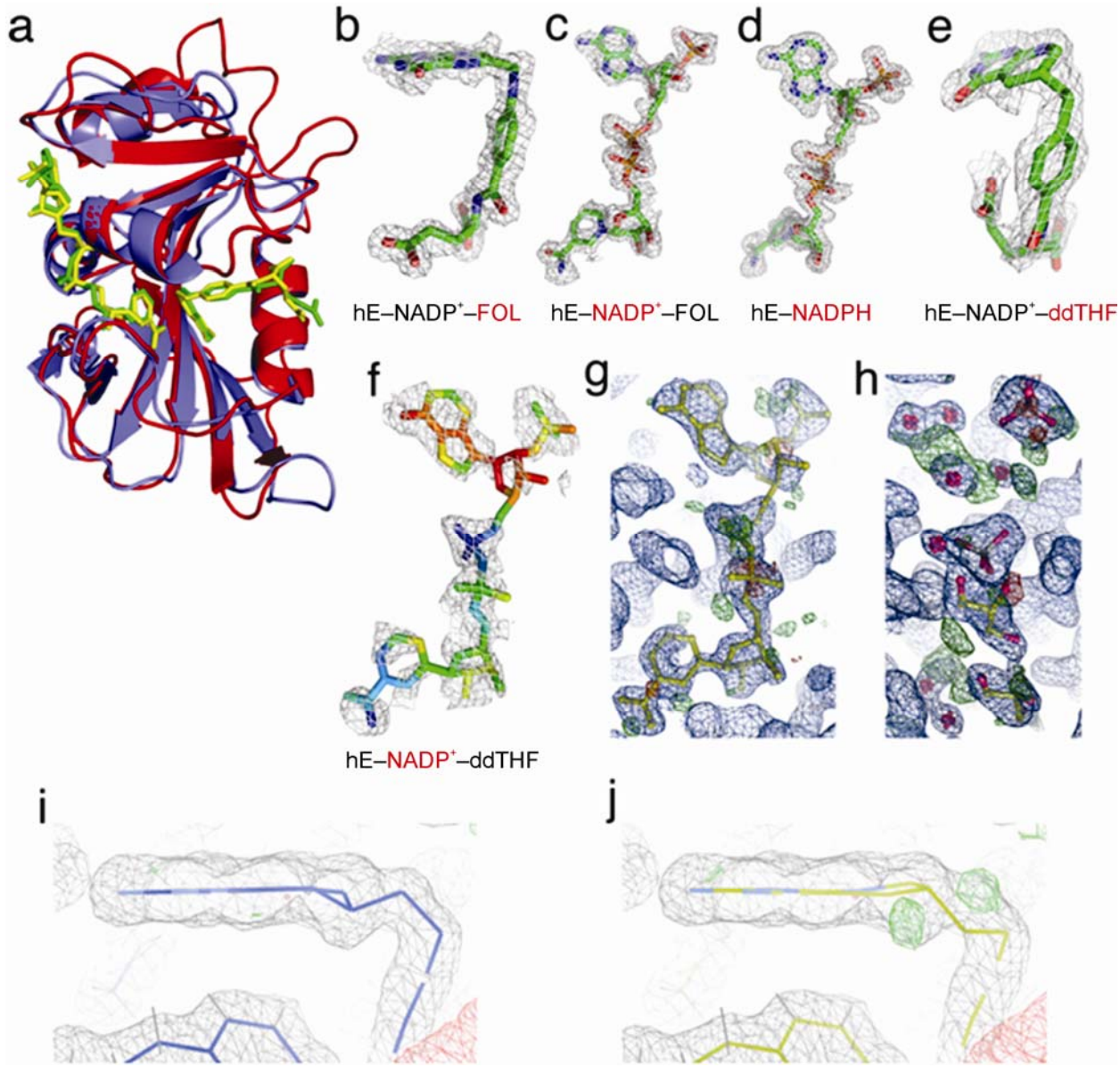
Divergent evolution of protein conformational dynamics in dihydrofolate reductase

Gira Bhabha, Damian C. Ekiert, Madeleine Jennewein, Christian M. Zmasek, Lisa M. Tuttle, Gerard Kroon, H. Jane Dyson, Adam Godzik, Ian A. Wilson and Peter E. Wright

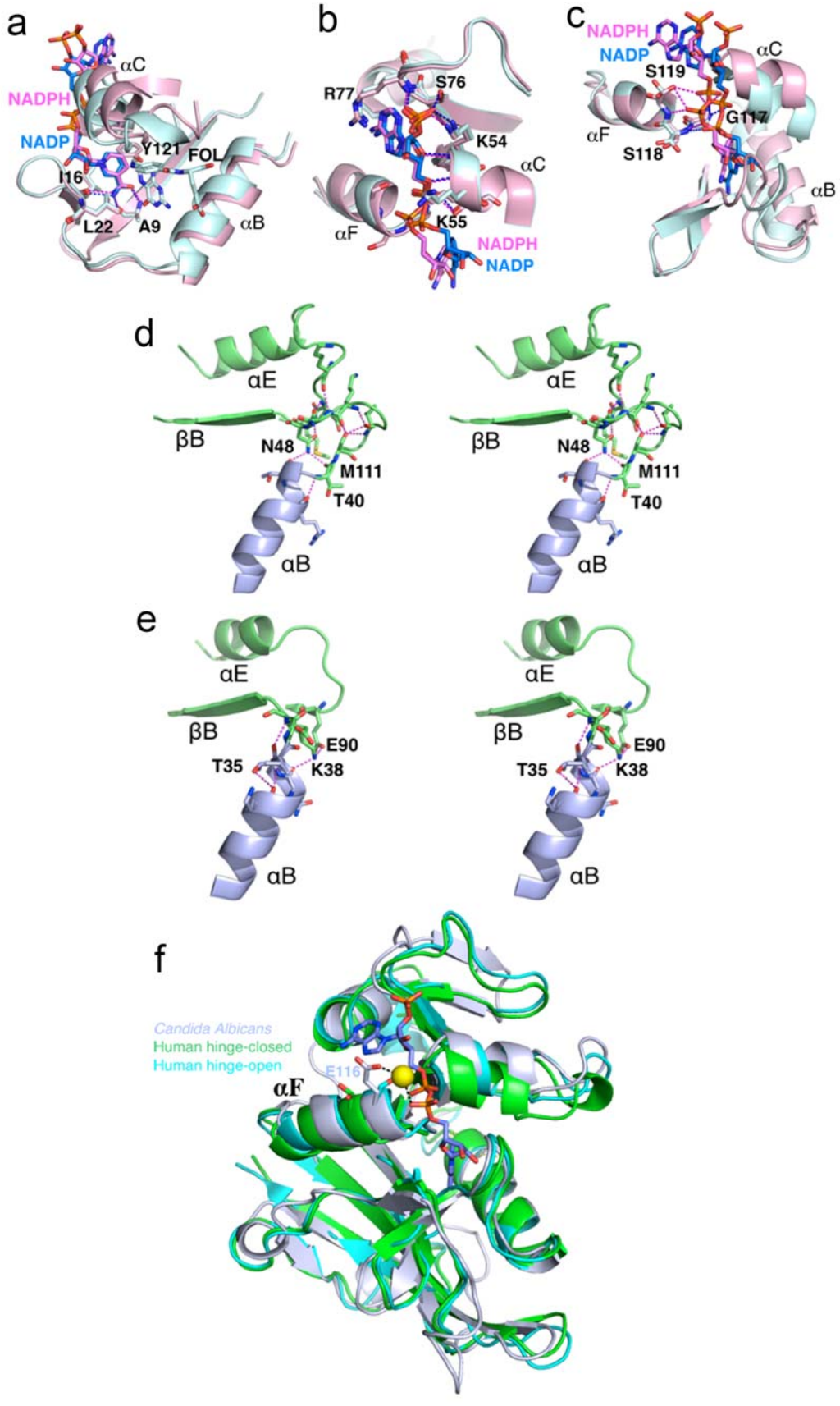
Supplementary Information:

Supplementary Figures

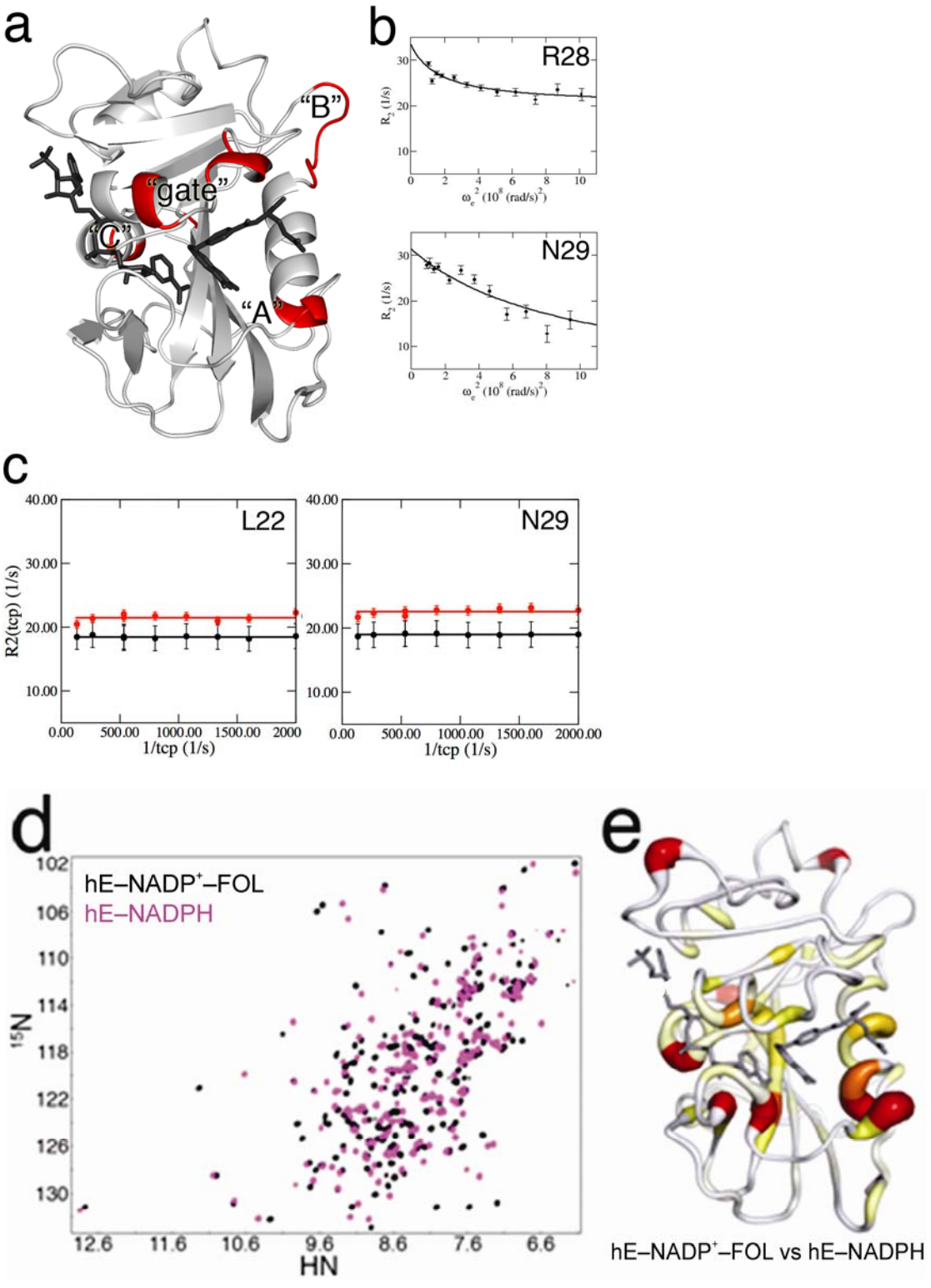


Supplementary Figure 1. X-ray structures and omit maps for ligands built into hDHFR structures. (**a**) Overlay of ecDHFR (PDB code 1RX2[1]) and hDHFR bound to $NADP^+$ and folate, showing that hDHFR and ecDHFR are highly conserved at the structural level. ecDHFR is shown in purple with green ligands, and hDHFR is shown in red with yellow ligands. (**b-f**) $2F_O$-F_C omit maps are shown for each ligand, contoured to σ = 1.0. (**b**) Folate from the hE–$NADP^+$–FOL structure. (**c**) $NADP^+$ from the hE–$NADP^+$–FOL structure. (**d**) NADPH from the hE–NADPH structure. (**e**) ddTHF from the hE–$NADP^+$–ddTHF structure. (**f**) $NADP^+$ from the hE–$NADP^+$–ddTHF structure, the $NADP^+$ is colored by B-factor, after final refinement. (**g-h**) As discussed in the text, the density for this $NADP^+$ ligand was discontinuous, indicating disorder and/or low occupancy. (**g**) Ligand occupancy was refined using PHENIX, and the ligand is

modeled in at an occupancy of ~0.8. $2F_O$-F_C map (σ = 1; blue) and F_O-F_C map (σ = 3; red, negative density; green positive density) after refinement with $NADP^+$ modeled into the hE–$NADP^+$–ddTHF structure. (**h)** $2F_O$-F_C map (σ = 1; blue) and F_O-F_C map (σ = 3; red, negative density; green positive density) after refinement with 2 phosphates, 2 glycerols and waters modeled into the NADP-binding site in the hE–$NADP^+$–ddTHF structure. Remaining positive density is observed in the adenine region and the nicotinamide region, showing that the density is not satisfied upon modeling in buffer molecules. **(i)** $2F_O$-F_C map (σ = 1; grey) and F_O-F_C map (σ = 3; red, negative density; green positive density) after refinement with 6S-ddTHF (blue sticks). **(j)** $2F_O$-F_C map (σ = 1; grey) and F_O-F_C map (σ = 3; red, negative density; green positive density) after refinement with 6R-ddTHF (yellow sticks).


a
αC
NADPH
NADP
Y121
FOL
I16
L22
A9
αB
b
R77
S76
K54
αC
αF
K55
NADPH
NADP
c
NADPH
NADP
αC
S119
G117
αF
S118
αB
d
αE
βB
N48
M111
T40
αB
αE
βB
N48
M111
T40
αB
e
αE
βB
E90
T35
K38
αB
αE
βB
E90
T35
K38
αB
f
Candida Albicans
Human hinge-closed
Human hinge-open
E116
αF

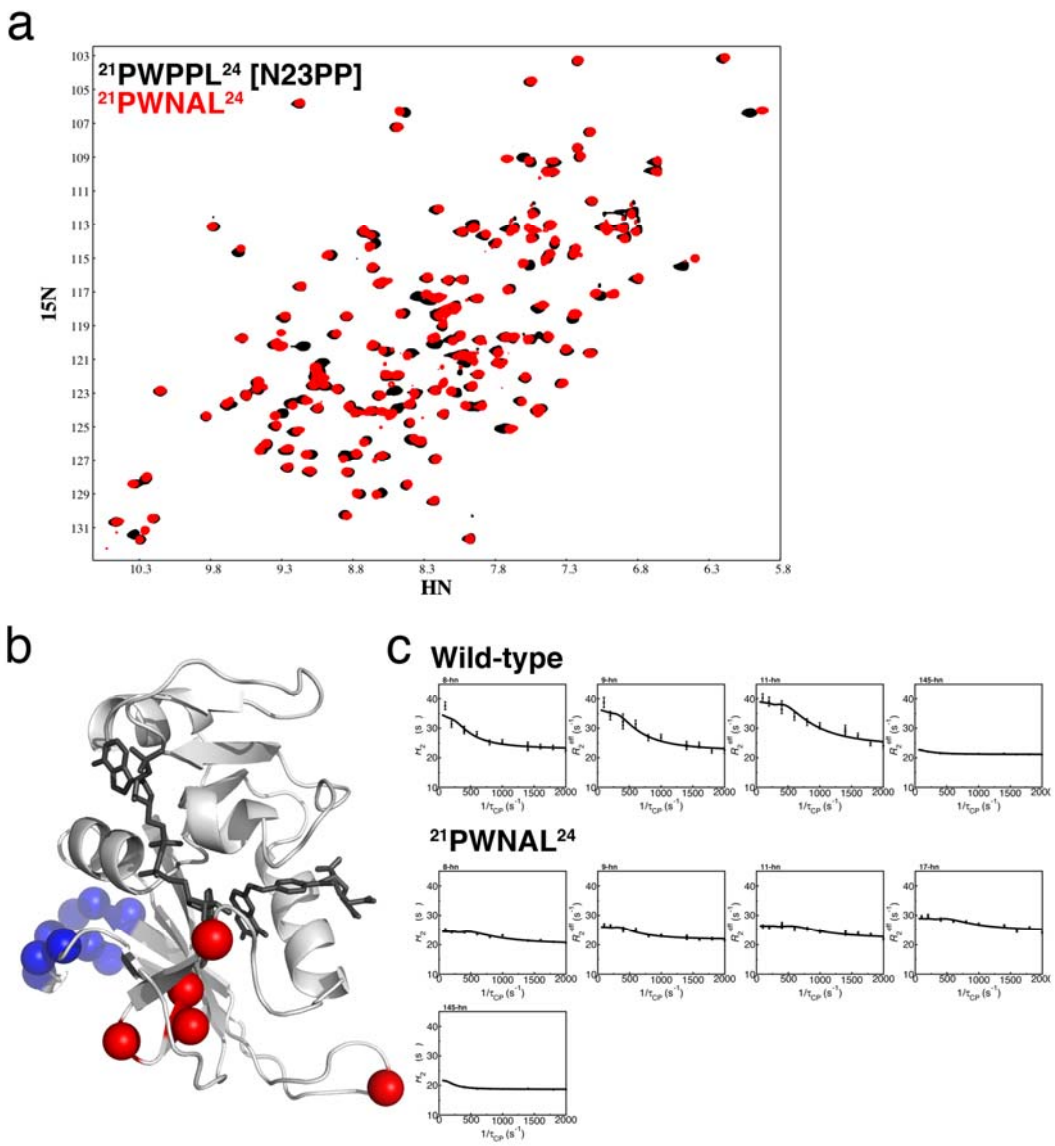
Supplementary Figure 2. Hydrogen bonding networks that support the hinge-open conformation. (a-c) Overlay of hE–$NADP^+$–FOL (cyan) and hE–NADPH (light pink), aligned on the "Met20" loop **(a)**, residues 71-114 **(b)** and helix αF **(c**). $NADP^+$ corresponding to hE–$NADP^+$–FOL is in blue, and NADPH corresponding to the hE–NADPH structure is in pink. Ligands and side chains are colored by element: nitrogen blue, oxygen red, phosphorus orange, carbon as described for each ligand or structure. Hydrogen bonds are shown as blue dashed lines for E–$NADP^+$–FOL and pink dashed lines for E–NADPH. Hydrogen bonds stabilizing the nicotinamide moiety **(a)** and the adenosine moiety **(b)** are present in both conformations. **(c)** Hydrogen bonds from the diphosphate groups of NADPH to Ser119 (pink dashed lines) stabilize the "hinge-open" conformation. These hydrogen bonds are broken in the "hinge-closed" conformation, as other hydrogen bonds are formed to folate (Supplementary Table 2). **(d-e)** Hydrogen bonds for Hinge 1 in hDHFR (d) and ecDHFR (e). Many hydrogen bonds stabilize Hinge 1 in human DHFR. Residues in the loop subdomain are colored purple, and residues in the adenosine-binding subdomain are coloured green. **(f). The hinge-open conformation is stabilized in *C. albicans* DHFR.** An overlay of caDHFR bound to NADPH (purple, PDB code: 1AI9[2]) with hE–$NADP^+$–FOL (green, hinge-closed) and hE–NADPH (cyan, hinge-open) show that the αF helix in cDHFR is aligned with that in hE–NADPH. The structures are aligned on the loop subdomain. Ser119, which stabilizes the hinge-open conformation in hDHFR, has been substituted with Glu116 in caDHFR, which can also stabilize the hinge-open conformation via a water-mediated hydrogen bond to NADPH.



Supplementary Figure 3. Supporting NMR data for hDHFR. (a-c) Numerous residues in hE–NADP$^+$–FOL show ^{15}N $R_{1\rho}$ dispersion at 280 K, but ^{15}N R_2 dispersion is not observed in hDHFR. (**a)** Selected residues showing ^{15}N $R_{1\rho}$ dispersion in regions A, B and C as well as a

region that likely “gates” NADP flux are shown in red on the hE–$NADP^+$–FOL structure. Ligands are shown as sticks. (**b)** Examples of $R_{1\rho}$ dispersion curves. Residue numbers are indicated in the top right for each curve. (**c**) Representative data for hE–$NADP^+$–FOL ^{15}N R_2 dispersion experiments at 500 MHz (black) and 800 MHz (red). These data were collected at 303 K. No dispersion was observed for data collected at 280 K, 292K, 298 K, or 309 K. **(d-e)** ^{15}N ^{1}H chemical shift differences in hDHFR binary and ternary complexes. (**d**) Overlay of hE–$NADP^+$–FOL (black) and hE–NADPH (magenta) ^{15}N HSQC spectra at pH 8.0 and a temperature of 300 K. (**e**) Weighted average ^{1}H, ^{15}N chemical shift differences between hE–$NADP^+$–FOL and hE–NADPH (shown in A), mapped onto the hDHFR structure. Chemical shift differences are colored onto the structure using a gradient of thickness and red to white, with red, thick regions representing the largest chemical shift differences between the two complexes; ligands are shown as dark gray sticks. Large changes are observed in helix αF and hinge 2. Chemical shift differences were calculated as follows:

$$\sqrt{\left(\Delta^{1}\mathrm{H}\right)^{2}+\left(\frac{\Delta^{15}\mathrm{N}}{5}\right)^{2}}$$



Supplementary Figure 4. Supporting NMR data for ^{21}PWNAL^{24} ecDHFR mutant. (a) The ^{21}PWNAL^{24} *E. coli* DHFR mutant is structurally similar to N23PP (or ^{21}PWPPL^{24}) ecDHFR. Overlay of ^{21}PWNAL^{24} ecDHFR ^{15}N HSQC spectrum (red) with N23PP ecDHFR ^{15}N HSQC spectrum (black), with both enzymes bound to $NADP^+$ and FOL, showing that resonances do not shift significantly. **(b-c)** ^{15}N R_2 dispersion is present, but considerably dampened, for a few active site residues in ^{21}PWNAL^{24} ecDHFR. (**b**) Residues showing ^{15}N R2 dispersion in relaxation compensated constant time CPMG experiments are plotted as spheres on the wild type ecDHFR–$NADP^+$–FOL structure (3QL0[3]). Dispersion profiles for C-terminal associated residues (blue) are unaffected by the mutation. All other residues exhibiting dispersion are located in the active site and are shown in red. **(c)**. ^{15}N R_2 dispersion curves for active site residues in the mutant, and corresponding curves for the wild type enzyme. The few residues that do show dispersion in the mutant have significantly reduced R_{ex} values.

a

	Region A	Conformational flexibility across h-transfer step?
E. coli (WT)	PWN-LP--AD	Yes
E. coli (^{21}PWPPL^{24})	PWPPLP--AD	Reduced
E. coli (^{21}PWNAL^{24})	PWNALP--AD	Reduced
B. anthracis	PWR-LP--SE	Yes
S. aureus	PWH-LP--ND	Yes
S. pneumoniae	PWH-LP--AE	Yes
V. cholerae	PWH-LP--AD	Yes
C. elegans	PWRIKK---D	Yes
D. rerio	PWHPIRLSNE	No
R. norvegicus	PWPLLR--NE	No
S. scrofa	PWPPLR--NE	No
B. taurus	PWPPLR--NE	No
H. sapiens	PWPPLR--NE	No

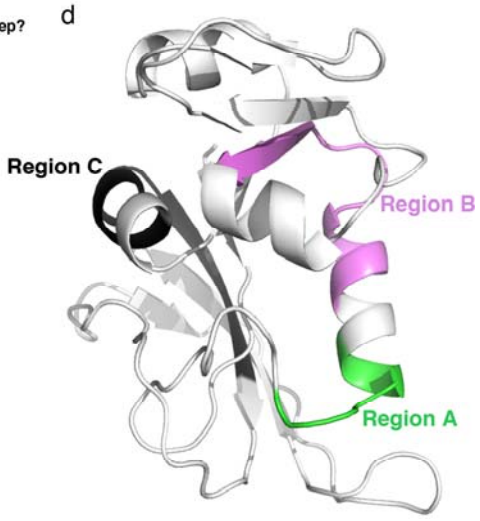
b

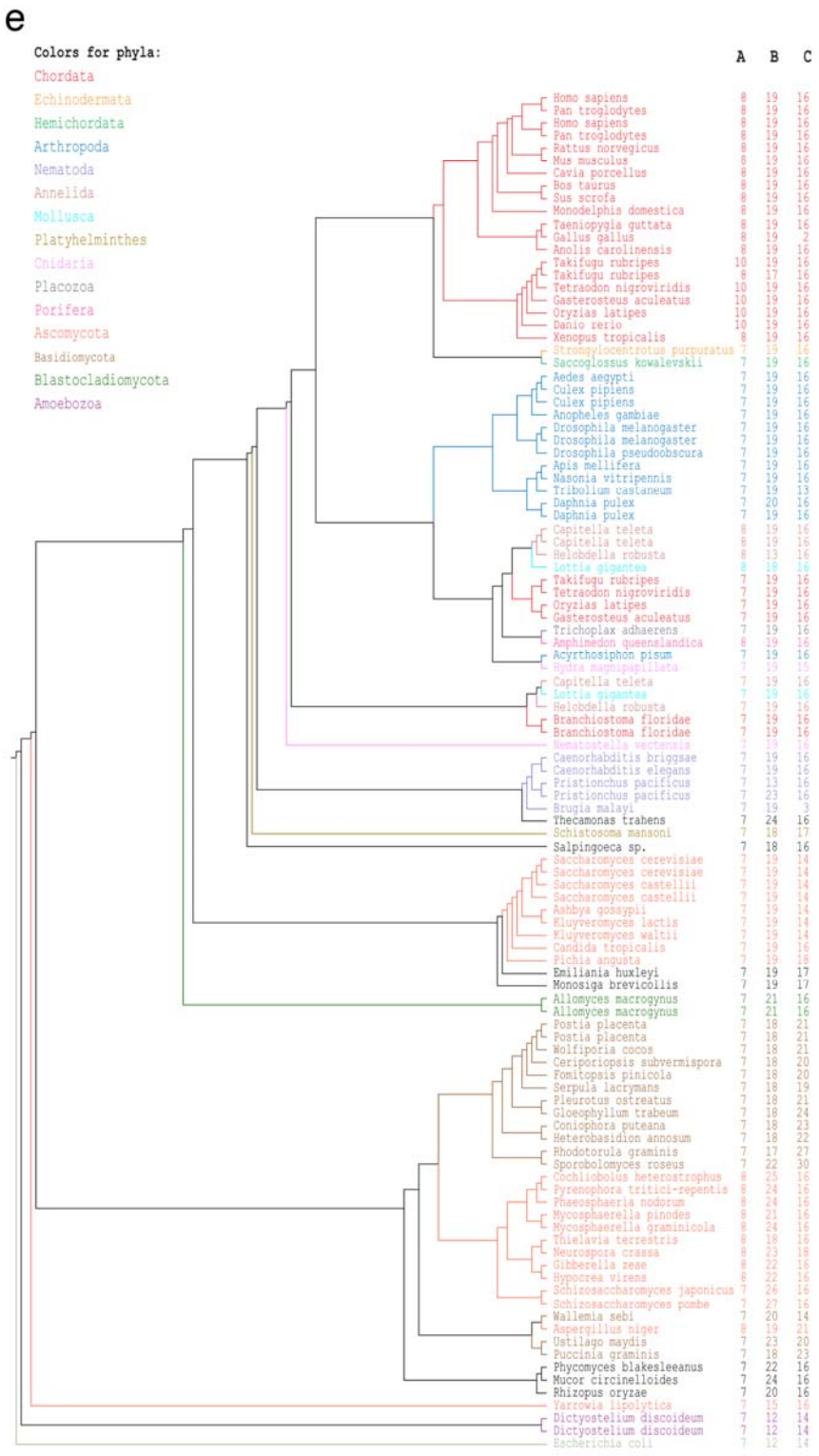
	Region B
E. coli (WT)	FKRNTLN-------KPVIM
E. coli (^{21}PWPPL^{24})	FKRNTLN-------KPVIM
E. coli (^{21}PWNAL^{24})	FKRNTLN-------KPVIM
B. anthracis	VKKTTMG-------HPLIM
S. aureus	VKKLSTG-------HTLVM
S. pneumoniae	FKETTLN-------HAILM
V. cholerae	FKRCTLG-------KPVVM
C. elegans	FASVTKNVSDQSKRNAVLM
D. rerio	FQKMTMTPSDEGKKNVVIM
R. norvegicus	FQRMTTTSSVEGKQNLVIM
S. scrofa	FQRMTTTSSVEGKQNLVIM
B. taurus	FQRMTTVSSVEGKQNLVIM
H. sapiens	FQRMTTTSSVEGKQNLVIM

c

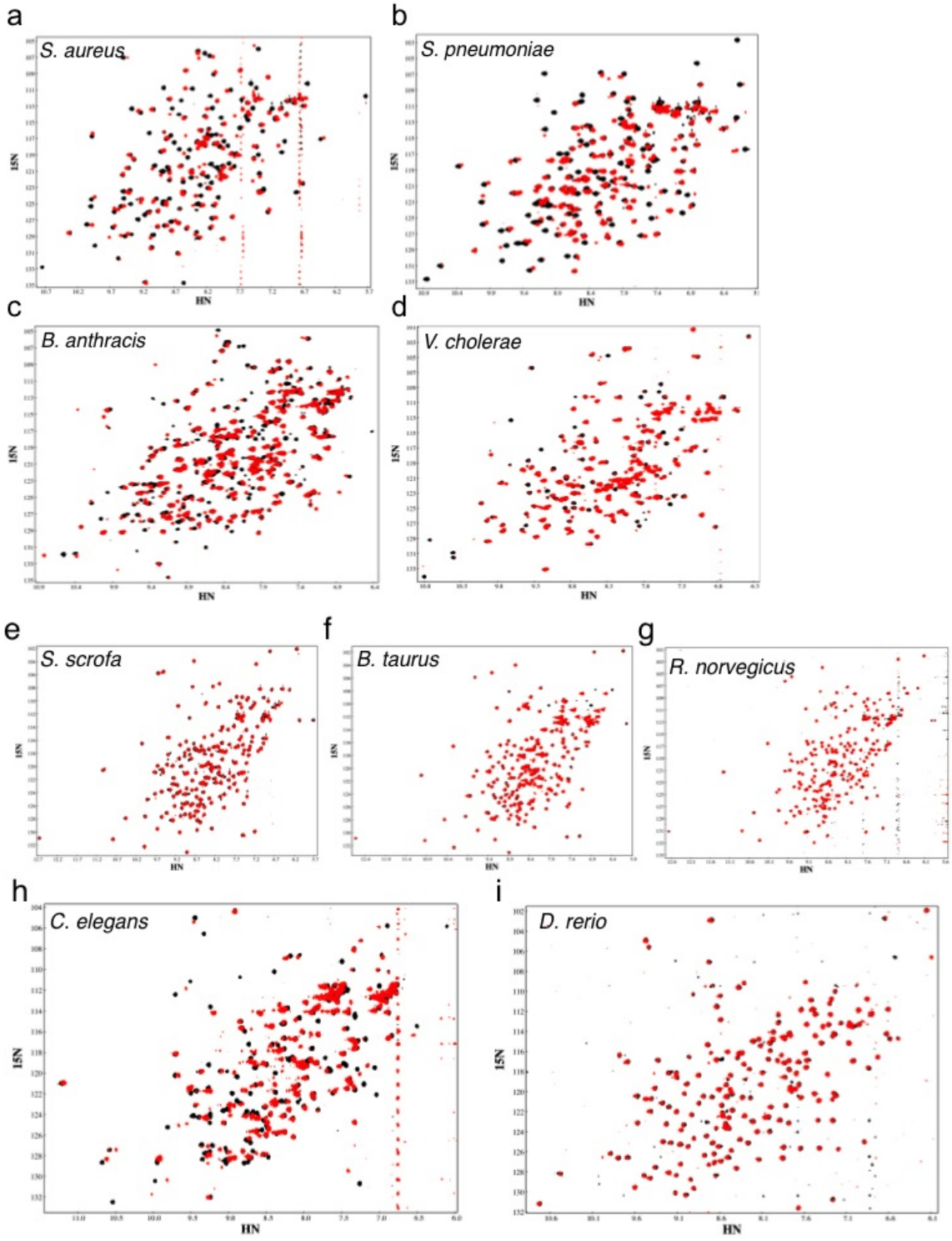
	Region C
E. coli (WT)	YEQFLP--KAQKLYLT
E. coli (^{21}PWPPL^{24})	YEQFLP--KAQKLYLT
E. coli (^{21}PWNAL^{24})	YEQFLP--KAQKLYLT
B. anthracis	YDLFLP--YVDKLYIT
S. aureus	FEEMID--KVDDMYIT
S. pneumoniae	FQAFEP--YLDEVIVT
V. cholerae	YAACLP--MAHKLYIT
C. elegans	YDLALRENLVDEIHLT
D. rerio	YKEVMERSGHRRLFVT
R. norvegicus	YQEAMNQPGHLRLFVT
S. scrofa	YKEAMNKPGHIRLFVT
B. taurus	YKEAMNKPGHVRLFVT
H. sapiens	YKEAMNHPGHLKLFVT

d





Supplementary Figure 5. Sequence analysis of DHFRs from different species. (a-d) Sequence alignments of DHFR enzymes for which NMR data were obtained. **(e)** Tree of eukaryotic DHFR sequences showing lengths of Regions A, B and C. *E. coli* DHFR is also shown for comparison.



Supplementary Figure 6. ^{15}N HSQC spectra for bacterial DHFRs (a-d) and eukaryotic DHFRs (e-i). Spectra show DHFRs in the E–NADP$^+$– FOL (black) or E–NADP$^+$ –THF (red) complexes. Spectra are shown for *S. aureus* (**a**), *S. pneumoniae* (**b**), *B. anthracis* (**c**), *V. cholerae* (**d**), *S. scrofa* (e), *B. taurus* (**f**), *R. norvegicus* (**g**), *C. elegans* (**h**), and *D. rerio* (**i**).

Supplementary Tables

Supplementary Table 1. **Ligand-protein hydrogen bonds in hDHFR**

hE–NADP$^+$–FOL			hE–NADPH		
Donor	**Acceptor**	**Distance**	**Donor**	**Acceptor**	**Distance**
A9-N	NAP-NO7	2.9	A9-N	NDP-NO7	3.0
K54-N	NAP-AO4*	3.0	K54-N	NDP-AO4*	3.0
K54-NZ	NAP-AOP2	2.8	K54-NZ	NDP-AOP2	2.5
K55-N	NAP-AO5*	3.1	K55-N	NDP-AO5*	3.2
T56-N	NAP-AO2	2.9	T56-N	NDP-AO2	3.0
T56-OG1	NAP-AO2	2.7	T56-OG1	NDP-AO2	2.6
S76-OG	NAP-AOP2	2.6	S76-OG	NDP-AOP2	2.5
R77-N	NAP-AOP1	2.7	R77-N	NDP-AOP1	2.8
G117-N	NAP-AO1	3.2	G117-N	NDP-AO1	3.1
			G117-N	NDP-AO2	3.1
S118-N	NAP-NO2	3.0	S118-N	NDP-NO5*	2.9
S118-OG	NAP-NO2	3.3			
			S119-OG	NDP-AO1	2.8
			S119-N	NDP-NO2	2.8
			S119-OG	NDP-NO2	3.1
			NDP-NN7	A9-O	2.8
NAP-AN1	R91-O	3.4			
NAP-NN7	A9-O	2.8			
NAP-NN7	I16-O	3.0	NDP-NN7	I16-O	3.2
NAP-AN7	S119-OG	3.2			
N64-ND2	FOL-O	2.9			
R70-NH1	FOL-O1	2.9			
R70-NH2	FOL-O2	2.8			
FOL-N3	E30-OE1	2.8			
FOL-NA2	E30-OE2	2.8			

Supplementary Table 2. **Table of hydrogen bonds in both hinge regions for human and *E. coli* DHFRs.** The extensive network of hydrogen bonds in the long hinge 1 of hDHFR supports the rigid body hinge-twisting motion that opens the active site.

HINGE 1

Human DHFR, E–NADP+–FOL			***E. coli* DHFR, E–NADP+–FOL**		
Donor	**Acceptor**	**Distance**	**Donor**	**Acceptor**	**Distance**
T39-N	Q35-O	2.9	L36-N	K32-O	2.9
T39-OG	Q35-O	2.7	D37-N	G56-O	2.8
T40-N	R36-O	3.1	K38-N	T35-O	3.2
T40-OG1	R36-O	2.7	K38-NZ	E90-OE1	2.8
S41-OG	V43-O	2.8	K38-NZ	N34-O	2.9
S41-OG	K46-O	3.2			
S41-OG	D110-OD2	3.5			
S42-N	D110-OD2	2.8			
S42-OG	D110-OD2	2.7			
V43-N	S41-OG	3.0			
K46-N	V43-O	3.2			
Q47-N	K108-O	3.0			
N48-ND2	T38-O	2.9			
N48-ND2	T40-O	3.0			
L49-N	M111-O	2.9			
R70-NH1	T38-OG1	2.9			
I71-N	N48-O	2.9			
D110-N	Q47-O	2.8			
M111-N	N48-OD1	2.9			
W113-N	L49-O	2.8			

HINGE 2

Human DHFR, E–NADP+–FOL			**E. coli DHFR, E–NADP+–FOL**		
H127-N	A124-O	3.0	K106-N	F103-O	3.0
H127-ND1	P128-O	2.8			
G129-N	D186-O*	2.8	A107-N	L104-O	3.1
L131-N	K184-O	3.1	R158-NE	A107-O	2.8
K184-NZ	H127-O	2.7			
K184-NZ	G129-O	2.7	R158-NE	A107-O	3.2
K184-N	L131-O	3.0			

*C-terminal residue

Supplementary Table 3. Complete list of fully sequenced genomes analyzed. An excel spreadsheet is also provided.

Species	Common Name	Lineage	Source	Download	Version	Assembly/date	File	Link
Metazoa								
Acyrthosiphon pisum	pea aphid	Ecdysozoa/Arthropoda/Insecta	Baylor	07.05.2010	Acyr_1.0 ()	NCBI Build 1	protein.fa.gz	ftp://ftp.ncbi.nlm.nih.gov/genomes/Acyrthosiphon_pisum/protein/
Aedes aegypti	Yellowfever mosquito	Ecdysozoa/Arthropoda/Insecta	Ensembl	12.07.2007	rel 45		Aedes_aegypti.AaegL1.45.pep.all.fa	http://www.ensembl.org/info/data/download.html
Amphimedon queenslandica	Sponge	Porifera	JGI	12.08.2010	v1	v1	Aqueenslandica_v1.pep.fasta	ftp://ftp.jgi-psf.org/pub/JGI_data/Amphimedon_queenslandica/annotation/
Anolis carolinensis	Green anole	Vertebrata/Amniota	Broad Institute of Harvard and MIT	07.05.2010	rel 57		Anolis_carolinensis.AnoCar1.0.57.pep.all.fa.gz	ftp://ftp.ensembl.org/pub/current_fasta/anolis_carolinensis/pep/
Anopheles gambiae	African malaria mosquito	Ecdysozoa/Arthropoda/Insecta	Ensembl	12.07.2007	rel 45		Anopheles_gambiae.AgamP3.45.pep.all.fa	http://www.ensembl.org/info/data/download.html
Apis mellifera	Honeybee	Ecdysozoa/Arthropoda/Insecta	HGSC at Baylor College of Medicine	16.11.2007	Release1 of the OGS		Amel_release1_OGS_pep.fa	http://genomes.tamu.edu/downloadFASTA.html
Bombyx mori	Silk moth	Ecdysozoa/Arthropoda/Insecta	Beijing Genomics Institute	16.11.2007			SW_ge2k_BGF.pep	http://silkworm.genomics.org.cn/jsp/download.jsp
Bos taurus	Cow	Vertebrata/Mammalia	Ensembl	21.05.2010	rel 57		Bos_taurus.Btau_4.0.57.pep.all.fa.gz	ftp://ftp.ensembl.org/pub/current_fasta/bos_taurus/pep/
Branchiostoma floridae	Florida lancelet, Amphioxus	Cephalochordata	JGI	12.07.2007	1	v.1.0 (March 2006)	Brafl1.fasta	http://genome.jgi-psf.org/Brafl1/Brafl1.home.html
Brugia malayi	Filarial nematode worm	Ecdysozoa/Nematoda	Wormbase	12.11.2009		9/29/2007	bma1.pep.fa.gz	ftp://ftp.wormbase.org/pub/wormbase/genomes/b_malayi/sequences/protein/
Caenorhabditis briggsae	Roundworm	Ecdysozoa/Nematoda	Wormbase	12.11.2009	206	8/29/2009	brigpep.WS206.fa.gz	ftp://ftp.sanger.ac.uk/pub/wormbase/cbriggsae/cb25.agp8
Caenorhabditis elegans	Roundworm	Ecdysozoa/Nematoda	Ensembl	12.07.2007	rel 45		Caenorhabditis_elegans.WB170.45.pep.all.fa	http://www.ensembl.org/info/data/download.html
Canis familiaris	Dog	Vertebrata/Mammalia	Ensembl	07.05.2010	rel 57		Canis_familiaris.BROADD2.57.pep.all.fa.gz	ftp://ftp.ensembl.org/pub/current_fasta/canis_familiaris/pep/
Capitella telata	Bristle worm	Lophotrochozoa/Annelida	JGI	28.08.2007	1	v.1.0 (October 2006)	CAP_FilteredModelsv1.0.aa.fasta	http://genome.jgi-psf.org/Capca1/Capca1.download.ftp.html
Cavia porcellus	Guinea pig	Vertebrata/Mammalia	Ensembl	07.05.2010	rel 57		Cavia_porcellus.cavPor3.57.pep.all.fa.gz	ftp://ftp.ensembl.org/pub/current_fasta/cavia_porcellus/pep/
Ciona intestinalis	Transparent sea squirt	Urochordata/Ascidiacea	Ensembl	07.05.2010	rel 57		Ciona_intestinalis.JGI2.57.pep.all.fa.gz	ftp://ftp.ensembl.org/pub/current_fasta/ciona_intestinalis/pep/
Ciona savignyi	Pacific transparent sea squirt	Urochordata/Ascidiacea	Ensembl	07.05.2010	rel 57		Ciona_savignyi.CSAV2.0.57.pep.all.fa.gz	ftp://ftp.ensembl.org/pub/current_fasta/ciona_savignyi/pep/
Culex pipiens	House mosquito	Ecdysozoa/Arthropoda/Insecta	Broad Institute of Harvard and MIT	18.01.2008		39434	culex_pipiens_3_proteins.fasta.gz	http://www.broad.mit.edu/annotation/genome/culex_pipiens/Downloads.html
Danio rerio	Zebrafish	Vertebrata/Euteleostei	Ensembl	07.05.2010	rel 57		Danio_rerio.Zv8.57.pep.all.fa.gz	ftp://ftp.ensembl.org/pub/current_fasta/danio_rerio/pep/
Daphnia pulex	Water flea	Ecdysozoa/Arthropoda/Crustacea	JGI	12.07.2007	1	v.1.0 (September 2006)	DAPPU_FrozenGeneCatalog_2007_07_03.aa.fasta	http://genome.jgi-psf.org/Dappu1/Dappu1.home.html
Drosophila melanogaster	Fruit fly	Ecdysozoa/Arthropoda/Insecta	Ensembl	07.05.2010	rel 57		Drosophila_melanogaster.BDGP5.13.57.pep.all.fa.gz	ftp://ftp.ensembl.org/pub/current_fasta/drosophila_melanogaster/pep/
Felis catus	Cat	Vertebrata/Mammalia	Ensembl	14.05.2010	rel 57		Felis_catus.CAT.57.pep.all.fa.gz	ftp://ftp.ensembl.org/pub/current_fasta/felis_catus/pep/
Gallus gallus	Chicken	Vertebrata/Amniota	Ensembl	07.05.2010	rel 57		Gallus_gallus.WASHUC2.57.pep.all.fa.gz	ftp://ftp.ensembl.org/pub/current_fasta/gallus_gallus/pep/
Gasterosteus aculeatus	Three-spined stickleback	Vertebrata/Euteleostei	Ensembl	07.05.2010	rel 57		Gasterosteus_aculeatus.BROADS1.57.pep.all.fa.gz	ftp://ftp.ensembl.org/pub/current_fasta/gasterosteus_aculeatus/pep/
Helobdella robusta	Californian leech	Lophotrochozoa/Annelida	JGI	22.09.07		v.1.0 (July 2007)	proteins.Helro1_FilteredModels3.fasta	http://genome.jgi-psf.org/Helro1/Helro1.download.ftp.html
Homo sapiens	Human	Vertebrata/Mammalia	Ensembl	07.05.2010	rel 57		Homo_sapiens.GRCh37.57.pep.all.fa.gz	ftp://ftp.ensembl.org/pub/current_fasta/homo_sapiens/pep/
Hydra magnipapillata	Freshwater polyp	Cnidaria	JCVI	07.05.2010	1 (1/26/2009)	NCBI Build 1	protein.fa.gz	ftp://ftp.ncbi.nlm.nih.gov/genomes/Hydra_magnipapillata/protein/
Ixodes scapularis	Black-legged tick	Ecdysozoa/Arthropoda/Chelicerata	VectorBase	19.05.2010	.1		iscapularis.PEPTIDES-IscaW1.1.fa.gz	http://iscapularis.vectorbase.org/GetData/Downloads?type=Geneset
Lottia gigantea	Owl limpet	Lophotrochozoa/Mollusca	JGI	29.07.2007	1	v.1.0 (08/24/2006)	Lotgi1_GeneModels_FilteredModels1_aa.fasta	http://genome.jgi-psf.org/Lotgi1/Lotgi1.download.ftp.html
Monodelphis domestica	Short-tailed gray opossum	Vertebrata/Mammalia	Ensembl	07.05.2010	rel 57		Monodelphis_domestica.BROADO5.57.pep.all.fa.gz	ftp://ftp.ensembl.org/pub/current_fasta/monodelphis_domestica/
Mus musculus	Mouse	Vertebrata/Mammalia	Ensembl	07.05.2010	rel 57		Mus_musculus.NCBIM37.57.pep.all.fa.gz	ftp://ftp.ensembl.org/pub/current_fasta/mus_musculus/pep/
Nasonia vitripennis	Parasitic wasp	Ecdysozoa/Arthropoda/Insecta	HGSC at Baylor College of Medicine	29.01.2008	.0	12 September 2007	protein.fa.gz	ftp://ftp.ncbi.nlm.nih.gov/genomes/Nasonia_vitripennis/protein
Nematostella vectensis	Starlet sea anemone	Cnidaria	JGI	12.07.2007	1		Nemve1FilteredModels1.fasta	http://genome.jgi-psf.org/Nemve1/Nemve1.home.html
Oikopleura dioica	Tunicate	Urochordata/Appendicularia	Genoscope/Sars centre/MPIMG	14.11.2008		26-Mar-2008	Oikopleura_dioica.SNAP.Od.fasta	http://www.genoscope.cns.fr/externe/Download/Projets/Projet_HG/data/annotation/
Ornithorhynchus anatinus	Duckbill platypus	Vertebrata/Mammalia	Ensembl	07.05.2010	rel 57		Ornithorhynchus_anatinus.OANA5.57.pep.all.fa.gz	ftp://ftp.ensembl.org/pub/current_fasta/ornithorhynchus_anatinus/pep/
Oryzias latipes	Japanese ricefish	Vertebrata/Euteleostei	Ensembl	07.05.2010	rel 57		Oryzias_latipes.MEDAKA1.57.pep.all.fa.gz	ftp://ftp.ensembl.org/pub/current_fasta/oryzias_latipes/pep/
Pan troglodytes	Chimpanzee	Vertebrata/Mammalia	Ensembl	07.05.2010	rel 57		Pan_troglodytes.CHIMP2.1.57.pep.all.fa.gz	ftp://ftp.ensembl.org/pub/current_pan_troglodytes/data/fasta/pep/
Pristionchus pacificus	Roundworm	Ecdysozoa/Nematoda	Wormbase	12.11.2009	205	7/30/2009	ppapep.WS205.fa.gz	ftp://ftp.wormbase.org/pub/wormbase/genomes/p_pacificus/sequences/protein/
Rattus norvegicus	Rat	Vertebrata/Mammalia	Ensembl	07.05.2010	rel 57		Rattus_norvegicus.RGSC3.4.57.pep.all.fa.gz	ftp://ftp.ensembl.org/pub/current_fasta/rattus_norvegicus/pep/
Saccoglossus kowalevskii	Acorn worm	Deuterostomia/Hemichordata	Baylor	29.10.2010	NCBI Build No 1 Ver 1	27 July 2010	protein.fa.gz (annotated proteins)	ftp://ftp.ncbi.nlm.nih.gov/genomes/Saccoglossus_kowalevskii/protein/
Schistosoma mansoni	Blood fluke	Lophotrochozoa/Platyhelminthes	TIGR	07.11.2007		5/14/2007	Smansoni_preliminary_annotation.pep.gz	ftp://ftp.tigr.org/pub/data/Eukaryotic_Projects/s_mansoni/preliminary_annotation
Strongylocentrotus purpuratus	Purple sea urchin	Echinodermata	HGSC at Baylor College of Medicine	05.09.2007		03/09/2007	STRPU_GLEAN3.fa	ftp://ftp.hgsc.bcm.tmc.edu/pub/data/Spurpuratus/fasta/Annotation
Taeniopygia guttata	Zebra finch	Vertebrata/Amniota	WashU/Ensembl	07.05.2010	rel 57		Taeniopygia_guttata.taeGut3.2.4.57.pep.all.fa.gz	ftp://ftp.ensembl.org/pub/current_fasta/taeniopygia_guttata/pep/
Takifugu rubripes	Japanese pufferfish	Vertebrata/Euteleostei	Ensembl	07.05.2010	rel 57		Takifugu_rubripes.FUGU4.57.pep.all.fa.gz	ftp://ftp.ensembl.org/pub/current_fasta/takifugu_rubripes/pep/
Tetraodon nigroviridis	Green puffer	Vertebrata/Euteleostei	Ensembl	07.05.2010	rel 57		Tetraodon_nigroviridis.TETRAODON8.57.pep.all.fa.gz	ftp://ftp.ensembl.org/pub/current_fasta/tetraodon_nigroviridis/pep/
Tribolium castaneum	Red flour beetle	Ecdysozoa/Arthropoda/Insecta	BeetleBase	06.11.2009	3.0	2/10/2009	TRICA_peptide.fa	ftp://bioinformatics.ksu.edu/pub/BeetleBase/3.0/Sequences/Tribolium_Official_Gene_Sequences/
Trichoplax adhaerens		Placozoa	JGI	01.10.2007		v1.0 (30 August 2006)	Triad1_best_proteins.fasta	http://genome.jgi-psf.org/Triad1/Triad1.download.ftp.html
Xenopus tropicalis	Western clawed frog	Vertebrata/Amphibia	Ensembl	06.11.2009	rel 57		Xenopus_tropicalis.JGI4.1.57.pep.all.fa.gz	ftp://ftp.ensembl.org/pub/current_fasta/xenopus_tropicalis/pep/
Choanoflagellates								
Monosiga brevicollis	Choanoflagellate	Choanoflagellida	JGI	12.07.2007	1	v.1.0 (July 2006)	Monbr1_best_proteins.fasta	http://genome.jgi-psf.org/Monbr1/Monbr1.home.html
Salpingoeca rosetta/Proterospongia sp. ATCC 50818		Choanoflagellida	BROAD	02.12.2010			salpingoeca_rosetta_1_proteins.fasta.gz	http://www.broadinstitute.org/annotation/genome/multicellularity_project/MultiDownloads.html
Filasterea								
Capsaspora owczarzaki		Ichthyosporea	BROAD	02.12.2010			capsaspora_owczarzaki_atcc_30864_1_proteins.fasta.gz	http://www.broadinstitute.org/annotation/genome/multicellularity_project/MultiDownloads.html
Fungi								
Acremonium alcalophilum		Ascomycota/Sordariomycetes	JGI	08.04.2011		1	Acral1_c3a_GeneCatalog_proteins_20100903.aa.fasta.gz	http://genome.jgi-psf.org/pages/dynamicOrganismDownload.jsf?organism=Acral1_c3a
Agaricus bisporus var. burnettii JB137-S8		Basidiomycota/Homobasidiomycetes	JGI	19.05.2010	2.0		Abisporus_varburnettii.v2.FilteredModels1.proteins.fasta.gz	http://genome.jgi-psf.org/Agabi_varbur_1/Agabi_varbur_1.download.ftp.html
Allomyces macrogynus		Blastocladiomycota	BROAD	02.12.2010			allomyces_macrogynus_atcc_38327_3_proteins.fasta.gz + allomyces_macrogynus_atcc_38327_mi	http://www.broadinstitute.org/annotation/genome/multicellularity_project/MultiDownloads.html
Ashbya gossypii		Ascomycota/Saccharomycotina	University of Basel	07.11.2007	AGD version 3		982.A_gossypii.fasta.gz	http://www.ebi.ac.uk/integr8/FtpSearch.do?orgProteomeId=982
Aspergillus flavus		Ascomycota/Eurotiales	BROAD	02.01.2008	2			http://www.broad.mit.edu/annotation/genome/aspergillus_group/MultiDownloads.html
Aspergillus niger	Black mold	Ascomycota/Eurotiales	BROAD	02.01.2008	1			http://www.broad.mit.edu/annotation/genome/aspergillus_group/MultiDownloads.html
Candida albicans WO-1		Ascomycota/Saccharomycotina	BROAD	02.11.2007	1 (05/07/2007)		candida_albicans_wo1_1_proteins.fasta	http://www.broad.mit.edu/annotation/genome/candida_albicans/Downloads.html
Candida tropicalis		Ascomycota/Saccharomycotina	BROAD	02.01.2008	3 (05/23/2006)		candida_tropicalis_3_proteins.fasta.gz	http://www.broad.mit.edu/annotation/genome/candida_tropicalis/Downloads.html
Ceriporiopsis subvermispora B		Basidiomycota/Homobasidiomycetes	JGI	02.12.2010	40514	v 1.0 (July 15, 2010)	Ceriporiopsis_subvermispora.proteins.fasta.gz	http://genome.jgi-psf.org/Cersu1/Cersu1.download.ftp.html
Chaetomium globosum	Soil fungus	Ascomycota/Sordariomycetes	BROAD	07.11.2007	1 (04/26/2006)		chaetomium_globosum_1_proteins.fasta	http://www.broad.mit.edu/annotation/genome/chaetomium_globosum/Downloads.html
Coccidioides immitis RS	Valley fever fungus	Ascomycota/Onygenales	BROAD	02.01.2008	2			http://www.broad.mit.edu/annotation/genome/coccidioides_group/MultiDownloads.html
Cochliobolus heterostrophus C5		Ascomycota/Dothideomycetes	JGI	01.07.2008	1	v1.0 (December 2007	CocheC5_1_GeneModels_FilteredModels1_aa.fasta.gz	http://genome.jgi-psf.org/CocheC5_1/CocheC5_1.download.ftp.html
Coniophora putean		Basidiomycota/Homobasidiomycetes	JGI	08.04.2011		1	Conpu1_GeneCatalog_proteins_20101116.aa.fasta.gz	http://genome.jgi-psf.org/pages/dynamicOrganismDownload.jsf?organism=Conpu1
Coprinus cinereus	Inky cap fungus	Basidiomycota/Homobasidiomycetes	Broad Institute of Harvard and MIT	01.11.2007	1 (05/04/2006)		coprinus_cinereus_1_proteins.fasta	http://www.broad.mit.edu/annotation/genome/coprinus_cinereus/Downloads.html
Cryptococcus neoformans A		Basidiomycota/Tremellomycetes	Broad Institute of Harvard and MIT	02.11.2007	1 (04/26/2006)		cryptococcus_neoformans_1_proteins.fasta	http://www.broad.mit.edu/annotation/genome/cryptococcus_neoformans/Downloads.html
Dichomitus squalens		Basidiomycota/Homobasidiomycetes	JGI	08.04.2011		1	Dicsq1_GeneCatalog_proteins_20101110.aa.fasta.gz	http://genome.jgi-psf.org/pages/dynamicOrganismDownload.jsf?organism=Dicsq1
Dothistroma septosporum NZE10 v1.0/Mycosphaerella pini		Ascomycota/Dothideomycetes	JGI	02.12.2010	1.0	v 1.0 (September 30, 2010)	Dotse1_GeneCatalog_proteins_20100818.aa.fasta.gz	http://genome.jgi-psf.org/Dotse1/Dotse1.download.ftp.html
Emericella nidulans		Ascomycota/Eurotiales	BROAD	14.07.2007	V.1 06/19/2006		aspergillus_nidulans_1_proteins.fasta	http://www.broad.mit.edu/annotation/genome/aspergillus_group/MultiHome.html
Encephalitozoon cuniculi	Microsporidian parasite	Microsporidia	Genoscope	29.01.2008	(7/24/2007)		11 files for 11 chromosomes	ftp://ftp.ncbi.nlm.nih.gov/genomes/Fungi/Encephalitozoon_cuniculi/
Fomitopsis pinicola SS1 v1.0		Basidiomycota/Homobasidiomycetes	JGI	16.12.2010	1 (12/15/2010)	v 1.0	Fompi1_GeneCatalog_proteins_20101006.aa.fasta.gz	http://genome.jgi-psf.org/Fompi1/Fompi1.download.ftp.html
Gibberella zeae	Wheat head blight fungus	Ascomycota/Sordariomycetes	BROAD	07.11.2007	3			http://www.broad.mit.edu/annotation/genome/fusarium_group/MultiDownloads.html
Gloeophyllum trabeum		Basidiomycota/Homobasidiomycetes	JGI	02.12.2010	1.0	v 1.0 (November 1, 2010)	Glotr1_1_GeneCatalog_proteins_20100921.aa.fasta.gz	http://genome.jgi-psf.org/Glotr1_1/Glotr1_1.download.ftp.html
Heterobasidion annosum	Root rot fungus	Basidiomycota/Homobasidiomycetes	JGI	27.11.2009	1	v 1.0 (May 2009)	Hannosum.FilteredModels1.proteins.fasta	http://genome.jgi-psf.org/Hetan1/Hetan1.download.ftp.html
Kluyveromyces lactis		Ascomycota/Saccharomycotina	Génolevures	01.11.2007	GL2: 2 (2006/08/17)		Klla-GL2r2.aa	http://cbi.labri.u-bordeaux.fr/Genolevures/download/KLLA_annot.php
Kluyveromyces waltii		Ascomycota/Saccharomycotina	BROAD,SGD	19.10.1910	(Wed Sep 21 2005)		kluyveromyces_waltii.SNAP.aa.gz	http://fungal.genome.duke.edu/annotations/kwal/aa/
Laccaria bicolor	Bicoloured deceiver	Basidiomycota/Homobasidiomycetes	JGI	12.07.2007	1	v.1.0 (March 2005)	LACBI_BestModelsv1.aa.fasta	http://genome.jgi-psf.org/Lacbi1/Lacbi1.home.html
Magnaporthe grisea	Rice blast fungus	Ascomycota/Sordariomycetes	BROAD	07.11.2007	5 (05/07/2007)		magnaporthe_grisea_5_proteins.fasta	http://www.broad.mit.edu/annotation/genome/magnaporthe_grisea/Downloads.html
Melampsora laricis-populina		Basidiomycota/Pucciniomycotina	JGI	17.11.2008	v1.0	v.1.0 (November 2007)	Mlaricis_populina.FilteredModels3.aa.fasta.gz	http://genome.jgi-psf.org/Mellp1/Mellp1.download.ftp.html
Mucor circinelloides		Mucoromycotina (Zygomycota)	JGI	07.05.2010	2.0		Mucor_circinelloides_v2_filtered_proteins.fasta.gz	http://genome.jgi-psf.org/Mucci2/Mucci2.download.ftp.html
Myceliophthora thermophila(=Sporotrichum thermophile)		Ascomycota/Sordariomycetes	JGI	08.04.2011		2	Spoth2_GeneCatalog_proteins_20101221.aa.fasta.gz	http://genome.jgi-psf.org/pages/dynamicOrganismDownload.jsf?organism=Spoth2
Mycosphaerella fijiensis		Ascomycota/Dothideomycetes	JGI	01.11.2007	1	v.1.0 (November 2006)	FilteredModelsv1.0.aa.fasta	http://genome.jgi-psf.org/Mycfi1/Mycfi1.download.ftp.html
Mycosphaerella graminicola		Ascomycota/Dothideomycetes	JGI	01.11.2007	1	v.1.0 (Nov 2005)	proteins.FilteredModels1.fasta	http://genome.jgi-psf.org/Mycgr1/Mycgr1.download.ftp.html
Neosartorya fischeri		Ascomycota/Eurotiales	BROAD	02.01.2008	1			http://www.broad.mit.edu/annotation/genome/aspergillus_group/MultiDownloads.html
Neurospora crassa		Ascomycota/Sordariomycetes	BROAD	14.07.2007	V.3 a. 7 05/07/2007		neurospora_crassa_7_proteins.fasta	http://www.broad.mit.edu/annotation/genome/neurospora/Downloads.html
Phaeosphaeria nodorum	Glume blotch fungus	Ascomycota/Dothideomycetes	Broad Institute of Harvard and MIT	01.11.2007	1 (3/14/2006)		stagonospora_nodorum_1_proteins.fasta	http://www.broad.mit.edu/annotation/genome/stagonospora_nodorum/Downloads.html
Phycomyces blakesleeanus		Mucoromycotina (Zygomycota)	JGI	28.08.2007	1	v.1.0 (September 2006)	Phybl1_best_proteins.fasta	http://genome.jgi-psf.org/Phybl1/Phybl1.download.ftp.html
Pichia angusta/Hansenula polymorpha NCYC 495 leu1.1/		Ascomycota/Saccharomycotina	JGI	02.12.2010	2	v.2.0 (November 17, 2010)	Hanpo2_GeneCatalog_proteins_20100927.aa.fasta.gz	http://genome.jgi-psf.org/Hanpo2/Hanpo2.download.ftp.html
Pleurotus ostreatus	Oyster mushroom	Basidiomycota/Homobasidiomycetes	JGI	07.05.2010	2.0		PleosPC9_1_GeneModels_Filteredmodels2_aa.fasta.gz	http://genome.jgi-psf.org/PleosPC9_1/PleosPC9_1.download.ftp.html
Pneumocystis carinii		Ascomycota/Taphrinomycotina	U Cincinnati	18.01.2008	Sep 21 2005		pneumocystis_carnii.SNAP.aa.gz	http://fungal.genome.duke.edu/annotations/pcar/aa/
Postia placenta		Basidiomycota/Homobasidiomycetes	JGI	01.11.2007	1	v.1.0 (September 2006)	Pospl1_FilteredModels2_proteins.fasta	http://genome.jgi-psf.org/Pospl1/Pospl1.download.ftp.html
Puccinia graminis	Black stem rust fungus	Basidiomycota/Pucciniomycotina	Broad Institute of Harvard and MIT	02.01.2008	2 (06/06/2007)		puccinia_graminis_f._sp._tritici_2_proteins.fasta.gz	http://www.broad.mit.edu/annotation/genome/puccinia_graminis/Downloads.html
Pyrenophora tritici-repentis		Ascomycota/Dothideomycetes	Broad Institute of Harvard and MIT	07.05.2008	1		pyrenophora_tritici-repentis_1_proteins.fasta.zip	http://www.broad.mit.edu/annotation/genome/pyrenophora_tritici_repentis.3/MultiDownloads.html
Rhizopus oryzae		Mucoromycotina (Zygomycota)	Broad Institute of Harvard and MIT	02.11.2007	3 (04/21/2006)		rhizopus_oryzae_3_proteins.fasta	http://www.broad.mit.edu/annotation/genome/rhizopus_oryzae/Downloads.html
Rhodotorula graminis strain WP1		Basidiomycota/Pucciniomycotina	JGI	02.12.2010	1.1	v 1.1 (November 17, 2010)	Rhoba1_1_GeneCatalog_proteins_20100913.aa.fasta.gz	http://genome.jgi-psf.org/Rhoba1_1/Rhoba1_1.download.ftp.html
Saccharomyces castellii	Yeast	Ascomycota/Saccharomycotina	Washington University	08.01.2008	(10/4/2003)		orf_trans.fasta.gz	ftp://genome-ftp.stanford.edu/pub/yeast/sequence/fungal_genomes/S_castellii/WashU/orf_protein/
Saccharomyces cerevisiae	Baker's yeast	Ascomycota/Saccharomycotina	Stanford University	16.07.2007	38844		yeast_nrpep.fasta	ftp://genome-ftp.stanford.edu/pub/yeast/sequence/GenBank
Schizosaccharomyces japonicus	Fission yeast	Ascomycota/Taphrinomycotina	Broad Institute of Harvard and MIT	14.07.2007	V.1 2/20/2007		schizosaccharomyces_japonicus_1_proteins.fasta	http://www.broad.mit.edu/annotation/genome/schizosaccharomyces_japonicus/Downloads.html
Schizosaccharomyces pombe	Fission yeast	Ascomycota/Taphrinomycotina	Sanger	16.07.2007	v 19 07/09/2007		pompep.fasta	ftp://ftp.sanger.ac.uk/pub/yeast/pombe/Protein_data/pompep
Septoria musiva SO2202/Mycosphaerella populorum		Ascomycota/Dothideomycetes	JGI	02.12.2010	1.0	v 1.0 (November 16, 2010)	Sepmu1_GeneCatalog_proteins_20100915.aa.fasta.gz	http://genome.jgi-psf.org/Sepmu1/Sepmu1.download.ftp.html
Serpula lacrymans	Dry rot fungus	Basidiomycota/Homobasidiomycetes	JGI	14.05.2010	2.0		Serpula_lacrymans_S7_3_v2.proteins.fasta.gz	http://genome.jgi-psf.org/SerlaS7_3_2/SerlaS7_3_2.download.ftp.html
Spizellomyces punctatus		Chytridiomycota	BROAD	02.12.2010			spizellomyces_punctatus_daom_br117_1_proteins.fasta.gz + spizellomyces_punctatus_daom_br117	http://www.broadinstitute.org/annotation/genome/multicellularity_project/MultiDownloads.html
Sporobolomyces roseus		Basidiomycota/Pucciniomycotina	JGI	21.08.2007	1	v.1.0 (Apr 2006)	Sroseus.FilteredModels1.aa.fasta	http://genome.jgi-psf.org/Sporo1/Sporo1.home.html
Thielavia terrestris		Ascomycota/Sordariomycetes	JGI	27.11.2009	1	v. 1.0 (September 2009)	Thite1_GeneModels_FilteredModels2_aa.fasta	http://genome.jgi-psf.org/Thite1/Thite1.download.ftp.html
Trametes versicolor		Basidiomycota/Homobasidiomycetes	JGI	08.04.2011		1	Trave1_GeneCatalog_proteins_20101111.aa.fasta.gz	http://genome.jgi-psf.org/pages/dynamicOrganismDownload.jsf?organism=Trave1
Tremella mesenterica	Jelly fungus	Basidiomycota/Tremellomycetes	JGI	27.11.2009	1	v.1.0 (June 2009)	Treme1_best_proteins.fasta	http://genome.jgi-psf.org/Treme1/Treme1.download.ftp.html
Trichoderma virens		Ascomycota/Sordariomycetes	JGI	01.01.2008	1	v 1.0	proteins.FM3_Trive1.fasta.gz	http://genome.jgi-psf.org/Trive1/Trive1.download.ftp.html
Uncinocarpus reesii		Ascomycota/Onygenales	BROAD	02.01.2008	2 (05/12/2006)		uncinocarpus_reesii_2_proteins.fasta.gz	http://www.broad.mit.edu/annotation/genome/uncinocarpus_reesii/Downloads.html
Ustilago maydis	Smut fungus	Basidiomycota/Ustilaginomycotina	Broad Institute of Harvard and MIT	02.11.2007	1 (04/26/2006)		ustilago_maydis_1_proteins.fasta	http://www.broad.mit.edu/annotation/genome/ustilago_maydis/Downloads.html
Wallemia sebi		Basidiomycota/Wallemiomycetes	JGI	02.12.2010	1.0	v 1.0 (October 11, 2010)	Walse1_GeneCatalog_proteins_20100910.aa.fasta.gz	http://genome.jgi-psf.org/Walse1/Walse1.download.ftp.html
Wolfiporia cocos MD-104 SS10		Basidiomycota/Homobasidiomycetes	JGI	02.12.2010	40514	v 1.0 (November 24, 2010)	Wolco1_GeneCatalog_proteins_20100915.aa.fasta.gz	http://genome.jgi-psf.org/Wolco1/Wolco1.download.ftp.html
Yarrowia lipolytica		Ascomycota/Saccharomycotina	Génolevures	01.11.2007	GL2: 2 (2006/05/22)		Yali-GL2r2.aa	http://cbi.labri.u-bordeaux.fr/Genolevures/download/YALI_annot.php
Apusozoa								
Thecamonas trahens		Apusozoa/Apusomonadidae	BROAD	02.12.2010			thecamonas_trahens_atcc_50062_1_proteins.fasta.gz	http://www.broadinstitute.org/annotation/genome/multicellularity_project/MultiDownloads.html
Amoebozoa								
Dictyostelium discoideum	Slime mold	Mycetozoa	dictyBase/HGSC at Baylor College of Medici	29.07.2007	39278.0875		dicty_primary_protein	http://dictybase.org/db/cgi-bin/dictyBase/download/blast_databases.pl
Dictyostelium purpureum QSDP1	Slime mold	Mycetozoa	JGI	19.09.2008	1.0	v1.0 (August 18 2008)	Dicpu1_best_proteins.fasta.gz	http://genome.jgi-psf.org/Dicpu1/Dicpu1.download.ftp.html
Entamoeba histolytica		Entamoebidae	TIGR	07.11.2007	(12/22/2004)		EHA1.pep	ftp://ftp.tigr.org/pub/data/Eukaryotic_Projects/e_histolytica/whole_genome_sequencing/EHA1.pep
Archaeplastida								
Aquilegia coerulea	Rocky mountain columbine	Viridiplantae/Embryophyta	JGI	11.04.2011	v6.0	8/26/10	Acoerulea_151_peptide.fa.gz	ftp://ftp.jgi-psf.org/pub/JGI_data/phytozome/v6.0/
Arabidopsis lyrata	Lyre-leaved rock-cress	Viridiplantae/Embryophyta	JGI	03.10.2008	v1.0	v1.0 (April 7, 2008)	Araly1_GeneModels_FilteredModels6_aa.fasta	http://genome.jgi-psf.org/Araly1/Araly1.download.ftp.html
Arabidopsis thaliana	Mouse-ear cress	Viridiplantae/Embryophyta	TAIR	14.07.2007	TAIR7		TAIR7_pep_20070425	ftp://ftp.arabidopsis.org/home/tair/Genes/TAIR7_genome_release/TAIR7_blastsets
Brachypodium distachyon	Purple false brome	Viridiplantae/Embryophyta	JGI	04.12.2011	v6.0	8/30/10	Bdistachyon_114_peptide.fa.gz	ftp://ftp.jgi-psf.org/pub/JGI_data/phytozome/v6.0/
Carica papaya	Papaya	Viridiplantae/Embryophyta	JGI	04.12.2011	v6.0	12/13/10	Cpapaya_113_peptide.fa.gz	ftp://ftp.jgi-psf.org/pub/JGI_data/phytozome/v6.0/
Chlamydomonas reinhardtii		Chlorophyta/Chlorophyceae	JGI	12.07.2007	3		Chlre3_1.GeneCatalogProteins.6JUL06.fasta	http://genome.jgi-psf.org/Chlre3/Chlre3.home.html
Chlorella sp. NC64A		Chlorophyta/Trebouxiophyceae	JGI	15.06.2008	1	1.0 (June 13, 2008)	Chlorella_NC64A.best_proteins.fasta.gz	http://genome.jgi-psf.org/ChlNC64A_1/ChlNC64A_1.download.ftp.html
Citrus clementina		Viridiplantae/Embryophyta	JGI	04.12.2011	v6.0	1/15/11	Cclementina_165_peptide.fa.gz	ftp://ftp.jgi-psf.org/pub/JGI_data/phytozome/v6.0/
Citrus sinensis	Sweet orange	Viridiplantae/Embryophyta	JGI	04.12.2011	v6.0	1/15/11	Csinensis_154_peptide.fa.gz	ftp://ftp.jgi-psf.org/pub/JGI_data/phytozome/v6.0/
Coccomyxa sp. C-169/Chlorella vulgaris C-169		Chlorophyta/Trebouxiophyceae	JGI	02.12.2010	2	2.0 (December 18, 2009)	Coccomyxa_C169_v2_filtered_proteins.fasta.gz	http://genome.jgi-psf.org/Coc_C169_1/Coc_C169_1.download.ftp.html
Cucumis sativis	Garden cucumber	Viridiplantae/Embryophyta	JGI	07.05.2010	1.0		Cucsa_v1.pep.fa.gz	http://genome.jgi-psf.org/cucumber/cucumber.download.ftp.html
Cyanidioschyzon merolae	Red alga	Rhodophyta	National Institute of Genetics, Japan	08.01.2008	(17-Sep-2005)		cds.fasta	http://merolae.biol.s.u-tokyo.ac.jp/download/
Eucalyptus grandis		Viridiplantae/Embryophyta	JGI	04.12.2011	v6.0	1/15/11	Egrandis_162_peptide.fa.gz	ftp://ftp.jgi-psf.org/pub/JGI_data/phytozome/v6.0/
Glycine max	Soybean	Viridiplantae/Embryophyta	JGI	09.11.2009	1.0	(12/17/2008)	Glyma1.pep.fa.gz	ftp://ftp.jgi-psf.org/pub/JGI_data/Glycine_max/Glyma1/annotation/
Manihot esculenta	Tapioca	Viridiplantae/Embryophyta	JGI	07.05.2010	1.1		cassava_v1.1.pep.filter.fa	http://genome.jgi-psf.org/cassava/cassava.download.ftp.html
Medicago truncatula	Barrel medic	Viridiplantae/Embryophyta	JGI	04.12.2011	v6.0	8/26/11	Mtruncatula_135_peptide.fa.gz	ftp://ftp.jgi-psf.org/pub/JGI_data/phytozome/v6.0/
Micromonas pusilla CCMP1545		Chlorophyta/Prasinophyceae	JGI	07.03.2008	2	v2.0(Feb, 2008)	MicromonasCCMP1545.FrozenGeneCatalog_20080206.aa.fasta.gz	http://genome.jgi-psf.org/MicpuC2/MicpuC2.download.ftp.html
Micromonas strain RCC299		Chlorophyta/Prasinophyceae	JGI	07.03.2008	2	v2.0 (Feb 2008)	MicromonasRCC299.FrozenGeneCatalog_20080206.aa.fasta.gz	http://genome.jgi-psf.org/MicpuN2/MicpuN2.download.ftp.html
Mimulus nasutus	Yellow monkey flower	Viridiplantae/Embryophyta	JGI	07.05.2010	1.0		Mimulus.guttatus.v1.0.peptide.fa.gz	http://genome.jgi-psf.org/mimulus/mimulus.download.ftp.html
Oryza sativa	Rice	Viridiplantae/Embryophyta	TIGR	14.07.2007	Release 5		osa_chr1_12_tigr.fasta	http://www.tigr.org/tdb/e2k1/osa1/data_download.shtml
Ostreococcus lucimarinus		Chlorophyta/Prasinophyceae	JGI	12.07.2007	2		O.lucimarinus.FM.aa.fasta	http://genome.jgi-psf.org/Ost9901_3/Ost9901_3.home.html
Ostreococcus sp. RCC809		Chlorophyta/Prasinophyceae	JGI	14.08.2008	1	v.1.0 (August 2008)	OstRCC809.FM1.aa.fasta.gz	http://genome.jgi-psf.org/OstRCC809_1/OstRCC809_1.info.html
Ostreococcus tauri		Chlorophyta/Prasinophyceae	JGI	22.09.2007	2		O.tauri.FM.aa.fasta	http://genome.jgi-psf.org/Ostta4/Ostta4.download.ftp.html
Physcomitrella patens ssp patens	Moss	Viridiplantae/Embryophyta	JGI	22.09.2007	1.1	v.1.1 (March 2007)	proteins.Phypa1_1.FilteredModels3.fasta	http://genome.jgi-psf.org/Phypa1_1/Phypa1_1.download.ftp.html
Populus trichocarpa	Western balsam poplar	Viridiplantae/Embryophyta	JGI	10.02.2007	1.1	v1.0 (June 2004)	proteins.Poptr1_1.JamboreeModels.fasta	http://genome.jgi-psf.org/Poptr1_1/Poptr1_1.download.ftp.html
Prunus persica	Peach	Viridiplantae/Embryophyta	JGI	04.12.2011	v6.0	8/26/10	Ppersica_139_peptide.fa.g	ftp://ftp.jgi-psf.org/pub/JGI_data/phytozome/v6.0/
Ricinus communis	Castor bean	Viridiplantae/Embryophyta	JGI	04.12.2011	v6.0	8/30/10	Rcommunis_119_peptide.fa.g	ftp://ftp.jgi-psf.org/pub/JGI_data/phytozome/v6.0/
Selaginella moellendorffii	Spikemoss	Viridiplantae/Embryophyta	JGI	21.12.2007	1	v1.0 (March 2007)	Selmo1_GeneModels_FilteredModels3_aa	http://genome.jgi-psf.org/Selmo1/Selmo1.download.ftp.html
Setaria italica	Foxtail millet	Viridiplantae/Embryophyta	JGI	04.12.2011	v6.0	8/30/10	Sitalica_155_peptide.fa.gz	ftp://ftp.jgi-psf.org/pub/JGI_data/phytozome/v6.0/
Sorghum bicolor	Sorghum	Viridiplantae/Embryophyta	JGI	21.04.2008	1.0	v1.0 (March 2007)	Sorbi1_GeneModels_FilteredModels6_aa.fasta.gz	http://genome.jgi-psf.org/Sorbi1/Sorbi1.download.ftp.html
Vitis vinifera	Grape	Viridiplantae/Embryophyta	Genoscope	23.11.2007	v1	(06-Sep-2007)	Vitis_vinifera_peptide_v1.fa	http://www.genoscope.cns.fr/externe/English/Projets/Projet_ML/data/annotation/
Volvox carteri f. nagariensis		Chlorophyta/Chlorophyceae	JGI	12.07.2007	1	v.1.0 (March 2007)	VOLCA.proteins.FilteredModels2.fasta.gz	http://genome.jgi-psf.org/Volca1/Volca1.home.html
Zea mays	Maize	Viridiplantae/Embryophyta	JGI	04.12.2011	v6.0	8/26/10	Zmays_121_peptide.fa.gz	ftp://ftp.jgi-psf.org/pub/JGI_data/phytozome/v6.0/
Chromalveolates								
Aureococcus anophagefferens		stramenopiles/Pelagophyceae	JGI	10.01.2007	1	v.1.0 (June 2007)	proteins.Auran1_FilteredModels3.fasta	http://genome.jgi-psf.org/Auran1/Auran1.download.ftp.html
Cryptosporidium hominis TU502		Alveolata/Apicomplexa/Coccidia	NIH/NIAID/ApiDB	07.11.2007	rel 3.0 (11-May-2007)		ChominisAnnotatedProtein.fsa	http://apidb.org/common/downloads/release3.0/CryptoPlasmoToxo_downloads/ChominisProteins/
Cryptosporidium parvum IOWA		Alveolata/Apicomplexa/Coccidia	NIH/NIAID/ApiDB	07.11.2007	rel 3.0 (11-May-2007)		CparvumAnnotatedProtein.fsa	http://apidb.org/common/downloads/release3.0/CryptoPlasmoToxo_downloads/CparvumProteins/
Emiliania huxleyi CCMP1516		Haptophyta	JGI	29.04.2008	V 1.0	v.1.0 (3 November 2006)	Emihu1_best_proteins.fasta.gz	http://genome.jgi-psf.org/Emihu1/Emihu1.download.ftp.html
Fragilariopsis cylindrus		stramenopiles/Bacillariophyta (diatoms)	JGI	27.11.2009	1	(September 30 2009)	Fracy1_GeneModels_FilteredModels1_aa.fasta	http://genome.jgi-psf.org/Fracy1/Fracy1.download.ftp.html
Paramecium tetraurelia		Alveolata/Ciliophora	Genoscope	31.07.2007	v1		Paramecium_peptide_v1.fa	http://www.genoscope.cns.fr/externe/Francais/Projets/Projet_FN/data/annotation/
Phaeodactylum tricornutum		stramenopiles/Bacillariophyta (diatoms)	JGI	12.07.2007	2	v.2.0 (August 2006)	Phatr2_geneModels_FilteredModels2_aa.fasta	http://genome.jgi-psf.org/Phatr2/Phatr2.home.html
Phytophthora infestans	Late blight agent	stramenopiles/Oomycetes (water molds)	Broad Institute of Harvard and MIT	14.01.2008	1		phytophthora_infestans_1_proteins.fasta.gz	http://www.broad.mit.edu/annotation/genome/phytophthora_infestans/Downloads.html
Phytophthora ramorum	Sudden oak death agent	stramenopiles/Oomycetes (water molds)	JGI	01.11.2007	1.1	v1.0 (April 2004)	proteins.FM_Phyra1_1.fasta	http://genome.jgi-psf.org/Phyra1_1/Phyra1_1.download.ftp.html
Phytophthora sojae	Root and stem rot agent	stramenopiles/Oomycetes (water molds)	JGI	10.02.2007	1.1	v1.0 (April 2004)	proteins.FM_Physo1_1.fasta	http://genome.jgi-psf.org/Physo1_1/Physo1_1.download.ftp.html
Plasmodium chabaudi AS		Alveolata/Apicomplexa/Aconoidasida	NIH/NIAID/ApiDB	07.11.2007	rel 3.0 5-3 (11-May-2007)		PchabaudiAnnotatedProteins_plasmoDB-5.3.fasta	http://apidb.org/common/downloads/release3.0/CryptoPlasmoToxo_downloads/PchabaudiProteins/
Plasmodium falciparum 3D7	Malaria parasite	Alveolata/Apicomplexa/Aconoidasida	NIH/NIAID/ApiDB	07.11.2007	rel 3.0 5-3 (11-May-2007)		PfalciparumAnnotatedProteins_plasmoDB-5.3.fasta	http://apidb.org/common/downloads/release3.0/CryptoPlasmoToxo_downloads/PfalciparumProteins/
Plasmodium yoelii yoelii		Alveolata/Apicomplexa/Aconoidasida	TIGR	29.01.2008	(7/23/2004)		PYA1.pep	ftp://ftp.tigr.org/pub/data/Eukaryotic_Projects/p_yoelii/annotation_dbs/
Tetrahymena thermophila		Alveolata/Ciliophora	TIGR	14.07.2007	8313006		TTA1_08302006.pep	ftp://ftp.tigr.org/pub/data/Eukaryotic_Projects/t_thermophila/annotation_dbs/interim_annotation_release_08313006/TTA1_08302006.pep
Thalassiosira pseudonana		stramenopiles/Bacillariophyta (diatoms)	JGI	30.07.2007	3	v3.0 (August 2006)	Thaps3_geneModels_FilteredModels2_aa.fasta	http://genome.jgi-psf.org/Thaps3/Thaps3.download.ftp.html
Theileria annulata Ankara C9		Alveolata/Apicomplexa/Aconoidasida	Sanger	12.08.2008	7/15/2005		TANN.GeneDB.pep	ftp://ftp.sanger.ac.uk/pub/pathogens/T_annulata/
Theileria parva Muguga	East coast fever parasite	Alveolata/Apicomplexa/Aconoidasida	TIGR/ILRI	29.01.2008	(7/25/2007)		3 files	ftp://ftp.ncbi.nlm.nih.gov/genomes/Protozoa/Theileria_parva/
Toxoplasma gondii ME49		Alveolata/Apicomplexa/Coccidia	NIH/NIAID/ApiDB	07.11.2007	rel 3.0 4-2 (11-May-2007)		TgondiiAnnotatedProteins_toxoDB-4.2.fasta	http://apidb.org/common/downloads/release3.0/CryptoPlasmoToxo_downloads/TgondiiProteins/
Excavata								
Giardia lamblia	Giardiasis agent	Metamonada/Diplomonadida	MBL/GiardiaDB	07.11.2007	1.0 (21-Sep-2007)		GlambliaAnnotatedProteins_GiardiaDB-1.0.fasta	http://www.giardiadb.org/common/downloads/release1.0/
Leishmania major Friedlin		Euglenozoa/Kinetoplastida	Sanger	08.11.2007	(5/11/2006)		GeneDB_Protein_database_110506	ftp://ftp.sanger.ac.uk/pub/databases/L.major_sequences/LEISHPEP/
Naegleria gruberi	Amoeba	Heterolobosea	JGI	12.07.2007	1	v.1.0 (August 2006)	Naegr1_best_poroteins.fasta	http://genome.jgi-psf.org/Naegr1/Naegr1.home.html
Trichomonas vaginalis G3		Metamonada/Parabasalidea	TrichDB	12.08.2008	1.0 (22-Aug-2007)		TvaginalisAnnotatedProteins_TrichDB-1.0.fasta	http://trichdb.org/common/downloads/release1.0/
Trypanosoma brucei	Sleeping sickness agent	Euglenozoa/Kinetoplastida	TIGR	07.11.2007	(7/24/2005)		TBA1.pep	ftp://ftp.tigr.org/pub/data/Eukaryotic_Projects/t_brucei/annotation_dbs/

Supplementary Table 4. **Sequence conservation of Hinge 1.** The first two residues shown are the bending residues in this hinge.

hDHFR	T	T	S	S	V	E	G	K	Q	N	L
%Identity	31.4	38.8	18.2	43.0	21.5	24.8	45.5	85.1	32.2	100.0	14.9
%conservation	**57.0**	**42.1**	**69.4**	**51.2**	**22.3**	**28.1**	**45.5**	**85.1**	**32.2**	**100.0**	**31.4**

Supplementary note

Expanded discussions

The human DHFR active site

For *E. coli* DHFR, a large chemical shift change in the ^{15}N dimension is observed for Ala6 between THF complexes and FOL complexes. This difference is the result of a hydrogen bond to Ile5, which is formed when THF is bound, but not when FOL is bound[1]. Based on the crystal structures, a similar hydrogen bond would be expected for hDHFR involving Ile7. However, there is no evidence for its formation from our NMR data. From the near identity of the E–$NADP^+$–FOL and E–$NADP^+$–THF spectra, we conclude that hDHFR does not discriminate as effectively as ecDHFR with respect to FOL and THF. Consistent with this notion, the dissociation constants for FOL, DHF and THF are very similar for hDHFR, while ecDHFR binds two orders of magnitude more tightly to THF than it does to FOL[4-7]. The ^{15}N HSCQ spectrum of hE–NADPH–THF shows significant broadening and some shifting of resonances compared with that of the hE–$NADP^+$–THF complex. However, we were unable to prepare a stable sample for this complex, and the ^{15}N HSQC spectrum began changing within a few hours. Therefore, reliable assignments could not be made, and we could not study this complex in detail. A chemical shift analysis of hE–THF and hE–$NADP^+$–THF shows that, in the product binary complex, the differences in chemical shifts can be attributed to the presence or absence of NADP, without additional changes in backbone conformation. Thus, X-ray structures and ^{15}N HSQCs show that the backbone conformation of hE–$NADP^+$–FOL, hE–$NADP^+$–THF and hE–THF are all very similar and in the "closed" conformation. The "closed" conformation in ecDHFR is stabilized by hydrogen bonds between the backbone amide and sidechain carboxylate of Asp122 and the backbone carbonyl and amide of Gly15 and Glu17, respectively. In the

"occluded" conformation, these hydrogen bonds are broken, and new ones are formed between backbone carbonyl and amide of Asn23 and the backbone amide and side-chain hydroxyl of Ser148, respectively[1]. In hDHFR, this hydrogen-bonding pattern is conserved for the observed "closed" conformation, but the hydrogen bonds stabilizing the "occluded" conformation cannot be formed as Asn23 of ecDHFR is replaced by two prolines in hDHFR. As mentioned in the main text, the active site of hDHFR appears to be better packed than that of ecDHFR. The key side chains that contribute to the efficient active site packing in the human enzyme are Leu22, Pro26, Phe31, Ile60 and Pro61. In ecDHFR, Leu22 is replaced by the more flexible Met20, Phe31 is substituted by a smaller Leu28, and Pro61 by the highly flexible Gly51, resulting in a more loosely packed active site.

Chemical shift analysis of hE–NADPH and hE–$NADP^+$–FOL

To assess the presence of the hinge motion in solution, we turned to NMR experiments. A comparison of the ^{15}N HSQC spectra of hE–NADPH and hE–$NADP^+$–FOL shows chemical shift differences that are consistent with the hinge movement observed in the crystal structures. As hE–NADPH is not stable at pH 6.5, the pH at which the chemical shift analysis was carried out for other complexes, spectra of both hE–NADPH and hE–$NADP^+$–FOL were collected at pH 8.0 to facilitate an analysis that is not influenced by pH-dependent chemical shift changes in either complex (Supplementary Fig. 3d,e). Several chemical shift changes are observed, of which only a subset reflect ligand (FOL) binding, as expected. Gly129, one of the hinge-bending residues shows a very large chemical shift difference in both ^1H and ^{15}N positions. Other residues in the region, Leu131 and Lys132 also show significant ^{15}N and ^1H chemical shift differences. These data are consistent with a conformational change in the region, and we conclude that the

difference in chemical shifts most likely reflects the hinge movement, with the hinge being closed in the hE–$NADP^+$–FOL complex, and open or averaging between open and closed in the hE–NADPH complex. In the region of hinge 1, the backbone amide cross peak of Gly45 is broadened beyond detection, and resonances of residues 42 and 44 are weak, indicative of flexibility in that region. We also observe a chemical shift difference for Asn107. Asn107 N forms a hydrogen bond with Pro103 CO, stabilizing a short 1.5 turn helix. In the ternary hE–$NADP^+$–FOL complex, Pro103 is slightly shifted out of register, and the distance between Asn107 N and Pro103 CO increases from 3.1 Å in hE–NADPH to 4.5 Å. The hydrogen bond between Asn107 N and Pro103 CO is, therefore, broken, and the secondary structure in this region is loosened in the ternary (hinge-closed) complex.

Hinge movements in hDHFR

A comparison of the hydrogen bonding patterns in the hE–$NADP^+$–FOL and hE–NADPH structures sheds considerable light on how these movements are driven by ligand binding and release in human DHFR. The NADP cofactor in both hE–NADPH and E–$NADP^+$–FOL is firmly anchored to the loop subdomain in the region of the nicotinamide ring, and to the adenosine binding subdomain in the region of the adenosine moiety. Thus, the hydrogen-bonding network to the protein is maintained at both ends of NADP. However, in the pyrophosphate region of NADP, the interactions with the αF helix differ between the binary hE–NADPH and ternary hE–$NADP^+$–FOL structures. The phosphate groups form hydrogen bonds to the hydroxyl group and backbone amide of Ser119 in hE–NADPH, and an additional hydrogen bond to the backbone amide of Gly117 (Supplementary Fig. 2a-c). Upon folate binding, several new hydrogen bonds are formed from FOL to both subdomains of the protein (Supplementary Table 1), pulling both

subdomains towards each other. In the hinge-closed state, the hydrogen bonds between NADPH and Ser119 are broken, allowing helix αF to slide 2.5 Å away from the active site as the rest of the adenosine binding subdomain closes in towards the loop subdomain. The sliding of helix αF introduces a twisting motion into the hinge movement, and allows sufficient space for the nicotinamide ring of $NADP^+$ to be rearranged slightly to avoid steric clash with the pterin ring of folate. In the absence of substrate, the NADP to Ser119 hydrogen bonds stabilize the hinge-open conformation, explaining why this conformation is not observed when a ligand is bound in the substrate-binding site. The hinge motions are, therefore, driven by ligand binding, and the two conformations are stabilized by hydrogen bonding interactions and hydrophobic packing interactions between the ligands and the enzyme.

Thr40, a classical helix C-cap and one of the bending residues in hinge 1, forms a hydrogen bond to the backbone CO of Arg36 in the substrate-binding helix, and also to the Asn48 side chain at the end of the hinge. Additionally, Asn48 forms a hydrogen bond to the backbone amide of Met111 in the adenosine-binding subdomain, and to Thr38, the last residue in the substrate binding helix preceding hinge 1. The bending of Thr39 and Thr40 results in a rigid body movement of a subset of residues in hinge 1 and the adenosine binding subdomain (Thr41-I114) moves as a rigid body; the length of this hinge provides the framework for the rigid body movement. The hinge 1 region in ecDHFR is 7 residues shorter than that of hDHFR, and forms a network of hydrogen bonds to both subdomains, restricting the extent of possible movements in this region; accordingly, hinge-bending is of much smaller magnitude in ecDHFR[1] than in the human enzyme (Supplementary Fig. 2a-c).

A comparison of the hE–NADPH complex with mouse DHFR (mDHFR, mE) in complex with NADPH shows that the hinge-open conformation is only observed in the hE–NADPH

structure, not in the mE–NADPH structure, although the sequences are highly conserved. The two structures are in different space groups, and both have lattice contacts in areas that could influence the stabilization of the hinge-open versus hinge-closed conformation in the crystal structures. Most importantly, the mE–NADPH structure (PDB code: 3D84[8]) has a glycerol molecule bound in the substrate-binding site, forming key hydrogen bonds to the substrate binding helix, adenosine binding domain and core β-sheet, which would be expected to stabilize the hinge-closed conformation in a similar manner as FOL.

Further discussion of PWNAL ecDHFR mutant

The ^{15}N HSQC spectrum of the ^{21}PWNAL^{24} E–NADP^{+}–FOL is very similar to that of the wild type enzyme, and even more so to the N23PP/S148A mutant, indicating that it is structurally similar (Supplementary Fig. 4a). Due to the very small amount of dispersion, which is detectable in ^{15}N R2 dispersion experiments only at the higher field, the data cannot be fit to obtain an accurate rate, but indicate the presence of limited flexibility in these residues on the μs-ms timescale, at a much faster rate (estimated to be ~3000 s^{-1}) than for the wild type enzyme. Motions are altered in key loop residues, such as Gly121 and His149, which show large fluctuations in wild type E–NADP^{+}–FOL, but no observable ^{15}N R_2 dispersion in the mutant (Supplementary Fig. 4b,c). C-terminal-associated residues maintain millisecond timescale dynamics similar to wild type and ^{21}PWPPL^{24} ecDHFR, with a rate (k_{ex}) of 600 (± 16) s^{-1}, and an excited state population of 3.2%.

Relationship between Regions A, B & C in DHFR

The length of Region 1 relates to conformational flexibility of the active site "Met20" loop, and the length and composition of Regions 2 and 3 are likely determinants of whether hinge movements can occur in the enzyme. Based on our sequence analysis, we find that the combination of lengths in these three regions show certain preferences. The vast majority of bacterial sequences contain 7 residues in Region A, 12 residues in region B, and 14 residues in Region C, and thus have flexible Met20 loops and short hinges that appear unable to support the large scale opening of the active site observed in hDHFR. All Region B sequences containing 19 residues exactly can be aligned accurately to the hDHFR sequence and structure, on the assumption that the flanking secondary structures (αB and βB) are conserved in length. Longer hinges are present in most eukaryotes, but lower eukaryotes retain flexible Met20 loops (short Region A), while higher eukaryotes contain an insertion in Region A, limiting the conformational change after hydride transfer. It seems likely that, while almost all prokaryotic DHFRs cannot accommodate the hinge movements, most eukaryotic DHFRs have already developed the length of hinge that would be required for the hinge movements, but only a few eukaryotic DHFRs have eliminated conformational flexibility influenced by Region A. In a minor group of bacterial sequences, the length of Region C (and therefore likely hinge 2) is similar to hDHFR, but the length of Region B (and likely hinge 1) is identical to the majority of bacterial sequences.

A comprehensive analysis of all DHFR structures in the PDB was carried out, paying special attention to loop and hinge conformations in Regions A, B and C. Although several structures of DHFR are available, analysis of these is not as insightful as one would expect for two main reasons: A) Most structures are of either the *E. coli* or human enzyme bound to an inhibitor and $NADP^+$ in a ternary complex. In the ternary complex, the active site loops are

preferentially closed, and do not give much insight into the conformations relevant to catalysis. B) An analysis of all structures shows most to be exactly the same as the "closed" conformation of ecDHFR. Very few structures differ significantly from this conformation. For the structures that shed insights on understanding the dynamic mechanism, we have included our observations below or in the main text.

It is interesting to note that, in a small subset of bacterial sequences, the length of Region B is increased by one residue compared with the consensus length for bacterial DHFRs. This subset represents thermophilic bacteria including *Thermatoga maritima* DHFR (tmDHFR), as well as a few enterobacteria. Since structures are available for tmDHFR, we can assess the differences between tmDHFR and ecDHFR further. tmDHFR is a dimer, in which the active site loop ("Met20 loop" assessed in region 1) forms part of the dimer interface, and is stabilized in an "open" conformation, which is different from the "closed" and "occluded" conformations of ecDHFR. The "open" conformation has been observed in some crystal structures of ecDHFR, wherein it is stabilized by lattice contacts, but is not a stable ground state in any intermediate complex for ecDHFR. While this conformation is conducive to NADP binding, it is not optimal for catalysis, as the "open" loop does not shield the active site from bulk solvent[1]. The additional residue in Region B of tmDHFR forms an extra turn in the substrate binding helix, and does not add to the length of the connecting loop. However, we note that the dynamic mechanism in tmDHFR is likely to be quite different from that of ecDHFR due to the dimerization interface of the enzyme at the active site, and the fact that the enzyme functions at much higher temperatures.

Gallus gallus (chicken) DHFR (cDHFR) is the vertebrate DHFR that is most divergent from hDHFR and for which a structure is available (PDB codes: 1DR1, ternary complex with thioNADP$^+$ and biopterin and 8DFR, binary complex with thioNADP$^+$)[9,10]. A comparison of both

cDHFR structures with hE–$NADP^+$–FOL shows that the αF helix superimposes exactly with hDHFR in the hinge-closed conformation. When the structures are aligned on the loop subdomain, only Gly53 and K68 in the cDHFR adenosine binding domain are in a slightly more "open" conformation than hE–$NADP^+$–FOL. In the presence of substrate or product, this region would be expected to form hydrogen bonds to the ligand, and close the active site. It is likely that helix αF cannot stabilize the hinge-open conformation, as Ser119 of hDHFR is substituted with Ala119 in cDHFR, and the hydrogen bonds between this residue at the N-terminus of helix αF forms stabilizing hydrogen bonds with NADPH in the binary hE–NADPH hinge-open conformation. There is no notable difference in the conformation of the binary cE–thio$NADP^+$ and ternary cE–thio$NADP^+$:biopterin, showing that the hinge-open conformation is not present, and providing further evidence that the interaction of the Ser119 residue in αF with NADP is important in maintaining the hinge-open conformation in hDHFR.

References

1. Sawaya, M.R. & Kraut, J. Loop and subdomain movements in the mechanism of *Escherichia coli* dihydrofolate reductase: crystallographic evidence. *Biochemistry* **36**, 586-603 (1997).
2. Whitlow, M. *et al.* X-ray crystallographic studies of *Candida albicans* dihydrofolate reductase. *J. Biol. Chem.* **272**, 30289-30298 (1997).
3. Bhabha, G. *et al.* A dynamic knockout reveals that conformational fluctuations influence the chemical step of enzyme catalysis. *Science* **332**, 234-238 (2011).
4. Fierke, C.A., Johnson, K.A., & Benkovic, S.J. Construction and evaluation of the kinetic scheme associated with dihydrofolate reductase from *Escherichia coli*. *Biochemistry* **26**, 4085-4092 (1987).
5. Appleman, J.R. *et al.* Atypical transient state kinetics of recombinant human dihydrofolate reductase produced by hysteretic behavior. Comparison with dihydrofolate reductases from other sources. *J. Biol. Chem.* **264**, 2625-2633 (1989).
6. Lewis, W.S. *et al.* Methotrexate-resistant variants of human dihydrofolate reductase with substitutions of leucine 22. Kinetics, crystallography, and potential as selectable markers. *J. Biol. Chem.* **270**, 5057-5064 (1995).
7. Tsay, J.-T. *et al.* Kinetic investigation of the functional role of phenylalanine-31 of recombinant human dihydrofolate reductase. *Biochemistry* **29**, 6428-6436 (1990).
8. Cody, V., Pace, J., & Rosowsky, A. Structural analysis of a holoenzyme complex of mouse dihydrofolate reductase with NADPH and a ternary complex with the potent and selective inhibitor 2,4-diamino-6-(2'-hydroxydibenz[b,f]azepin-5-yl)methylpteridine. *Acta Crystallogr. D. Biol. Crystallogr.* **64**, 977-984 (2008).
9. McTigue, M.A., Davies, J.F.I., & Kraut, J. Crystal structure of chicken liver dihydrofolate reductase complexed with NADP+ and biopterin. *Biochemistry* **31**, 7264-7273 (1992).
10. McTigue, M.A., Davies, J.F., Kaufman, B.T., & Kraut, J. Crystal structures of chicken liver dihydrofolate reductase: binary thioNADP+ and ternary thioNADP+.biopterin complexes. *Biochemistry* **32**, 6855-6862 (1993).