## Supporting Information

# Structure of (E)-4-hydroxy-3-but-2-enyl 4-diphosphate reductase, the terminal enzyme of the non-mevalonate pathway 

Ingo Rekittke ${ }^{* \#}$, Jochen Wiesner ${ }^{* \#}$, Rene Röhrich ${ }^{*}$, Ulrike Demmer ${ }^{\dagger}$, Eberhard Warkentin ${ }^{\dagger}$, Weiya Xu ${ }^{\S}$, Kathrin Troschke ${ }^{*}$, Martin Hintz*, Joo Hwan No ${ }^{\infty}$, Evert C. Duin ${ }^{\S}$, Eric Oldfield ${ }^{\infty \boldsymbol{\infty} \|}$, Hassan Jomaa ${ }^{* \ddagger}$ and Ulrich Ermler ${ }^{\dagger \ddagger}$

*Institut für Klinische Immunologie und Transfusionsmedizin, Justus-Liebig-Universität Giessen, Langhansstrasse 7, 35385 Giessen, Germany; ${ }^{\dagger}$ Max Planck Institute für Biophysik, Max-von-Laue-Straße 3, D-60438 Frankfurt am Main, Germany; ${ }^{\text {§ }}$ Department of Chemistry and Biochemistry, 179 Chemistry Building, Auburn University, Alabama 36849; ${ }^{\circ}$ Center for Biophysics and Computational Biology, 607 South Mathews Avenue, Urbana, Illinois 61801; and ${ }^{\text {T}}$ Department of Chemistry, University of Illinois at UrbanaChampaign, 600 South Mathews Avenue, Urbana, Illinois 61801

Date received: August XX, 2008
Running Title: x-ray of LytB
${ }^{\ddagger}$ To whom correspondence may be addressed. E-mail: eo@chad.scs.uiuc.edu; hassan.jomaa@uniklinikum-giessen.de and ulrich.ermler@mpibp-frankfurt.mpg.de.

## Supporting methods and materials

Work under oxygen exclusion. All work requiring anaerobic conditions was carried out in a tent (Coy Laboratories, Inc., Grass Lake, USA) under an atmosphere of $\mathrm{N}_{2} / \mathrm{H}_{2}$ $(95 \% / 5 \%)$. Residual $\mathrm{O}_{2}$ inside the tent was removed with palladium catalysts. Buffers were degassed in an ultrasonic bath by passing a stream of helium through the liquid.

Plasmid construction. A synthetic gene encoding the LytB protein of A. aeolicus in the preferred codon usage of $E$. coli was produced by custom synthesis (Eurofins Medigenomix):
ccATGGTGGATATCATTATTGCCGAACATGCCGGTTTCTGTTTCGGTGTCAAAC GTGCAGTTAAGCTGGCGGAAGAGTCCCTGAAAGAGAGCCAAGGCAAAGTCT ATACCCTGGGCCCGATCATTCATAATCCGCAAGAAGTGAACCGCCTGAAAAA CCTGGGCGTGTTTCCATCTCAGGGTGAAGAATTTAAAGAAGGCGATACCGTC ATTATCCGCTCACATGGCATTCCGCCGGAAAAAGAAGAAGCGCTGCGCAAAA AAGGCCTGAAGGTGATTGACGCGACGTGCCCTTACGTGAAAGCGGTTCACGA AGCGGTGTGTCAACTGACGCGTGAAGGTTACTTTGTGGTCCTGGTCGGCGAG AAAAATCACCCGGAAGTGATCGGTACGCTGGGCTATCTGCGCGCGTGCAACG GCAAGGGTATTGTCGTTGAAACTCTGGAGGATATCGGTGAGGCACTGAAACA TGAACGTGTGGGTATCGTTGCGCAAACGACCCAGAATGAAGAATTTTTTAAG GAAGTGGTGGGTGAAATTGCCCTGTGGGTGAAAGAGGTTAAAGTTATCAATA CCATCTGTAACGCAACCTCGCTGCGCCAGGAATCTGTTAAAAAACTGGCCCC AGAGGTTGACGTCATGATTATTATTGGTGGCAAGAACAGTGGCAACACCCGC CGTCTGTATTACATCTCAAAAGAACTGAATCCGAATACTTATCACATTGAAAC TGCCGAGGAGCTtCAGCCGGAATGGTTCCGCGGCGTGAAGCGTGTTGGCATTT CGGCAGGTGCGTCCACCCCTGATTGGATCATCGAACAGGTCAAAAGCCGTAT CCAGGAAATCTGCGAGGGTCAGCTGGTCAGCAGTagatct

The synthetic gene was cloned into the pQE60 vector (Qiagen) issuing a T5 promoter/lac operator element, using Nco I and Bgl II restriction sites. After this construct proved to be unstable in E. coli cells (presumably due to leaky expression), the gene was re-cloned into the tetracycline inducible pASK-IBA33-plus vector (IBA, Göttingen, Germany). The
coding sequence was amplified by polymerase chain reaction (PCR) using the oligonucleotides Aaeo_LytBsynth_pASK_FP (5’-ATG GTA GGT CTC AAA TGG TGG ATA TCA TTA TTG CCG AAC ATG-3') and Aaeo_LytBsynth_pASK_RP (5'-ATG GTA GGT CTC AGC GCT GCT GAC CAG CTG ACC CTC GC-3'). The PCR product was first ligated in the pCR2.1-TOPO-vector (Invitrogen) and then sub-cloned into pASK-IBA33-plus (IBA, Göttingen, Germany) using BsaI restriction sites, in-frame with the vector-derived C -terminal $\mathrm{His}_{6}$ stretch. The resulting plasmid (pASK-AaeoLytB) provided stringent repression and was stably propagated in E. coli cells.

Protein purification and $\mathbf{F e} / \mathbf{S}$ cluster reconstitution. E. coli TOP 10 cells (Invitrogen) were transformed with pASK-AaeoLytB and grown in LB broth supplemented with 150 $\mu \mathrm{g} \mathrm{ml}^{-1}$ ampicillin and $300 \mu \mathrm{M} \mathrm{FeCl}_{3}$ at $37{ }^{\circ} \mathrm{C}$ until an $\mathrm{OD}_{600}$ of 0.6 . Induction was achieved with $200 \mu \mathrm{~g} \mathrm{l}{ }^{-1}$ anhydrotetracycline for 15 h at $30^{\circ} \mathrm{C}$. The cells were harvested by centrifugation ( $17700 \mathrm{~g}, 15 \mathrm{~min}, 4^{\circ} \mathrm{C}$ ) and stored at $-30^{\circ} \mathrm{C}$ until use. All subsequent purification steps were carried out under oxygen exclusion. The cell pellet ( 15 g ) was resuspended in $200 \mathrm{ml} 100 \mathrm{mM} \mathrm{NaCl}, 30 \mathrm{mM}$ Tris- $\mathrm{HCl}(\mathrm{pH} 7.5), 10 \mathrm{mM}$ imidazole, and disintegrated by ultrasonic treatment at $0^{\circ} \mathrm{C}$. After centrifugation ( $75600 \mathrm{~g}, 25 \mathrm{~min}, 4^{\circ} \mathrm{C}$ ) the supernatant was filtered through a $0.22 \mu \mathrm{~m}$ filter (Millipore). Using a Pharmacia FPLC device equipped with a Waters 996 Photodiode Array Detector and a Waters 474 Scanning Fluorescence Detector, the $\mathrm{His}_{6}-\mathrm{LytB}$ fusion protein was loaded on a TALON superflow immobilized cobalt column ( 50 ml bed volume; BD Clonetech, Heidelberg, Germany) at a flow rate of $10 \mathrm{ml} \mathrm{min}^{-1}$. After two washing steps with 10 mM and 20 mM imidazole, respectively, each 15 min , the protein was eluted with 150 mM imidazole. The

LytB peak fraction was subjected to in vitro reconstitution of the Fe/S cluster, basically as described ${ }^{1}$. The protein solution was first adjusted to $1.5 \mathrm{mM} \mathrm{Na}_{2} \mathrm{~S}, 1.4 \mathrm{mM}$ cysteine and 3.5 mM dithiothreitol (DTT). Then, $\mathrm{FeCl}_{3}$ was added from a 150 mM stock solution to a concentration of 1.5 mM . The mixture was gently agitated, and the addition of $\mathrm{FeCl}_{3}$ repeated after 20 min and 40 min . In order to remove the precipitate that formed, the mixture was centrifuged $(2440 \mathrm{~g}, 10 \mathrm{~min})$ and the supernatant filtered through a $0.22 \mu \mathrm{~m}$ filter. The filtrate was diluted $1: 10$ with 30 mM Tris- $\mathrm{HCl}(\mathrm{pH} 7.5)$ and loaded on a Source 15 Q strong anion exchange chromatography column ( 20 ml bed volume; Amersham) using a Waters 600S HPLC device. The column was developed with a linear gradient of $0-500 \mathrm{mM} \mathrm{NaCl}$ in 1 h at a flow rate of $5 \mathrm{ml} \mathrm{min}^{-1}$. The homogeneity of the eluted LytB fractions was analyzed by size exclusion chromatography on a Superdex 200 analytical grade $10 / 300$ column (Amersham) with $150 \mathrm{mM} \mathrm{NaCl}, 30 \mathrm{mM}$ Tris- $\mathrm{HCl}(\mathrm{pH}$ 7.5) as eluent, at a flow rate of $0.4 \mathrm{ml} \mathrm{min}^{-1}$. Fractions containing the LytB dimer with no apparent amounts of higher aggregates were combined and concentrated using Centricon ultrafiltration units with 30 kDa cut-off (Millipore). The concentration of the final LytB preparation was adjusted to $20 \mathrm{mg} \mathrm{ml}^{-1}$ with $40 \mathrm{mM} \mathrm{NaCl}, 30 \mathrm{mM}$ Tris- $\mathrm{HCl}(\mathrm{pH} 7.5)$. Protein concentrations were determined with the BCA Protein Assay Kit (Pierce).

Crystallization and data collection. Crystallization was performed at a temperature of $18^{\circ} \mathrm{C}$ under exclusion of oxygen using the hanging drop vapor diffusion method. Initial screening experiments with the Hampton and Jena Bioscience crystallization kits resulted in a hit in drop C6 of the JBScreen Classic Nr. 4 kit. Best crystals grew by mixing $1.5 \mu \mathrm{l}$ enzyme solution and $1.5 \mu 1$ reservoir solution composed of $10 \%(w / w)$ PEG $8000,0.1 \mathrm{M}$

Tris- $\mathrm{HCl}\left(\mathrm{pH} 8\right.$ ). Their space group is $P 2_{1}$ with unit cell parameters of $60.3 \AA, 87.6 \AA$, $72.1 \AA$ and $95.1^{\circ}$, best compatible with two monomers in the asymmetric unit $\left(\mathrm{V}_{\mathrm{M}}=2.3\right.$ $\left.\AA^{3} / \mathrm{Da}\right)$. X-ray measurements were performed at the PXII beamline of the Swiss Light Source (Villigen, Switzerland). Native data were collected at $1.65 \AA$ resolution. A peak, inflection and remote data set at $2.5 \AA$ resolution was recorded at the iron edge to solve the phase problem by means of the multiple anomalous dispersion (MAD) method. Data processing was performed by using $\mathrm{HKL}^{2}$ and $\mathrm{XDS}^{3}$.

Phase determination and refinement. The coordinates of the two [3Fe-4S] clusters in the asymmetric unit were calculated from the anomalous information of the peak data set using SHELXD $^{4}$, and further refined using SHARP $^{5}$. Phases were subsequently calculated with SHARP at $2.6 \AA$, resolution including the peak, inflection and remote data, and improved by solvent flattening assuming a solvent content of $52 \%{ }^{6}$. The quality of the electron density was sufficient to identify the individual iron atoms in the cluster, and to improve thereby the phases. Partial model building of the two molecules in the asymmetric unit allowed the definition of non-crystallographic symmetry operators. Twofold molecular averaging within DM resulted in a clearly interpretable electron density map of lobes $A$ and $B^{7}$ and the model could be built in automatically ${ }^{8}$. The more flexible and variable lobe C was modelled manually using $\mathrm{O}^{9}$ and $\mathrm{COOT}^{10}$. Refinement was started with $\mathrm{CNS}^{11}$ and brought to convergence with REFMAC5 ${ }^{12}$. Except for residues 290 and 291 at the C-terminal end, the complete polypeptide chain was visible in the electron density. Moreover, a glycerol molecule was identified at the surface of lobe B at the same site in both monomers. The quality of the data was assessed by using the
programs CNS and PROCHECK ${ }^{13}$. Figures were produced with PYMOL (DeLano Scientific).

Ligand docking. HMBPP was docked to LytB using the Glide program ${ }^{14}$. The diphosphate was constrained to hydrogen bond to His42 in the imidazolium form, while the HMBPP 4-OH was modeled as the alkoxide, since this RO-group is formally analogous to the HO-group thought to bind to $\mathrm{Fe}(4)$ in aconitase. The fourth ion was positioned by using homology with the $\mathrm{Fe}_{4} \mathrm{~S}_{4}$ cluster in a HiPIP protein, PDB file 1HPI.

## Supporting x-ray crystallographic data

Table S1. Data acquisition and structure refinement statistics for $\mathrm{LytB}^{\mathrm{a}}$

| Data set | Native <br> high resolution | peak | MAD <br> Inflection | Remote |
| :---: | :---: | :---: | :---: | :---: |
| Data collection |  |  |  |  |
| Wavelength | 0.9828 | 1.7374 | 1.7402 | 1.716 |
| Space group | $P 2_{1}$ | $P 2_{1}$ | $P 2{ }_{1}$ | $P 2_{1}$ |
| Resolution ( $\AA$ ) | $\begin{gathered} 30.0-1.65 \\ (1.65-1.7) \end{gathered}$ | $\begin{aligned} & 30.0-2.1 \\ & (2.1-2.14) \end{aligned}$ | $\begin{gathered} 30.0-2.5 \\ (2.5-2.6) \end{gathered}$ | $\begin{aligned} & 30.0-2.8 \\ & (2.8-2.87) \end{aligned}$ |
| Cell axis $(\AA)$ <br> $\left({ }^{\circ}\right)$ | $\begin{gathered} \text { 60.4,87.8,72.4 } \\ 95.1 \end{gathered}$ | $\begin{gathered} 60.3,87.6,72.1 \\ 95.1 \end{gathered}$ | $\begin{gathered} 60.4,87.5,72.1 \\ 95.0 \end{gathered}$ | $\begin{gathered} 60.0,87.0,71.7 \\ 94.9 \end{gathered}$ |
| Completeness (\%) | 90.9 (58.3) | 99.6 (97.9) | 99.7 (100.0) | 99.5 (99.8) |
| $\mathrm{R}_{\text {sym }}$ (\%) | 4.6 (34.5) | 6.6 (37.5) | 5.5 (20.7) | 6.2 (18.9) |
| I/sigI | 15.3 (3.0) | 13.2 (3.3) | 19.1 (7.8) | 17.0 (7.8) |
| Redundancy | 3.5(2.3) | 3.6 (2.8) | 3.6 (3.6) | 3.5 (3.5) |
| Refinement |  |  |  |  |
| No. mon. a.u. | 2 |  |  |  |
| Resolution (Á) | 20.0-1.65 |  |  |  |
| No. reflections | 77992 |  |  |  |
| $\mathrm{R}_{\text {working }}, \mathrm{R}_{\text {free }}$ (\%) | 19.4, 22.8 |  |  |  |
| No. atoms protein [ $3 \mathrm{Fe}-4 \mathrm{~S}$ ] clusters water + glycerol | $\begin{gathered} 4589 \\ 14 \\ 561 \end{gathered}$ |  |  |  |
| B-factors protein [3Fe-4S] clusters Water+glycerol | $\begin{aligned} & 35.1 \\ & 28.2 \\ & 44.2 \end{aligned}$ |  |  |  |
| R.m.s deviation bond lengths ( $\AA$ ) bond angles ( ${ }^{\circ}$ ) | $\begin{gathered} 0.017 \\ 1.45 \\ \hline \end{gathered}$ |  |  |  |

${ }^{a}$ All data were measured under cryogenic conditions. $\mathrm{R}_{\text {sym }}$ and completeness for the last resolution shell are given in parentheses. $5 \%$ of the data were set aside for $\mathrm{R}_{\text {free }}$ calculation.


Figure S1: Full alignment of 224 LytB proteins obtained by using the JPRED3 ${ }^{15}$ program. The QUERY sequence (at the top of the output) is LytB from Aquifex aeolicus.


Figure S1 (continued).


Figure S1 (continued).


Figure S1 (continued).


Figure S1 (continued).


Figure S1 (continued).


Figure S1 (continued).


Figure S1 (continued).


Figure S1 (continued).

UniRef90_A1SFK8 UniRef90-Q3J5X0 UniRef90-A1AXS1 UniRef90_Q5FFL5 UniRef90_Q2S9U2 UniRef90_Q6NI36 UniRef90_Q5L5D3 UniRef90_P57247 UniRef90-A3DDC0 UniRef90 Q2DMA niRef90 Q2GDX3 UniRef90 A0XAV2 UniRef90-A0WA76 UniRef90_Q3CHQ UniRef90-A1HTW2 UniRef90-Q6AL80 UniRef90_Q8D2R2 UniRef90_Q1EXF2 UniRef90_Q0VSD8 UniRef90_A1A2C9
 UniRef90-Q9Z6P2 UniRef90 ${ }^{-} 4 \mathrm{MZ17}$ UniRef90-Q11KC7 UniRef90-Q3A3D3 UniRef90-Q8F313 UniRef90_Q9RSG0 UniRef90_Q3AAU Q1K0U3 UniRef90_AOQ0L7 UniRef90_A0RQC4 niRef90-Q2AHH UniRef90-Q3C6J0 UniRef90-Q187C6 UniRef90-Q30S81 UniRef90 ${ }^{-}$Q2RYD UniRef90_UPI0000DB0174 UniRef90_A1IAW9 UniRef90_Q9BJX6 UniRef90_Q8K9Z4
 UniRef90 P54473 UniRef90 Q4HN50 UniRef90 Q7M8Y6 UniRef90 Q2D3W2 UniRef90-Q1NXJ4 UniRef90-Q9KD37 UniRef90-Q895G2 UniRef90-A3ERPg UniRef90-Q8DK29 UniRef90-P58675 UniRef90_P65186 UniRef90_Q5N249 UniRef90_A0V1C5 UniRef90_Q67QZ8 UniRef90-Q3W86 UniRef90-Q1KYM5 UniRef90-Q5KX24 UniRef90-Q81LU9 UniRef90-A0YN97 UniRef90-Q7V4T7 UniRef90-Q0AXA2 UniRef90-Q7VDS UniRef90_A2TGW4 UniRef90_AOLJ95 niRef90-A41FR8 UniRef90 ${ }^{-}$Q191X6 UniRef90_Q4C6T2 niRef90_Q9X781 UniRef90_Q7VN UniR 90 -IniRef90-Q72G65 MiR10-Q7V329 niRef90-A2BUP5 UniRef90_Q7NG74 UniRef90_Q2WJ12 niRef90 UPIOO00D73330 UniRef90_Q10WA8 UniRef90 Q013S6 UniRef90-P58676 IniRef90-Q6W3K2


Figure S1 (continued).



Figure S1 (continued).

|  | WLS T TLIVIDE TLE TVIDAlUERFPDUP P D DI |  |
| :---: | :---: | :---: |
| UniRef90_A1SFK8 | WL S Q T T L S V\|D E T L E TVVDAl R E R F P D|L| P P S D D| | C Y A |
| UniRef90-A1URX2 | FVTQTTLSVEDTAGILDVLQKRFPALESPAT | CYATTNRQ |
| UniRef90_Q3J5X0 | YIT Q T TLSI D D T A A V A AL RERFPAI PRKED | CYATTNRQ |
| UniRef90-A1AXS1 | Y TTQ T T L SI D D TKHIVD Y L K K F P T I P K K D N | CYATQNRQ |
| UiRef90_Q5FFL5 | YVT Q T T L SV D D T KVI I D K L K K K F P NI K G P D D I | CYATQNRQ |
| niRef90-Q2S9U2 | YV T Q T T S V D D ARVI D ALR Q T F P I P R K D | CYATQNRQ |
| iRef90-Q6N | WLS Q T TLSVDETMTIVKKLHERFAHLPPSDD | CYATQNRQ |
| Ref90-Q5L | FVTEQT T S L D DVAEVTQALK A R Y P HI DP S S SV | C Y A TQ QRQ |
| Ref90-P57247 |  | CYATTNRQ |
| iRef90-A3DDC0 | VVA Q T TI T NEKWLEVNTALKKKFKNILIKFD T | C S A T SRRQ |
| niRef90_Q26TT7 | YVT QTTLSLDDTAEMIA ILRERFPALPHKED | C Y ATTNRQ |
| iRef90_Q2DMA0 | VVAQTTQSFENLKNVVDACLTKGGEI RVFHT | C D A T $A$ VRQ |
| UniRef90_Q2GDX3 | YVTQSTLALEYTSTMINKIKEKFPNVVGR | CYATQNRQ |
| niRef90_A0XAV2 | FLS QT TLSV D T VEII GVLRRRFPQIPPRSED | CYATSNRQ |
| niRef90-A0WA76 | VVAQTTQSFDNLKSVVTECLRRGGEVRVYNT | CDATAVRQ |
| iRef90_Q3CHQ0 | AVA Q T T D E K H WK DI L EVI KL K VK L I F F N | CDATQKRQ |
| 90_A1HTW2 | VVA QTTFASDQFNTIVDILKTKCREI EVVSRT | CTATDQRQ |
| Ref90-Q6 | FVTQT TLAEND\EQVLVVL R R R F P L L Q G K S N | CFATQNRQ |
| Ref90-Q8 |  | CYATENRQ |
| f90-Q1 | 1 VAQTTI STSLLYTELTAMI EQKSNNVLK F NT | CSATRDRQ |
| 90-Q0 | FVTTQTTLSVDDTARVIDAL RQRFPAIJPKRED | CYATTNRQ |
| 90-A1A2C9 | LLT Q T TLSVDETAGTI TALKARFPEMPPNSD | C Y A T S NRQ |
| niRef90_Q8RA76 | AVSQTTLIEKHWED L EVLKSKAKELVY F NT | CNATQKRQ |
| niRef90_UPI0000382 | WI SQTTLSVDETMETVRRLREKFPTLPPSSDD | CYATQN |
| UniRef90-Q9Z6P2 | Y I T Q T T L SLD DVQEI S SALL K R Y P SIIP S S S | CYATTNRQ |
| UniRef90_Q4FPE1 | YV T Q T T L SV D D TI I KA L K K K F P K $\square$ KEPFK | CYATTNRQ |
| niRef90_Q4MZ17 | YV T Q T T L SL D D CKEl KEALVKK Y P P I E P S G S | CYATTNRQ |
| Ref90_Q11KC7 | YVSQT TLSLEDTAEI 1 S ALERRFDNI PAAES | CYATTNRQ |
| Ref90-Q3A3D3 | $V A Q T T Q S F K N L Q Q I A D \\| C L G K S K E V R \square F N$ | CDATSVRQ |
| Ref90_Q8F313 | YLMQT TLSVADTKNIID Q | CYATTERQ |
| Ref90-Q9RSG0 | $V L T Q T T L S V D D T R K T V D I L K A R F P A V P P S E D$ | CYATKNRQ |
| Ref90-Q3AAU0 | $V 1$ AQTTQKSENFWQVVEVIKKKGGEVVPFNT | CHATRTRQ |
| Ref90-Q1K0U3 | $L \vee A Q T T Q S F D N F S E I V Q E L L E N S K E L R \square F N T$ | CDATAVRQ |
| iRef90_A0Q0L7 | IVSQTTEKQENFNKVIDKI TPLCKSLANFNT | CNATKTRQ |
| 90-A0RQC4 |  | CNATLENQ |
| Q1Q304 | I VAQTTQDRHLFKRAIDEIKRRYPNSESFET | C S STYKRQ |
| Q2 | LI AQT TKSPASFQK V V S Y L P EVKELR | CRTTETRQ |
| Ref90-Q3 | $1 \vee V Q T T M P 1 V Q Y E E L S T L I S T K S N Q V E Q F N$ | CHATKERQ |
| Ref90_Q187C6 | L V A Q T 1 NPELY | CSATKTRQ |
| Ref90_Q30S81 | $\checkmark$ A Q T RKVEDYLEI A N Y L I PRYKEVRVF | CNATFENQ |
| iRef90_Q2RYD1 | YV T Q T T L SVDDTREVIEALKLRFPAIVGPD D | CYATQNRQ |
| niRef90-Q2LUA5 | $V \vee A Q T T Q N V E E Y A A I V R R I R D R F P E T V V F N T$ | CDSTEQRQ |
| UiRef90_UPI0000D | LVSQTTRKVEKFMQI ANYLMLRVKEVRVFNT | CNATFENQ |
| niRef90_A1IAW9 | IVAQTTQSTRLFDEVGAWAAAHHPHYQVFNT | CPSTKNRQ |
| iRef90_Q9BJX6 | YVTQT T L SMDDCALI VQKLKNKFPHIIPS GS | CYATTNRQ |
| Ref90-Q8K9Z4 |  | CYATTNRQ |
| niRef90_UPI0000670 | WLS QTTLSVDETMETVRRLRQKFPQDPPSDD | CYATQNRQ |
| iRef90_P54473 | VTNQT TMSQWDVHDIMELVKEKYPHVEYHQE | Q VR Q |
| Ref90-Q4HN50 | $V \mathrm{~V}$ SQTTKKIEKFMEIVNFLMLKVKEVRVFNT | CDATFKNQ |
| Ref90-Q7M8Y6 | $L V S Q T T K K I E D F L G V A S F L V Q R C A E V R I F N T$ | CNATFDNQ |
| iRef90_Q2D3W2 | $V \vee A Q T T Q P L S N F Q A L L E A L Q N Q G V E V D A R N T$ | CHATGERQ |
| iRef90_Q1NXJ4 | $V \vee S Q T T Q D E E M F Q L I T A G V L S R F P G G K V F N$ | CDSTHKRQ |
| Ref90_Q9KD37 | I TN QT TMSQWDVSEIMKKAMEKYPQAEVHN | $C L A T Q V R Q$ |
| Ref90_Q895G2 | $V$ VSQT TEKKEHWISVLSKVVNECREVVAFN | CNATEVRQ |
| iRef90-A3ERP9 |  | CDATEKRQ |
| iRef90-Q8DK29 | ANQT TMLK GETEQI GKLFERLNEHFMSFNT | CDATQERQ |
| Ref90-P58675 | $V \mathrm{~V}$ S Q TEKKETWNKVLNEIVRASKEIVAFNT | CSATDVRQ |
| Ref90-P65186 | LVSQTTKM T P KLLQI A S Y L VERCTEVRIF NT | CNATSYNQ |
| niRef90_Q5N249 | I ANQT TMLKSETEAI GKLLLKKYDHFLAFNT | CDATQERQ |
| iRef90-A0V1C5 | V VAQTTEI REKWEKI I L R L NKRFENVLKFD T | CNATDKRQ |
| iRef90_Q67QZ8 | $V T N Q T T L S Q W D T Q A L M E Q V K A R W P Q T E I Y N E$ | C L A TQQRQ |
| 90_Q3W867 | YLMQTTLAVDEAEEVVDALRDRFPAVPGSAD | CYATSNRQ |
| 90-Q5 | $V A N Q T T M L K S E T E M I G K L F E E L N N H F M S F N T$ | C D A Q R Q |
| Ref90-Q1KYM5 | I A NQT TMLKGETEDI GKLVEKTNDHFDSFN | CDATQERQ |
| iRef90_Q5KX24 | $V T$ NQTTMSQWDVADI MAKVKEKYPHVEMHK | $L A T Q L R Q$ |
| Ref90_Q81LU9 | $V T N Q T T M S Q W D V Q H L M E D I S K K F P T A E F H K$ | $C L A T Q V R Q$ |
| niRef90-A0YN97 | I ANQTTMLKSETEQI GKLFEQLTEHFQSFNT | CDATQERQ |
| Ref90-Q7V4T7 | $V A N Q T T M L K S E T E E I G R L F E E L N K H F L S F N T$ | CDATEERQ |
| iRef90_Q0AXA2 |  | CSATRSRQ |
| iRef90_Q7VDS2 | $V A N Q T T M L K S E T E E I G R L F E E L N E H F L A F N T$ | CDATEERQ |
| niRef90_A2TGW4 | A NQTTMLK GETEAI GKLMMQKYNHFMSFNTI | CDATQERQ |
| UniRef90_A0LJ95 | $L \vee A Q T T Q N E E V F Q E I Q E E F L K K Y P K G K V K N T$ | CDSTHERQ |
| iRef90-A0ZJU9 | I ANQTTMLKDETEQI GKLFEELNQHFQSFNT | CDATQERQ |
| iRef90-Q41FR8 | $V T$ NQTTMSQWDVQALMEHVRKKYPHVEVHNE | C N A T Q VRQ |
| niRef90-Q191X6 | $V L A Q T T Q P A A N F H G I V D E L K T H T H A L K V H N T$ | CNATEERQ |
| niRef90_Q4C6T2 | $V A N Q T T M L K S E T E Q 1$ GKLFEQLNQHFMSFNT | CDATQERQ |
| niRef90_Q9X781 | WLS QT T L SVDE TMRIVERLRQRFPKLPPSDD | CYATQNRQ |
| iiRef90_Q7VJV5 | LVSQTTKQPEHFGKI AALLVSLVNECR】FNT | CSATFENQ |
| iRef90-Q5NGK4 | YATQTTLSVDETQGI IQ AL K CKYPNTPKKED | CYATQNRQ |
| iRef90-Q2RIU7 |  | CQATHQRQ |
| Ref90-Q72G65 | VLT Q T T L S D D TLA TIA I L K K RFPKVPSRKDL | CYATQNRQ |
| iRef90_Q7V329 | $V A N Q T T M L K S E T E E \\| G K V F E K L N S H F L A F N T I$ | CDATEERQ |
| UniRef90-A2BUP5 | $V A N Q T T M L K S E T E E I G K F F E K T N D H F L A F N T$ |  |
| niRef90-Q7NG | I A NQT TMLKGETERI GKLFERLADHFMSHDT | CDATQERQ |
| UniRef90_Q2WJ12 | $V \mathrm{~V}$ SQTTEKMKNWEETLKNLS YAREILAFNT | C A A TDVRQ |
| UniRef90-UPI0000D73330 | YI T Q T TLS MDD TRAVVA AL R QRYPQIV GPD T | $C Y A V Q N R Q$ |
| UniRef90-Q10WA8 | IA NQTTMLKTETEEI GKLFEQLN AHYQNFNT | CDATQERQ |
| UniRef90-Q013S6 | I ANQTTMLKGETEAI GKLFMEKFQHFVVMDT | CDATQERQ |
| UniRef90-Q2XTB5 | I ANQTTMLKGETADI GKLMMQKYNHFVSFNT I | CDATQERQ |
| UniRef90_P58676 | VTNQTTMSKWDVADLMHYIKGKFPKAI QHQEI | CMATQVRQ |
| UniRef90-Q6W3K2 | $V \mathrm{~V}$ A Q T T RKPADFLEIVNSLILKHKEVRVFNTI | CNATFENQ |
|  |  |  |

Figure S1 (continued).


Figure S1 (continued).

UniRef90_A1SFK8 UniRef90 Q3J5X0 UniRef90_A1AXS1 UniRef90_Q5FFL5 UniRef90_Q2S9U2 UniRef90_Q6NI36 UniRef90_Q5L5D3 UniRef90_P57247 UniRef90-A3DDC0 UniRef90 Q2DMA niRef90 Q2GDX3 niRef90 A0XAV2 UniRef90 ${ }^{-}$A0WA76 UniRef90-Q3CHQ UniRef90-A1HTW2 UniRef90-Q6AL80 UniRef90_Q8D2R2 UniRef90_Q1EXF2 niRef90_Q0VSD8 UniRef90_A1A2C9 UR101 UniRef90-Q9Z6P2 UniRef90-Q4MZ17 UniRef90-Q11KC7 UniRef90 ${ }^{-}$Q3A3D3 UniRef90-Q8F313 UniRef90_Q9RSG0 niRef90_Q3AAU Q1K0U3 UniRei90_AOQUL7 UniRef90_A0RQC4 UniRei90-Q1Q304 UniRef90-Q3C6J0 UniRef90 Q187C6 IniRef90-Q30S81 UniRef90-Q2RYD UniRef90-UPI0000DB0174 niRef90_A1IAW9 UniRef90_Q9BJX6 UniRef90_Q8K9Z4俗
niRei90_P54473 niRef90 Q4HN50 UniRef90_Q7M8Y6 UniRef90-Q1NXJ4 UniRef90-Q9KD37 UniRef90 Q895G2 UniRef90-A3ERP9 UniRef90-Q8DK29 UniRef90 - P58675 UniRef90-P65186 UniRef90_Q5N249 UniRef90_A0V1C5 niRef90_Q67QZ8 niRef90_Q3W867 UniRef90 Q1KYM5 UniRef90-Q5KX24 UniRef90-Q81LU9 UniRef90-A0YN97 UniRef90-Q7V4T7 UniRef90_Q0AXA2 UniRef90_Q7VDS2 niRef90_A2TGW4 UniRef90_A0LJ95 UniRef90-Q41FR8 UniRef90 ${ }^{-}$Q191X6 UniRef90-Q4C6T2 UniRef90 ${ }^{-}$Q9X781 UniRef90 Q7VJV IniRef90-Q5NGK UniRef90-Q2RIU7 UniRef90-Q72G65 MiR10-Q7V329 niRef90-A2BUP5 UniRef90_Q7NG7 UniRef90_Q2WJ12 niRe90 UPIOO00D73330 UniRef90_Q10WA8 UniRef90 Q013S6 UniRef90 P58676 niRef90 Q6W3K2


$V$ Q R A D DI D F D R L GTLLSVI SLISAGASAPEVIVVDEII TAFARK
$V$ Q R A D DI D F D R L GTLLSVI SLISAGASAPEVIVVDEII TAFARK
$V$ QRAT DI D WEALLQGATAVGVAAGASAPEVLVVDEVIAAFAA
$V$ QRAT DI D WEALLQGATAVGVAAGASAPEVLVVDEVIAAFAA


I D S Y K E: $1 \begin{array}{ll}\text { D L } & \text { D L L } \\ \text { D GAHE I } & \text { QP D WF }\end{array}$
I D S Y K E: $1 \begin{array}{ll}\text { D L } & \text { D L L } \\ \text { D GAHE I } & \text { QP D WF }\end{array}$




V $\operatorname{N}$ S S P D H I S D E ILL H Y S G D I A V T A G A S T P E HI I Q S C I
V $\operatorname{N}$ S S P D H I S D E ILL H Y S G D I A V T A G A S T P E HI I Q S C I


$V$ QRAADI D WTVVFS S I A S L GVVTA GA S A P EVVLVEEVI: DA
$V$ QRAADI D WTVVFS S I A S L GVVTA GA S A P EVVLVEEVI: DA
I E MA Q QL N P R WF E GV GKV GVT A GA S T P K WL I DEVV E Q I E K
I E MA Q QL N P R WF E GV GKV GVT A GA S T P K WL I DEVV E Q I E K
I DNYNEVEPDWFEGVDRVGISSGASAPEVLVQDLVSFLKK
I DNYNEVEPDWFEGVDRVGISSGASAPEVLVQDLVSFLKK
QRAAEI DFGWLDGVRTLGITAGASAPEILVRELVDT AT
QRAAEI DFGWLDGVRTLGITAGASAPEILVRELVDT AT
I
I




I D TY K E I K K S W L
I D TY K E I K K S W L


I DEPSSQI EA S WL K GKA A V GVVTA GA SAP E DLVVQQVI A T L K N
I DEPSSQI EA S WL K GKA A V GVVTA GA SAP E DLVVQQVI A T L K N


$V$ D H A K E $\quad$ D P A W WL
$V$ D H A K E $\quad$ D P A W WL
I N N P E D I D T N IV N H S G DI A MT A GA STP E DVVVQACTR K L S S
I N N P E D I D T N IV N H S G DI A MT A GA STP E DVVVQACTR K L S S
I H S D S E I P F E KL E N C K VI GIIS S GA S A P E I L V E N F I N E I K S
I H S D S E I P F E KL E N C K VI GIIS S GA S A P E I L V E N F I N E I K S


















I E T A GE: D D F R QL K S Y N K
I E T A GE: D D F R QL K S Y N K
I E T K D E L D V N TLL K K Y K N L G il T A G A S T P N Wi I E E V V T F
I E T K D E L D V N TLL K K Y K N L G il T A G A S T P N Wi I E E V V T F
E T:K D E L D V N T L K K Y K N L G G T A G A S T P N Wi: E E V V T F L E N
E T:K D E L D V N T L K K Y K N L G G T A G A S T P N Wi: E E V V T F L E N
I $A$ A D A E A
I $A$ A D A E A
I
I

| $I V$ | $E$ | $T$ | $S$ | $E$ | $E$ | $L$ | $D$ | $T$ | $N$ | $A$ | $L$ | $A$ | $K$ | $S$ | $D$ | $H$ | $I$ | $G$ | $T$ | $A$ | $G$ | $A$ | $T$ | $P$ | $N$ | $W I$ | $I$ | $K$ | $K$ | $I$ | $R$ | $E$ | $L$ |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $L$ | $N$ | $T$ | $V$ | $H$ | $D$ | $L$ | $D$ | $Q$ | $Q$ | $I$ | $L$ | $K$ | $N$ | $V$ | $N$ | $K$ | $I$ | $A$ | $L$ | $T$ | $S$ | $A$ | $A$ | $S$ | $T$ | $P$ | $E$ | $Q$ | $E$ | $T$ | $Q$ | $K$ | $F$ |
| $V$ | $N$ | $L$ | $L$ |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |


| $I V$ | $E$ | $T$ | $S$ | $E$ | $E$ | $L$ | $D$ | $T$ | $N$ | $A$ | $L$ | $A$ | $K$ | $S$ | $D$ | $H$ | $I$ | $G$ | $T$ | $A$ | $G$ | $A$ | $T$ | $P$ | $N$ | $W I$ | $I$ | $K$ | $K$ | $I$ | $R$ | $E$ | $L$ |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $L$ | $N$ | $T$ | $V$ | $H$ | $D$ | $L$ | $D$ | $Q$ | $Q$ | $I$ | $L$ | $K$ | $N$ | $V$ | $N$ | $K$ | $I$ | $A$ | $L$ | $T$ | $S$ | $A$ | $A$ | $S$ | $T$ | $P$ | $E$ | $Q$ | $E$ | $T$ | $Q$ | $K$ | $F$ |
| $V$ | $N$ | $L$ | $L$ |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

II $\operatorname{E}$ S S F F S D D I I K
II $\operatorname{E}$ S S F F S D D I I K






V E S A R E L R K E WF L G H E IVVGL T A G A S T P D A I I E E V V K K R R M N E
V E S A R E L R K E WF L G H E IVVGL T A G A S T P D A I I E E V V K K R R M N E












$V E S V E D V Q L D W L K G V R T V G I S A G A S A P P L L V E E$
$V E S V E D V Q L D W L K G V R T V G I S A G A S A P P L L V E E$
I D S G D R I GP G W L A G KI T V GVV T S G A S T P D K V V E E V M K K I L A
I D S G D R I GP G W L A G KI T V GVV T S G A S T P D K V V E E V M K K I L A






$\begin{array}{llllll:lllllllllll}1 & Q & K & A & S & E & K & P & E & W & L & D & G & V & S & K & I \\ G & G & T \\ I & D & T & P & E & R & G & E & D & D & L & E & G & K & I & N & V \\ G & T\end{array}$
$\begin{array}{llllll:lllllllllll}1 & Q & K & A & S & E & K & P & E & W & L & D & G & V & S & K & I \\ G & G & T \\ I & D & T & P & E & R & G & E & D & D & L & E & G & K & I & N & V \\ G & T\end{array}$



Figure S1 (continued).



 $\begin{array}{llllllllll}L V R L V & A\end{array}$



 $M K I N L D R T S G F C D G V Q G T V H M A E E K M A Q S G E \quad L Y C L G B V$
 M N V I RARTA





 $M 1 P$ EI DS GSGFCFGVVVNAI RHAEKQLLEKSSDKLYCLGDD $M K I V V A K N D G F C F G V E R A I R T V E E L L D E G G K V V T D G E I$ $M Q V I R A K T A G T C W G V D L A L R K L I E N K E K A N T R F V M Y G L$


 $M Q I V V A E G \square G F C F G V E N A V S K A K E L L K K$ G L S SVVWTDDDM




Figure S1 (continued).

lupas_21
lupas_14
lupas_28
jnet
conf
sol25
sol5
sol0
jhmm
jpssm

Figure S1 (continued).


lupas_21
lupas_14
lupas_28
jnet
conf
sol25
sol5
sol0
jhmm
jpssm


Figure S1 (continued).


## lupas_21 <br> lupas-28 <br> jnet <br> conf sol25 <br> sol5 sol0 <br> jhmm <br> jpssm




Figure S1 (continued).

| UniRef90_Q97109 UniRef90_Q2l11 UniRef90_A2XLP6 UniRef90-Q73NQ6 UniRef90-Q8A625 UniRef90_Q3VW18 UniRef90_Q2JVA7 UniRef90_Q15I91 UniRef90_Q8RI52 UniRef90_Q11YY5 UniRef90_Q8KF UniRef90_Q4JIR7 UniRef90_Q72G08 UniRef90-Q3B241 UniRef90-Q4AMW0 UniRef90-Q14P56 UniRef90-Q44PB7 UniRef90-QOYRX2 UniRef90 A1ESC1 UniRef90_P21864 UniRef90_Q3VMP5 UniRef90_Q7MWK6 UniRef90-Q9X1F7 UniRef90-Q2ZKH6 UniRef90-Q3AT23 UniRef90_Q2S1D9 UniRef90_A1GMA0 UniRef90_Q5K294 UniRef90_A0HQ83 UniRef90_Q0PH57 UniRef90_Q4PKE9 UniRef90-O83558 UniRef90_Q89AV1 UniRef90_Q7MUK6 |  |
| :---: | :---: |

lupas_21
lupas_14
lupas_28
jnet
conf
sol25
sol5
sol0
jhmm
jpssm


Figure S1 (continued).

lupas_21
lupas_14
lupas_28
jnet
conf
sol25
sol5
sol0
jhmm
jpssm


Figure S1 (continued).


! E E DE E DE E R WLR NVATVVGVCGATSTPMWHDEKVADHDEK I S SPEVVVVPDMLVPPASIGGICGATSTPRWQMEEVABSHIMA
 E TVD
I
E R P Q D L







lupas_21
lupas_14
lupas_28
jnet
conf
sol25
sol5
sol0
jhmm
inssm

Figure S1 (continued).
UniRef90_Q2l111
UniRef90_Q73NQ6 UniRef90_Q8A625 UniRef90 Q3VW18
UniRef90_Q2JVA7
UniRef90_Q3GFY7
UniRef90_Q15I91
UniRef90_Q8RI52
UniRef90_Q11YY5 UniRef90_Q8KFN9 UniRef90 Q316F5
UniRef90_Q4JIR7
UniRef90 Q72G08
UniRef90_Q8EWR9
UniRef90_Q3B241
UniRef90_Q4AMW0 UniRef90 Q14P56
UniRef90_A1BDJ5
UniRef90_Q44PB7
UniRef90-Q0YRX2
UniRef90_A1ESC1
UniRef90-P21864
UniRef90-Q3VMP5
UniRef90-Q7MWK6
UniRef90-Q9X1F7
UniRef90 Q1MQE5
UniRef90_Q2ZKH6
UniRef90_Q1ATX3
UniRef90_Q3AT23
UniRef90-Q2S1D9
UniRef90_A1GMA0
UniRef90 Q5K294
UniRef90_A1ZJU0
UniRef90-A0HQ83
UniRef90_Q0PH57
UniRef90_Q4PKE9
UniRef90 O83558
UniRef90 Q89AV1
UniRef90_Q7MUK6
LK R QELGEA
LK R QELGEA
$\bar{L} \quad \mathrm{KD}$
$\bar{L} \quad \mathrm{KD}$
LREQ
LREQ
$\square \mathrm{YK}$
$\square \mathrm{YK}$
TFGD
TFGD
VA GVE P
VA GVE P
${ }_{\square}^{\square} Y$ A G
${ }_{\square}^{\square} Y$ A G
R Y N H D P
R Y N H D P
NFAQ
NFAQ
$\stackrel{L}{M}$
$\stackrel{L}{M}$
$\stackrel{R}{A} R G \square R \nabla E E A$
$\stackrel{R}{A} R G \square R \nabla E E A$
TYANS
TYANS
WF P D
WF P D
GG
GG
L K A L E
L K A L E
$\stackrel{L}{N} \underset{\sim}{E}$ G G
$\stackrel{L}{N} \underset{\sim}{E}$ G G
NFSPEVI
NFSPEVI
G G A
G G A
lupas 21
upas_14
lupas_28
jnet
02 268888899
conf
sol25
sol5
sol0
jhmm

Figure S1 (continued).

## References

(1) Tong, W.-H.; Jameson, G. N. L.; Huynh, B. H.; Rouault, T. A. Proc. Natl. Acad. Sci. U.S.A. 2003, 100, 9762-9767.
(2) Otwinowski, Z.; Minor, W. Methods Enzymol. 1996, 276, 307-326.
(3) Kabsch, H. J. Appl. Cryst. 1993, 26, 795-800.
(4) Schneider, T. R.; Sheldrick, G. M. Acta Crystallogr. 2002, D58, 1772-1779.
(5) De la Fortelle, E.; Bricogne, G. Methods Enzymol. 1997, 276, 472-494.
(6) Abrahams, J. P.; Leslie, A. G. W. Acta Crystallogr. 1996, D52, 30-42.
(7) Cowtan, K. D. Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography 1994, 31, 83-91.
(8) Terwilliger, T. C. Acta Crystallogr. 2003, D59, 38-44.
(9) Jones, T. A.; Zou, J. Y.; Cowan, S. W.; Kjeldgaard, M. Acta Crystallogr. 1991, A47, 110-119.
(10) Emsley, P.; Cowtan, K. Acta Crystallogr. 2004, D60, 2126-2132.
(11) Brünger, A.; Adams, P. D.; Clore, G. M.; Delano, W. L.; Gros, P.; GrosseKunstleve, R.; Jiang, J.-S.; Kuszewski, J.; Nilges, M.; Pannu, N. S.; Read, R. J.; Rice, L. M.; Simonson, T.; Warren, G. L. Acta Crystallogr. 1998, D54, 905-921.
(12) Murshudov, G. N.; Vagin, A. A.; Dodson, E. J. Acta Crystallogr. 1997, D53, 240255.
(13) Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. J. Appl. Cryst. 1993, 26, 283-291.
(14) Glide 4.5 ; Schrodinger, LLC: New York, NY, 2007
(15) Cuff, J. A.; Clamp, M.E.; Siddiqui, A. S.; Finlay, M.; Barton, G. J. Bioinformatics 1998, 14, 892-893.

