Protein Tunnels: The Case of Urease Accessory Proteins

Published Version:
Protein Tunnels: The Case of Urease Accessory Proteins / Musiani, Francesco; Gioia, Dario; Masetti, Matteo; Falchi, Federico; Cavalli, Andrea; Recanatini, Maurizio; Ciurli, Stefano. - In: JOURNAL OF CHEMICAL THEORY AND COMPUTATION. - ISSN 1549-9618. - STAMPA. - 13:5(2017), pp. 2322-2331. [10.1021/acs.jctc.7b00042]

Availability:
This version is available at: https://hdl.handle.net/11585/586983 since: 2020-01-03

Published:
DOI: http://doi.org/10.1021/acs.jctc.7b00042

Musiani, F.; Gioia, D.; Masetti, M.; Falchi, F.; Cavalli, A.; Recanatini, M.; Ciurli S. (**2017**) *J. Chem. Theory Comput.*, 13:2322-2331. “Protein tunnels: the case of urease accessory proteins”.

Protein tunnels: the case of urease accessory proteins

Francesco Musiani,[†] *Dario Gioia,*[‡] *Matteo Masetti,*[‡] *Federico Falchi,*[§] *Andrea Cavalli,*[‡,§] *Maurizio Recanatini,*[‡] *and Stefano Ciurli*[†,*]

[†]Laboratory of Bioinorganic Chemistry, Department of Pharmacy and Biotechnology, University of Bologna, Viale G. Fanin 40, I-40127, Bologna, Italy.

[‡]Laboratory of Computational Medicinal Chemistry, Department of Pharmacy and Biotechnology, University of Bologna, Via Belmeloro 6, I-40126, Bologna, Italy.

[§]Compunet, Istituto Italiano di Tecnologia, Via Morego 30, I-16163, Genova, Italy.

ABSTRACT

Transition metals are both essential micronutrients and limited in environmental availability. The Ni(II)-dependent urease protein, the most efficient enzyme known to date, is a paradigm for studying the strategies that cells use to handle an essential, yet toxic, metal ion. Urease is a virulence factor of several human pathogens, in addition to decreasing the efficiency of soil organic nitrogen fertilization. Ni(II) insertion in the urease active site is performed through the action of three essential accessory proteins: UreD, UreF, and UreG. The crystal structure of the UreD-UreF-UreG complex from the human pathogen *Helicobacter pylori* (*Hp*UreDFG) revealed the presence of tunnels that cross the entire length of both UreF and UreD, potentially able to deliver Ni(II) ions from UreG to apo-urease. Atomistic molecular dynamics simulations performed on the *Hp*UreDFG complex in explicit solvent and at physiological ionic conditions demonstrate the stability of these protein tunnels in solution and provide insights on the trafficking of water molecules inside the tunnels. The presence of different alternative routes across the identified tunnels for Ni(II) ions, water molecules and carbonate ions, all involved in urease activation, is highlighted here, and their potential role in the urease activation mechanism is discussed.

INTRODUCTION

Transition metal ions are needed by all living organisms as essential micronutrients.[1-2] Their essentiality, coupled with their limited environmental availability and toxicity, has stimulated all life forms to develop mechanisms for selective metal ions accumulation and utilization.[3] Accordingly, all organisms possess metal homeostasis networks that ensure the availability and the correct localization of metal ions in metallo-proteins and sub-cellular compartments.[4] In parallel, intracellular metal trafficking mechanisms maintain the concentration of free metal ions in the cytoplasm under the physiological limits.[5-7] Despite their biological and biophysical relevance, metal trafficking processes in the cell and inside the proteins are still poorly understood.[5-6]

Ni(II) ions are essential for the survival of several and often deadly pathogenic strains of *Helicobacter*, *Staphylococcus*, *Clostridium*, *Vibrio*, *Mycobacterium*, *Yersinia*, *Escherichia*, *Proteus*, *Ureaplasma*, *Klebsiella*, *Pseudomonas*, *Corynebacterium*, *Providencia*, *Morganella*, and *Cryptococcus*. These ureolytic bacteria rely on the activity of the enzyme urease (urea amidohydrolase; EC 3.5.1.5) to colonize and survive the host organism.[8-9] Urease catalyzes urea degradation to yield ammonia and bicarbonate and causing a local pH increase to values suitable for bacterial survival. This is especially true in the case of *Helicobacter pylori*, a spiral-shaped neutrophilic bacterium able to survive in the highly acidic gastric niche.[10] The Center for Disease Control and Prevention estimates that approximately two-thirds of the world's population harbors *H. pylori*, whose infection can lead to stomach cancer and cause gastric mucosa-associated lymphoid tissue lymphoma.[11] In 1994, the International Agency for Research on Cancer classified *H. pylori*, uniquely among bacteria, as a class-I carcinogen in humans. The importance of this bacterium has been emphasized by the award of the Nobel Prize in Medicine to Marshall and Warren in 2005, for their studies on the link between *H. pylori* infection with gastritis and peptic ulcers. Moreover, the rising antibiotic resistance that affects the most commonly used *H. pylori* eradication treatments requires the identification of new drug targets.[12-13]

The knowledge of the activation mechanisms that lead from the inactive apo-urease to its active holo-form, with the insertion of two essential Ni(II) ions in the active site of the enzyme, is thus essential for the development of new drugs able to eradicate *H. pylori* infections. In particular, it is fundamental to understand, at the molecular detail, the structure-function relationships for the proteins that play key roles in this mechanism. The available crystal structures of ureases from several bacteria and higher plants show a typical quaternary structure formed by a functional minimal trimeric assembly.[8-9] Each monomer is in turn composed by a single chain in ureases from higher plants, as in the cases of jack bean (*Canavalia ensiformis*)[14] and pigeon pea (*Cajanus cajan*),[15] by two chains in the case of *H. pylori*,[16] and by three chains in the cases of *Sporasarcina pasterurii* and *Klebsialla aerogenes* [8-9]. The minimal trimeric assembly eventually forms dimers in higher plants or nearly spherical tetramers in *H. pylori* (Fig. 1A) [16]. Each trimeric assembly hosts three conserved active sites, each containing two Ni(II) ions (Fig. 1B).[9] Urease is produced in vivo in an inactive form, and its activation necessitates the carbamylation of a key lysine in the active site to render this residue able to bind and bridge the two Ni(II) through the carbamate moiety, and the delivery of Ni(II) into the active site, two steps that apparently require GTP hydrolysis and CO_2 uptake.[8] This activation process is carried out through the action of four accessory proteins, named UreD (called UreH in *H. pylori*), UreF, UreG, and UreE.[8] UreD appears to be the first protein that binds apo-urease, although little is known about its functional properties.[17] UreF is proposed to bind the urease-UreD complex through a direct interaction with UreD,[18] facilitating the formation of the complex between apo-urease and the UreD-UreF-UreG complex (UreDFG hereafter).[19] UreG is a GTPase proposed to couple GTP hydrolysis to the process of urease activation, and its putative role has been proposed to be linked to the formation, in the presence of CO_2, of carboxyphosphate, an excellent carbamylation agent for the conserved metal-binding lysine in the enzyme active site.[19] UreG is the first reported case of an intrinsically disordered enzyme,[20] which can retain enzymatic activity because of the rigidity of the GTP binding site region while possessing disordered regions involved in protein-protein recognition mechanisms, prodromal to a disorder-to-order transition.[21] UreF has also been proposed to act as a GTPase-activating protein (GAP) to regulate the folding and the function of UreG.[22] Finally, UreE is known to be the metallo-chaperone[23] in charge of delivering and transferring Ni(II) ions to the apourease-UreDFG complex in a GTP-dependent activation process.[24] The urease activation mechanisms proposed so far involve either the sequential binding of UreD, UreF, and UreG to apo-urease[8] (Fig. 1C), or the direct interaction of a preformed UreDFG protein complex to the inactive form of the enzyme, to build a pre-activation complex that prepares apo-urease for nickel binding. Ni(II) ions were suggested to be directly delivered by UreE to apo-urease in the final step of the process[24] (Fig. 1C). However, a recent study indicated the occurrence of a preliminary step of Ni(II) translocation from UreE to UreG when the latter is not yet bound to the UreF and UreD accessory proteins.[25] According to this scheme, UreG subsequently separates from UreE to join the preformed UreD-UreF assembly and form the Ni(II)-bound UreDFG complex. The latter would finally interact with apo-urease completing the insertion of two Ni(II) ions into the enzyme following GTP hydrolysis and through activation by carbonate[25] (Fig. 1C).

The recent publication of the apo-UreDFG crystal structure from *H. pylori* (*Hp*UreDFG, Fig. 1D),[26] devoid of metal ions, represents a crucial breakthrough for the understanding of Ni(II) ions delivery to the apo-urease active site. The *Hp*UreDFG structure features a central core composed by *Hp*UreF in the same dimeric form as previously observed for *Hp*UreF itself.[27] A monomeric chain of *Hp*UreD is bound to each *Hp*UreF monomer in the same arrangement found in the UreD-UreF crystal structure from *H. pylori* (*Hp*UreDF).[28] In the *Hp*UreDFG complex, the *Hp*UreG dimer is bound to the *Hp*UreDF complex interacting with a large concave region formed on the *Hp*UreF dimer surface and flanked, on each side, by one *Hp*UreD monomer. In the crystal structure of this super-complex, a GDP molecule is bound to each *Hp*UreG monomer. A recent study on the metal-binding properties of recombinant *Hp*UreF, determined using site-directed mutagenesis and isothermal titration calorimetry, indicated that His229 and Cys231 in *Hp*UreF are involved in Ni(II) binding in vitro, and are critical for urease activation in vivo.[29] This result prompted a detailed analysis of the structure of the *Hp*UreDFG complex, which revealed the presence of a large cavity at the interface between *Hp*UreF and *Hp*UreG, containing several internal water molecules interconnected through a network aligned along the long horizontal axis of the UreD-UreF$_2$-UreD portion of the structure (Fig. 1E).[29] This internal water network encompasses two nearly identical and symmetric tunnels going from the central cavity in the complex and exiting near the *Hp*UreD C-terminal, passing through *Hp*UreF in the vicinity of His229 and Cys231, located at the interface between *Hp*UreF and *Hp*UreD.[29] This observation prompted the proposal for a role of UreF in the metal ion transport through these tunnels during urease activation.[29] In particular, this hypothesis entails that Ni(II) ions, known to bind to a conserved Cys-Pro-His (CPH) motif on the surface of the *Hp*UreG dimer[8, 30-31] facing *Hp*UreF in the *Hp*UreDFG complex structure, can proceed through the tunnels in order to reach the apo-urease active site.[29] This hypothesis has subsequently received support by in vivo assays using site-directed mutagenesis coupled with bioinformatics and atomistic molecular dynamics (MD) simulations on UreD from *K. aerogenes* (*Ka*UreD), of.[32]

Here, we use atomistic 200 ns-long MD simulations in explicit solvent to extend the investigation to the structural fluctuations and the stability of the full *Hp*UreDFG complex, focusing the analysis on the dynamic behavior of the internal tunnels that encompass the entire protein super-complex. Our findings indicate that the tunnels are persistent during the investigated time of the MD simulation, further suggesting alternative routes for Ni(II) and carbonate ions during the urease activation process. The results of this study represent the starting point for atomistic simulations of Ni(II) permeation through the tunnels in the full *Hp*UreDFG complex by using enhanced sampling algorithms.

MATERIALS AND METHODS

The *Hp*UreDFG crystal structure from *H. pylori* strain 26695 (PDB code: 4HI0)[26] was used. The most probable protonation state of titratable amino acids, and the tautomeric state of histidine residues at neutral pH, were assigned through the Protein Preparation wizard tool of the Schrödinger suite 2015.4.[33] The complex was embedded into a truncated octahedron water box using an 11-Å buffer zone of solvent around the protein complex. The resulting system consisted of ca. 337,000 atoms. The Amber ff99SB force fields[34] for the protein and the TIP3P model[35] for water were used, while known parameters were applied to the GDP molecule found in the crystal structure.[36] The systems were neutralized by adding Na^+ and Cl^- ions using the *genion* program of the GROMACS 4.6.2 package.[37-39] Analogously, additional Na^+ and Cl^- ions were placed in the water box to achieve the physiological ionic strength (150 mM). A total of 350 Na^+ and 316 Cl^- ions were added. The system was energy-minimized and then equilibrated at 300 K and 1 atm by performing 1 ns of gradual annealing using GROMACS 4.6.2. The geometry optimization was performed in four cycles. In the first two cycles, which comprised 800 steps of steepest descent followed by 3000 steps of conjugate gradient, the water molecules were relaxed while the protein was constrained using a harmonic potential with a force constant of 1,000 J mol^{-1} $Å^{-2}$. In the third and in the fourth cycles the procedure was repeated without applying any constraint. During this equilibration phase, positional constraints were applied on the protein atoms and on GDP (force constant of 1,000 J mol^{-1} $Å^{-2}$). The temperature and pressure was controlled using a Berendsen thermostat and barostat,[40] respectively. An integration step of 1 fs was used, and the structures were sampled every 0.1 ps. Periodic boundary conditions (PBC) were applied. The Particle Mesh Ewald (PME) method was used to calculate electrostatic interactions.[41] The cut-off values for the real part of the electrostatic interactions and for the van der Waals interactions were set to 10 Å. In the 200 ns-long production run, the temperature and pressure coupling was made using a Nose-Hoover thermostat[42-43] and a Parrinello-Raman barostat,[44-45] respectively. The calculations were performed using the Eurora and the PLX supercomputers at CINECA (Italy).

The program CAVER 3.0[46] was used to calculate all pathways departing from the *Hp*UreG CPH motif region within 2,000 superimposed MD snapshot of the *Hp*UreDFG complex. All water molecules were removed before the tunnel calculation. The starting point of the tunnel search was calculated as the average position between the *Hp*UreG Cys66 Sγ atoms from each *Hp*UreG monomer. The tunnel search was performed using a probe of 0.9-Å radius. This probe was selected for two reasons: i) it is the CAVER 3.0 default value for the determination of molecular tunnels using molecular dynamics simulations,[46] and ii) it provides a reasonable agreement with the shape of the water density inside the tunnels (see below). The GDP molecules were considered in the computation in order to avoid bias due to the *Hp*UreG GTP/GDP solvent exposed cavity. Each tunnel is determined as an ensemble of beads of radii selected to fit the tunnel diameter. The identified tunnels were clustered by hierarchical average link by calculating the pairwise distances (i.e. dissimilarities) among the centers of the beads for all the computed tunnels.

*Hp*UreD residue conservation was analyzed using the ConSurf server.[47] The server calculates conservation on the basis of a PSI-BLAST analysis[48] that retrieved the best 150 UreD sequences with more than 15% and less than 90% sequence identity with respect to *Hp*UreD. Residue conservation was mapped on the *Hp*UreD structure found in the *Hp*UreDFG complex (PDB code: 4HI0)[26]. Molecular graphics and analyses were performed using the UCSF Chimera package[49] and VMD.[50]

Solvent molecules passing through or in the vicinity of the tunnel bottleneck characterized by *Hp*UreD residues Arg95 and Asp140 (see Results and Discussion section below) were filtered out from the remaining part of the solvent by selecting the water molecules passing within 3.2 Å of the tunnel-facing atoms of Arg95 and Asp140 along the entire trajectory. The volumetric density map of the selected water molecules was then created using the VolMap plugin in VMD. VolMap replaced each oxygen atom of the selected water molecules with a normalized Gaussian distribution of width corresponding to the oxygen atomic radius. The superimposition of all the trajectory frames were used to compute the resulting density map. An in-house tcl script was used to calculate the time spent inside the tunnels by the selected water molecules.

RESULTS AND DISCUSSION

To gain a deeper understanding of the structural and dynamic behavior of the *Hp*UreDFG complex, we ran one 200 ns-long MD simulation in explicit solvent, using an atomistic force field. The root-mean-square-deviation (RMSD) of the Cα atoms of the complex from the initial conformation stabilizes after ca. 25 ns and then oscillates around 2.0 Å for the remaining time of the simulation (Fig. 1-SI in the Supplementary Information). The RMSD of each protein forming the complex stabilizes after a short period (less than 25 ns) and remains stable around values of ca. 1.0, 1.0 and 1.5 Å for *Hp*UreF, *Hp*UreD, and *Hp*UreG, respectively (Fig. 2-SI). This observation, together with the constant secondary structure content of the complex (Fig. 3-SI) indicates that the simulation was long enough to relax and equilibrate each protein within the *Hp*UreDFG complex. In particular, the *Hp*UreG dimer remains well-folded and stable along the simulation (Fig. 4-SI), especially in those regions that were predicted to be intrinsically unfolded through bioinformatics predictions[20] and MD calculations conducted on the *Hp*UreG model structure before the release of the *Hp*UreG crystal structure.[21] Apparently, the interaction between *Hp*UreF and the predicted intrinsically unfolded region of *Hp*UreG (residues 38-94, 128-137, and 155-176) stabilizes the latter in the simulation time scale, consistently with the hypothesis that UreF acts as a GTPase activator (GAP) for UreG.[22]

Analysis of the tunnels. Following the assessment of the overall structure stability of the HpUreDFG complex, the tunnels departing from the cysteine residues in the conserved CPH motif of the *Hp*UreG dimer, and able to pass through the entire complex in order to eventually transport the Ni(II) ions needed for urease activation to *Hp*UreG to *Hp*UreD, were then examined. The software CAVER 3.0 was used for the analysis of 2,000 snapshots from the 200 ns-long trajectory of *Hp*UreDFG. In each snapshot, all possible pathways with a bottleneck radius equal or larger than 0.9 Å were identified, leading to a set of ca. 58,000 pathways along the simulation. These pathways were clustered, using the average-link algorithm based on the pairwise distances of the pathways, to yield 2,081 clusters. Considering the symmetry of the *Hp*UreDFG complex and the position along the major vertical axis of the starting point of the tunnels (Fig. 2A), in principle each cluster of tunnels could have a symmetrically corresponding tunnel. Moreover, considering further the symmetry of the system, each side of the *Hp*UreDFG complex along the vertical axis can be considered independent from the other side. In other words, by running 200 ns of simulation on the *Hp*UreDFG structure, we sampled a total of 400 ns of the tunnels dynamics. The analysis of the cluster of tunnels revealed that the largest portion of the tunnels departing from the CPH motif passes through only *Hp*UreG itself or *Hp*UreF (Fig. 2A). On the other hand, it is possible to identify five pairs of symmetric clusters of tunnels passing through both *Hp*UreF and *Hp*UreD (Table 1). Tunnels 1-3 are present in about 30% of the analyzed frames, while tunnels 4 and 5 are closed for the large part of the simulation time. We thus concentrate on tunnels 1-3 (Fig. 2B-D, Table 1 and 2), while tunnels 4 and 5 are reported in the SI (Fig. 5-SI and Table 1-SI). From Table 1, it appears that the tunnels going through the *Hp*UreDF monomer located on the left in Fig. 2B-D are more present and active during the MD trajectory as compared to the tunnel on the *Hp*UreDF monomer on the right. The analysis of the tunnels' bottlenecks (i.e. the region of each tunnel showing the smallest tunnel radius, Table 2 and Fig. 6,7,8-SI) revealed that the *Hp*UreD residues Arg95 and Glu140 are among those more frequently found in a single bottleneck. Glu140 was also identified in a recent study on *Ka*UreD (Asp142 in *K. aerogenes* numeration) as one of the crucial residues in urease maturation.[32] The analysis of the distance between *Hp*UreD Arg95 Cζ and Glu140 Cδ shows that the side chains of these two residues are at different distances in the left and in the right *Hp*UreDF monomer, respectively, as shown in Fig. 3A. In particular, in the *Hp*UreD monomer on the right side of the *Hp*UreDFG complex, the Arg95 Cζ - Glu140 Cδ distance is less than 6 Å for the 95% of the simulation time, while in the monomer on the left side this happens only for only 21% of the time. Consistently, *Hp*UreD Arg95 and Glu140 side chains are involved in the formation of a H-bond for 98% and 30% of the simulation time in the right and in the left *Hp*UreDF monomer, respectively. When *Hp*UreD Glu140 is not forming a H-bond with Arg95, it can form a H-bond with the Oη atom of the *Hp*UreD Tyr138 residue. The analysis of the H-bonds formed by the side chains of *Hp*UreD Tyr138 and Glu140 showed that one H-bond is present between these residues for 54% and 5% of the simulation time in the left and in the right *Hp*UreDF monomer, respectively. A conservation analysis conducted using PSI-BLAST[48] showed that Glu140 is conserved in 50% of the *Hp*UreD most similar sequences, while in the remaining cases this residue is conservatively mutated with an aspartate. Arg95 is present in only

1% of the cases, while in 96% of the sequences it is mutated with a glutamine, which is nevertheless able to form a H-bond with the residue in the position of Glu140 through its –NH_2 group. Moreover, the residue in position 94 is not largely conserved, and is present as a tyrosine in 36% of the cases, and as an arginine in 17% of the cases. Finally, in *K. aerogenes* UreD and only in the 3% of the considered sequences, position 96 is again mutated with an arginine.[32] In the case of the model structure of *Ka*UreD used for MD simulations,[32] position 96 does not correspond to an arginine because of possible misalignment between the structure of *Ka*UreD and *Hp*UreD, for which the crystal structure has been experimentally determined. Tyr138 is present in 47% of the cases, while in the remaining sequences it is mostly mutated with a glycine. On the other hand, *Hp*UreD Ser139 is frequently mutated with a tryptophan residue that can ensure the formation of an H-bond with Glu140 through the nitrogen atom in the indole group. Taken together, these results show that the formation of the Glu140-Tyr138 H-bond opens the tunnels passing through *Hp*UreF and *Hp*UreD (Fig. 3B), while the presence of the Arg95-Asp140 salt bridge closes the tunnels (Fig. 3C). In the present simulation, the Arg95-Asp140 bottleneck in the left side of the structure is open for ca. 95% of the simulation time, but only for ca. 70% of time on the right side. No correlation has been identified between the open/close behaviour of the tunnels on the left with the respect of the tunnels on the right, suggesting a completely stochastic situation.

Tunnels 1, 2 and 3 coincide from the starting CPH motif at the interface of the *Hp*UreG dimer until they reach the *Hp*UreD Arg95-Asp140 bottleneck (Fig. 2B-D). In this region, these three most populated tunnels split to reach different regions on the *Hp*UreD surface. Tunnels 1 and 2 emerge in highly conserved region of the *Hp*UreD surface, while tunnel 3 flows into a less conserved zone (Table 2). *Hp*UreD is composed by 17 β-strands and 2 α-helices. The structure is characterized by two mixed strand β-sheets with β-strands β1, β2, β5, β8, β11, β13, and β14 forming β-sheet I and β-strands β3, β4, β6, β7, β9, β10, β12, and β15-β17 forming β-sheet II (Fig. 4A). Helix α1 and α2 are packed against anti-parallel β-strands β15- β17 to form an α/β-like motif located at the C-terminus of β-sheet II. The C-terminal regions of both β-sheets I and II constitute the interacting region with *Hp*UreF. Tunnel 1 passes through β-strand β9, β10 and β12 in β-sheet II and emerges in the space between helix α2 and β-strand β6 (Table 2 and Fig. 4A). Tunnel 1 is characterized by a second bottleneck located at the *Hp*UreF-*Hp*UreD interface and involving Ala233, Val235, Lys237 and Asp238 on the *Hp*UreF side together with *Hp*UreD Asp174, Thr176, Tyr197 and Asn199 residues. In the vicinity of this bottleneck, the two *Hp*UreF residues His229 and Cys231 were experimentally found to be involved in Ni(II) binding events critical for urease activation in vivo.[29] Indeed, all the most populated tunnels pass through this region, and *Hp*UreD Asp174 was among the identified residues in *Ka*UreD to be crucial for urease activation,[32] while Asn199 mutation resulted only in a small but recognizable reduction of urease activation. The mouth of tunnel 1 remains open for all the simulation time and is stabilized by the formation of one salt-bridge between the side chains of *Hp*UreD Arg76 and Glu251 (Fig. 4B). Tunnel 2 passes through the space between β-sheets I and II and emerges in the vicinity of β-strands β1, β3, β4, and β6 (Table 2 and Fig. 4A). Also in this case, the exit mouth of tunnel 2 remains open for all the simulation time and is stabilized by a dense H-bonding network involving *Hp*UreD Ser54, Asp61, Gln63, and Glu83 (Fig. 4C). Among these residues, *Hp*UreD Asp61 and Glu83 are in the list of crucial residues for urease activation identified in a mutagenesis/MD study on *Ka*UreD,[32] while Ala52, Met57, Gln63, and Lys84 mutations did not alter the *K. aerogenes* urease activity with respect of the level of active enzyme obtained using the wild type *Ka*UreD. Moreover, the exit mouth of tunnel 2 is located in a highly conserved surface region of *Hp*UreG, and this region has been proposed to act as the UreD/urease binding interface.[32] Finally, tunnel 3 passes through the space between β-sheets I and II, and crosses the entire length of the *Hp*UreDF complex exiting between β-strands β1 and β2 (Table 2 and Fig. 4A). The tunnel mouth is open for the largest part of the simulation time even if some hydrophobic residues (*Hp*UreD Ile24, Leu35, and Ala37) can close at times the access to the bulk of the solvent (Fig. 4D). Indeed, these three residues are also found among the most frequent bottleneck residues for tunnel 3 (Table 2 and Fig. 8-SI). As for the residues identified in this region (Table 2), Leu35 mutation in *Ka*UreD didn't produce large changes in urease activation levels with respect to the wild type, while the mutation of Tyr40 resulted in a small reduction.[32] All the remaining mutated *Ka*UreD residues that didn't show any relevant activity on urease maturation (Asp44, Ala46, Glu47, His86, Asn87, Phe109, Thr126, Arg146, Glu151, Lys161, Ser163, Glu167, and Arg213 in *Hp*UreD numeration) are not involved in any of the tunnels identified in the present study.

These results shows that the tunnels observed in the *Hp*UreDFG crystal structure and passing through each *Hp*UreDF monomer are stable in the simulation time scale and can open and close in the nanoseconds – tens of nanoseconds time scale. The three most present tunnels, observed during the simulation, feature openings in different regions of the *Hp*UreD structure. The exit of tunnel 2 is found in a conserved region proposed to be in contact with apo-urease[32] and thus tunnel 2 should be the preferred pathway for Ni(II) ions during urease maturation. However, the role of tunnel 1 and 3 has not been elucidated yet. Even when the *Hp*UreD Arg95-Glu140 bottleneck is closed, the tunnels mouths are open for the largest part of the simulation time and can theoretically be navigated by water molecules coming from the tunnels or from the bulk of the solvent. In order to gather additional information on the role of the three tunnels passing through the *Hp*UreDF monomers, we analyzed the dynamic behavior of water molecules present inside the tunnels.

Water molecules trafficking inside the tunnels. The presence and persistence of the identified tunnels passing through each *Hp*UreDF dimer in the simulation time scale is not enough to demonstrate the effective passage of water molecules together with Ni(II) ions through the tunnels during the urease activation process. We thus focused our attention on the water molecules passing at H-bond distance from the *Hp*UreD Arg95-Asp140 bottleneck of tunnels 1-3. This analysis resulted in 511 out of the initial 108,896 water molecules. The density of water molecules calculated from the trajectory of these 511 specimens retraces with very good agreement the shape of the tunnels (Fig. 5A). We then concentrated only on those water molecules that are able to enter in the tunnels from the bulk of the solvent or that escape from the tunnels during the simulation. For this analysis we developed an in-house algorithm built on the following assumptions/approximations: i) a total number of fourteen overlapping spheres of radius equal to 15 Å were used to encompass the water molecules density map contoured at 0.025 oxygen atoms $Å^{-3}$ (Fig. 12-SI). This step allowed us to unambiguously define the edges of the tunnels in a computationally efficient way; ii) a water molecule entering the tunnels was defined as a molecule moving from the bulk of the solvent (i.e. from outside the spheres defined above) to the interior of at least one of the spheres and spending no less than 5 consecutive ps inside the tunnels; vice versa, a water molecule exiting the tunnels was defined when the same movement occurred in the opposite direction; iii) in order to avoid artifacts due to the way the spheres were defined, only water molecules with a persistence inside the tunnels greater than 1 ns were considered. This analysis allowed us to identify a subset of 370 water molecules moving from or into the tunnels during the course of the simulation and spending an average time of ca.

52 ns inside the tunnels. Ca. 30% of these water molecules spent less than 10 ns inside the tunnels, but a considerable amount of them (ca. 20%) remained inside the tunnels for more than half of the simulation time (Fig. 13-SI). The visual inspection of these 370 water molecules revealed that the majority of them explore only a small portion of the tunnels. On the other hand, 46 water molecules were able to cover almost one half of the tunnels inside almost one of the monomers composing the *Hp*UreDFG complex (Fig. 14-SI). Among these water molecules, only two passed in the proximity of the Arg95-Asp140 bottleneck on the right side, possibly due to the persistent closure of that gateway during the simulation. On the other hand, in the remaining 44 cases, 13 water molecules were able to pass from *Hp*UreF to *Hp*UreD or vice versa. Fig. 5B-E reports selected examples of such water molecules, able to traverse a large part of tunnel 1. These water molecules started their journey from the inside of the protein complex (Fig. 5B,C), or come from the bulk of the solvent and enter the tunnels at the *Hp*UreG-*Hp*UreF interface (Fig. 5D) or from the mouth of tunnel 1 (Fig. 5E). Among the selected solvent molecules, there are cases of water molecules entering/exiting the tunnel from the mouth of tunnel 2 (6 cases) and tunnel 3 (3 cases), demonstrating the feasibility of these routes for the trafficking of water molecules within the *Hp*UreDFG protein supercomplex.

CONCLUSIONS

The *Hp*UreDFG complex and the tunnels observed in the crystal structure appear to be stable in solution in the hundreds of nanoseconds time scale. The traffic of water molecules inside the tunnels that start from the proposed Ni(II) binding site located at the *Hp*UreG-*Hp*UreF interface and pass through *Hp*UreF and *Hp*UreD can be regulated by the opening of two main bottlenecks, the first located at the *Hp*UreF-*Hp*UreD interface (bottleneck 1), which remains always open, and the second in the core of *Hp*UreD (bottleneck 2). The latter appears to be of particular importance because it is able to control the flow of water molecules as well as Ni(II) ions. Indeed, Ni(II) ions can travel through the tunnels either in their hydrated form, or directly establishing bonds with the atoms found on the tunnels' walls, as reported for other ions,[51-53] Ni(II) ions have a smaller radius (0.83 Å)[54] with respect to water molecules, but their movement can be stopped by the closure of bottleneck 2. The latter can further regulate the access to three different tunnel openings on the *Hp*UreD surface. In particular, the exit of tunnel 2 is in the vicinity of *Hp*UreD Asp61 and Glu83, two crucial residues for urease activation,[32] suggesting that this route is favorable for Ni(II) ions trafficking from *Hp*UreG to the apo-urease reaction site (see Scheme 1). On the other hand, tunnel 1 is the most stable along the simulation and its opening mouth is also located in a highly conserved region on the *Hp*UreD surface. Considering that i) Ni(II) insertion in the urease reaction site should be subsequent to the carbamylation of a conserved lysine residue, ii) a carbonate/bicarbonate ion cannot pass through the bottlenecks of the tunnels inside the *Hp*UreDFG complex because of size constraints, and iii) the opening of tunnel 1 shows the largest diameter with respect of the exits of tunnels 1 and 3, we put forth the hypothesis of the possible involvement of the terminal part of tunnel 1 in the carbonate/bicarbonate transport inside the reaction site of apo-urease through the terminal part of tunnel 2 (see Scheme 1).

The present study highlights the presence of potential distinct routes for the traffic of Ni(II) ions, water molecules, and carbonate ions, all necessary for the activation of urease through incorporation of the metal ions and the carbamylation of the conserved lysine residue that is essential for Ni(II) binding and positioning in the active site. The identified bottlenecks can represent potential new targets for drugs aimed at eradicating infections by ureolytic human pathogens as alternative to the currently used antibiotic treatments. The present results, together with the development of a multi-site model of Ni(II) ions currently underway in our laboratories, pave the way to the atomistic simulations of Ni(II) permeation through the *Hp*UreDFG tunnels.

FIGURES

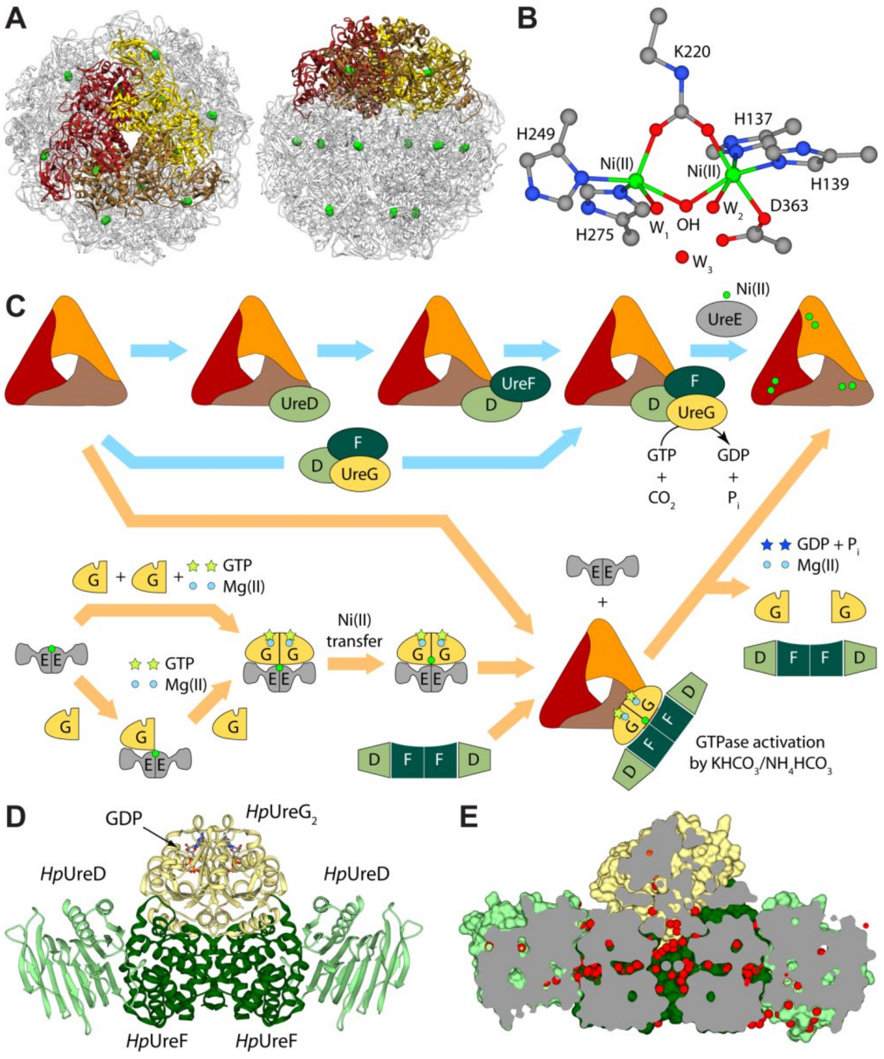


Figure 1. (**A**) Ribbon diagram of urease from *H. pylori* (PDB code: 1E9Z). Ribbon colors highlight the chains composing the trimer of oligomers constituting the minimal quaternary structure of urease. Ni(II) ions are reported as green spheres. The right panel is rotated by 90° around the horizontal axis *vs*. the left panel. (**B**) Coordination geometry of the Ni(II) ions in native urease active site (source *Sporosarcina pasteurii*, PDB code 4CEU). Color scheme: nickel, green; carbon, gray; nitrogen, blue; oxygen, red. (**C**) Schematic representation of the proposed mechanisms for urease activation. (**D**) Ribbon diagram and (**E**) longitudinal section of

the solvent-excluded surface of the apo *Hp*UreDFG crystal structure (PDB code 4HI0). *Hp*UreD, *Hp*UreF, and *Hp*UreG chains are colored as in panel (**C**). Water molecules are depicted as red spheres, GDP is reported as balls-and-sticks and colored according to atom type.

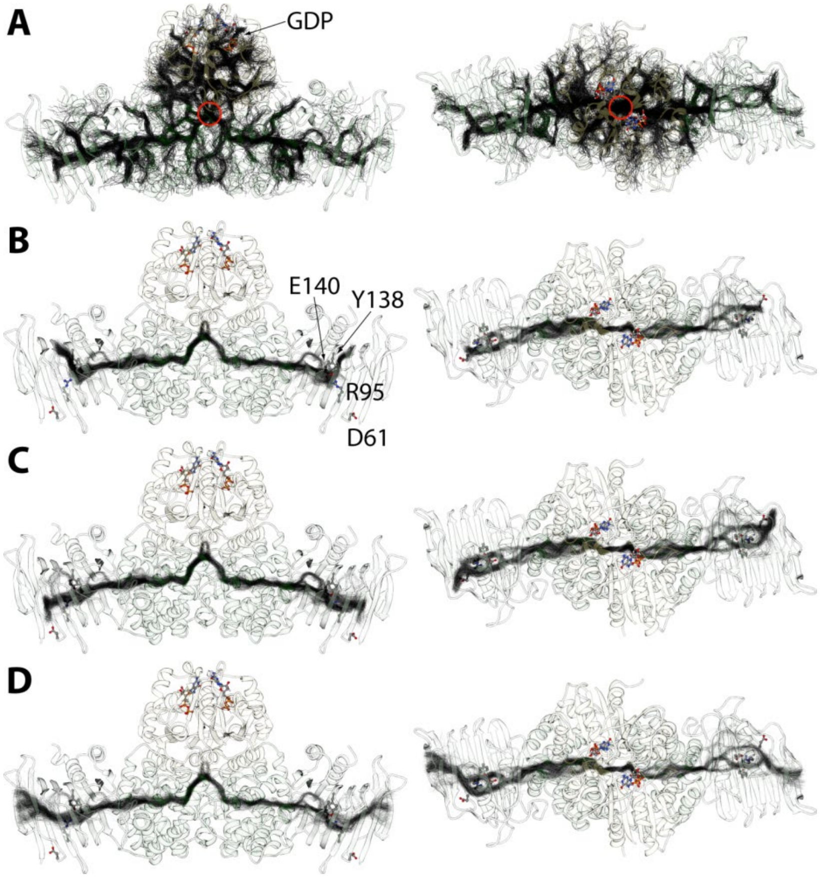


Figure 2. Ribbon diagram of the *Hp*UreDFG complex and tunnels identified throughout the MD simulation by CAVER 3.0, all depicted in one frame as the tunnel centerlines. In the right panels, the *Hp*UreDFG complex is rotated by 90° around the horizontal axis with respect to the orientation in the left panels. In panel (**A**) all the tunnels identified in the protein complex are reported (only one frame per ns was considered for clarity reasons). The red circles identify the starting point position of the tunnels. Tunnels 1-3 are depicted in panels (**B**-**D**), respectively (see Table 1 and 2). Residues cited in the text are reported as balls-and-sticks.

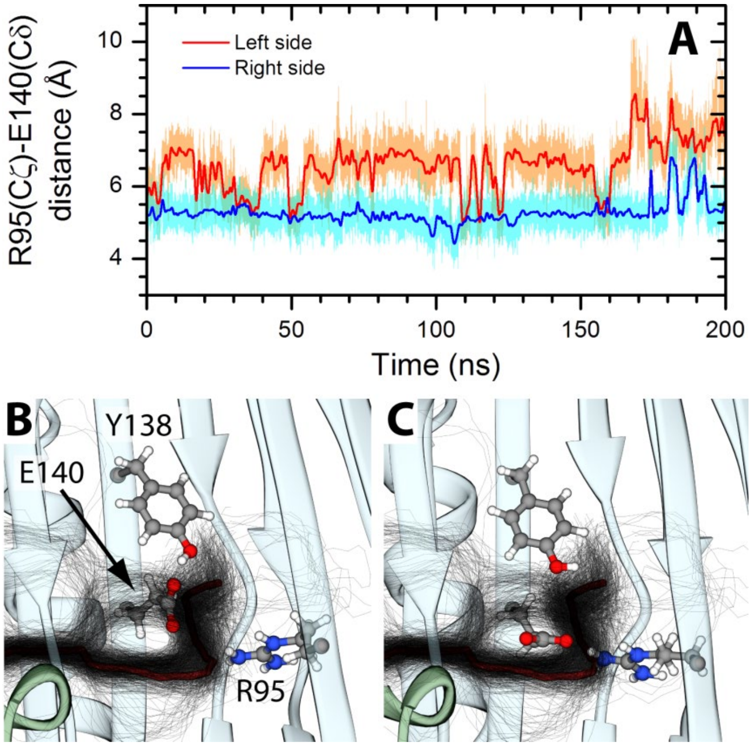


Figure 3. (**A**) *Hp*UreD Arg95 Cζ and Glu140 Cδ distance plotted as a function of time. The orange and light blue lines represent the effective sampling during the simulation, while the red and blue lines have been obtained by applying a Fast Fourier Transform filter in order to cut-off noise. (**B**, **C**) Selected snapshot showing *Hp*UreD Arg95, Tyr138 and Glu140 in different arrangements during the simulation. In panel (**B**), tunnel 1 (black lines) is open, while in (**C**) it is closed.

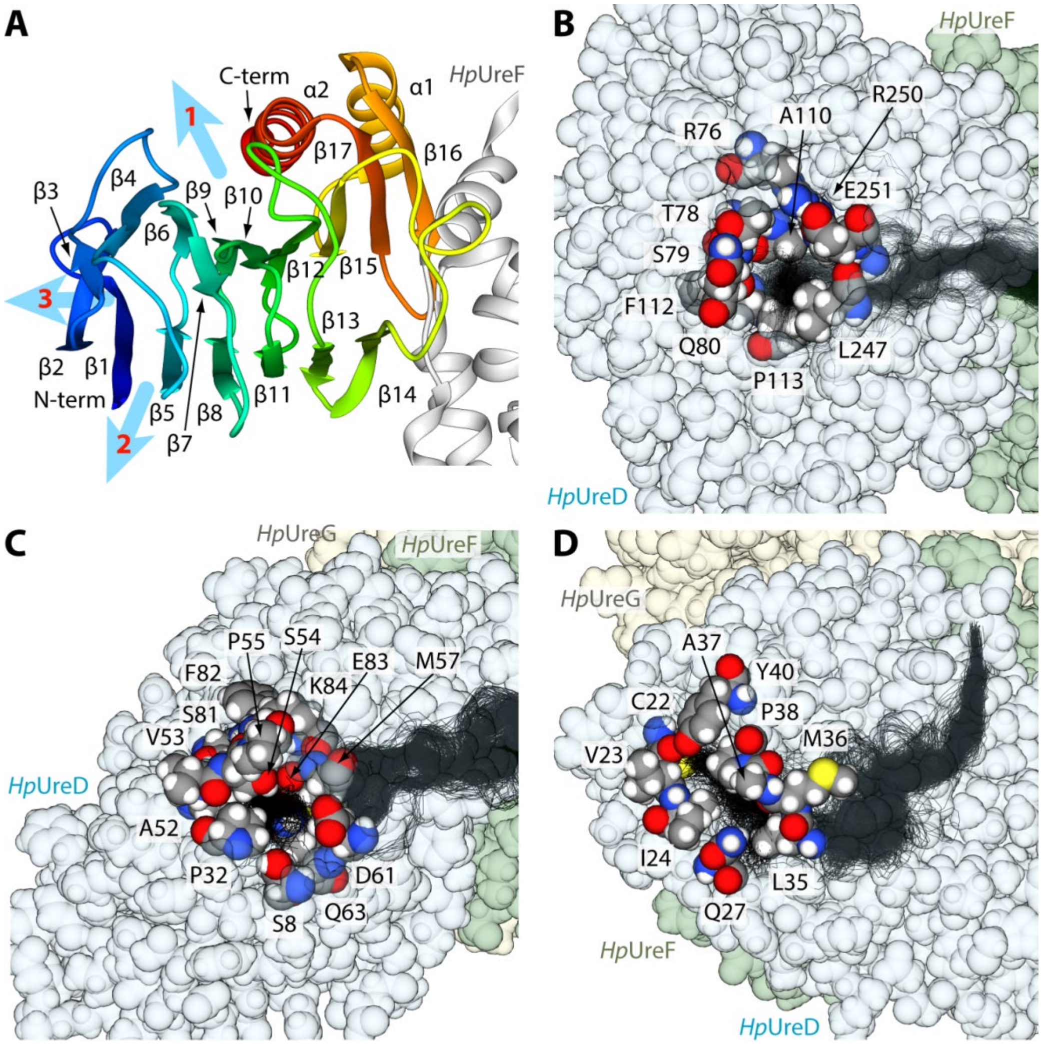


Figure 4. (**A**) Ribbon diagram of *Hp*UreD as found in the *Hp*UreDFG complex. The ribbons are colored from blue in the proximity of the N-terminal to red at the C-terminus. Tunnel openings are indicated by light blue arrows. Details of tunnel 1 (**B**), 2 (**C**) and 3 (**D**) exits. *Hp*UreD, *Hp*UreF and *Hp*UreG atoms are reported as spheres and are colored in light blue, light green and light yellow, respectively. Residues located next to each tunnel exit are colored according to atom type. The tunnels identified throughout the MD simulation by CAVER 3.0 are depicted as the tunnel centerlines.

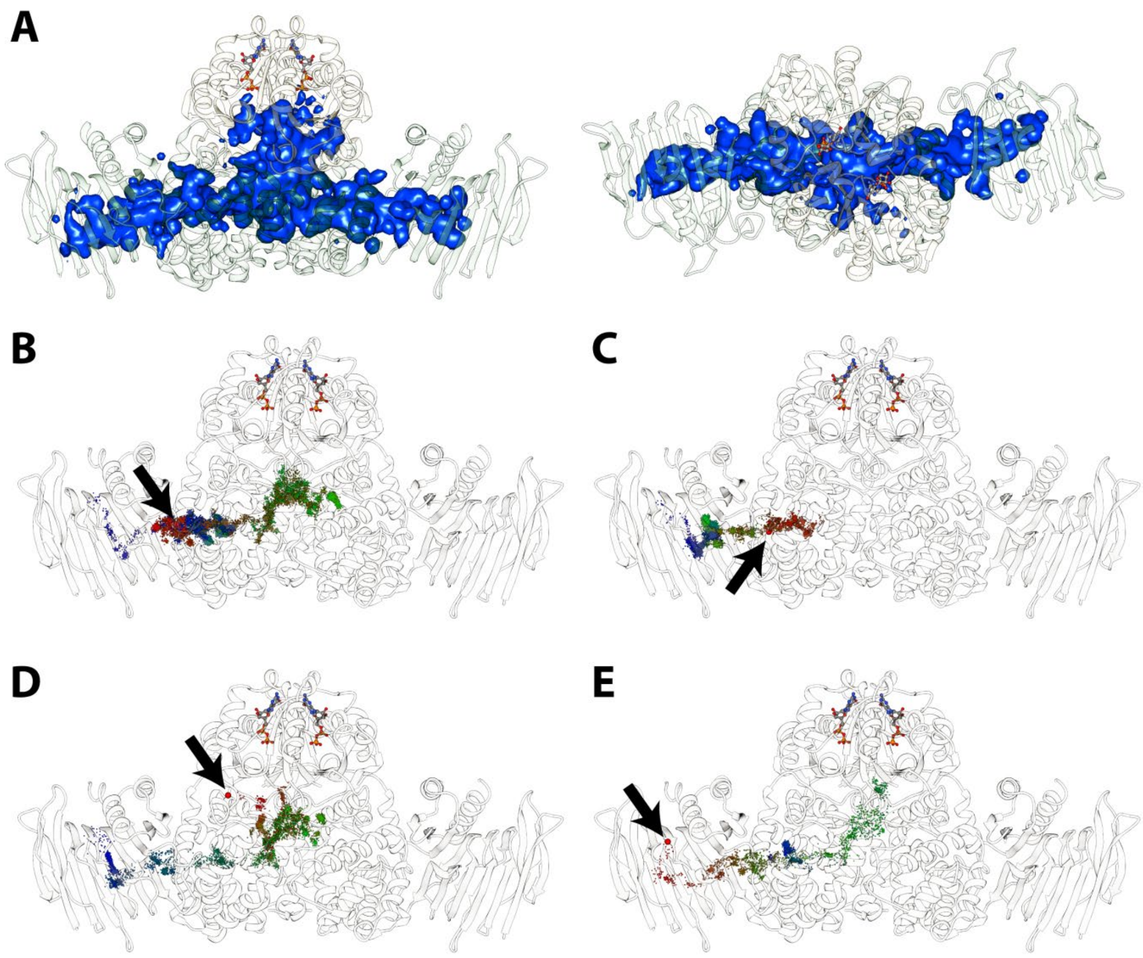
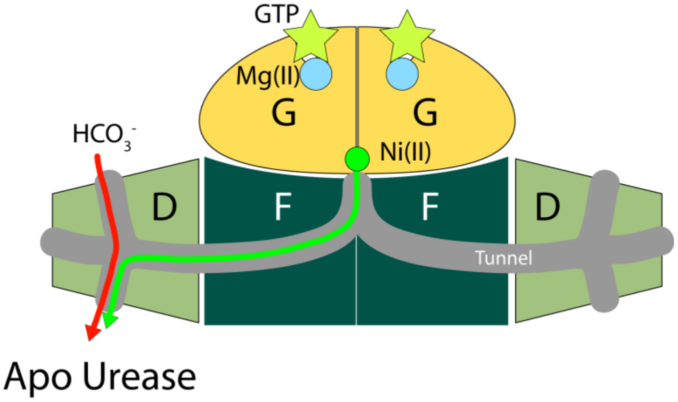


Figure 5. (**A**) Density isosurface of the water molecules found inside the *Hp*UreDFG tunnels contoured at 0.025 oxygen atom Å^{-3} (blue). In the right panel, the *Hp*UreDFG complex is rotated by 90° around the horizontal axis with respect to the orientation in the left panel. (**B**-**E**) Trajectory of four selected water molecules inside the *Hp*UreDFG tunnels. In each frame, the water molecule is reported as a sphere colored from red to green and finally to blue, accordingly to the simulation time. The starting frame is indicated by a red sphere of larger radius with respect to the others, and by a black arrow. The *Hp*UreDFG complex is reported as white ribbons and GDP is depicted as balls-and-sticks.

SCHEMES



Scheme 1. Schematic representation of the pathways followed by Ni(II) ions (green arrow) and by the carbonate/bicarbonate ion (red arrow) inside the *Hp*UreDFG tunnels.

TABLES

Table 1. Analysis of the tunnels identified by CAVER 3.0 and passing through both *Hp*UreF and *Hp*UreD during the MD simulation (Fig. 2B-D and 5-SI).

Tunnel #	Cluster # (left side / right side)	# of snapshots (left side / right side)	Total snapshots	Average bottleneck radius (left side / right side) (Å)	Maximum bottleneck radius (left side / right side) (Å)	Average length (left side / right side) (Å)
1	17 / 63	1745 / 534	2279 (57%)	1.00 ± 0.08 / 0.96 ± 0.07	1.28 / 1.32	78 ± 6 / 80 ± 7
2	35 / 72	1457 / 678	2135 (53%)	0.97 ± 0.06 / 0.94 ± 0.04	1.28 / 1.14	86 ± 6 / 89 ± 7
3	73 / 135	831 / 331	1162 (29%)	0.95 ± 0.05 / 0.93 ± 0.03	1.26 / 1.16	95 ± 6 / 98 ± 7
4	105 / 119	228 / 103	331 (8%)	0.94 ± 0.04 / 0.94 ± 0.05	1.15 / 1.22	77 ± 7 / 69 ± 6
5	157 / 737	269 / 20	289 (7%)	0.94 ± 0.03 / 0.92 ± 0.02	1.11 / 0.97	97 ± 7 / 115 ± 14

Table 2. Relevant regions in tunnels 1-3 identified by CAVER 3.0 and passing through both *Hp*UreF and *Hp*UreD during the MD simulation (Fig. 2B-D).

Tunnel #	***Hp*UreD residues (and conservation[a]) at tunnel exit**	**Most frequent bottleneck residues**
1	Arg76 (8), Thr78 (9), Ser79 (9), Gln80 (9), Ala110 (6), Phe112 (9), Pro113 (9), Leu247 (8), Arg250 (6), Glu251 (3)	*Hp*UreF: Ala233, Val235 *Hp*UreD: Arg95, Leu114, Glu140, Ile141, Ile142, Thr160, Asp174, Thr176, Tyr197, Asn199, Lys237
2	Ser8 (8), Pro32 (9), Ala52 (9), Val53 (9), Ser54 (9), Pro55 (9), Met57 (7), Asp61 (9), Gln63 (9), Ser81 (8), Phe82 (9), Glu83 (7), Lys84 (9)	*Hp*UreD: Phe33, Ala52, Gln63, Glu83, Arg95, Pro111, Phe112, Leu114, Glu140
3	Cys22 (9), Val23 (6), Ile24 (6), Gln27 (5), Leu35 (7), Met36 (9), Ala37 (9), Pro38 (8), Tyr40 (9)	*Hp*UreD: Leu12, Ile24, Gln27, Leu35, Met36, Ile48, Ile77, Arg95, Pro111, Glu140

[a] Conservation calculated by the ConSurf server.[47] The score goes from 0 (variable) to 5 (average) to 9 (highly conserved).

ASSOCIATED CONTENT

Supporting information. RMSD and RMSF of the *Hp*UreDFG complex along the simulation; diagrams and relevant regions of tunnels 4 and 5; analysis of bottleneck residues found in tunnels 1-5; multiple sequence alignment of UreD sequences and ConSurf analysis; distribution of the time spent inside the tunnels by water molecules; and trajectory of selected water molecules inside the *Hp*UreDFG tunnels.

AUTHOR INFORMATION

Corresponding Author

* Phone: +39-051-209-6204; Fax: +39-051-209-6203; E-mail: stefano.ciurli@unibo.it.

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENT

FM was supported by CIRMMP (Consorzio Inter-universitario di Risonanze Magnetiche di Metallo-Proteine) and the University of Bologna. This project was granted 48,000 core hours on the Eurora and PLX supercomputer at CINECA (Italy, ISCRA project HP10CV9TCE).

REFERENCES

1. Finney, L. A.; O'Halloran, T. V., Transition metal speciation in the cell: insights from the chemistry of metal ion receptors. *Science* **2003,** *300* (5621), 931-936.

2. Martinez-Finley, E. J.; Chakraborty, S.; Fretham, S. J. B.; Aschner, M., Cellular transport and homeostasis of essential and nonessential metals. *Metallomics : integrated biometal science* **2012,** *4* (7), 593-605.

3. Sigel, A.; Sigel, H.; Sigel, R. K. O., *Interrelations between Essential Metal Ions and Human Diseases*. Springer: Netherlands, 2013; Vol. 13.

4. Ma, Z.; Jacobsen, F. E.; Giedroc, D. P., Coordination Chemistry of Bacterial Metal Transport and Sensing. *Chem. Rev.* **2009,** *109* (10), 4644-4681.

5. Ba, L. A.; Doering, M.; Burkholz, T.; Jacob, C., Metal trafficking: from maintaining the metal homeostasis to future drug design. *Metallomics : integrated biometal science* **2009,** *1* (4), 292-311.

6. Higgins, K. A.; Carr, C. E.; Maroney, M. J., Specific metal recognition in nickel trafficking. *Biochemistry* **2012,** *51* (40), 7816-7832.

7. Musiani, F.; Zambelli, B.; Bazzani, M.; Mazzei, L.; Ciurli, S., Nickel-responsive transcriptional regulators. *Metallomics : integrated biometal science* **2015,** *7* (9), 1305-1318.

8. Zambelli, B.; Musiani, F.; Benini, S.; Ciurli, S., Chemistry of Ni^{2+} in Urease: Sensing, Trafficking, and Catalysis. *Acc. Chem. Res.* **2011,** *44* (7), 520-530.

9. Maroney, M. J.; Ciurli, S., Nonredox nickel enzymes. *Chem. Rev.* **2014,** *114* (8), 4206-4228.

10. Roesler, B. M.; Rabelo-Goncalves, E. M.; Zeitune, J. M., Virulence factors of *Helicobacter pylori*: a review. *Clin. Med. Insights Gastroenterol.* **2014,** *7*, 9-17.

11. Kusters, J. G.; van Vliet, A. H.; Kuipers, E. J., Pathogenesis of *Helicobacter pylori* infection. *Clin. Microbiol. Rev.* **2006,** *19* (3), 449-490.

12. Zhang, M., High antibiotic resistance rate: a difficult issue for *Helicobacter pylori* eradication treatment. *World J. Gastroenterol.* **2015,** *21* (48), 13432-13437.

13. Hu, Y.; Zhang, M.; Lu, B.; Dai, J. F., *Helicobacter pylori* and antibiotic resistance, a continuing and intractable problem. *Helicobacter* **2016,** *21* (5), 349-363.

14. Balasubramanian, A.; Ponnuraj, K., Crystal structure of the first plant urease from jack bean: 83 years of journey from its first crystal to molecular structure. *J. Mol. Biol.* **2010,** *400* (3), 274-283.

15. Balasubramanian, A.; Durairajpandian, V.; Elumalai, S.; Mathivanan, N.; Munirajan, A. K.; Ponnuraj, K., Structural and functional studies on urease from pigeon pea (*Cajanus cajan*). *Int. J. Biol. Macromol.* **2013,** *58*, 301-309.

16. Ha, N. C.; Oh, S. T.; Sung, J. Y.; Cha, K. A.; Lee, M. H.; Oh, B. H., Supramolecular assembly and acid resistance of *Helicobacter pylori* urease. *Nat. Struct. Biol.* **2001,** *8* (6), 505-509.

17. Steyert, S. R.; Rasko, D. A.; Kaper, J. B., Functional and phylogenetic analysis of ureD in Shiga toxin-producing *Escherichia coli*. *J. Bacteriol.* **2011,** *193* (4), 875-886.

18. Chang, Z.; Kuchar, J.; Hausinger, R. P., Chemical cross-linking and mass spectrometric identification of sites of interaction for UreD, UreF, and urease. *J. Biol. Chem.* **2004,** *279* (15), 15305-15313.

19. Soriano, A.; Hausinger, R. P., GTP-dependent activation of urease apoprotein in complex with the UreD, UreF, and UreG accessory proteins. *Proc. Natl. Acad. Sci. USA* **1999,** *96* (20), 11140-11144.

20. Zambelli, B.; Stola, M.; Musiani, F.; De Vriendt, K.; Samyn, B.; Devreese, B.; Van Beeumen, J.; Turano, P.; Dikiy, A.; Bryant, D. A.; Ciurli, S., UreG, a chaperone in the urease assembly process, is an intrinsically unstructured GTPase that specifically binds Zn2+. *J. Biol. Chem.* **2005,** *280* (6), 4684-4695.

21. Musiani, F.; Ippoliti, E.; Micheletti, C.; Carloni, P.; Ciurli, S., Conformational fluctuations of UreG, an intrinsically disordered enzyme. *Biochemistry* **2013,** *52* (17), 2949-2954.

22. Salomone-Stagni, M.; Zambelli, B.; Musiani, F.; Ciurli, S., A model-based proposal for the role of UreF as a GTPase-activating protein in the urease active site biosynthesis. *Proteins* **2007,** *68* (3), 749-61.

23. Carter, E. L.; Flugga, N.; Boer, J. L.; Mulrooney, S. B.; Hausinger, R. P., Interplay of metal ions and urease. *Metallomics : integrated biometal science* **2009,** *1* (3), 207-221.

24. Soriano, A.; Colpas, G. J.; Hausinger, R. P., UreE stimulation of GTP-dependent urease activation in the UreD-UreF-UreG-urease apoprotein complex. *Biochemistry* **2000,** *39* (40), 12435-12440.

25. Yang, X.; Li, H.; Lai, T. P.; Sun, H., UreE-UreG complex facilitates nickel transfer and preactivates GTPase of UreG in *Helicobacter pylori*. *J. Biol. Chem.* **2015,** *290* (20), 12474-12485.

26. Fong, Y. H.; Wong, H. C.; Yuen, M. H.; Lau, P. H.; Chen, Y. W.; Wong, K. B., Structure of UreG/UreF/UreH complex reveals how urease accessory proteins facilitate maturation of Helicobacter pylori urease. *PLoS Biol.* **2013,** *11* (10), e1001678.

27. Lam, R.; Romanov, V.; Johns, K.; Battaile, K. P.; Wu-Brown, J.; Guthrie, J. L.; Hausinger, R. P.; Pai, E. F.; Chirgadze, N. Y., Crystal structure of a truncated urease accessory protein UreF from *Helicobacter pylori*. *Proteins* **2010,** *78* (13), 2839-2848.

28. Fong, Y. H.; Wong, H. C.; Chuck, C. P.; Chen, Y. W.; Sun, H.; Wong, K. B., Assembly of preactivation complex for urease maturation in *Helicobacter pylori*: crystal structure of UreF-UreH protein complex. *J. Biol. Chem.* **2011,** *286* (50), 43241-43249.

29. Zambelli, B.; Berardi, A.; Martin-Diaconescu, V.; Mazzei, L.; Musiani, F.; Maroney, M. J.; Ciurli, S., Nickel binding properties of Helicobacter pylori UreF, an accessory protein in the nickel-based activation of urease. *J. Biol. Inorg. Chem.* **2014,** *19* (3), 319-334.

30. Bellucci, M.; Zambelli, B.; Musiani, F.; Turano, P.; Ciurli, S., Helicobacter pylori UreE, a urease accessory protein: specific Ni(2+)- and Zn(2+)-binding properties and interaction with its cognate UreG. *Biochem. J.* **2009,** *422* (1), 91-100.

31. Merloni, A.; Dobrovolska, O.; Zambelli, B.; Agostini, F.; Bazzani, M.; Musiani, F.; Ciurli, S., Molecular landscape of the interaction between the urease accessory proteins UreE and UreG. *Biochim. Biophys. Acta* **2014,** *1844* (9), 1662-1674.

32. Farrugia, M. A.; Wang, B.; Feig, M.; Hausinger, R. P., Mutational and computational evidence that a nickel-transfer tunnel in UreD is used for activation of *Klebsiella aerogenes* urease. *Biochemistry* **2015,** *54* (41), 6392-6401.

33. Sastry, G. M.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W., Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J. Comput. Aid. Mol. Des.* **2013,** *27* (3), 221-234.

34. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C., Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* **2006,** *65* (3), 712-725.

35. Jorgensen, W. L.; Chandrasekhar, L.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983,** *79* (2), 926-935.

36. Meagher, K. L.; Redman, L. T.; Carlson, H. A., Development of polyphosphate parameters for use with the AMBER force field. *J. Comput. Chem.* **2003,** *24* (9), 1016-1025.

37. Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R., GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **1995,** *91* (1-3), 43-56.

38. Lindahl, E.; Hess, B.; van der Spoel, D., GROMACS 3.0: a package for molecular simulation and trajectory analysis. *Journal Mol. Model.* **2001,** *7* (8), 306-317.

39. Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C., GROMACS: Fast, flexible, and free. *J. Comput. Chem.* **2005,** *26* (16), 1701-1718.

40. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R., Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984,** *81* (8), 3684-3690.

41. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G., A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995,** *103* (19), 8577-8593.

42. Hoover, W. G., Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **1985,** *31* (3), 1695-1697.

43. Nosé, S., A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* **2002,** *100* (1), 191-198.

44. Nosé, S.; Klein, M. L., Constant pressure molecular dynamics for molecular systems. *Mol. Phys.* **1983,** *50* (5), 1055-1076.

45. Parrinello, M.; Rahman, A., Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981,** *52* (12), 7182-7190.

46. Chovancova, E.; Pavelka, A.; Benes, P.; Strnad, O.; Brezovsky, J.; Kozlikova, B.; Gora, A.; Sustr, V.; Klvana, M.; Medek, P.; Biedermannova, L.; Sochor, J.; Damborsky, J., CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures. *PLoS Comput. Biol.* **2012,** *8* (10), e1002708.

47. Celniker, G.; Nimrod, G.; Ashkenazy, H.; Glaser, F.; Martz, E.; Mayrose, I.; Pupko, T.; Ben-Tal, N., ConSurf: Using Evolutionary Data to Raise Testable Hypotheses about Protein Function. *Isr. J. Chem.* **2013,** *53* (3-4), 199-206.

48. Altschul, S. F.; Madden, T. L.; Schaffer, A. A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. J., Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **1997,** *25* (17), 3389-3402.

49. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E., UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004,** *25* (13), 1605-1612.

50. Humphrey, W.; Dalke, A.; Schulten, K., VMD: Visual molecular dynamics. *J. Mol. Graphics* **1996,** *14* (1), 33-38.

51. Ceccarini, L.; Masetti, M.; Cavalli, A.; Recanatini, M., Ion Conduction through the hERG Potassium Channel. *PloS one* **2012,** *7* (11).

52. Corry, B.; Thomas, M., Mechanism of ion permeation and selectivity in a voltage gated sodium channel. *J. Am. Chem. Soc.* **2012,** *134* (3), 1840-1846.

53. Sauguet, L.; Poitevin, F.; Murail, S.; Van Renterghem, C.; Moraga-Cid, G.; Malherbe, L.; Thompson, A. W.; Koehl, P.; Corringer, P. J.; Baaden, M.; Delarue, M., Structural basis for ion permeation mechanism in pentameric ligand-gated ion channels. *Embo J.* **2013,** *32* (5), 728-741.

54. Shannon, R., Revised effective ionic radii and systematic studies of interatomic distances in halides and chalcogenides. *Acta Crystallogr. A* **1976,** *32* (5), 751-767.

Protein tunnels: the case of urease accessory proteins

Francesco Musiani,[†] Dario Gioia,[‡] Matteo Masetti,[‡] Federico Falchi,[§] Andrea Cavalli,[‡,§]
Maurizio Recanatini,[‡] and Stefano Ciurli[†,*]

SUPPLEMENTARY INFORMATION

[†] Laboratory of Bioinorganic Chemistry, Department of Pharmacy and Biotechnology, University of Bologna, Viale G. Fanin 40, I-40127, Bologna, Italy.

[‡] Laboratory of Computational Medicinal Chemistry, Department of Pharmacy and Biotechnology, University of Bologna, Via Belmeloro 6, I-40126, Bologna, Italy.

[§] Compunet, Istituto Italiano di Tecnologia, Via Morego 30, I-16163, Genova, Italy.

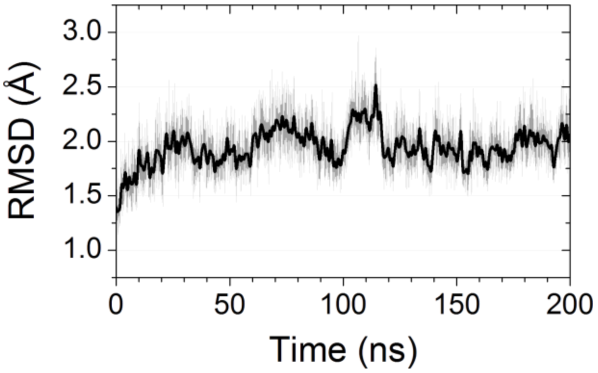


Figure 1-SI. Calculated root mean square deviations (RMSD) of the *Hp*UreDFG Cα from the initial X-ray structure plotted as a function of time. The grey line represent the effective sampling of RMSD during the simulation, the black line has been obtained by applying a Fast Fourier Transform filter in order to cut-off noise.

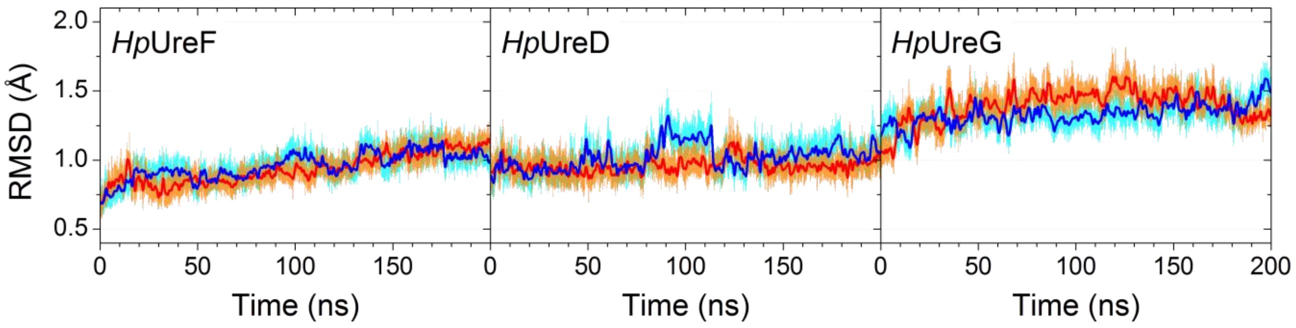


Figure 2-SI. Calculated RMSD of the monomeric *Hp*UreF (left panel), *Hp*UreD (central panel) and *Hp*UreG (right panel) Cα from the initial X-ray structure plotted as a function of time. The RMSD plots of the different monomers of the same protein are in light blue and orange. The blue and red lines have been obtained by applying a Fast Fourier Transform filter in order to cut-off noise.

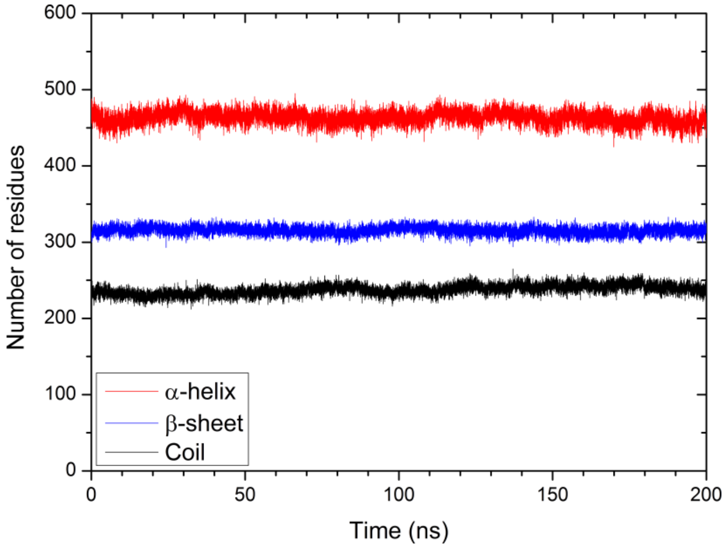


Figure 3-SI. Secondary structure content of the *Hp*UreDFG complex plotted as a function of time.

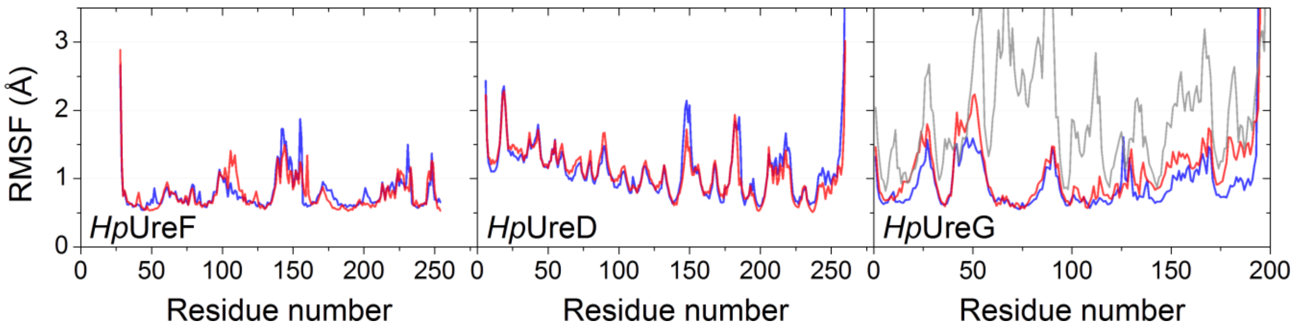


Figure 4-SI. Calculated root mean square fluctuations (RMSF) of the monomeric *Hp*UreF (left panel), *Hp*UreD (central panel) and *Hp*UreG (right panel) Cα during the simulation. The RMSF plots of the different monomers of the same protein are blue and red. In the right panel, the gray line represent the average RMSF obtained from replica exchange MD simulations performed on the *Hp*UreG model structure (*1*).

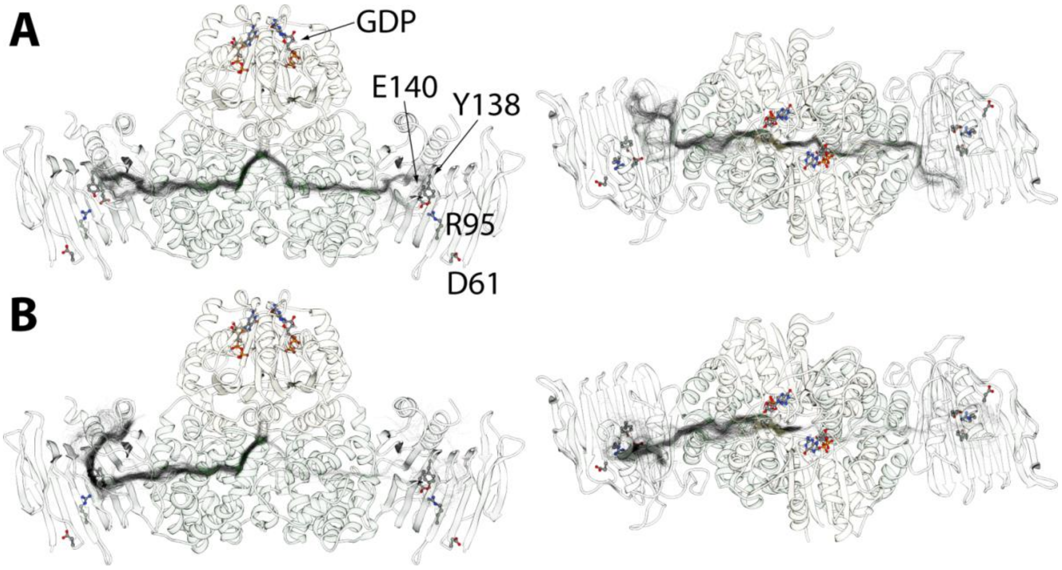


Figure 5-SI. Ribbon diagram of *Hp*UreDFG complex and tunnels 4 (**A**) and 5 (**B**) identified throughout the MD simulation by CAVER 3.0 all depicted in one frame as the tunnel centerlines (see Table 1 and 1-SI). In the right panels, the *Hp*UreDFG complex is rotated by 90° around the horizontal axis with respect to the orientation in the left panels.

Table 1-SI. Relevant regions in tunnels 4 and 5 (Table 1) identified by CAVER 3.0 and passing through both *Hp*UreF and *Hp*UreD during the MD simulation (Fig. 5-SI).

Tunnel #	***Hp*UreD residues (and conservation[a]) at tunnel exit**	**Most frequent bottleneck residues**
4	Ser132 (2), Ser133 (3), Ser134 (8), Gln135 (4), Gln166 (7), Asn204 (4), His233 (7)	*Hp*UreF: Leu113 *Hp*UreD: Leu130, Leu136, Ile164, Gln166, Ile171, Tyr172, Tyr173, Val201, Val203, His233
5	Ile216 (7), Glu217 (1), Ser219 (1), Val222 (1), Asp223 (1)	*Hp*UreD: Ser139, Ile141, Leu198, Leu200, Val236, Leu246, Leu249, Arg250, Ile253,

[a] Conservation calculated by the ConSurf server (*2*). The score goes from 0 (variable) to 5 (average) to 9 (highly conserved).

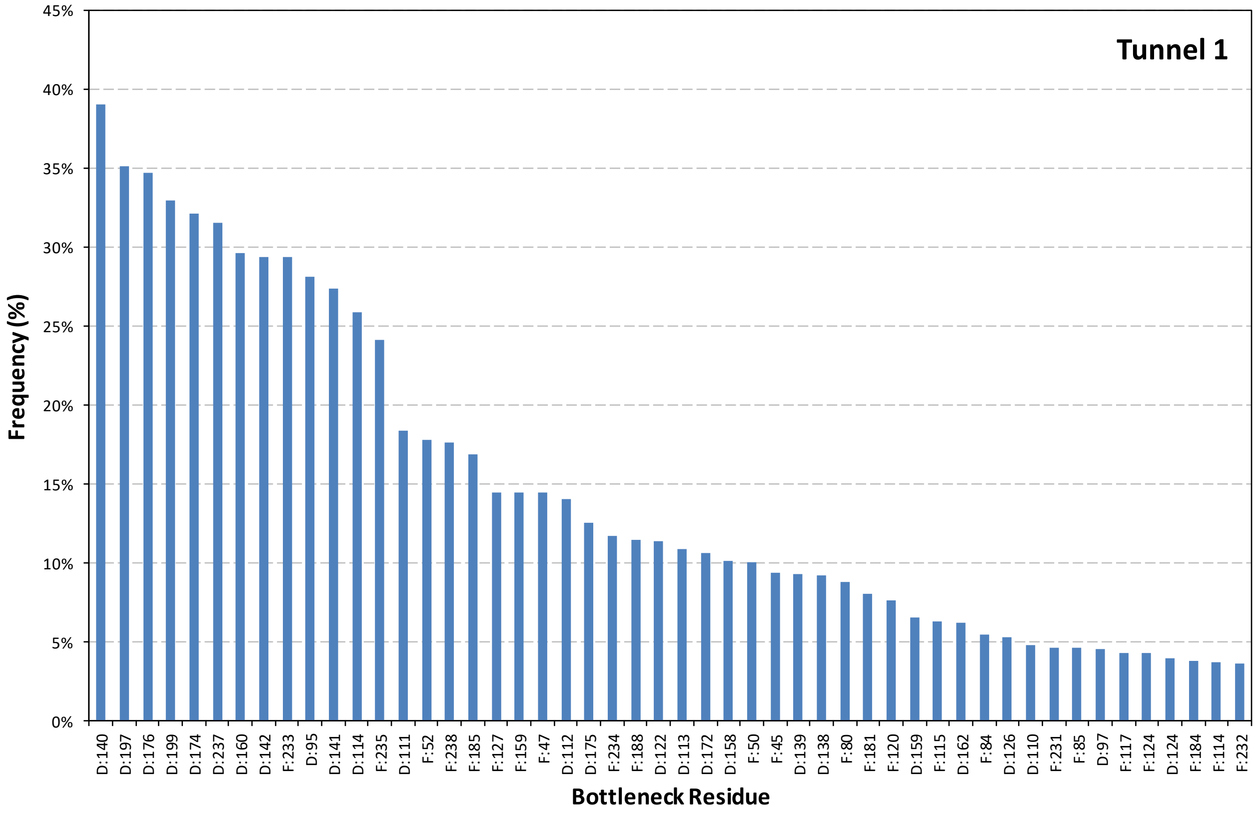


Figure 6-SI. Analysis of bottleneck residues found in tunnel 1 (Table 1).

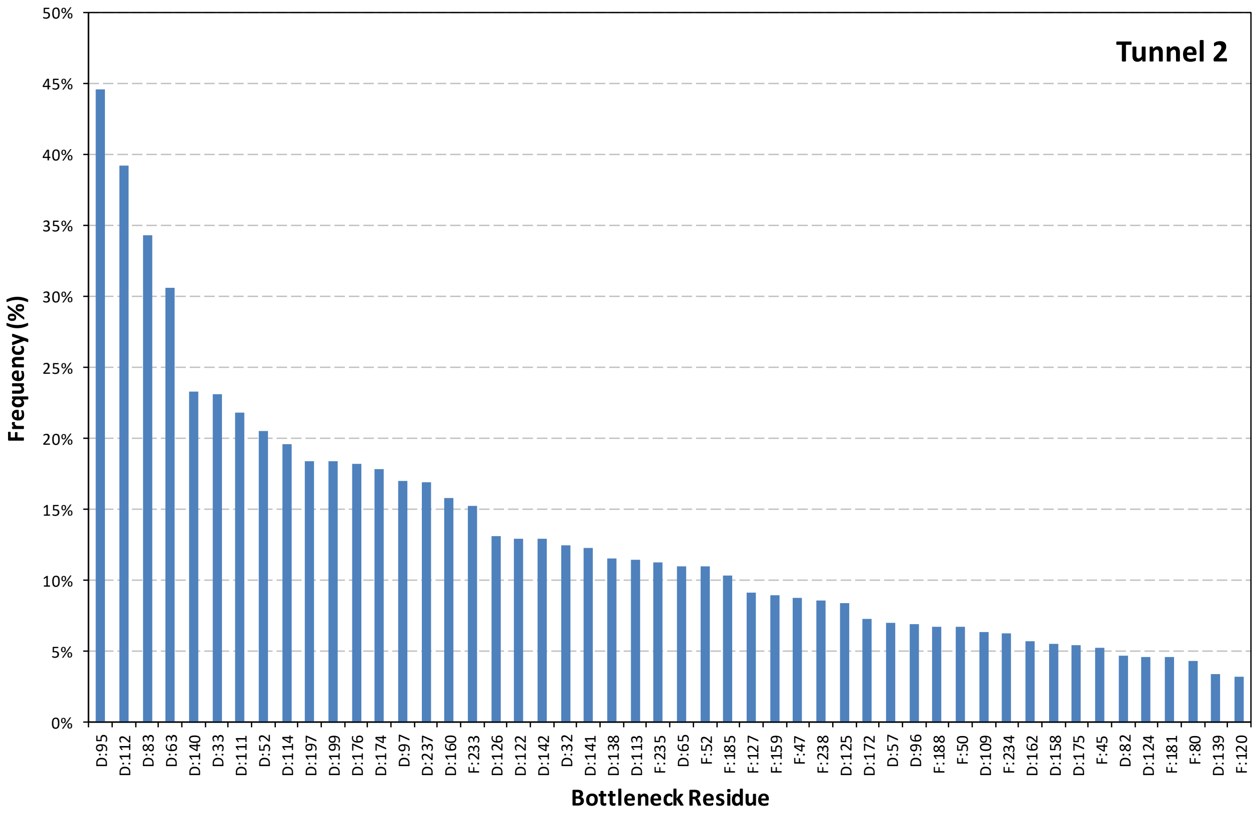


Figure 7-SI. Analysis of bottleneck residues found in tunnel 2 (Table 1).

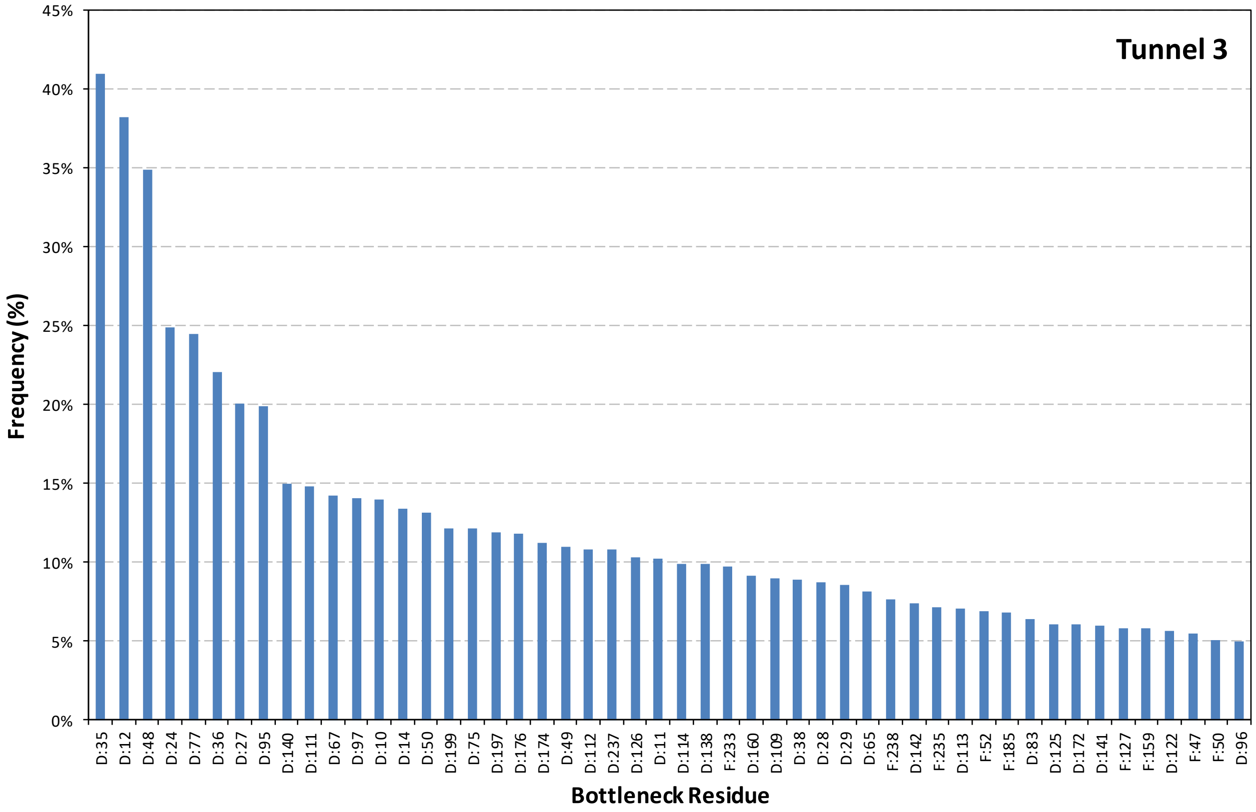


Figure 8-SI. Analysis of bottleneck residues found in tunnel 3 (Table 1).

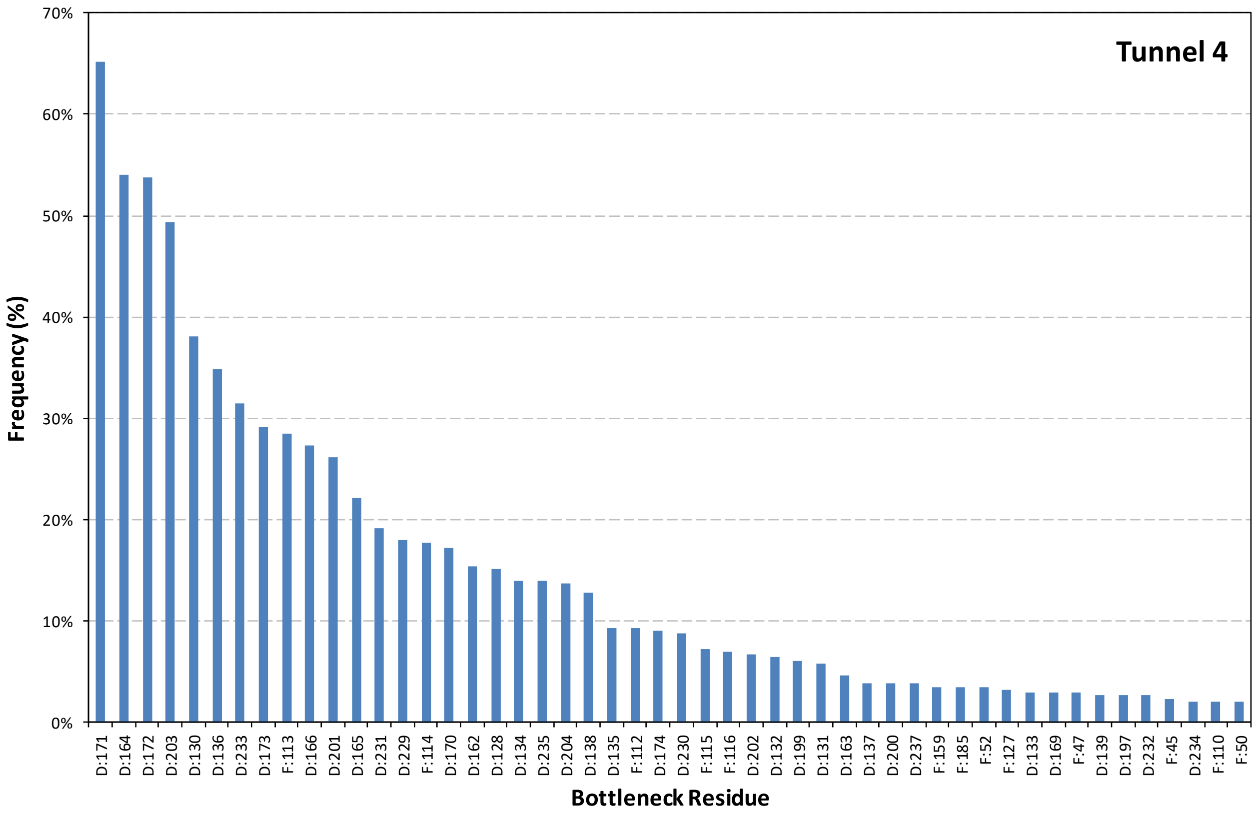


Figure 9SI. Analysis of bottleneck residues found in tunnel 4 (Table 1).

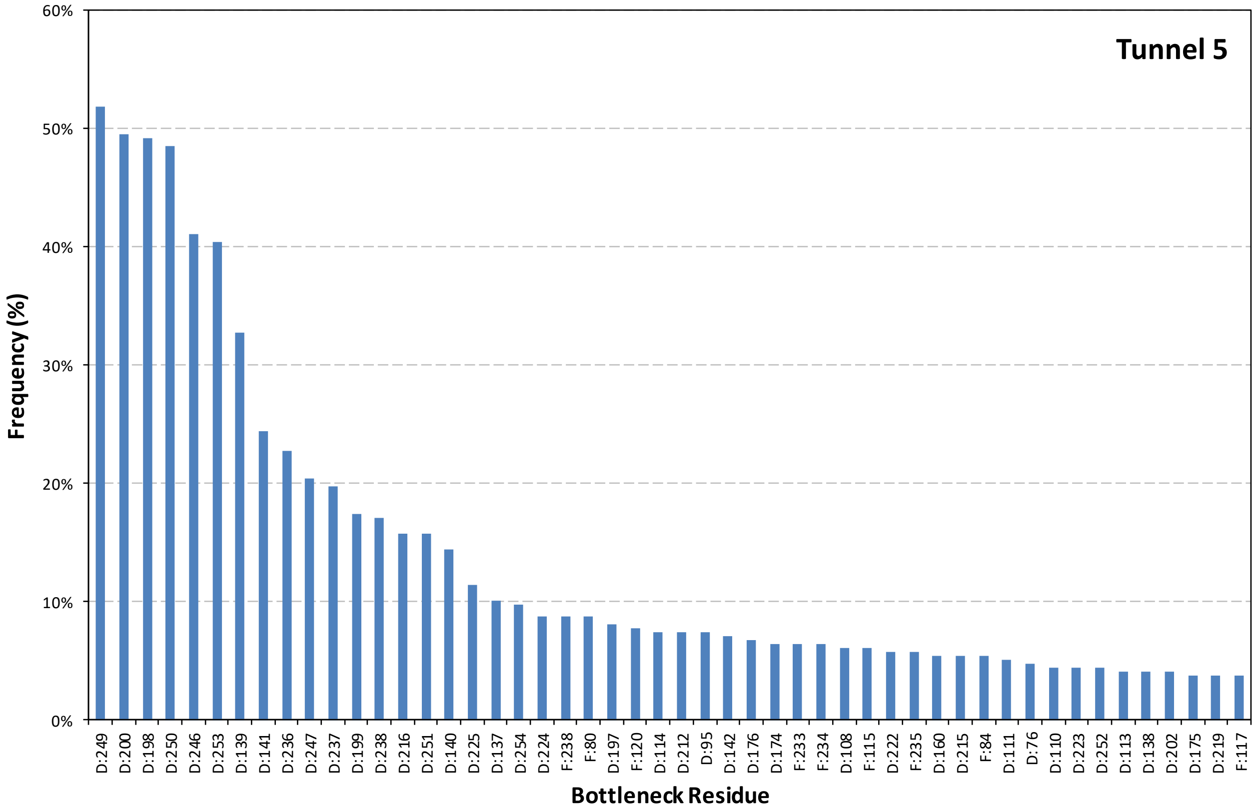


Figure 10-SI. Analysis of bottleneck residues found in tunnel 5 (Table 1).

Figure 11-SI. Multiple sequence alignment obtained by using PSI-BLAST and considering the 150 UreD best sequences with more than 15% and less than 90% sequence identity with respect to *Hp*UreD.

```
                                 1       10        20        30        40               50
                                 |        |         |         |         |                |
HpUreD                        ---MNTYAQESKLRLKTKIGADGRCVIEDNFFTPPFKLMAPFY------PKD-DLAEIML
UniRef90_A0A0K9H6B6_4_266     --------WTGILRLEAE-DRKGKTVAKNVYFQGAFKVMRPIY------HDDSGQPCYYI
UniRef90_A0A0K9GXZ4_4_265     --------WTGTLRLDVE-ERQGKTVAKNVYFQGALKVMRPVY------HDDSGQACYYI
UniRef90_A0A0A3IXZ5_4_264     --------WTGILSLDLE-NRNGKTVAKRAYFQGALKVMRPIY------HDDSGQVCYYL
UniRef90_K9ZPZ7_7_273         VNSPIDKNWHGRLNLVYA-KRQDSTQLIYNHHQAPFNIQRPFY------PEGQEVCHSVI
UniRef90_A0A0C1XDA3_13_272    -------GWHGKLNLVYA-DRLGTTALISNSHQAPLKVQRPFY------PEGQQICHSVI
UniRef90_UPI000379D7E3_4_266  --------WTGSLSLELE-DRNGKTVAKRVYFQGAFKVMRPIY------TDDSGQVCYYL
UniRef90_UPI0002DF3930_13_277 -------SWHGKLDLLYA-NRQGITQLIHAHHQAPLKVQRPFY------PEGKAVCHSVI
UniRef90_A0A0M0W0K1_1_266     -----MNDWTGSLSLDLE-DRNGKTVAKRVYFQGAFKVMRPIY------ADDSGQVCYYL
UniRef90_UPI000422C885_4_265  --------WTGTLRLDVE-VRQGKTVAKNVYFQGALKVMRPIY------HDNSGQACYYI
UniRef90_B4W160_8_270         --------WQGSLELVYA-NDQGKTRLVRDRITSPLKVQRPFY------PEGQGVCHTVV
UniRef90_Q8YQZ4_10_269        -------GWHGKLNLVYA-DRSNSTQLIYNHQQAPLKVQRPFY------PEGEKVCHSVI
UniRef90_A0A0P1BUZ9_3_269     VNSPIDKTWHGKLNLVYA-HRLNSTQLIHSHHQAPLKIQKPFY------PEGEKICHSVI
UniRef90_A0A0M1JRC4_17_282    ---VRQAGWQGILNLVYA-NHQGKTQVTDSYMKAPLKIQRPFY------PEGETICHSVV
UniRef90_UPI0007108B5C_4_266  --------WTGVLRLGAE-ERNGKTVAKNVYFQGAYKVMRPIY------HDESGQVCYYI
UniRef90_UPI0006A76A92_4_265  --------WTGDLSLDLE-NRNGKTVARNVYFQGAFKVMRPIY------HDNSGQVCYYL
UniRef90_W4ETC9_4_265         --------WTGVLSLDLE-ERRGKTVAKNVYFQGALKVMRPIY------HDNSGQVCYYL
UniRef90_A0A0D6KKC6_12_271    -------TWHGKLNLVYA-DRKNATQLIYNHQQAPLKVQRPFY------PEGEKVCHSVI
UniRef90_A0A0D8ZYE2_11_271    --------WHGSLNLVYA-QHQGKTQVIHAQVKAPLKVQRPFY------PE-DGVCHSVV
UniRef90_B2IT63_10_269        -------GWHGKLNLVYA-DRQGATQLIYNQQQAPLKVQRPFY------PEAEKVCHSVI
UniRef90_UPI000308766C_11_271 ------TGWYGKLSLVYA-HRQNTTTLIHNQNQAPLKVQRPFY------PEGQQVCHSVI
UniRef90_K9R6Q0_10_280        ----SSQSWHGKLNLVYN-HSQNKTALIQSRNQAPLKVQRPFY------PEGQSVCHSVI
UniRef90_A0A127D3L2_4_264     --------WTGVLCLDVE-DRNGKTVAKNVYFQGALKVMRPVY------HDDSGQACYYI
UniRef90_UPI00028931D6_4_266  --------WTGVLRLDAE-DRNGKTVAKNVYFQGAFKVMRPIY------HDDSGQACYYI
UniRef90_A0A0M0ENP6_4_266     --------WTGTLSLDLE-ERKGKTVAKRVYFQGALKVMRPIY------FDDSGQVCYYL
UniRef90_A0A0S3PHU6_16_275    -------GWHGKLDLVYA-DRSNSTGLIYNHQQAPLKVQRPFY------PEGDKICHSVI
UniRef90_A0A081NYG6_3_264     --------WTGYLRLAVA-PKHGKTIATDLYYENAFKLTRPLY------PDESGQPHYYI
UniRef90_A0A0K9GPB1_4_266     --------WTGELRLDLE-ERLGKTVAKNVYFQGALKVMRPVY------HDDSGQVCYYI
UniRef90_UPI000717378B_2_265  ------TAWTGELSLSLE-DRGGKTVAKRVYFQGALKVMRPVY------HDDSGKVCYYL
UniRef90_K9W1V1_14_278        -----KSSWHGSLDLKFA-CRDGGSQVIKSQGKAPLKVQRPFY------PEGREVCHSVI
UniRef90_K9U4Q1_17_272        -------TWHGNLDIVYA-LRNGKTQPISDRVQAPLKVQRPFY------PEG-DICHTAI
UniRef90_A0A0M2SWM6_4_265     --------WTGELSLDLE-NRNGKTVAGNVFFQGAFKVMRPIY------HDDSGQVCYYL
UniRef90_UPI0003652A48_11_275 ---LNPTDWHGILQLGFE-YRQGKTQLIRNQGQAPLKVQRPFY------PEGDEVCHSVI
UniRef90_UPI00047A812C_16_286 -------GWHGKLNLVYA-HSQSGTQLIFNRNQAPLKVQRPFY------PEGQEVCHSVI
UniRef90_K9TKA4_12_277        ------QGWHGSLELVYA-QSGNATQMVSAKATAPLKIQRPFY------PEGEGVCHSVI
UniRef90_M7NKH7_1_269         ---MTVTQWTGELNLSLE-DRAGKTVARNVYFQGAFKVMRPIY------HDDSGKVCYYL
UniRef90_K6DR39_4_265         --------WTGVLHLDAE-DRKGKTVAKKVYYQGAFKVMRPIY------HDNSGQACYYL
UniRef90_A0A0B4RFS1_1_265     -----MGQWTGELSLDIE-ERKGKTVAKNVYFQGAFKVMRPVY------LEDSGKVCYYL
UniRef90_A0A168N9T6_4_266     --------WTGSLSLDLE-DRLGKTVAKRVYFQGAFKVMRPIY------ADDSGQVCYYL
UniRef90_A0A139X4D9_7_274     -------GWHGKLDLVYA-QCQGKTTLIHNQNIAPLKVQRPFY------PEGQEICHSVI
UniRef90_UPI00030AB192_16_277 --------WHGNLHLVYA-DRQNQTQLIFNQNQAPLKVQRPFY------PEGQKVCHSVI
UniRef90_K7W9H0_9_269         ------QGWHGLLNLVYA-NRQDSTQLIYNHHQAPLKVQRPFY------PEGQEICHSVI
UniRef90_UPI00034A5E0C_10_263 -------GWHGQLNLVYA-DRHNSTQLIYNHHQAPLKVQRPFY------PEGEKVCHSII
UniRef90_A0A0A0E4Q3_4_265     --------WTGVLHLNAE-DRKGKTVAKNVYYQGAFKVMRPIY------HDDSGRACYYL
UniRef90_UPI000717263C_4_266  --------WTGSLSLDLE-DRKGKTVAKRVYFQGALKVMRPIY------LDDSGQVCYYL
UniRef90_K9QPZ4_10_269        -------GWHGKLNLVYA-DRLGATELIYNHQQAPLKIQRPFY------PEGERVCHSVI
UniRef90_UPI0002D28249_14_276 -----NQGWYGNLNLVYA-RTQGETQLIHSQSQAPLKVQRPFY------PEGKNVCHSVI
UniRef90_F9DU20_4_266         --------WTGVLDLAME-NRQGRSVAKSVYFQGAFKVMRPVY------LNNYSYPCYYL
UniRef90_UPI00047933CA_3_266  -------AWTGLLHLNAE-DRRGKTISKNVYFQGAFKVMRPIY------HDESGQPCYYL
UniRef90_W7RFE8_4_266         --------WTGVLRLDAE-EKNGKTIAKNVYFQGAYKVMRPIY------HDESGQACYYI
UniRef90_UPI0007441C40_4_265  --------WTGDLSLDLE-CRNGRTVARNVYFQGAFKVMRPIY------HDNSGQVCYYL
UniRef90_A0A0T7BRT9_7_268     -------SWHGKLHLTYG-SHGGKTQLVSSQNQAPLKVQRPFY------PEGEAVCHSII
UniRef90_UPI0006A78357_4_266  --------WTGVLHLGAE-ARNGKTVANKVYFQGAFKVMRPIY------HDDSGQVCYYI
UniRef90_UPI0007C7BB46_4_266  --------WTGVLRLSME-KKRDKTVTKQVYFQGAFKMLRPHY------LDNSGQACYFL
UniRef90_UPI000472611B_4_265  --------WTGVLSLDAE-VRNGKTVGKDVYFQGAYKVMRPVY------HDDSGQACYYI
UniRef90_UPI0002ACBA0D_11_270 --------WHGNLNLVYA-QHQGKTQVIHSQMKAPLKVQRPFY------PEG-GVCHSVV
UniRef90_A0A0C2KR59_11_269    ------TGWHGKLNLVYA-HRQNTTALIHNQNQAPLKVQRPFY------PEGQEVCHSVI
UniRef90_K9WE93_4_277         ----SSSGWHGSLQLVYA-HDHNGTQLTHAQVQAPLKVQRSFY------PEGSAVCHSVV
UniRef90_A0A0M0SP30_16_277    --------WYGNLDLVYA-YRQNQTQLIFNQNQAPLKVQRPFY------PEGQEVCHSVI
UniRef90_K9Q9K1_10_268        -------AWHGKLNLIYA-DRQNSTQLIYNHHQAPLKVQRPFY------PEGEKVCHSVI
UniRef90_A0A0S3TTX4_17_278    --------WHGNLHLVYA-HHQNQTQLIFNQNQAPLKVQRPFY------PEGKEVCHSVI
UniRef90_K9VQF3_56_323        -----QTAWHGRLNLAYA-NRSGATQIIHNQMQAPLKVQRPFY------PEGKDVCHSVI
UniRef90_W1SM28_4_265         --------WTGALNLEVE-DRKGKTVAKKVYFQGAFKVMRPIY------HDDSGHACYYL
UniRef90_A0A0C1N935_15_285    -------GWHGKLDLVYA-QRQGTTALIYNQNQAPLKVQRPFY------PEGQEICHSVV
UniRef90_A0A139SK67_11_275    ----DSSGWHARLALDFA-RQGPRTVLVGNRHTGPLRVQKPLH------PEGEAVCHAIV
UniRef90_UPI00041C7CDC_4_265  --------WTGVLRLDLE-ERLGKTVAKNVYFQGALKVMRPVY------HDDSGQVCYYI
UniRef90_K9XHH8_15_275        -------SWHGSLNMVYT-CVDGATTVTHQQMQAPLKVQRPFY------PEGAEVCHSVI
UniRef90_UPI0002E7D901_4_265  --------WTGVLELDVE-NRGGRTVADNIYFQGAFKVMRPVY------LNGYSHPCYYL
UniRef90_A0ZB05_10_269        -------TWHGKLNLVYA-DRLNSTQLIYSHNQAPLKVQRPFY------PEGAEICHSVI
```

```
UniRef90_D4ZSS4_5_270         ----NPSEWHGILNLTYS-QNHGITQVVDKYTTAPYKIQRPFY------PPGEEICHSVA
UniRef90_A0A0F5YFY1_6_271     -----STGWQGILRLKYA-KDNNSTKLIEAYSQAPLKIQRSFY------PEGKDICHNVI
UniRef90_K9T9Y9_59_322        -------GWHGSLDLVYS-CERDTTYLTHALVKAPLKVQRPFY------PEGKSVCHSVI
UniRef90_K9VAD2_10_273        ----NSHIWHGKLNLIYA-KRQNKTQLIHAHHQAPLNIQRPFY------PEGEEICHSVA
UniRef90_W7Z4J7_4_265         --------WTGVLRLEAE-ERNGKTVAKRVYFQGAFKVMRPIY------HDDSGQACYYI
UniRef90_Q47G52_13_278        -----SPSWHAELHLGFA-RAGERTVLRENRHRGPLRVQKALY------PEGEAVCQTIV
UniRef90_UPI0004024FB0_3_269  INSPIDKSWHGKLNLVYA-NCQNSTQLVYNHHQAPFKVQRPFY------PEGEEICHTVI
UniRef90_K8GMD3_12_265        -----SSSWQGKLTLSYG-FVDKETRILKSQVQAPLKVQRPFY------PEGSEICHSVI
UniRef90_A0A0S3UB12_16_269    --------WHGILRLDFE-NRRGITQLSRNQGQAPLKVQRPFY------PEGGGVCHCVV
UniRef90_A0A0F7D4R9_4_265     --------WTGILELDVE-NRQGRTVADNIYFQGAFKVMRPVY------FGKNSYPCYYL
UniRef90_A0YQS4_5_271         ----NLTGWQGILRLKYA-KDNNSTKLIETYSQTPLKIQRSFY------PEGKEICHNVI
UniRef90_A0A0Q8RCL2_14_276    ----SHARWQAQLTLGFV-SDRDTTRLMERTHRGPLRVQKPLY------PEGGQVCHAIV
UniRef90_UPI0005625009_5_271  ---VNAQGWQALLKLQFA-KRAERTRIVSRERRGPLAVQRPFY------PEG-GVCHTYL
UniRef90_UPI000305E365_16_284 --------WHGNLHLIYA-DRQNQTQLIFNQNQAPLKVQRPFY------PEGQKVCHSVI
UniRef90_UPI0002E38242_11_269 -----TPSWHAELELGYA-RFGATTRPVQRRHKGPLRVQKHLY------AEGPEVCQHII
UniRef90_G8Q511_12_270        -----TPSWHAELDLGYA-RFGDCTRPVQRRHKGPLRVQKHLY------AEGPEVCQHII
UniRef90_A0A168IUL4_4_266     --------WTGYLRLAVQ-RKKETTIPSEVFHQGALKVTPAIY------LDDPQQPCFYL
UniRef90_UPI0004799FB1_14_281 ---VAAPTWHAELHLGFA-HRAGRTVLRENRHRGPLRVQKALY------PEGDGVCQALL
UniRef90_Q3KIS7_11_269        -----TPSWHAELELAYA-RFGDCTRPTRRRHLGPLRVQKHLY------AEGPEVCQHII
UniRef90_K9XZ09_12_271        -------GWYGSLELVYA-HYEGKTKILHSFHQAPLKLQSPFY------PEGQQICHSVI
UniRef90_A0A090RU86_25_288    ------FGWKARLSLGFV-NRGDKTVLKHRSQQGPLAIQRPLY------PEG-NPCHTYL
UniRef90_A0A0M5LWB1_11_269    -----TPSWHAELELGYG-RFGDSTRPTLRRHFGPLRVQKHLY------AEGPEVCQHII
UniRef90_A0A0C2I8A1_11_269    -----TPSWHAELELGYA-RFGDCTRPVTRRHLGPLRVQKHLY------AEGPEVCQHII
UniRef90_H3SNI3_3_265         -------SWTGQLELKIE-HRGDRSAASRQYHQGAYKIARPIY------PDRSGQVHYYV
UniRef90_A0A0Q0XPT8_11_269    -----TPSWHAELALGYG-RFGDSTRPTLRRHLGPLRVQKHLY------AEGPEVCQHII
UniRef90_K4ZJ70_4_265         --------WTGVLRLNAE-ERNGKTVARDVYFHGAFKVMRPIY------HDDSGQACYYI
UniRef90_A0A085V951_11_269    -----TPSWHAELELGYG-RFGDSTRPTQRRHKGPLRVQKHLY------AEGPEVCQHII
UniRef90_A0A098SV87_11_271    -----TPSWHAELELAYG-RFGDSTRPTQRRHKGPLRVQKHLY------AEGPQVCQHII
UniRef90_A0A0A1GEL6_9_267     -----TPHWNAELELGYA-RFEGGTRPVLRRHSGPLRVQKHLY------PEGPEVCQHII
UniRef90_I4CP19_9_267         -----TPHWHAELELGYA-LCAGATRPVLRRHNGPLRVQKHLY------PEGPDVCQHII
UniRef90_A0A0W0P2S5_11_269    -----TPSWHAELELAYA-HAGGATRPVLRRHQGPLRVQKHLY------AEGPDVCQHII
UniRef90_C3K5A6_11_269        -----TPSWHAELELGYA-RFGDTTRPVMRRHLGPLRVQKHLY------AEGPEVCQHII
UniRef90_A0A078LV36_9_267     -----TPEWHAELELGYA-RFGDCTRPVQRRHSGPLRVQKHLY------AEGPEVCQHII
UniRef90_J3GGT8_11_269        -----TPSWHAELELGYA-RFGESTRPVQRRHKGPLRVQKHLY------AEGPEVCQHII
UniRef90_A5L5M9_32_304        -------GWQASLNLTFV-DRGDKTVLKNRQQSGPLAVQRPLY------PDG-ETCHTYL
UniRef90_A0A0J6GPD8_11_270    -----TPSWDAELELGYG-RFGESTRPTLRRHAGPLRVQKHLY------AEGPQVCQHII
UniRef90_Q4KJ05_11_269        -----TPSWHAELELAYA-RCGATTRPVLRRHLGPLRVQKHLY------AEGPEVCQHII
UniRef90_A0A0V7ZQT1_2_281     ------------LNLTYA-SRQGKTVIVEQQNQAPLKVQRPLYPEEQTDPEGQKICHSII
UniRef90_A0A075PF54_11_269    -----TPSWHAALELGYA-RFGATTRPVLRRHLGPLRVQKHLY------AEGPEVCQHII
UniRef90_A0A0D5Y774_11_269    -----TPSWHAELELGYA-RYGDSTRPVLRRHQGPLRVQKHLY------AEGPEVCQHII
UniRef90_A0A0Q5EC30_11_271    -----TPSWHAELELGYA-RFGDSTRPVQRRHLGPLRVQKHLY------AEGPEVCQHII
UniRef90_I4N4V9_11_269        -----APSWHAELELGYG-RFGDTTRPTLRRHLGPLRVQKHLY------AEGPEVCQHII
UniRef90_D4TUH6_7_266         -------SWHGKLELVYA-QRQNSTQLMFSHNQAPLKVQRPFY------PEGEKICHSVI
UniRef90_UPI00048AE6EC_1_266  -----MNTWTGNLQLKIE-NKKGKSIPKDIYFQGAFKLMRPKY------FDDSGQPCYFI
UniRef90_UPI0007398261_11_285 --------WQGRLALTFC-DRQGETYLSRCFVQAPLKVQRPFY------PEGRGVCHGVM
UniRef90_B0C790_12_273        -------SWHGRLSLTYE-KKAHQTQVQQSYHQAPLNLQRPFY------PEG-PVCHSVM
UniRef90_A0A066UMR7_26_298    -------GWQANLNLTFA-DRGDKTVLKHRYQLGPLAVQRPLY------PDG-KTCHSYL
UniRef90_A0A0Q9XWI2_4_265     --------WTGILQLDVE-NRQGKSVTKNLHFQGAFKIMRPVY------HQQSKHPCYYI
UniRef90_A0A0S7ZTJ9_14_276    ---MQANGWHAQLDLDFA-QRESRTVLARRQHFGPLVVQKPFY------PEG-AVCHVYI
UniRef90_A6SZ04_26_286        ---------QARLSLAFT-DDAGTTRMTERSHFGPLRVQKTLY------PEHPAVCHAII
UniRef90_U3H3U5_9_267         -----TPSWHAELELGYA-RRDERTIPVLRRHLGPLRVQKHLH------AEGPEVCQHII
UniRef90_A0A073CY52_6_272     ---INPSQWQGILELDYQ-KINNSTQLVKAYSQAPLKIQRSFY------PEGQEICHSII
UniRef90_U6ZYX1_11_270        -----TPSWHAELELAYA-CTGNATRPVLRRHQGPLRVQKHLY------AEGPQVCQHII
UniRef90_UPI00067CF5D4_26_284 ---------VARLRLGFS-DDAGVTRMTERSHFGPLRVQKPLY------PEHPSICHAII
UniRef90_A0A0K2BGW7_26_288    ---------KARLTLGFA-DDAGTTRMIERSHFGPLRVQKPLY------PEHPAVCHAVI
UniRef90_A0A011QEK6_40_302    -----RPGWQARLALAFE-RRGETSALVRREHFGPLRVQKALY------PEGPDVCHAIL
UniRef90_G4T117_15_272        ----SRQGWQAELRLGFA-KNESRTVLRHRAHRGPLTVQRPFY------PEG-DVCHLYL
UniRef90_Q87VP5_11_271        -----TPSWHAELELGYG-RFYDCTRPTQRRHKGPLRVQKHLY------AEGPEVCQHII
UniRef90_A0A089YS31_11_269    -----TPSWHARLELGYA-RFGDSTRPTLRRHLGPLRVQKHLY------AEGPEVCQHII
UniRef90_UPI0004174C9A_9_267  -----TPHWQAELELGYA-RIGGATRPVLRRHSGPLRVQKHLY------AEGPEVCQHIV
UniRef90_UPI000379F3E4_11_286 -------GWQGSLNLIYA-NRTGTSQLVGERVGAPLKVQRPFY------PEGAAVCHSVI
UniRef90_A4VQU8_9_267         -----TPHWQAELELGYT-RIGDATRPVLRRHSGPLRVQKHLY------PEGSEVCQHII
UniRef90_A0A0M3V4J1_7_297     ----TATAWQGKLNLVYE-NCQNSTQLIYNHHQAPLKVQRPFY------PEGEQVCHSVI
UniRef90_A0A098ESZ3_4_266     --------WTGILRLDAE-DRYGKTVARNVYFQGALKVMRPIY------HDNSGQACYYI
UniRef90_UPI000345DD51_13_277 ----NDSAWHARLTLGFA-DDAGTTRLVERSHAGPLRVQKPLY------PEGGAVCHAIV
UniRef90_UPI00047D06E7_9_270  -----SSGWKAKLELAYV-ARPERTVLARSSRRGPLAVQRAFY------PED-GVCHSYV
UniRef90_I3BUX5_1_262         ---MAASGWQAELSLGFA-CRGGKTVLAERRQRGPLAVQRPFY------PEG-DVCHAYV
UniRef90_UPI000780B9AB_11_271 -----TPSWHAELELGYG-RFGDSTRPIQRRHKGPLRVQKHLY------AEGPEVCQHII
UniRef90_UPI000255752C_11_273 -----TPSWHAELELGYA-RDTDTTRPVMRRHKGPLRVQKHLY------AEGPQVCQHII
UniRef90_UPI000484E5E1_11_269 -----TPSWHAELQLGYA-RFGDSTRPILRRHSGPLRVQKHLY------AEGPQVCQHII
UniRef90_A0A0D6AS13_1_264     ---MSGQNWQGKINLVYE-YQEGKTKIKSAYHQAPLKIQRSFY------PEGDSICHSVI
UniRef90_UPI00034A425D_18_281 --------WHGKLELDFA-NRDGATHVKHSYSQAPWKLQRPFY------PEGDRICHSVL
UniRef90_UPI00046A7B93_11_269 -----TPSWHAELELAYA-RFGATTRPVLRRHLGPLRVQKHLY------AEGPQVCQHII
UniRef90_I3YAM1_6_262         ------PGWNARLALGFA-ERAGRTRLVERHQRGPLAVQRPFH------PEG-APCHCYL
UniRef90_A0A0D9AIT7_9_267     -----TPHWNAELELGYA-RFDHVTRPVLRRHSGPLRVQKHLY------PEGPEVCQHII
UniRef90_UPI0003FDB5F5_4_265  --------WTGVLDLVME-NRLGRSVAKSVYFQGAFKVMRPVY------FNKNSYPCYYL
UniRef90_K9SBL5_6_267         ----TTQPWCGTLDLVYA-QRQGITQPIHNLALAPLKIQRPFH------PEG-AVCHSVL
UniRef90_UPI00040AC544_17_274 ----SAQGWEAKLALGFA-RQHGKTVLAHRRHYGPLTVQRPFY------PEG-GVCHVYI
```

```
UniRef90_A6D6Q9_31_296        ----TQFGWKASLDLTFI-DRGDKTVLKHRSQQGPLAIQRPLY------PEG-NPCHTYL
UniRef90_B8HW54_12_274        -------SWQGNLDLKFA-RKQDTTQLIHCLGKAPLKLQRPFY------PEGPQICHGVI
UniRef90_A0A0J6H3B0_11_269    -----TPHWLAELELGYG-RFGDSTRPTLRRHCGPLRVQKHLY------AEGPQVCQHII

                             51       60        70        80        90                100
                              |        |         |         |         |                 |
HpUreD                        LAVSPGMMRGDAQDVQLNIGPNCKLRITSQSFEKIHNTEDGF--------ASRDMHIVVG
UniRef90_A0A0K9H6B6_4_266     LNPGGGYLDGDRYRMQIALDKQAKMTLTTQSATKVYKTPNNY--------VYQETEISLK
UniRef90_A0A0K9GXZ4_4_265     LNPGGGYLDGDRYNLQFSLKEKTKLTLTTQSATKVYKTPNQH--------AYQETEFILK
UniRef90_A0A0A3IXZ5_4_264     LNPGGGYLDGDRYRMEISVDAGAEVILTTQSATKVYKTPKSL--------AYQETEITLK
UniRef90_K9ZPZ7_7_273         LHTAGGIVGGDRLSSDIHLEKDSQALITTAAAGKVYRSNGLP--------AKQTVNIQIG
UniRef90_A0A0C1XDA3_13_272    LHTAGGVVGGDRLSLNFHLQPNTQALITTAAASKIYRSNGTQ--------AKQSVNIQVD
UniRef90_UPI000379D7E3_4_266  LNPGGGYLDGDRYKMDISADEGSKVTLTTQSATKVYKTPKNY--------AYQETVIRLK
UniRef90_UPI0002DF3930_13_277 LHTAGGVVGGDKLSCNFQLQPESQVLITTAAAGKIYRSNGRQ--------ATQNINIEVS
UniRef90_A0A0M0W0K1_1_266     LNPGGGYLDGDRYKMQISADEGSKVTLTTQSATKIYKTPKSH--------AYQETEINLK
UniRef90_UPI000422C885_4_265  LNPGGGYLDGDRYKLQFSLKEQAKLTLTTQSATKVYKTPNQH--------AYQETEFFLK
UniRef90_B4W160_8_270         LHTAGGIVGGDRLSQTIHLQEDSQALITTAAASKIYRSNGQR--------ANQRIHIHVE
UniRef90_Q8YQZ4_10_269        LHTAGGVVGGDRLSYNLHLQPNAQALITTAAAGKVYRSDGLQ--------ARQTIEIKID
UniRef90_A0A0P1BUZ9_3_269     LHTAGGVVGGDRLSYNLHLQPHTQALITTAAASKIYRTNGLQ--------ARQNINIQID
UniRef90_A0A0M1JRC4_17_282    LHTAGGIVGGDRLAQNFHLRENAKALITTAAASKIYRSNGNN--------AQQTINIKVD
UniRef90_UPI0007108B5C_4_266  LNPGGGYLDGDRYQMKISLEKQAKLTLTTQSATKIYKTPNSH--------AYQEAEFNLK
UniRef90_UPI0006A76A92_4_265  LNPGGGYLDGDRYRMKISAAENSKVTLTTQSATKVYKTPTDH--------VYQETEISLK
UniRef90_W4ETC9_4_265         LNPGGGYLDGDRYKMEISANEGAKVTLTTQSATKVYKTPKSF--------AYQETTISLK
UniRef90_A0A0D6KKC6_12_271    LHTAGGVVGGDRLSLNFHLQPHAQALITTAAASKIYRSNGLQ--------AKQIIDIKVD
UniRef90_A0A0D8ZYE2_11_271    LHTAGGIVGGDRNTLSFHLQPQSQALITTAAASKIYRSNGSS--------ASQNVQIQVD
UniRef90_B2IT63_10_269        LHTAGGMVGGDRLSSNIHLQPQAQALITTAAASKIYRSNGLQ--------ARQTIQMQVD
UniRef90_UPI000308766C_11_271 LHTAGGIVGGDRLSCNFHLQPNAQALITTAAASKIYRSNGTQ--------AQQNIEIRVD
UniRef90_K9R6Q0_10_280        LHTAGGIVGGDRLSSDFHLQPDSKALITTAAANKIYRSNGLQ--------ARQNIDIKID
UniRef90_A0A127D3L2_4_264     LNPGGGYLDGDRYKIQITLEKQARLTLTTQSATKVYKTPNTH--------AYQETEIILQ
UniRef90_UPI00028931D6_4_266  LNPGGGYLDGDRYQLKISLEKQAKLTLTTQSATKIYKTPKKH--------AYQETEINLK
UniRef90_A0A0M0ENP6_4_266     LNPGGGYLDGDRYKMEISADEGSKVTLTTQSATKVYKTPKGY--------AYQETQIHLQ
UniRef90_A0A0S3PHU6_16_275    LHTAGGVVGGDRLSYNFHLQPNAQALITSAAAGKIYRSNGLL--------AKQTINIKID
UniRef90_A0A081NYG6_3_264     LNPGGGYVDGDRYKLDIELAEDAKLLLTTQSSTKVYKTRNTA--------PVQDMEIRMK
UniRef90_A0A0K9GPB1_4_266     LNPGGGYLDGDRYHLHIALEEQARVTLTTQSATKVYKTPQSY--------AYQETEIFLK
UniRef90_UPI000717378B_2_265  LNPGGGYLDGDRYRMEIAVEEGAKVTLTTQGATKVYKTPNDH--------VYQETEISLK
UniRef90_K9W1V1_14_278        LHTGGGVVGGDRLSLNFHLQPNTHALITSAAAGKVYRSNGLE--------ARQTVEMRVE
UniRef90_K9U4Q1_17_272        LHTAGGIVGGDCLSINLQLQPRSQALVTTTAASKIYRSNGLQ--------ARQIVEIQID
UniRef90_A0A0M2SWM6_4_265     LNPGGGYLDGDRYRMKISAGEDSKVTLTTQSATKVYKTPKDH--------VYQESEILLK
UniRef90_UPI0003652A48_11_275 MHTAGGIVGGDRLTFDFHLASGSQALITTPAASKIYRTNGRE--------AHQVIRVDVA
UniRef90_UPI00047A812C_16_286 LHTAGGVVGGDRLSYHLHLQPNAQALITTAAAGKIYRSNGTT--------ARQTIEIKVD
UniRef90_K9TKA4_12_277        LHTAGGIVGGDSLGQSIHVQENAHALITTAAAAKIYRSTGEV--------ARQSIKINVD
UniRef90_M7NKH7_1_269         LNPGGGYLDGDRYRMDITAEPDARVTLTTQGATKVYKTPKDH--------AYQETTMTLK
UniRef90_K6DR39_4_265         LNPGGGYLDGDSYQMKISLAENARMTLTTQGATKVYKTPNKY--------AYQESDISLK
UniRef90_A0A0B4RFS1_1_265     LNPGGGYLDGDRYRMKVSASTGSKVTLTTQSATKVYKTPKGY--------AYQETEIYLE
UniRef90_A0A168N9T6_4_266     LNPGGGYLDGDRYKMEISADEGSKVTLTTQSATKVYKTPTSY--------AYQEAEINLK
UniRef90_A0A139X4D9_7_274     LHTAGGVVGGDRLSYKVHLQPKAQALITTAAASKIYRSNGAQ--------ARQNIDIQVD
UniRef90_UPI00030AB192_16_277 LHTAGGVVGGDRLSYDFHLQPHAQALITTATAGKIYRTNGMT--------AKQMIEIKVD
UniRef90_K7W9H0_9_269         LHTAGGIVGGDRLSSHIHLQPDTNALITTAAAGKIYRSNGLP--------ARQTVNIQVD
UniRef90_UPI00034A5E0C_10_263 LHTAGGIVGGDRLSSKIHLQPDAQAVITTAAASKIYRSNGLP--------ATQTINLKID
UniRef90_A0A0A0E4Q3_4_265     LNPGGGYLDGDRYQMKIGLGENAKLTLTTQGATKVYKTPNRF--------AYQESEISLQ
UniRef90_UPI000717263C_4_266  LNPGGGYLDGDRYKLEFTADEGSKVTLTTQSATKVYKTPKSH--------AYQETTIHLK
UniRef90_K9QPZ4_10_269        LHTAGGVVGGDRLSTKIHLQPHTQAVITTAAAGKIYRSNGLQ--------ARQTIDIQID
UniRef90_UPI0002D28249_14_276 LHTAGGIVGGDRLSCNFHLEPHAQALITTAAASKIYRTNGLR--------ARQTINIQID
UniRef90_F9DU20_4_266         LNPGGGYLDGDRYRMEITLDEQAQLTLTTQSATKVYKTPTRQ--------VYQESVFHMK
UniRef90_UPI00047933CA_3_266  LNPGGGYLDGDRYKMKISLEKNARVTMTTQSATKVYKTPKNH--------AYQEAEFLLE
UniRef90_W7RFE8_4_266         LNPGGGYLDGDRYQMKISLKKQARLTLTTQSATKIYKTPHTH--------AYQETEFNLR
UniRef90_UPI0007441C40_4_265  LNPGGGYLDGDRYRMKISAAENSKVTLTTQSATKVYKTPKDH--------VYQESEIFLK
UniRef90_A0A0T7BRT9_7_268     LHTAGGVVGGDRLSTNIHLQPQSQVLLTTAAASKIYKSNGLQ--------ARQDVTIHID
UniRef90_UPI0006A78357_4_266  LNPGGGYLDGDRYQMKISLEKQARLTLTTQGATKIYKTPNKH--------AYQETEITLK
UniRef90_UPI0007C7BB46_4_266  INPGGGYVDGDRYRMEITLEEQAELILTTQSATKVYRTPHTP--------VLQENEIILK
UniRef90_UPI000472611B_4_265  LNPGGGYLDGDRYQMQVTVQEEARLTLTTQSATKIYKSPRSF--------SYSEMEFTLK
UniRef90_UPI0002ACBA0D_11_270 LHTAGGIVGGDRNNLSFHLQPHSQSLITTATASKIYRSNGLL--------AKQNIQMQVD
UniRef90_A0A0C2KR59_11_269    LHTAGGVVGGDRLSCNFHLQRNAQALITTAAASKIYRSNGTQ--------ARQSIEITVD
UniRef90_K9WE93_4_277         LHTAGGIVGGDRLSQTIHLSPHAHALITTAAASKIYGSKGKSRACPQGQPAKQTIQIQID
UniRef90_A0A0M0SP30_16_277    LHTAGGVVGGDRLSYDFHLQPDAQALITTATAGKIYRSNGMI--------AKQVIEIKVE
UniRef90_K9Q9K1_10_268        LHTAGGVVGSDRLSSHFHLQPNTQALITTAAASKIYRSNGLQ--------ARQTIDIQID
UniRef90_A0A0S3TTX4_17_278    LHTAGGVVGGDRLSYNFHLQPHAQALITTATAGKIYRSNGTI--------AKQKIEIKVD
UniRef90_K9VQF3_56_323        LHTAGGVVGGDRLSGHFHLQPNAKALITTAAAGKIYRSSGLE--------SQQNIDIQLD
UniRef90_W1SM28_4_265         LNPGGGYLDGDRYQMKLSLLEKAKLTLTTQSATKVYKTPNNF--------AYQEAEISLQ
UniRef90_A0A0C1N935_15_285    LHTAGGVVGGDRLSCKVHLQPKAQALITTAAASKIYRSNGAQ--------ARQNIEIQID
UniRef90_A0A139SK67_11_275    LHPPAGIVGGDSLAIAVDLAAGAHALLTTPGAGKWYRSGGPS--------ASLTQTITVG
UniRef90_UPI00041C7CDC_4_265  LNPGGGYLDGDRYHLQISLEKQARLTLTTQSATKVYKTPYSY--------AYQETEIFLK
UniRef90_K9XHH8_15_275        LHTAGGVVGGDKLSLNFHLQQNAHTLITTAAASKIYRSNGWE--------ARQNIQVQVD
UniRef90_UPI0002E7D901_4_265  LNPGGGYLDGDTYRMKVTLAEDSRLTLTTQSSTKVYKTPVSY--------AYQETEFHLE
UniRef90_A0ZB05_10_269        LHTAGGVVGGDRLSSNIHLQTNAQALITTSAASKIYGTNGLQ--------ARQNIDIQVD
UniRef90_D4ZSS4_5_270         LHTAGGMVGGDRLSQNLHLQADTKVLLTTAAASRVYRSTGKT--------ASQNVKIKLE
UniRef90_A0A0F5YFY1_6_271     LHTAGGIVGGDRLLQEIHLQPKTQAVITTASASKIYRSSGKQ--------AKQTVKIKVD
UniRef90_K9T9Y9_59_322        LHTAGGIVEGDSLSQTIRLRENANTLITTAAAGKVYRSNGKL--------AKQIVRIKIE
```

```
UniRef90_K9VAD2_10_273        LHTAGGVVGGDKLSYDVHLQDNSQALITTAAASKIYRSNGYQ--------AKQDIKIKLD
UniRef90_W7Z4J7_4_265         LNPGGGYLDGDRYQMKISLNERARLTLTTQSATKIYKTPNQL--------AYHETEINLK
UniRef90_Q47G52_13_278        LHPPSGIAGGDHLAISAEVGEGSHAQLTTPGAGKWYRSGGAE--------ASQRVAFTVG
UniRef90_UPI0004024FB0_3_269  LHTAGGILGGDRLTSDIHLQPQTNALITTAAASKIYRSQGLP--------ARQIVNIQVN
UniRef90_K8GMD3_12_265        LHTAGGVVGGDRLFLNLELQPQAHALITTAAAGKIYRSNGLE--------AQQIVTVKVA
UniRef90_A0A0S3UB12_16_269    MHTAGGIVGGDRLSFEFRLAADSRSLITTPAASKIYRTNGRE--------AHQTIQIDIE
UniRef90_A0A0F7D4R9_4_265     LNPGGGYLDGDTYRMRVSLGEDSRLTLTTQSSTKVYKTPKSY--------AYQETEFHLK
UniRef90_A0YQS4_5_271         LHTGGGIVGGDRLLQEIHLQPQTQALITTASASKIYRSSGKQ--------AKQTINIQVD
UniRef90_A0A0Q8RCL2_14_276    VHPPGGVVGGDELRIDASVGQNAGALLTTPGAAKWYKANGHI--------SRQDVHLQAG
UniRef90_UPI0005625009_5_271  LHPPGGVVGGDSLDIQVHVESGAQALITTPGATKFYRSGGRL--------ATQIQTLSVA
UniRef90_UPI000305E365_16_284 LHTAGGVVGGDRLSYNFHLQPHAQALITTATAGKIYRSNGMT--------AKQIIEIKVD
UniRef90_UPI0002E38242_11_269 VHPPGGIAGGDRLDIRASVERDAWAQLTSPGAAKWYRATGS---------ASQTLSLKVA
UniRef90_G8Q511_12_270        VHPPGGIAGGDRLDISAHVGPDAWAQLTSPGAAKWYRAAGP---------AYQQLSLSVA
UniRef90_A0A168IUL4_4_266     MSFGGGYVGGDRYKLEIHLGEQAQMLLTTQSATKIYKTINRP--------AMQEMNIVLE
UniRef90_UPI0004799FB1_14_281 LHPPSGIAGGDHLAIDIAVDDNSHAQITTPGAGKWYRSGGAA--------ASQTVTLNVA
UniRef90_Q3KIS7_11_269        VHPPGGIAGGDRLDISARVAQGAWAQITSPGAAKWYRAAGP---------AYQSLNLHVA
UniRef90_K9XZ09_12_271        LHTAGGIVGGDRLSQKIQVQPNAHSLITTAAASKIYRSNGRE--------AKQILTIEIE
UniRef90_A0A090RU86_25_288    LHPPGGVVGGDTLQIKAKAERGASVLITTPGATKFYRSENKY--------AKQSQILSVE
UniRef90_A0A0M5LWB1_11_269    VHPPGGIAGGDRLNICASVGPDAWAQLTSPGAAKWYRAAGP---------AYQSLTLNVA
UniRef90_A0A0C2I8A1_11_269    VHPPGGIAGGDRLAISARVEPAAWAQLTSPGAAKWYRATGP---------AYQTLDLKVA
UniRef90_H3SNI3_3_265         MNPGGGYVGGDRYRMELELGEGASSLMTTQSSTKIYRTPKEP--------VFQLTRIALE
UniRef90_A0A0Q0XPT8_11_269    VHPPGGIAGGDRLHINAHVGPDAWAQLTSPGAAKWYRAAGP---------AYQTVELSVA
UniRef90_K4ZJ70_4_265         LNPGGGYLDGDRYLMQVSLEEGARLTLTTQAATKIYKTPKKP--------AYQEVEIRLK
UniRef90_A0A085V951_11_269    VHPPGGIAGGDRLDISASVGPNAWAQLTSPGAAKWYRAAGP---------AYQQLDLQVA
UniRef90_A0A098SV87_11_271    VHPPGGIAGGDRLDIDVSVGANAWAQLTSPGAAKWYRAAGP---------AYQQLELSVA
UniRef90_A0A0A1GEL6_9_267     VHPPGGIAGGDRLDISVAVGSGAWAQLTSPGAAKWYRAGGP---------AFQNVHLRVE
UniRef90_I4CP19_9_267         VHPPGGIAGGDRLDISASVAAGAWAQLTSPGAAKWYRAVGP---------AFQDLRLRVE
UniRef90_A0A0W0P2S5_11_269    VHPPGGIAGGDRLAIHARVDTGAWAQLTSPGAAKWYRANGP---------ASQRLELQVA
UniRef90_C3K5A6_11_269        VHPPGGIAGGDRLDISAHVGAGAWAQLTSPGAAKWYRAGGP---------AYQQLDLCVE
UniRef90_A0A078LV36_9_267     VHPPGGIAGGDRLDISAHAGRNAWVQLTSPGAAKWYRSSGP---------ASQTLNLHVE
UniRef90_J3GGT8_11_269        VHPPGGIAGGDRLDISASVGTNAWAQITSPGAAKWYRAAGP---------AYQKLDLRVA
UniRef90_A5L5M9_32_304        LHPPGGVVGGDTLNINVNLEHGAHALITTPGATKFYRSNNKY--------AKQKQTLRVE
UniRef90_A0A0J6GPD8_11_270    VHPPGGIAGGDRLNIRASVGPGAWAQLTSPGAAKWYRAAGP---------AYQTVELKVA
UniRef90_Q4KJ05_11_269        VHPPGGIAGGDRLAISARVDSGAWAQLTSPGAAKWYRATGP---------ASQTLNLQVA
UniRef90_A0A0V7ZQT1_2_281     LHTAGGVVGGDRLSCNFDLQPNSQALITTAAASKIYRSNGSQ--------ARQKIEINVD
UniRef90_A0A075PF54_11_269    VHPPGGIAGGDRLDITAHLAQGAWAQLTSPGAAKWYRASGP---------AYQQLALTVE
UniRef90_A0A0D5Y774_11_269    VHPPGGIAGGDRLQIRASVDRDAWAQLTSPGAAKWYRATGP---------AYQTLELTIA
UniRef90_A0A0Q5EC30_11_271    VHPPGGIAGGDRLAISAQVGPQAWAQLTSPGAAKWYRAAGP---------AFQHLDLHVE
UniRef90_I4N4V9_11_269        VHPPGGIAGGDRLDINVNVGADAWAQLTSPGAAKWYRAAGP---------AYQTLNLSVA
UniRef90_D4TUH6_7_266         LHTAGGVVAGDRLSSNIHLQSETDVLITTAAASKIYRSNGLY--------AKQTVSIKID
UniRef90_UPI00048AE6EC_1_266  LNPGGGYLDGDRYRMDLNLEEKAELLLTTQAATKVYKTPNQS--------VIQETNISMG
UniRef90_UPI0007398261_11_285 LHTAGGIVGGDRLSTTLQLEANAHALLTTATAGKVYRSNGQE--------AQQEVQISLA
UniRef90_B0C790_12_273        MHTAGGMVGGDRLSINVTLQPQTHALLTTTSAGKVYRSNGHG--------AQQTVQCQLD
UniRef90_A0A066UMR7_26_298    LHPPGGVVGGDTLNIDISAESGAHTLITTPGATKFYRSNAKY--------AKQKQMLHVA
UniRef90_A0A0Q9XWI2_4_265     LNPGGGYLNGDTYRMEVSVQEEAKATLTTQSATKIYKTPSKH--------AYQETEIHLK
UniRef90_A0A0S7ZTJ9_14_276    IHPPGGVVGGDTLSINVTSCERSQTLITTPAANKFYRSSGPV--------ARLEQILTLK
UniRef90_A6SZ04_26_286        VHPPGGIVGGDQLTITARVGDRAHALLTTPGAGKWYRANGQL--------SQQQVSLEVG
UniRef90_U3H3U5_9_267         VHPPGGIAGGDRLDLSARVGDGAWAQLTSPGAAKWYRARGP---------AFQTLALRVE
UniRef90_A0A073CY52_6_272     LHTAGGMVGGDRLSQTINLQPETQVLLTTPAASKIYRSSGET--------AQNTINIEIQ
UniRef90_U6ZYX1_11_270        VHPPGGIAGGDRLAIRARVDSGAWAQLTSPGASKWYRAASP---------ASQLLELAVE
UniRef90_UPI00067CF5D4_26_284 VHPPGGIVGGDELHIDATLGDQAHALLTTPGAGKWYRANGNL--------SHQYVTLQAA
UniRef90_A0A0K2BGW7_26_288    VHPPGGILGGDVLNISARVGDNAHALLTTPGAGKWYRANGFV--------SQQQVSLTAT
UniRef90_A0A011QEK6_40_302    LHPPGGIAGGDSLEISLHVGAGAHALLTTPGAGKWYRSGGRP--------ARQSLSVRVG
UniRef90_G4T117_15_272        LHPPGGVVAGDRLTIEVKAEHGSHALITTPAAGKFYRSEGDC--------ASQSVTMTIE
UniRef90_Q87VP5_11_271        VHPPGGIAGGDRLDISVNVGAHAWAQLTSPGAAKWYRAASP---------AFQQLELHVQ
UniRef90_A0A089YS31_11_269    VHPPGGIAGGDRLDISAQVQANAWAQLTTPGAAKWYRALSP---------AYQSTELHVA
UniRef90_UPI0004174C9A_9_267  VHPPGGIAGGDRLAIRAAVGDGAWTQLTSPGAAKWYRAAGP---------AFQTLELRVG
UniRef90_UPI000379F3E4_11_286 LHTAGGVVGGDRLCFDITLHPDTRALITTAAAGKIYRTNGLE--------AKQTALIKIA
UniRef90_A4VQU8_9_267         VHPPGGIAGGDRLDISVTLGPGAWAQLTSPGAAKWYRAASP---------AFQDLQLRVE
UniRef90_A0A0M3V4J1_7_297     LHTAGGVVGGDHLSSHFHLKPNTQALITTAAASKIYRSNGLQ--------ARQTIDIQID
UniRef90_A0A098ESZ3_4_266     LNPGGGYLDGDRYRMQFLLEENAKLTLTTQSATKVYRTPRLH--------AYQETEIVMK
UniRef90_UPI000345DD51_13_277 VHPPGGVVGGDQLSISARVGPAAHALLTSPGAAKWYKANGKV--------SRQQVYLQAQ
UniRef90_UPI00047D06E7_9_270  LHPPGGVVGGDELIIDVTVEAEAHALLTTPGATKFYRSAGPF--------AEQIQQFKVE
UniRef90_I3BUX5_1_262         LHPPGGVVGGDALHLHFNVDAAAHALLTTPGATKFYRSAGMQ--------AAQHQQFHIS
UniRef90_UPI000780B9AB_11_271 VHPPGGIAGGDRLDISASVGPQAWAQLTSPGAAKWYRAASP---------AYQKLELKVA
UniRef90_UPI000255752C_11_273 VHPPGGIAGGDRLDIDIHLGENAWAQLTSPGAAKWYRAAGP---------AYQQLNIRVE
UniRef90_UPI000484E5E1_11_269 VHPPGGIAGGDRLDISASVERDAWAQLTSPGAAKWYRANSP---------AYQQLEVTVA
UniRef90_A0A0D6AS13_1_264     LHTAGGIVGGDILSQTIHLSSNSQVFITTPAATKIYRTQEKK--------AFQEIIINLE
UniRef90_UPI00034A425D_18_281 LHTAGGMVGGDRLSAEINLAENTHALITTAAAAKIYRSNGLI--------AQQSTRIQIA
UniRef90_UPI00046A7B93_11_269 VHPPGGIAGGDQLAISARVDNGAWAQLTSPGAAKWYRANGP---------AGQTLALQVA
UniRef90_I3YAM1_6_262         LHPPGGLVGGDRLEIAVSVDSGAHALLTTPGATKLYRSNGAE--------AHQTQHLRVA
UniRef90_A0A0D9AIT7_9_267     VHPPGGIAGGDRLDISATVGPGAWAQLTSPGAAKWYRAAGP---------AFQNLQLRVE
UniRef90_UPI0003FDB5F5_4_265  LNPGGGYLDGDRYQMKVTLGERAMLTLTTQSSTKVYRTPTKP--------VYQETIFHMK
UniRef90_K9SBL5_6_267         LHTAGGIVGGDRLTIRARLEPESHALLTTAAASKLYRSNGFE--------AQQRVHLDVG
UniRef90_UPI00040AC544_17_274 LHPPGGIVAGDHLSLEIDAEAGSHALITTPAAGKFYRSAGGL--------ARQDVSITVA
UniRef90_A6D6Q9_31_296        LHPPGGVVGGDTLQISVHAIQGAKSLITTPGATKFYRSESKY--------AKQKQTLHVA
UniRef90_B8HW54_12_274        LHTAGGVVGGDRLSLDIHLDTDTQVLLTQAAASKIYRSEGLE--------AHQQVRITLE
UniRef90_A0A0J6H3B0_11_269    VHPPGGIAGGDQLDIRASVGPNAWAQLTSPGAAKWYRAAGP---------AYQRIELTVA
```

```
                                    110       120       130       140       150       160
                                     |         |         |         |         |         |
HpUreD                        ENAFLDFAPFPLIPFENAHFKGNTTISLRSSSQLLYSEIIVAGRVARNELFKFNRLHTKI
UniRef90_A0A0K9H6B6_4_266     EGSYLEYIPDPLIAYRHARYKQKNIIRMEKGATFLYSDILTPGWSPEGDKFSYDTLQLIN
UniRef90_A0A0K9GXZ4_4_265     KGSYLEYIPDPLIMYRNARYKQKNVIRMEQGATLIYSDIITPGWSPDGEQFSYDQLQLMN
UniRef90_A0A0A3IXZ5_4_264     KGSLLEYLPDPLIAYQNAHYKQKNVIHMETGATFIYTDILTPGWSPNGEKFSYNTVQLVT
UniRef90_K9ZPZ7_7_273         ANACLEYLPQETILFNGAVYRQDLKVKLDTNSSFIGWEITRLGRSARGEKFLEGEMRSHT
UniRef90_A0A0C1XDA3_13_272    AGACLEWFPQETIVFNGAIYRQDLRVELATKASYLGWEITRFGRSARGEKFVQGEWRNHT
UniRef90_UPI000379D7E3_4_266  KGSYLEYLPDPLIAYQNAHYKQKTVVHIERGATFLYTDILTPGWSPNGEKFTYDTVQLMT
UniRef90_UPI0002DF3930_13_277 TGATLEWLPQETIVFDGAIYRQDTRINLAENANFIGWEITRFGRSARGEKFLHGNWKSNT
UniRef90_A0A0M0W0K1_1_266     KGSYLEYLPDPLIAYQHAHYKQKTVVHMEPGATFLYTDILTPGWSPSGERFSYDTIQLVT
UniRef90_UPI000422C885_4_265  KGSYLEYIPDPLIMYRDAQYVQKNVIRMEKGATLIYSDILTPGWSPDGDQFSYDRLRLLN
UniRef90_B4W160_8_270         AGACLEWLPQETIVFNGADYQQQMTVELAPGASWLAWEITRFGRTARKERFLQGDWRSHT
UniRef90_Q8YQZ4_10_269        AGACLEWLPQETILFNGAIYRQDLRVELATGANFLGWEITRFGRSARGEKFYQGEWRSHT
UniRef90_A0A0P1BUZ9_3_269     AGACLEWLPQETILFNGAIYRQDLRVELAPGASWLGWEITRFGRTARGEKFVQGEWRSHT
UniRef90_A0A0M1JRC4_17_282    NDACLEFIPQETIVFNQALYRQDLTVELAPGASFFGWEITRFGRSARGEKFLAGAWRSNT
UniRef90_UPI0007108B5C_4_266  EGSYLEYIPDPLIGYQDARYKQKNMIHMEKGCTLLYSDIITPGWSPGGEKFSYDKLQLIN
UniRef90_UPI0006A76A92_4_265  KGSYLEYLPDPLIAYENAHYKQKNVVYLESGSTFLYTDILTPGWSPSGEKFSYDTIQLKN
UniRef90_W4ETC9_4_265         RGSYLEYLPDPLIAYENARYKQKNVIHMESGATFLYTDILTPGWSPSGTRFSYDTVQLVT
UniRef90_A0A0D6KKC6_12_271    AGACLEWLPQETILFNSAIYRQDLRVELATEASWIGWEITRLGRSARGEKFLQGEWRSHT
UniRef90_A0A0D8ZYE2_11_271    ANACLEWLPQETIVFNGAIYRQDLRVELAPGASWLGWEITRFGRTARGERFLEGEWKSHT
UniRef90_B2IT63_10_269        PGACLEWLPQETILFNDAIYRQDLRVELATGASWLGWEITRFGRSARGEKFLQGEWRSHT
UniRef90_UPI000308766C_11_271 SGACLELLPQETIVFNGAIYRQDLRVELATKASWIGWEITRFGRSARGEKFLQGEWRSHT
UniRef90_K9R6Q0_10_280        KNACLEWLPQETIVFSGASFRQDIKIELADDANFIGWEITRFGRTARQEKFLTGEWRSNT
UniRef90_A0A127D3L2_4_264     SGCYLEYIPDPLIMYQDARYKQKNVIRMEKGATFLYSDILTPGWSPDGDRFSYNRLQFIN
UniRef90_UPI00028931D6_4_266  EGSYLEYIPDPLIGYKDARYKQKNVIHMEKGTTLLYSDIITPGWSPDGERFSYDILQLIN
UniRef90_A0A0M0ENP6_4_266     KGSYLEYLPDPLIAYENAHYKQKNVVHMESGATFIYTDILTPGWSPEGKKFSYNTIQLVT
UniRef90_A0A0S3PHU6_16_275    AGACLEWLPQETILFNGAIYRQDLRVKLATGANFLGWEITRFGRSARGEKFYQGEWRSHT
UniRef90_A0A081NYG6_3_264     KGSFLEYVPDPLIAYRHAQYKQKTTIRMERGSTLIFSDIVTPGWSPDGELFSYDKLRLKT
UniRef90_A0A0K9GPB1_4_266     AGSYLEYIPDPLIAYQHARYKQKNVIHMEKGATFLYSDIITPGWSPNGERFSYDKVQLIN
UniRef90_UPI000717378B_2_265  KGSYLEYLPDPLIAYENAHYKQKNVIHIEEGATFLYTDIITPGWSPSGDKFSYKTIQLIN
UniRef90_K9W1V1_14_278        AGACLEWLPQENIIFNDANYRQDLRVELATDATWMGWEITRLGRTARGEQFLQGNWRSHT
UniRef90_K9U4Q1_17_272        EGACLEWLPQETIVFNGANYRQDLRVELAPGASWLGWEITRFGRSARGERFLQGEWRSYT
UniRef90_A0A0M2SWM6_4_265     KGSYLEYLPDPLIAYENAHYKQKNVVRMESGSTFLYTDILTPGWSPSGEKFSYDTIQLKT
UniRef90_UPI0003652A48_11_275 EGACLEWLPLDSIVFNQAIYRQTMQINLAQGANWLGWEITRFGRSARGEKFVEGNWRSRT
UniRef90_UPI00047A812C_16_286 AGACLEWIPQETIVFNSALYRQDLRVELTTGASYLSWEITRFGRSARGEKFLQGEWRSNT
UniRef90_K9TKA4_12_277        SGAICEWLPQESIIFNGAIYRQDLRIELAPDARFLLWEINRFGRSARGETFVQGEWRSQT
UniRef90_M7NKH7_1_269         AGSYLEYLPDPLIAYKDAAYRQKNIVRMESGATFLYTDILTPGWSPDGGQFAFDSVRLIN
UniRef90_K6DR39_4_265         AGSYLEYIPDPLIAYQNARYKQKNVIRMDQTATFLYSDILTPGWSPEGERFSYTTVQLLN
UniRef90_A0A0B4RFS1_1_265     KGSYLEFLPDPLIAYENANYRQYNTVRLETGATFLYTDILTPGWSPSGKKFTYHSVQLVN
UniRef90_A0A168N9T6_4_266     KGSYLEYLPDPLIAYENAHYKQRNIVHMERGATFLYTDILTPGWSPSGEKFSYDTIQLVT
UniRef90_A0A139X4D9_7_274     AGASLEWLPQETIVFNGAIYRQDLRVELATGANWIGWEITRFGRSARGEKFCLGEWRSHT
UniRef90_UPI00030AB192_16_277 DGACLEWLPQETIVFDGALYRQDINVKLATTASYIGWEITRFGRSARGEKFLQGEWRSHT
UniRef90_K7W9H0_9_269         SHACLEYLPQETILFNGGIYRQDLRVELATDASYLAWEITRFGRSARGEKFVQGEMRSHT
UniRef90_UPI00034A5E0C_10_263 TGACLEWLPQETILFNDGIYRQDLRVELATGASFLGWEITRFGRTARGEKFYSGEWRSHT
UniRef90_A0A0A0E4Q3_4_265     AGSYLEYFPDPLIAYQNAHYKQKNVIKMHPSATLLYSDIITPGWSPEGERFSYKSVQLLN
UniRef90_UPI000717263C_4_266  KDSYLEYLPDPLIAYENAHYKQKNVVHMDRGATFLSTDILTPGWSPSGGKFSYDTIQLVS
UniRef90_K9QPZ4_10_269        AGACLEWLPQETILFNGAIYRQDLRVELATGANFIGWEITRFGRSARGEKFYQGEWRSHT
UniRef90_UPI0002D28249_14_276 ANACLEWLPQETIVFNSALYRQDLRVELATGASYLGWEITRFGRTARGEKFVQGEWRSHT
UniRef90_F9DU20_4_266         KDSYLEYLPDALIAYKDAKYYQKNIVHMEKGATLLYSDILTPGWSPEGEKFSYDMLRLKT
UniRef90_UPI00047933CA_3_266  AGSYLEYIPDPLIAYQHAKYKQKNVIRMDKTATLLYSDIITPGWSPEGEQFSYETIQLLN
UniRef90_W7RFE8_4_266         EGSYLEYIPDPLIGYQDARYKQKNVVYIEKGCTLLYSDIVTPGWSPSGEKFTYNMLQFIN
UniRef90_UPI0007441C40_4_265  KGSYLEYLPDPLIAYENAHYKQKNIVHMESGSTFLYTDILTPGWSPSGENFSYDTIQLKT
UniRef90_A0A0T7BRT9_7_268     SHACLEYLPQETIVFNSANYRQDVRIHLAVDASFLGWEITRFGRTARNEQFIQGEWRNYT
UniRef90_UPI0006A78357_4_266  EGSFLEYLPDPLIGYRDARYKQKNVIHMEKGSTLVYSDIITPGWSPEGXPFSYDMLQLIS
UniRef90_UPI0007C7BB46_4_266  KGSVLEYIPDPLIAYRDAKYKQNTIIRMERGATLFYSDILTPGWSPDGEWFRYDLLQLKN
UniRef90_UPI000472611B_4_265  NGSYLEYIPDPLIGYRDARYKQKTVVHMEKGATLLYSDIITPGWSPDGEKFSYHTLQLKT
UniRef90_UPI0002ACBA0D_11_270 TDACLEWLPQETIVFDGAIYRQDLQVELAPGAKWLGWEITRFGRTAKGERFLHGDWKSHT
UniRef90_A0A0C2KR59_11_269    AGACLEWLPQETIVFNGAIYRQDLRVELATEGSFLGWEINRFGRSARGEIFATGEWRSHT
UniRef90_K9WE93_4_277         QAACLEWLPQETIVFNGAIYQQDLRVELAPGASWLGWEITRFGRSARGERFLQGEWRSHT
UniRef90_A0A0M0SP30_16_277    DGACLEWLPQETIVFDGALYQQDLRVNLATKASYIGWEITRFGRSARGEKFLQGEWRSHT
UniRef90_K9Q9K1_10_268        AGACLEFLPQETILFNGADYRQDLRVELATGACFLGWEITRCGRSARGEKFLEGKWRSHT
UniRef90_A0A0S3TTX4_17_278    DHAYLEWLPQETIVFDGALYRQENIVKLATNSHYLGWEITRFGRSARGEKFLQGEWRSHT
UniRef90_K9VQF3_56_323        TGANLEWLPQETIVFDGAIYRQNLRVELAPTARILLWEITRFGRSARGENFLSGEWRSHT
UniRef90_W1SM28_4_265         PGSYLEYLPDPLIAYQNAKYKQKNIIRMDKTATFLYADIVTPGWSPDGERFSYQTVQILN
UniRef90_A0A0C1N935_15_285    AGACLEWLPQETIIFNGAIYRQDVRVELSAAANWMGWEITRFGRSARGEKFYTGEWRSHT
UniRef90_A0A139SK67_11_275    DGAVCEWLPQETLVFDGARGGQTTEIELAADARFVGVELLCFGRTASGERFAHGDFAMRT
UniRef90_UPI00041C7CDC_4_265  AGSYLEYFPDPLIAYQDARYKQKNIIRMEKGATLLYSDIITPGWSPNGKRFSYDKVQLIN
UniRef90_K9XHH8_15_275        SNACLEWFPQETIVFNGAIYRQDLRVELAPGATWTGWEITRFGRSARGEKFLQGNWRSHT
UniRef90_UPI0002E7D901_4_265  KGSYLEYLPDPLIAYKDAKYVQKNVVHMERGSTLLYSDIVTPGWSPEGKEFSYDTLRLKN
UniRef90_A0ZB05_10_269        AGACLEWLPQETIVFNGAIYRQDLRVELAPGGSWLGWEITRFGRTARGEKFLQGEWRSHT
UniRef90_D4ZSS4_5_270         KGAYLEYLPRETIIFNGAIYRQDLRVELAPEATWLGWEITRFGRSARGENFGQGEWRSHT
UniRef90_A0A0F5YFY1_6_271     ENACCEYLPRETIIFNGAIYRQDLQVELGQNATWLGWEITRFGRTARGEKYTQGEWRSCT
UniRef90_K9T9Y9_59_322        SGACLEWVPQETIIFNGAIYQQDLRVELAPGASWLGWEITRLGRSARGEKFLEGNWRSHT
UniRef90_K9VAD2_10_273        SNSYLEWLPQETIIFNDAEYSQKLRVEVGENATFMGWEITRFGRSARGEKFTAGEFRSYT
UniRef90_W7Z4J7_4_265         KGSYLEYITDPLIGYQDARYKQKTVINMEKGAAFLYSDIITSGWSPNGKQFSYDLLQLIN
UniRef90_Q47G52_13_278        EGATLEWLPQETIVFDGARARMETQVDLAADSRYIGWDILCLGRVAAGERFEKGRFDLFL
UniRef90_UPI0004024FB0_3_269  DHSHLEYLPQETILFNGAIYRQDLRVELGKNANYLGWEIMRFGRSARGEKFLGGEMRSHT
```

```
UniRef90_K8GMD3_12_265        PNACLEWLPQETIVFAQAQYRQQMQIDLAPGALWIGWEITRFGRTARGEKFFAGEWRSRT
UniRef90_A0A0S3UB12_16_269    NGACLEWLPLDAIVFNQASFRQEMMINLAPGAEWIGWEITRFGRSARGERFVEGNWRSQT
UniRef90_A0A0F7D4R9_4_265     KNSYLEYLPDALIAYKDAKYIQKNVVYMEKGATLLYSDIVTPGWSPEGDAFSYDTLRLKS
UniRef90_A0YQS4_5_271         SEACCEYLPRETIVFNGAIYRQDLYVELAQNATWLGWEITRFGRTARGEQFTAGEWRSCT
UniRef90_A0A0Q8RCL2_14_276    DSATLEWLPQETIFFDAAHVRLDTAIELGTDATYIGTEILCFGRTASGESFDSGMVEQKT
UniRef90_UPI0005625009_5_271  KGGMLEWMPQENIFFPDANARIETHIALEPGAAFIGWDIQCLGRPVNDEPFDIGSMASAT
UniRef90_UPI000305E365_16_284 EHACLEWLPQETIIFDGALYQQDLRVNLATEASYIGWEITRFGRTARGEKFLQGEWRSHT
UniRef90_UPI0002E38242_11_269 AGATLEWLPQETIVFSAAQAELATRIELEGDARLFYWDIVALGRPASGERFDRGHFQAHL
UniRef90_G8Q511_12_270        PGATLEWLPQETIVFSAAQAELSTSIDLQGDGRLFYWDMVALGRPASGERFDLGHFQSRL
UniRef90_A0A168IUL4_4_266     KGSYLEYMPDPIIAYEHAKYLQNTIVHMNPGSAMIYGDIITPGWSPDGKWFRYHTLQIKT
UniRef90_UPI0004799FB1_14_281 AGATLEWLPQESIVFDGAIARMDTHVTLAADSRYLGWDILCLGRAAAGERFDHGRFDLHC
UniRef90_Q3KIS7_11_269        DGATLEWLPQETIVYSAAQAELTTSIELEGDARLFYWDVVALGRPASGERFDLGHFQAHL
UniRef90_K9XZ09_12_271        PGATLEYLPQENIVFSGAIYRQDLRIQLAPGANWIGWEINRFGRSARGEQFIAGEWRSYT
UniRef90_A0A090RU86_25_288    KGARLEWLPQENIFFPDAHVRMDTQVHLEKDAQFLGWEMHCFGRPALNEGFSSGHLVGKT
UniRef90_A0A0M5LWB1_11_269    AGATLEWLPQETIVFSGARAELTTRIELEGDARLFYWDMVALGRLASGERFEHGHFQSQL
UniRef90_A0A0C2I8A1_11_269    AGATLEWLPQETIVFSAAQAELRTTIDLEGDARLFYWDMVALGRPASGERFDRGHFHSQL
UniRef90_H3SNI3_3_265         AGSYLEWLPDSVIAYRDSRYRQQTDIRMHSSAALILGEIVTPGWSPDGEHFSYDEITLKT
UniRef90_A0A0Q0XPT8_11_269    AGATLEWLPQETIVFSAAQAELSTQIELQGDARLFYWDMVALGRMASGERFECGHFQSRL
UniRef90_K4ZJ70_4_265         KGSYLEYITDPIIGYEHARYKQKTVIWMEKGANLLYSEIVTSGWSPDGRQFSYDLLQFMN
UniRef90_A0A085V951_11_269    PGATLEWLPQETIVFSAAQAELTTRIELQGDAKLFYWDVVALGRPASGERFEGGHFQAQL
UniRef90_A0A098SV87_11_271    SGGTLEWLPQETIVFSEAKAELTTRINLQGDARLFYWDVVALGRPASGERFDQGHFQAHL
UniRef90_A0A0A1GEL6_9_267     AGATLEWLPQETIVYCAAQAELSTSIELEGDARLFYWDIIALGRPASGERFDAGHFQAQL
UniRef90_I4CP19_9_267         TGATLEWLPQETIVYSAAQAELSTVIELEGDARLFYWDMVALGRPAAGERFDAGHFQARL
UniRef90_A0A0W0P2S5_11_269    PGATLEWLPQETIVFSAAQAELTTCIELQGDARVCYWDIIALGRPASGERFAQGYLQAHL
UniRef90_C3K5A6_11_269        AGATLEWLPQETIVFSAAQAELTTRIELHGDARLFYWDVVALGRPASNERFDLGHFQSHL
UniRef90_A0A078LV36_9_267     AGGTLEWLPQETIVYSAAQAELKTCIDLEADARLFYWDMVALGRPAATERFDDGYFQAQL
UniRef90_J3GGT8_11_269        AGATLEWLSQETIIFSDAQAELSTRIDLEGDARLFYWDVVALGRPASGERFDLGHFQAHL
UniRef90_A5L5M9_32_304        KGARLEWMPQENIFFPNAHVRLDTEIRLEKGAQFWGWEMHCFGRPAQNEGFEHGHLVGKT
UniRef90_A0A0J6GPD8_11_270    AGATLEWLPQETIVFSAARAELSTRIELEGDARLFYWDMVALGRQASDEHFDLGHFQSQL
UniRef90_Q4KJ05_11_269        PGATLEWLPQETIVFSAAQAELTTRIELQGDARLFYWDIIALGRPASGERFEQGHFQAHL
UniRef90_A0A0V7ZQT1_2_281     ERACLEWLPQETIVFNAADYRQDLRVDLANNASFLGWEINRFGRTARGEKFVQGNWRSHT
UniRef90_A0A075PF54_11_269    AGATLEWLPQETIVFSAAQAELRTRIDLEGDARLFYWDVVALGRPASGERFDLGHFQSHL
UniRef90_A0A0D5Y774_11_269    AGATLEWLPQETIVFSAAQAELSTRIELQGDARLFYWDVVALGRPASDERFEQGHFHSRL
UniRef90_A0A0Q5EC30_11_271    PGATLEWLPQETIVFSQAQAQLETCINLHGDARLCYWDVVALGRPASGERFAQGYFQSQL
UniRef90_I4N4V9_11_269        PGATLEWMPQETIVYSAAQAELQTHIELQGDARLVYWDMVALGRPAADERFARGYFQSHL
UniRef90_D4TUH6_7_266         RGSCLEYLPQETIVFNGGRYRQDVRIELGEGSSFIGWEISRLGRTARGEKFLEGEMLSHT
UniRef90_UPI00048AE6EC_1_266  KDSTLEYVPDPIIAYRSADYIQHNNIHMESGATLIYSDILTPGWSPDGSLFSYDNIQVKN
UniRef90_UPI0007398261_11_285 SGACLEWLPQETIVFNGAEFRQQLRVDLEPGALWMGWEIARLGRSARGERFVCGNWRSHT
UniRef90_B0C790_12_273        TNAILEWLPLGTIVFDQAQFRQTLQVELGPGAIFCGWDLTRFGRSARGERFMQGDWRSHT
UniRef90_A0A066UMR7_26_298    KDARLEWMPQENIFFPDAHVRLDTEIYLEKGAQFWGWEMHCFGRPAQNEGFEQGHLVGKT
UniRef90_A0A0Q9XWI2_4_265     QGSYLEYLPDPLIAYKDACYVQKNIVRMEKGAEFLYTDILTPGWSPEGKHFSYESLRLIN
UniRef90_A0A0S7ZTJ9_14_276    QHASLEWLPQETILFNGCEVSATTCINLSDNSRFIGWEISCLGRPASGEAFDYGRVRQRF
UniRef90_A6SZ04_26_286        ADAALEWLPQETIFFNEADVRLEHNVTLAADARYIGGEILCFGRTASGETFDSGRVAQRT
UniRef90_U3H3U5_9_267         AGATLEWLPQETIVFSGAQAELSTRIDLAADARLFYWDVVALGRPAAGEHFASGHFQAQL
UniRef90_A0A073CY52_6_272     EQAYLEFIPREIIIFNGAIFSQNLRVNLDPNACYLGWEITRFGRTARGEIFNQGQWKSCT
UniRef90_U6ZYX1_11_270        PGATLEWLPQETIVFSAAQAELTTRIELHGDARLCYWDIVALGRPASGERFVQGHFQAHL
UniRef90_UPI00067CF5D4_26_284 AGARLEWLPQETIFFNEADVRLEHKVTLAVDASYIGGEILCFGRTASGESFTRGSVGQRT
UniRef90_A0A0K2BGW7_26_288    AGAALEWLPQETIFFNDADVRMEHSVDLAADACYIGGEILCFGRTASGESFNSGRVSQCT
UniRef90_A0A011QEK6_40_302    EGGVVEYLPQETIVFDGAEAQMQTQVELAAGALFCGWEILCLGRTASGERFRHGRLQLAT
UniRef90_G4T117_15_272        ENAVLEWLPQETIVYEGARFTSETTIKIDTGSRFIAWEVIVLGRPASGEGFELGEALLNW
UniRef90_Q87VP5_11_271        PGATLEWLPQETIVFSNAQAELTTRIELHGDARLCYWDVVALGRPASGERFEHGHFQSHL
UniRef90_A0A089YS31_11_269    AGATLEWLPQETIVYSAAQAELDTRIHLQGDARLFYWDIVAVGRPAAAERFDQGHFQSRL
UniRef90_UPI0004174C9A_9_267  AGATLEWLPQETIVYAGAQPELDTRISLEDDARLFYWDMVALGRPASGERFDSGRLRARL
UniRef90_UPI000379F3E4_11_286 GGACLEWLPQETIVYNNAIYRQDLRVELEPGASILIWEITRFGRSARGEQFLQGFWRSHT
UniRef90_A4VQU8_9_267         AGATLEWLPQETIVYSAAQAELNTRIELHGDARLFYWDIVALGRPAAGERFDAGHFQARL
UniRef90_A0A0M3V4J1_7_297     AGACLELLPQETILFNGADYRQDLRVELAPGACFLAWEITRFGRSARGEKFLQGKWRSHT
UniRef90_A0A098ESZ3_4_266     KGSYLEYLPDPLIAYRNAKYLQKNIFRMEKGATLFYTDIITPGWSPDGKLFSYSMLQLIN
UniRef90_UPI000345DD51_13_277 AGASVEWLPQESIFFDQARVALEQTVVLEADASYIGCEILCLGRRASGETFNSGKIAQRT
UniRef90_UPI00047D06E7_9_270  EGGALEWLPQETIFFPGACAGLKTQIQLHPKASYLGWEVHCLGRPTNQETFDSGELLFQT
UniRef90_I3BUX5_1_262         DG-CLEWLPQENIFFPSANALLSSEVHLHGTAQYLGWEIHCLGRPVIGETFAAGRALFKT
UniRef90_UPI000780B9AB_11_271 AGGTLEWLPQETIVFSEAKAELTTRIDLECDARLLFWDVVALGRPASGERFEQGHFQAHL
UniRef90_UPI000255752C_11_273 AGATLEWMPQETIVYSAAHAELVSNIELIGDGKLFYWDIVALGRPASDERFDLGHFQAEL
UniRef90_UPI000484E5E1_11_269 AGATLEWLPQETIVFSAAQAELNTSIDLQGDARLFYWDVIALGRPASDERFEHGHFQSQL
UniRef90_A0A0D6AS13_1_264     NDTYLEYLPQETIVFNKCQYQQKLRVNLRENATWLGWEIIRFGRSARGEIFTHGQWLNQT
UniRef90_UPI00034A425D_18_281 AGAYLEWLPQDSIIFDGAIYNQHLHVELAPQATWCGWEVCRYGRTARGESFLSGQVRSHT
UniRef90_UPI00046A7B93_11_269 PGATLEWLPQETIVFSAAQAELSTQVDLQGDARLFYWDVVALGRPASGEHFAHGHFQSHL
UniRef90_I3YAM1_6_262         ADGILEWLPQENILFAGTHARLATQVDLDPGARFIGWEVQALGRPANGERLTTGTADLSL
UniRef90_A0A0D9AIT7_9_267     PGATLEWLPQETIVYSAAQAELEMSIELQGDARLLYWDIVALGRPASGERFDAGHFQAQL
UniRef90_UPI0003FDB5F5_4_265  KDSYLEYLPDALIAYKDASCYQKNSIYMEKGATLLYSDILTPGWSPEGEKFSYDMLRLKT
UniRef90_K9SBL5_6_267         AGACVEWLPQESIVFDGAVYRQDLRVELAPDAWWFGWELTRLGRSARGERFVQGNWRSHT
UniRef90_UPI00040AC544_17_274 EGAALEWLPQETIIYEGARLRSTVRVELASKARFIGWEILSLGRPACGERFDEGLADLDW
UniRef90_A6D6Q9_31_296        SGSRLEWLPQENIFFPDAHVRLDTEIHLERGAQFLGWEMHCFGRPALNEGYSAGHLLGKT
UniRef90_B8HW54_12_274        PGAWLEWLPRETIVFNQAHYHQDLRVELAPGAVWLGWEITRLGRTARGEQFLQGHWRSHT
UniRef90_A0A0J6H3B0_11_269    AGATLEWLPQETIVFSSAQAELSTRIELQGDARLFYWDMIALGRQASGERFERGHFQSHL
```

```
                                    170       180       190       200           210
                                     |         |         |         |             |
HpUreD                        SILQDEKPIYYDNTILDPKTTDLNNMCMFDGYTHYLNLVLVNCPIEL----SGVRECIEE
UniRef90_A0A0K9H6B6_4_266     ELYVDDELAVYDHIKLNPRNQDMNTIGFMEGFSHLGSMIVVGEQTDS----DLLDRLHEA
UniRef90_A0A0K9GXZ4_4_265     EVYMDDELVMYDHIKLNPAEQDMKAIGFMEGYTHLGSMIVIGEQTNA----DLLDRLYNS
UniRef90_A0A0A3IXZ5_4_264     EIYLDGQLGVFDHIKLTPGKHSISGIGFMEGYTHLGSMIVVSELTNN----SLLDELYEM
UniRef90_K9ZPZ7_7_273         EIWQNGIPQWIDRQILPGSEEVFHSPHGLAGYPVVGSLVWVGSSVSR----EIIEKARSL
UniRef90_A0A0C1XDA3_13_272    EIWQQGKPLWIDRQWLPGNEEVFHSPHGLAGQPITGNLVWVGGVVSA----EIVDKARNL
UniRef90_UPI000379D7E3_4_266  EIYWDGKLGVFDHIKLMPQQQTISGLGFMEGYTHLGSMIAVSEQTEP----EYLDELYEA
UniRef90_UPI0002DF3930_13_277 EVWQQGKPLWVDRQWLPGSEEIFYSPHALNGQPVVGTFIYIGSTVSP----EIIEKASSY
UniRef90_A0A0M0W0K1_1_266     EIYWNGRLGVFDHLKLVPQQQKISGLGFMEGYTHLGSMIAISEQMDA----ALLDELYDA
UniRef90_UPI000422C885_4_265  EIYMDDELVVYDHIKVNPAEQDMQSIGFMEGYTHLGSMFVIGEQTNA----ELLDKLHHL
UniRef90_B4W160_8_270         EIWQQGHPLWIDRQWLPGGEAVLDSPHGLAGEPIVGTLIWMGPPVSS----EIIDNARSL
UniRef90_Q8YQZ4_10_269        EIWQQGVPLWIDRQWLPGNDAVFHSPHGLAGQPIVGSLVWLGSPIST----EIIEKARNL
UniRef90_A0A0P1BUZ9_3_269     EIWQQGVPLWIDRQWLPGKEEVFHSPHGLAGQPIAATLTWVGTSVTE----EILDKARNL
UniRef90_A0A0M1JRC4_17_282    EVWQAGEPLWIDRQCLFGSEEMFHSPNALKGYPLVGTLFWIGQPVSA----DAIASARNF
UniRef90_UPI0007108B5C_4_266  EIYMENELVVYDHIKLSPATQNINGLGLMEGYSHLGSMIVVDEKTDH----DLLDRLYEA
UniRef90_UPI0006A76A92_4_265  EIYLDGKLGVYDHIKLVPGKSPMTGLGFMEGFTHLGSMIAVSEDTTN----ELLDELCEI
UniRef90_W4ETC9_4_265         EIHLDGKLGMFDHIKLTPHNQNLSGLGFMEGYTHLGSMVAVSENTDN----ALLDELYET
UniRef90_A0A0D6KKC6_12_271    EIWQQGVPLWIDRQYLPGSEEVFHSPHGLAGQPIVGSLVWVGNPVDS----EILAKARNL
UniRef90_A0A0D8ZYE2_11_271    EVWQQGKPLWIDRQWLPATETIINSHHGLNGQPIVGSFAWIGQPVTK----EIVEQARNL
UniRef90_B2IT63_10_269        EIWQQSVPLWIDRQCLRGSEDIFHSPHGLAGKPIVGSLVWVGGAVSA----EIVEKTRSL
UniRef90_UPI000308766C_11_271 EIWQQGVPLWIDRQLLPGSEEVFHSPHGLAGQPLVGSLVYVGQEVSP----ELVEKVRSL
UniRef90_K9R6Q0_10_280        EIWQNNKPLWIDRQYLPGSEEVFHSPHALAGKPIVGTLIYIGKPVSP----EIVQKIRTL
UniRef90_A0A127D3L2_4_264     EIYMDDELVAYDHIQLNPAAQNIEKIGFMEGFSHLGSMLVVGEQTSP----DLLDRLYQA
UniRef90_UPI00028931D6_4_266  EIYMENELVVYDHIKLNPAMQNMEKLGLMEGYSHLGSLIVIDEKANH----ALLDRLYQA
UniRef90_A0A0M0ENP6_4_266     EVYLNGKLGVFDHIKLMPEKQTLSGLGFMEGFTHLGSMIAVSEYTNN----SLLDELYEV
UniRef90_A0A0S3PHU6_16_275    EVYQEGVPLWIDRQWLPGNEGVFHSSHGLNGQPIVGSFVWVGSAVTE----EFVQKARDL
UniRef90_A0A081NYG6_3_264     EIYVDGDLAVYDQVRLNPSEQDLTGIGLLEGYTHFGSLIVVGEQMTS----DFLDQLYEA
UniRef90_A0A0K9GPB1_4_266     EIYMDDELISFDHIKLNPATQNIEALGFMEGFSHLGSMMVISEQMSP----DLLDRLYHA
UniRef90_UPI000717378B_2_265  EIYLDGLLGVFDHIKLTPSQQSLTDIGFMEGYTHLGSMIVIGNQTNN----HLLDELYEM
UniRef90_K9W1V1_14_278        EIWRQGDPLWIDRQWLPGGENIINSPHDLAGYPVIASFAFVGKAVSK----DLIEKARNC
UniRef90_K9U4Q1_17_272        EVWQQGQPLWIDRQWLPGEEAILNSPHGLAGHSIVASLTWIGCEVSP----ELVTKCRDV
UniRef90_A0A0M2SWM6_4_265     EIYLDGKLGVFDHIKLSPEDNSVEGLGFMEGFTHLGSLMAVSEETDN----QLLDELYDK
UniRef90_UPI0003652A48_11_275 EVWQAGKPIWIDRQWMPGSEENFASPHGLAGCPVVGSFAWVGQVVTP----ELVEKAREL
UniRef90_UPI00047A812C_16_286 EIWQQGKPLWIDRQWVPGSEEIFYSPHGLAGQPVVGSLVWVGSAVSV----EIIEKARNI
UniRef90_K9TKA4_12_277        EIWQQGRPLWIDRQHLQGSESAVSSNSALAGFPIVATLAWIGDPVTP----ELVQEARSL
UniRef90_M7NKH7_1_269         EIYVDGEPAVFDNIRLSPSDQKVGGLGFMEGYTHLGSMIAIGEQTTD----DLIDILHDL
UniRef90_K6DR39_4_265         EIYMDDELVVFDHIKLQPSEQNIRGLGFMEGYSHLGSMIVVSEQSNS----SFLDQLYST
UniRef90_A0A0B4RFS1_1_265     KIYVDGDLAVFDHIKLSPGEQNISGVGMMEGYSHIGSMIVIGEQTTK----ELLDELYEA
UniRef90_A0A168N9T6_4_266     EIYWDGKLGVFDHIKLVPKQQNITGLGFMEGYTHLGSMIAVSDQMDN----VLLDALYEM
UniRef90_A0A139X4D9_7_274     EVWQMGVPLWIDRQWLPGSEDVFHSPHGLAGQPVTGSLVYIGQEVSQ----ELVHKARTL
UniRef90_UPI00030AB192_16_277 EIWQQDKPLWIDRQWLPGSEEVFHSPHGLAGQPIVGSLVYIGQKISP----ELVHQARNL
UniRef90_K7W9H0_9_269         EIWQNGIPLWIDRQIVPGSEEVFHSPHGLRDNPVVGSFVGVGFPISP----EIINQARSL
UniRef90_UPI00034A5E0C_10_263 EIWQHNIPLWIDRQWLPGNQDVFHSPHGLSGKPIVGTFVWVGDVVSA----EIVETARNL
UniRef90_A0A0A0E4Q3_4_265     EIYMDGELVVYDHLHLKPALQNIRGLGFMEGYTHLGSMFVVCKQMNP----SFLDQLYSL
UniRef90_UPI000717263C_4_266  EIYVDGQIGVFDHIKLVPKQQSVSGLGFMEGYTHLGSMLVVSEYTNN----AFIDELYEA
UniRef90_K9QPZ4_10_269        EIWQEGVPLWIDRQYLPGSEEVFHSPHGLSGQPIAGNFIYLGSPVSK----ETIEKARSI
UniRef90_UPI0002D28249_14_276 EIWQQGKPLWIDRQWLPGSDKIFHSPHGLNGQAIAGSLVWVQGAVSQ----DMIEKARDL
UniRef90_F9DU20_4_266         EIYMEDELVVFDHIKLHPASQNMNGLGFMEGYTHLGSFIVVGEKMNE----DLLDRLHET
UniRef90_UPI00047933CA_3_266  EIYMDDELVVYDHIKLNPGQQNLRGIGFMEGFSHLGSMIVIGEQSNS----VLLDQLYSV
UniRef90_W7RFE8_4_266         EIYMENELAVYDHIKLCPAKQNIDSLGLMEGYSHLGSMIVINENVDH----DLLDRLYQA
UniRef90_UPI0007441C40_4_265  EIYLDGVLGVFDHIKLAPEKNSIAGLGFMEGYTHLGSMIAVSENTDN----GLLDMLYET
UniRef90_A0A0T7BRT9_7_268     EIWRDNQPLWIDRQYLPGDMAVFHSPHGLNSQAIAGSFIYLGKQISP----QLITQIRQM
UniRef90_UPI0006A78357_4_266  EIYMENELVVYDHIKLNPSIRNMEELGLMEGYSHLGSMIVIDEKSNQ----DLLDRLYQA
UniRef90_UPI0007C7BB46_4_266  EIYLDEKLIMFDHLKLQPDEAGVDGIGYMEGYTHLGSMIVMSEQVNE----DVLQRLYSV
UniRef90_UPI000472611B_4_265  EIYMDEDLVVFDHIKLSPATQDIEGLGFMEGFSHMGSMIVIGEKSNA----ALLDLVYQE
UniRef90_UPI0002ACBA0D_11_270 EVWQQGLPLWIDRQWLPAGEKIIDSPHGLAGLPIVGSLAWIGQPVEP----EIVEKARVL
UniRef90_A0A0C2KR59_11_269    EIWQQGVPLWIDRQWLPGRVEVFHSPHGLAGQPLAGSLVYVGQEVSS----DLVEKARSL
UniRef90_K9WE93_4_277         EVWQQGRPLWIDRQWLPGEEQVLDSPHGLAGKPIVASLAWVGQAVSP----EMIEKARLI
UniRef90_A0A0M0SP30_16_277    EIWQQNKPLWIDRQSLPASEEVFHSPHGLAGQPIVGSLVYIGQETSP----ELVNKARNL
UniRef90_K9Q9K1_10_268        EIWQQGVPLWIDRQFLPGNTDIFHSPHGLFGQPIVGSLLWLGHPVST----EIIEQVRSL
UniRef90_A0A0S3TTX4_17_278    EIWQQGKPLWIDRQWLPGREEVFHSFHGLAGQPIVGSLVYIGREISP----EIVEKARNL
UniRef90_K9VQF3_56_323        EVWQENSPLWIDRQLLKGGEKMLESPHGLAGKPVVATLAWVGEPVTA----EFVEKVRDL
UniRef90_W1SM28_4_265         EVYVDNELVVYDHIKLNPASQNMNGLGFMEGFSHLGSMLVVGKQTNS----SLLDQLYSA
UniRef90_A0A0C1N935_15_285    EVWQMGVPLWIDRQWLPGREDVFHSPHGLAGQPVTGSLVYIGQEVSH----ELVHKARSL
UniRef90_A0A139SK67_11_275    RIRRDGQTLWLERGRVTGNSPLLASPIGLAGQPVVATMWVVAPQVNE----GLRDACRAI
UniRef90_UPI00041C7CDC_4_265  EIYLDDELVSLDHIKLHPAKQHMEALGFMEGYSHLGSMMVVGEQMTP----ELLDRLYYA
UniRef90_K9XHH8_15_275        EVWQQQRPLWIDRQQLRPDVKVIDSPHGLAGKSIIGSFVWIGQPVSA----DVVEKVRML
UniRef90_UPI0002E7D901_4_265  EIYMDDELVVFDHIKLQPARQNMGGLGFMEGYTHLGSFMVIGEQTDD----DLLDSLYEI
UniRef90_A0ZB05_10_269        EIWQQGVPLWIDRQYLPGSEAVFHSPHGLAGQPIAGSLVWVGSDISA----EFLAKARSL
UniRef90_D4ZSS4_5_270         EIWQNEKPLWIDRQWLPGGETILESPHGLGGWPVVATLTWVGEPVSK----ETLNHVRML
UniRef90_A0A0F5YFY1_6_271     EIWQNGKPLWIDRQWLPASEEILTSPHGLAGQAIVGTLAWVGHSVSE----EMLKEIRQL
UniRef90_K9T9Y9_59_322        EIWQQGQPLLIDRQWLPAGEAIINSSLGLAGQPIVASLIWIGKPVSQ----NIIERAQTL
UniRef90_K9VAD2_10_273        EIWHLGKPLWIDRQWVPGCEEVFHSHHGLGGNAIAGTLIWLKNPVSS----EIITQVRNL
UniRef90_W7Z4J7_4_265         EIYLDGELVVYDHIKLDPLHVNITSLGFMEGFTHLGSMIVVGEQVNH----DLLDQLYEA
UniRef90_Q47G52_13_278        QVNRDQRPIWIERGGFDGSDPMLISPAGWAGATVCGTLLCAFPEWPMQAS-ALLEACRKI
UniRef90_UPI0004024FB0_3_269  EIWQNDIPLWIDRQIIPGNEEVFYSPHGLAGNPVVGTLVWVGNSVSG----EMIDKARSL
```

```
UniRef90_K8GMD3_12_265        EVWQQGRPLWIDRQWLPGNEETFSSPHGLANCPVVGSFAVIGQTVNP----DLIEKIRAT
UniRef90_A0A0S3UB12_16_269    EVWQSGKPIWIDRQWLPGSEATFNSPHGLAGCPVVASFAWIGKTVDS----ECVEKARSL
UniRef90_A0A0F7D4R9_4_265     EIYMDGELVVFDHVKLQPEQQNMTGLGFMEGYTHLGSFIVIGEKTDD----ALIDRLYEV
UniRef90_A0YQS4_5_271         EIWQNGKPLWIDRQWLPANVEILTSPHGLAGQAIVGTLAWVGHSVSE----EMLKEIRQL
UniRef90_A0A0Q8RCL2_14_276    SIRRDGKLLWFEQGMLAAGSAAMTSPLGLAGNTVCATLIAVGKPLAA----SMMATLREA
UniRef90_UPI0005625009_5_271  HISIDGELVLIDQLRTEG-RALLDAAAGLRGYPMQASLFIVPGEACRVSLTDLLEHIRES
UniRef90_UPI000305E365_16_284 EIWQQDKPLWIDRQWLPGSEEIFHSLHGLAGQPIVGSLIYIGQEISP----EIVEKARSL
UniRef90_UPI0002E38242_11_269 DIRRDGQPLWHERQRIEGDDGLLDSPIGLDGQPVFATLLVTGE-IDS----ELLERCRNL
UniRef90_G8Q511_12_270        DIRRDGQLLWHERQRIVGGDGLLDSPIGLGGDPVFATLLVTGE-IDS----ELLEQCRSL
UniRef90_A0A168IUL4_4_266     QVYLEDKLVVFDHLQLRPANDPMEGIGLLEGYTHLGSMIVIGERTDP----ELIERLSDS
UniRef90_UPI0004799FB1_14_281 RIERAGKPLWIERGSVAGGDAMLHSPAGWAGATVCGTLLCSFPELPQQAA-ALLEALRTL
UniRef90_Q3KIS7_11_269        DIRRDGRLLWHERQRIVGADGLLDSPIGLDGHPVFATLLVTGE-IDA----ELLERCRSL
UniRef90_K9XZ09_12_271        EVWQNNRLLWIDRQWLPGEAELINSINGLAGKPVIGTFSYLGKPVDK----EILEKINNL
UniRef90_A0A090RU86_25_288    EIYLDGQRLLTEGLNVRGEDKLLKD-KGLLGYQMMGTLYISID--DE----DFYQLVQSL
UniRef90_A0A0M5LWB1_11_269    DIRRDGQLLWHERQRISGGDGLLDSPIGLDGNPVFATLLVTGE-ISA----QLLEICRNL
UniRef90_A0A0C2I8A1_11_269    DIRRDGQLFWHERQRIRGGDGLLDSPIGLDGQPVFATLLVTGE-IDS----ELLERCRSV
UniRef90_H3SNI3_3_265         MIEMDGVPVLFDHLRLRPGEQPIHGLGRMDGHTHIGSLYVVGPLATR----AFIEELAEK
UniRef90_A0A0Q0XPT8_11_269    DIRRDGQLLWHERQRIVGDDGLLDSPIGLDGHPVFATLLVTGE-IST----ELLHTCRSL
UniRef90_K4ZJ70_4_265         EIYLDGVLAIHDHIKLDPAAHRMTSLGFMEGYSHLGSMMMISEQVDQ----DFLDRLYLE
UniRef90_A0A085V951_11_269    DIRRDGTLLWHERQRVIGGDGLLDSPIGLDGKTVFGTLLVTGE-IES----ELLEACRSL
UniRef90_A0A098SV87_11_271    DIRRDGRLLWHERQRVVGNDGLLDSPIGLDGKPVFATLIVTGD-IDP----ELMERCGEL
UniRef90_A0A0A1GEL6_9_267     NISRDGELLWHERQRVTGNDGLLDSPIGLDGHPVFATMIVSGE-ISA----ELLERCRAL
UniRef90_I4CP19_9_267         DIRRDGQLIWHERQRIAGGDGLLDSPIGLDGHSVFATLIVSGE-IDA----ELMERCREL
UniRef90_A0A0W0P2S5_11_269    DIRRDGRPLWHERQRIEGDDGLLESPVGLDGHPVFATLLITGQ-AED----DLLERCRAL
UniRef90_C3K5A6_11_269        DIRRDGQLLWHERQRIVGADGLLDSPIGLDGQPVFATLLVTGD-IEP----ELLEHCRAL
UniRef90_A0A078LV36_9_267     DIRRDGKLLWHERQRISGADGLLDSPIGLNGYPVFATLIATGE-IDA----DLLERCREL
UniRef90_J3GGT8_11_269        DIRRDGQLLWHERQRIVGNDGLLDSPIGLDGQPVFATLLVTGD-IDS----ELLEKCRLL
UniRef90_A5L5M9_32_304        EIYLDNQRLLTEGFDFHGGDKLMIN-MGLLDYAMMGTFYLTSN--EK----QDLELVQSL
UniRef90_A0A0J6GPD8_11_270    DIRRDGRLLWHERQRIVGGDGLLHSPIGLDGHPVFATLLVTGD-INP----ELLETCRSL
UniRef90_Q4KJ05_11_269        DLYRDGRPLWHERQRIGGGDGLLDSPIGLDGQPVFATLLVTGE-IDS----ELLERCRAL
UniRef90_A0A0V7ZQT1_2_281     EIYQEGIPLWIDRQWLPGSEEIFYSPNGLAGQAVIGSLIFVGKTISK----DIVEQVRSL
UniRef90_A0A075PF54_11_269    DIRRDGQLLWHERQRIVGDDGLLDSPIGLDGQPVFATLLVTGE-IDS----ELLEQCRAL
UniRef90_A0A0D5Y774_11_269    DIRRDGQLLWHERQRLVGGDGLLDSPIGLDGQPVFATLLASGE-IDA----ELLERCRSL
UniRef90_A0A0Q5EC30_11_271    EIRRDGRLLWHERQRLEGGDGLLDSPIGLDGKTVFATLLMTGE-AGS----ELLETCRSL
UniRef90_I4N4V9_11_269        DIRRDGELLWHERQRVVGDDGLLDSPMGLDGHTVFATLLVTAD-ISP----ELLEQCRGF
UniRef90_D4TUH6_7_266         EIWQGEVPLWIDRQHIPGGVEAFYNPHSLKGNPVIGSFVCVGLPISE----ERIEKSRSG
UniRef90_UPI00048AE6EC_1_266  QIFMNGRRVVFDHLRLKPNSQDIGGIGLMEGFTHVGSMIVICDQVSK----EFMERIKEH
UniRef90_UPI0007398261_11_285 QVWQAGRLIWVDPQWVAGGSEMMESLHGLAGYPVIASFALLGHPVSG----ELVERVRSL
UniRef90_B0C790_12_273        EIWQQGAPLWIDRQWLPGQPDIWESPHGLAGQPVVGSFLWVGQGVEP----NLVQTARDL
UniRef90_A0A066UMR7_26_298    EIYLDNKKLLTEGFNFHGGDKLMIN-MGLLNYPMMGTFYITAD--EP----QDLELVQSL
UniRef90_A0A0Q9XWI2_4_265     EIYLNDEIVVFDHLHLSPRNKAMNGLGHMEGYTHLGSLIAISEKVND----AFIETLYNT
UniRef90_A0A0S7ZTJ9_14_276    ELYKHEKPLFIERALLEGGQPTLAAHWGLQSFTVTATMIAYP--ADK----AVLELARKS
UniRef90_A6SZ04_26_286        SIRRGGKLLWFEQGALQARSTSMHSPLSLAGYTVCATLIAVGKTMNG----AFLNELREE
UniRef90_U3H3U5_9_267         DIRRDGRALWHERQRISGDDGLLDSPIGLAGQPVFATLIATGQ-LDP----ELLERCREL
UniRef90_A0A073CY52_6_272     EIWQNGCPVWIDRQGFIANEEILNSPHGLGGKPVIATLTWVGQPVSE----DIVKNIRQL
UniRef90_U6ZYX1_11_270        DIRRDGRPLWHERQRISGGDGLLDSPVGLDGQPVFATLLVTGE-IDA----QLLERCRAL
UniRef90_UPI00067CF5D4_26_284 SIRRDGKLLWFEQGVLHAHAASMHSPLSLAGYTVCATLIAVGKTMTP----AFLQTLREE
UniRef90_A0A0K2BGW7_26_288    SIRRGGKLVWFEQGTLRAGTSSMTSPLALAGFTVSATLIAVGLPINA----AFLSELREQ
UniRef90_A0A011QEK6_40_302    RIERQGRPLWLERGRLLGASSWLDAAPGLAGFPVSASLLLAGRAVEP----EWLAACRAL
UniRef90_G4T117_15_272        RIVLGDQPIYLERLRLDA--QAFAARWGLSRHSSCGTLFAYP--ASA----EVLEIVRNV
UniRef90_Q87VP5_11_271        DIRRDGTLLWHERQRIIGADGLLDSPIGLDGKTVFATLLLTGD-VDS----DLLEVCRSL
UniRef90_A0A089YS31_11_269    EIRRDGELLWHERQRIVGDDGLLDSAVGLGGKPVFATLLVTGE-IDP----ALLERCRGL
UniRef90_UPI0004174C9A_9_267  DIRRDGRLLWHERQRVEGGDALLDSPIGLAGQPVFATLIASGE-IDA----DLLERCRQL
UniRef90_UPI000379F3E4_11_286 EIWQQGRPLWIDRQQLIGGEETLNSPHGLAGKPVVGSLAWIGVRVTP----ELVEKARSL
UniRef90_A4VQU8_9_267         DIRRDDRLIWHERQRIAGADALLDSPIGLDGRSVFATLIASGE-LDA----DLLERCRGL
UniRef90_A0A0M3V4J1_7_297     EIWQQGVPLWIDRQFLPGNPEIFHSPHGLFGQPIVGSLIWLGHPVSS----EIIEKARSL
UniRef90_A0A098ESZ3_4_266     EIYMEDELAAFDHIKLVPSVQNISGLGFLENHTHLGSMIVIGEQANR----EFLDHLQNE
UniRef90_UPI000345DD51_13_277 QIHRDGKLVWWEQGALAGGGEMLRSPLGLDGHSVCATLLAVGKVLPA----AALASLREA
UniRef90_UPI00047D06E7_9_270  DIQRDGEPLYKEHLQLDNPWD-LQGMAGLSGYPVMATMLALP--AGN----DALELARQA
UniRef90_I3BUX5_1_262         ALYRDGKPLLLDRLLIQGERD-LQLAAGLRGNPVFATLLATP--ATP----ELLEQARSL
UniRef90_UPI000780B9AB_11_271 DIRRDGKLLWHERQRITGNDGLLDSPVGLGGKPVFATLIVTGE-VDA----ELMERCRAL
UniRef90_UPI000255752C_11_273 NIRRDGELLWYERQRVVGDDGLLDSPIGLDGYPVFATLLLSAE-IAP----ELLEQCREL
UniRef90_UPI000484E5E1_11_269 NIRRDGQLLWHERQRIVGNDGLLDSPIGLDGQPVFATLLVTGE-IDG----ELLERCRAL
UniRef90_A0A0D6AS13_1_264     EVWYKDKPLWIDRQLLQGESNLFSAINGLANKPVIGNLTLISAKINQ----DIIREIREL
UniRef90_UPI00034A425D_18_281 EVWQQGRPLWIDRQRLSGGADTIHSPHALAGQPVVANLAFIGQVIPT----EIVEQARTL
UniRef90_UPI00046A7B93_11_269 DIRRDGKPLWHERQRIVGGDGLLDSPIGLDGHPVFATLLATGE-ISA----ELLEHCRAL
UniRef90_I3YAM1_6_262         AIQRAGRPLLLDRLRIDAGTG-LDGPAGLRGFAVTGTLVATG--IDR----EDLAAVRGL
UniRef90_A0A0D9AIT7_9_267     NITRDGEWLWHERQRVTGDDGLLDSPIGLDGNPVFATLIVTGE-IDA----ELMERCREL
UniRef90_UPI0003FDB5F5_4_265  EIYMENQLVAFDHIKLHPASQHMNELGFMEGYTHLGSLIVVGEKTSD----ALLDRLYET
UniRef90_K9SBL5_6_267         EVWRQGQPLWIDRQWLPGNPAWLDSPHGLHGRSLVGSFAVVGQPIAP----ELVAEARAL
UniRef90_UPI00040AC544_17_274 QVYCRDQPLLLERLKLDA--RAFAARWGLQGCSACGTLFAYP--AGA----ETLAAVQQL
UniRef90_A6D6Q9_31_296        EVYLDKQLLLTEGLNVRGNDKLLKS-RGLLNYQMMGTVYVSID--DE----DFYQLVQTL
UniRef90_B8HW54_12_274        EVWQQGTPLWIDPQWLPGGEELLKSYHGLSGYPVVGTLVLIGQAAEA----ELIAQIRQL
UniRef90_A0A0J6H3B0_11_269    DIRRDGQLLWHERQRIVGGDGLLDSPIGLDGNPVFATLLVTGD-VSP----ELLDACRSL
```

```
                              220               230
                               |                   |
HpUreD                        SE----------------------------------------GVDGAVSETASSHLCVKA
UniRef90_A0A0K9H6B6_4_266     IHV--NTN----------------------------------EYRIGLSLLPVPGFTIRV
UniRef90_A0A0K9GXZ4_4_265     MSH--DEN----------------------------------EYRMGLSMLPVKGFTIRV
UniRef90_A0A0A3IXZ5_4_264     IQE--EST----------------------------------DFKFGLSRLPVSGLSIRI
UniRef90_K9ZPZ7_7_273         VT----------------------------Q-----------NNLTGVSRLQ-NGFLCRY
UniRef90_A0A0C1XDA3_13_272    WC----------------------------G-----------KGEAGVTRLE-NGFLCRY
UniRef90_UPI000379D7E3_4_266  IRS--EES----------------------------------DFSFGISHLAIPGFCIRI
UniRef90_UPI0002DF3930_13_277 TQ-----------------------HSVLST-----------QYSFGVTRLE-HGFLCRY
UniRef90_A0A0M0W0K1_1_266     VMT--VEA----------------------------------DFTFGISEVSIPGFCLRI
UniRef90_UPI000422C885_4_265  MDQ--NEQ----------------------------------DYKFGLSMLSVKGFSIRV
UniRef90_B4W160_8_270         WT-------------------------ADQR-----------QGEAGVTQTQAQGLLCRY
UniRef90_Q8YQZ4_10_269        GN----------------------------T-----------QGEAGVTSLE-NGFLCRY
UniRef90_A0A0P1BUZ9_3_269     WH----------------------------G-----------SGQVGATRLE-NGFLCRY
UniRef90_A0A0M1JRC4_17_282    WY-------------------------NRKG-----------EGEAGVTQIL-NGLVCRY
UniRef90_UPI0007108B5C_4_266  VPN--QTK----------------------------------EYKIGLSLLPIPGFTVRV
UniRef90_UPI0006A76A92_4_265  IHS--EEQ----------------------------------NFKFGISRLAIPGLSIRI
UniRef90_W4ETC9_4_265         IQI--EEA----------------------------------DFKVGISRLAVPGFSIRI
UniRef90_A0A0D6KKC6_12_271    WD----------------------------G-----------AGEVGVSRLQ-HGLLCRY
UniRef90_A0A0D8ZYE2_11_271    WQ-------------------------PTNC-----------HSLTGVTRLP-TGLLCRY
UniRef90_B2IT63_10_269        WN----------------------------G-----------EGEVGASRLQ-HGLLCRY
UniRef90_UPI000308766C_11_271 WN----------------------------G-----------KGEAGVTRLS-CGLLCRY
UniRef90_K9R6Q0_10_280        FI--------------------PPSPPLPIP-----------PSSQGVTRIE-NGLLCRY
UniRef90_A0A127D3L2_4_264     ING--NTD----------------------------------RYKIGLSLLSVPGFTVRI
UniRef90_UPI00028931D6_4_266  IDA--NTK----------------------------------EYKVGLSFLSIPGFTVRV
UniRef90_A0A0M0ENP6_4_266     IQI--EES----------------------------------SFKFGISKLAVPGFSIRI
UniRef90_A0A0S3PHU6_16_275    WD----------------------------G-----------VGEVGVTQLQ-NGFLCRY
UniRef90_A0A081NYG6_3_264     MDA--ETV----------------------------------PCRMGLSMLPVSGFSVRV
UniRef90_A0A0K9GPB1_4_266     LNS--KTN----------------------------------AYEIGLSLLPVKGFTLRV
UniRef90_UPI000717378B_2_265  MKG--IES----------------------------------NFTFGLSRLAIPGFGIRV
UniRef90_K9W1V1_14_278        WQ-------------------------AGEY-----------QGESGVTTLL-EGMLCRY
UniRef90_K9U4Q1_17_272        TC----------------------NVSTTIP-----------ITNYGVTRLP-HGLLCRY
UniRef90_A0A0M2SWM6_4_265     IYS--EQA----------------------------------NIKFGLSKLAIPGLSIRV
UniRef90_UPI0003652A48_11_275 WA---------------------------GS-----------SGEIGVTRLS-IGLLCRY
UniRef90_UPI00047A812C_16_286 WQ-----------------PIPPSPSLPLPL-----------SSQIGVTRLE-HGFLCRY
UniRef90_K9TKA4_12_277        WE-----------------------GRSSSS-----------EGEAGVTRLT-HGLLCRY
UniRef90_M7NKH7_1_269         LAGTLEED----------------------------------QVKFGISRLAVPGFSIRI
UniRef90_K6DR39_4_265         LSK--STY----------------------------------ECKVGLSLLSVPGFTIRV
UniRef90_A0A0B4RFS1_1_265     ILS--QEA----------------------------------DVKFGLSELVVSGLSIRI
UniRef90_A0A168N9T6_4_266     IQA--EEA----------------------------------DFSFGISNLAIPGFSLRI
UniRef90_A0A139X4D9_7_274     FL----------------------------S---PTPHSLLPTPSVGVTRLS-CGLLCRY
UniRef90_UPI00030AB192_16_277 WK-------------------------PTLT-----------HSQIGVTRLE-HGLLCRY
UniRef90_K7W9H0_9_269         II----------------------------Q-----------NSDAGVTRLQ-HGFLCRY
UniRef90_UPI00034A5E0C_10_263 WN----------------------------G-----------EGETGVTLLT-HGFLCRY
UniRef90_A0A0A0E4Q3_4_265     LDK--NTL----------------------------------DYKVGLSLLPVSGFTARV
UniRef90_UPI000717263C_4_266  IQL--EQA----------------------------------DFKAGISKLAVSGFSIRI
UniRef90_K9QPZ4_10_269        FT----------------------------P-----------HALIGVTRLE-NGFLCRY
UniRef90_UPI0002D28249_14_276 WH----------------------------G-----------EGEVGVTRLE-HGFLCRY
UniRef90_F9DU20_4_266         IQK--EAG----------------------------------DFAFGLSKLAVPGFTIRI
UniRef90_UPI00047933CA_3_266  IEM--NTN----------------------------------DYKVGLSSLSVPGFTIRV
UniRef90_W7RFE8_4_266         IDT--QTK----------------------------------EYKIGLSLLPIPGFSVRV
UniRef90_UPI0007441C40_4_265  IHS--EQS----------------------------------NFKFGLSRLAIPGLSIRV
UniRef90_A0A0T7BRT9_7_268     WG--------------------------ENS-----------PSAVGVTALE-HGFLCRY
UniRef90_UPI0006A78357_4_266  IDP--NAK----------------------------------EYKVGLSALPIPGFTIRI
UniRef90_UPI0007C7BB46_4_266  IGK--DLD----------------------------------ECKAGLSLLIEGGFTLRI
UniRef90_UPI000472611B_4_265  LHE--NST----------------------------------DFKIGLSALSVPGFTLRV
UniRef90_UPI0002ACBA0D_11_270 FP--------------------------NNS-----------SSQGGVTRLP-MGLLCRY
UniRef90_A0A0C2KR59_11_269    WR----------------------------G-----------EGQAGVTRLS-CGLLCRY
UniRef90_K9WE93_4_277         WA-------------------------TQER-----------QGEAGVTQLM-SGLLCRY
UniRef90_A0A0M0SP30_16_277    GK-------------------------PTLS-----------HSQIGVTRLE-HGLLCRY
UniRef90_K9Q9K1_10_268        FT----------------------------------------KHLTGVTQLE-HGLLCRY
UniRef90_A0A0S3TTX4_17_278    WQ-------------------------LPII-----------HPQIGVTRLE-HGLLCRY
UniRef90_K9VQF3_56_323        PS----------------------EATIYPG-----------NSTVGVTRIP-NGLLCRY
UniRef90_W1SM28_4_265         IDM--NSD----------------------------------EYKLGLSILSVPGLTIRV
UniRef90_A0A0C1N935_15_285    FL----------------------------SSTAPNPSFSIPSPQVGVTRLS-CGLLCRY
UniRef90_A0A139SK67_11_275    LPE---------------------------------------VGAGGVTLLPGGVLLARW
UniRef90_UPI00041C7CDC_4_265  LNP--KTD----------------------------------AYDVGFSLLPVEGFTLRV
UniRef90_K9XHH8_15_275        ST---------------------------VD-----------QGETGVTRLT-TGLLCRY
UniRef90_UPI0002E7D901_4_265  VDG--MEG----------------------------------DFEFGISRLTIPGFTIRI
UniRef90_A0ZB05_10_269        WD----------------------------G-----------SGEVGMTRLQ-NGFLCRY
UniRef90_D4ZSS4_5_270         WG-------------------------EHQS-----------EGEAGATQLL-SGLLCRY
UniRef90_A0A0F5YFY1_6_271     WN-------------------------PPDT-----------LGEVGVTQLL-SGLLCRY
UniRef90_K9T9Y9_59_322        WS-------------------------DKQY-----------LGEAGVTQTQARGLLCRY
UniRef90_K9VAD2_10_273        WQ----------------------------G-----------KGEVGVTRLE-HGFLCRY
UniRef90_W7Z4J7_4_265         IHE--DTN----------------------------------EYKVGLSLLPVAGLIVRV
UniRef90_Q47G52_13_278        ----------------------------------VPAD----GAQHGLSALP-GVLIARY
UniRef90_UPI0004024FB0_3_269  II----------------------------E-----------NNFSGVTRLQ-QGFLCRY
```

```
UniRef90_K8GMD3_12_265        WN---------------------------GA-----------VGQTGITRLQ-SGILCRY
UniRef90_A0A0S3UB12_16_269    WA---------------------------GT-----------SGEIGVTRLP-LGLLCRY
UniRef90_A0A0F7D4R9_4_265     VHA--ESG----------------------------------DFEFGISKLTVPGFTLRI
UniRef90_A0YQS4_5_271         WN-------------------------HSDT-----------LGEVGVTQLL-SGFLCRY
UniRef90_A0A0Q8RCL2_14_276    DVS---------------------------------------DGAFGVTQMK-SVIVARY
UniRef90_UPI0005625009_5_271  IS--------------------------------SVDS---IALEVGATQVD-GVLVVRV
UniRef90_UPI000305E365_16_284 WQ------------------LPITHHPLPIT-----------HPQIGVTRLE-HGLLCRY
UniRef90_UPI0002E38242_11_269 ------------------------------------ST----AVRGDLSQLP-GLLVARC
UniRef90_G8Q511_12_270        ------------------------------------GH----DVRGDLTQLP-GLVVARC
UniRef90_A0A168IUL4_4_266     LKC--CTS----------------------------------SAHIGLSTLMVPGFSLRV
UniRef90_UPI0004799FB1_14_281 ----------------------------------TPGD----GARHGITAPP-GLLIARY
UniRef90_Q3KIS7_11_269        ------------------------------------GH----EVRGDLTQLP-GLLVARC
UniRef90_K9XZ09_12_271        RK-------------------------TREN-----------KGEFGVTELM-SGLLCRY
UniRef90_A0A090RU86_25_288    LT-----------------NMQQE----------NKKG----AVLIAASQLE-NLLVIRA
UniRef90_A0A0M5LWB1_11_269    ------------------------------------PG----PVRADLTQLP-GLLVARC
UniRef90_A0A0C2I8A1_11_269    ------------------------------------PA----RVRGDLTQLP-NLLVARC
UniRef90_H3SNI3_3_265         LDL--NRM----------------------------------EGCIGLSELIIPGFGVRM
UniRef90_A0A0Q0XPT8_11_269    ------------------------------------PN----PVRGDLTQLP-GLLVARC
UniRef90_K4ZJ70_4_265         IDG--KYE----------------------------------DCRIGLSLLPVSGLMARV
UniRef90_A0A085V951_11_269    ------------------------------------PA----QVRGDLTQLP-GLLVARC
UniRef90_A0A098SV87_11_271    ----------------------------------AEHS----PVRGDLSQLP-GLIIARC
UniRef90_A0A0A1GEL6_9_267     ------------------------------------PN----RVRGDLTQLP-GMLVGRC
UniRef90_I4CP19_9_267         ------------------------------------PS----RVRGDLTQLP-GLVVARC
UniRef90_A0A0W0P2S5_11_269    ------------------------------------QH----PVRGDLTQLP-GLLLARC
UniRef90_C3K5A6_11_269        ------------------------------------PH----TVRGDLTQLP-GLLVARC
UniRef90_A0A078LV36_9_267     ------------------------------------PG----RVRGDLTQLP-GLLVGRC
UniRef90_J3GGT8_11_269        ------------------------------------PN----DVRGDLTQLP-GLLVARC
UniRef90_A5L5M9_32_304        LL-----------------SITQQASQQSDSSKISSDP----TLILGATQIE-GLIVVRA
UniRef90_A0A0J6GPD8_11_270    ------------------------------------PN----PVRGDLTQLP-GLLVARC
UniRef90_Q4KJ05_11_269        ------------------------------------AH----PVRGDLSQLP-GLLVARC
UniRef90_A0A0V7ZQT1_2_281     WK---------SNPHMTLSRHAALTPIPNSK-----------SPNQGVTRLN-NGFLCRY
UniRef90_A0A075PF54_11_269    ------------------------------------PH----AVRGDLTQLP-GLLVARC
UniRef90_A0A0D5Y774_11_269    ------------------------------------GH----AVRGDLTQLP-GLLVARC
UniRef90_A0A0Q5EC30_11_271    ----------------------------------NMPN----PVRGDLTQLP-GLIVARC
UniRef90_I4N4V9_11_269        ------------------------------------KG----KVRGDLSQLP-GLLVARC
UniRef90_D4TUH6_7_266         IA----------------------------N-----------GWDAGVTRLE-QGILCRY
UniRef90_UPI00048AE6EC_1_266  YSE--DDQ----------------------------------KYKMGISSLVIPGFSVRI
UniRef90_UPI0007398261_11_285 WE------------ALPEQEVRSQHGVPRNV-----------LTQVGVTRLM-SGLLCRY
UniRef90_B0C790_12_273        WQ-------------------------PTTD-----------GAEMGVTRLP-LGLVCRY
UniRef90_A0A066UMR7_26_298    LL-----------------SITQQASLQPVHFGSSSKS----SLIMGATQIE-GLIVIRA
UniRef90_A0A0Q9XWI2_4_265     LEQ--MEG----------------------------------NFKIGISRLATEGLSIRI
UniRef90_A0A0S7ZTJ9_14_276    T---------------------------------AAHT----TALCSATLVD-EVLVCRY
UniRef90_A6SZ04_26_286        SS--------------------------------ALAQ----GGRSGATQMK-QVLVARY
UniRef90_U3H3U5_9_267         ------------------------------------PS----AVRGDLTQLP-GLLVARC
UniRef90_A0A073CY52_6_272     WS-------------------------QRET-----------SSQAGVTQLI-SGLLCRY
UniRef90_U6ZYX1_11_270        ------------------------------------EH----PVRGDLSQLP-GLLLARC
UniRef90_UPI00067CF5D4_26_284 TA--------------------------------VAAE----DGRSGATQMK-QVLVARY
UniRef90_A0A0K2BGW7_26_288    TG--------------------------------ALTRDS--NDRTGATQMK-QVLVLRY
UniRef90_A0A011QEK6_40_302    ----------------------------------PVAD----GLLTGVTALP-ELLVARC
UniRef90_G4T117_15_272        IGE---------------------------------------APGRGVTRID-DLLICRA
UniRef90_Q87VP5_11_271        ----------------------------------SMPS----PVRGNLTQLP-GLIVARC
UniRef90_A0A089YS31_11_269    ------------------------------------TT----RVRGDLSQLP-GLLVARC
UniRef90_UPI0004174C9A_9_267  ------------------------------------DT----PVRGDLTQLP-GLLVARC
UniRef90_UPI000379F3E4_11_286 RT------------PLSLDGSGFGGEAARGV-----------HHQAGATRLT-DGLLCRY
UniRef90_A4VQU8_9_267         ------------------------------------PS----RVRGDLTQLP-GLIVARC
UniRef90_A0A0M3V4J1_7_297     FTPYLALSEKLRAGVPPVEQTSVTQHSALST-----------QHLVGVTQLE-HGLLCRY
UniRef90_A0A098ESZ3_4_266     IQN--KSP----------------------------------AAAIGISMLPIPGFTVRI
UniRef90_UPI000345DD51_13_277 ----------------------------------GLDG----AGKFGVTQMK-GVLSARY
UniRef90_UPI00047D06E7_9_270  CK--------------------------------AFGG----EGYTAPTLLD-DVLVVRY
UniRef90_I3BUX5_1_262         C---------------------------------AEAG----MGTAGATLFN-GVLVVRY
UniRef90_UPI000780B9AB_11_271 ----------------------------------AEHS----PVRGDLSQVP-GLIIARC
UniRef90_UPI000255752C_11_273 ------------------------------------AADNRLELRGDLSQLP-GILVARC
UniRef90_UPI000484E5E1_11_269 ------------------------------------DH----PVRGDLTQLP-GLLVARC
UniRef90_A0A0D6AS13_1_264     IE-------------------------SKFT-----------NLIICITTLQ-QGLLCRY
UniRef90_UPI00034A425D_18_281 VK----------------------TAIKGEM-----------QGEFGVTRLE-QGIICRY
UniRef90_UPI00046A7B93_11_269 ------------------------------------EH----PVRGDLTQLP-GLLVARC
UniRef90_I3YAM1_6_262         A---------------------------------ATDS----DILWGATLLD-DLLVARC
UniRef90_A0A0D9AIT7_9_267     ------------------------------------SS----RVRGDLTQLP-GLVIGRC
UniRef90_UPI0003FDB5F5_4_265  IQQ--EAG----------------------------------EFTFGLSKLAVPGFTIRI
UniRef90_K9SBL5_6_267         WQ----------------------------G-----------RGEVGVTRLM-AGMLCRY
UniRef90_UPI00040AC544_17_274 IGD---------------------------------------ATGRGVTRMD-DLLVCRA
UniRef90_A6D6Q9_31_296        LD-----------------NMQQE----------HNEG----AVLIAASQLE-NLLVIRA
UniRef90_B8HW54_12_274        WN-------------------------AEKF-----------EGETGVTRLM-SGLLCRY
UniRef90_A0A0J6H3B0_11_269    ------------------------------------PH----PVRGDLTQLP-GLLVARC
```

```
                              240       250       260  265
                               |         |         |    |
HpUreD                        LAKGSEPLLHLREKIARLVTQTTTQKV
UniRef90_A0A0K9H6B6_4_266     LANSTQTVERLFTECHRIISEEWFNK-
UniRef90_A0A0K9GXZ4_4_265     LANKTQTIERLFTECHQIISEAWFN--
UniRef90_A0A0A3IXZ5_4_264     LAHSTQLIERIFDNCHRLIKKSWF---
UniRef90_K9ZPZ7_7_273         RGNSTSEVRNWFTNVWQILRVSLLNR-
UniRef90_A0A0C1XDA3_13_272    RGSSTSEVRNWFIDVWQLLRMSFSSR-
UniRef90_UPI000379D7E3_4_266  QANATQTIERIFNQCHTIISKKWNHR-
UniRef90_UPI0002DF3930_13_277 RGSSTSEVRNWFTAAWQLLRQSLLER-
UniRef90_A0A0M0W0K1_1_266     QANTTQTIERIINQCHAILSEKWNHR-
UniRef90_UPI000422C885_4_265  LANKTQTIERIFSECHQLISEDWFN--
UniRef90_B4W160_8_270         RGSSTTEVRNWFTQVWQCLRLTYLGR-
UniRef90_Q8YQZ4_10_269        RGASTSEVRNWFTSVWQLLRGEFFSR-
UniRef90_A0A0P1BUZ9_3_269     RGSSTSEVRNWFTSVWELLRVSFLHR-
UniRef90_A0A0M1JRC4_17_282    RGNSTSEVRSWFIDVWHLLRLSYLG--
UniRef90_UPI0007108B5C_4_266  LANGTQTIERIFSEFHNIISQEWFNK-
UniRef90_UPI0006A76A92_4_265  LANSTQLIERIFNNCHRIISEKWNN--
UniRef90_W4ETC9_4_265         LANSTQLIERIFNSCHKIISEKWTN--
UniRef90_A0A0D6KKC6_12_271    RGASSSEVRNWFTSVWQMLRVNFLSR-
UniRef90_A0A0D8ZYE2_11_271    RGSSTTEVRHWFTNVWQLIRSSYLER-
UniRef90_B2IT63_10_269        RGSSTSEVRNWFIDVWQLLRVSFLNR-
UniRef90_UPI000308766C_11_271 RGSTTSEVRNWFTAVWQLLRQSFLAR-
UniRef90_K9R6Q0_10_280        RGDSTAKVRNWFISVWQLLRISFLNR-
UniRef90_A0A127D3L2_4_264     LANSTQEIEKIFTECHRIISQEWF---
UniRef90_UPI00028931D6_4_266  LANNTQAIENIFSEFHHIISLEWFNK-
UniRef90_A0A0M0ENP6_4_266     LAHSTQVIERILDQCQKIISEKWNNR-
UniRef90_A0A0S3PHU6_16_275    RGLSTSEVRNWFTVVWQLLRVSFLHR-
UniRef90_A0A081NYG6_3_264     LASSTQDIERLFAACQRLVREQWLG--
UniRef90_A0A0K9GPB1_4_266     LANLTQTIEQLFTECHCMISEEWFQK-
UniRef90_UPI000717378B_2_265  LANSTQLIERIFANCHKDITQKWFG--
UniRef90_K9W1V1_14_278        RGYSTLEARNWFIRVWELLRLAYLGK-
UniRef90_K9U4Q1_17_272        RGSSSIEVRNWFTSVWQ----------
UniRef90_A0A0M2SWM6_4_265     LANSTQIIERIFNDCHKIISRKLNH--
UniRef90_UPI0003652A48_11_275 RGHSSSEARRWFLAVWQLIRVSYFQR-
UniRef90_UPI00047A812C_16_286 RGSSTAEVRHWFIGVWQLLRMSFLNR-
UniRef90_K9TKA4_12_277        RGSSTPEVRNWFSEVWQLLRLSFIGR-
UniRef90_M7NKH7_1_269         LGRSTGRIEKMLNACHKRISEEWLG--
UniRef90_K6DR39_4_265         LANSTQMIEKIFSEIHQMISREWFQ--
UniRef90_A0A0B4RFS1_1_265     LANSTQVIEKIINECHRIIHERWFG--
UniRef90_A0A168N9T6_4_266     QANATQTIERILNQCHTIISEKWNNR-
UniRef90_A0A139X4D9_7_274     RGSGTTEVRNWFTSAWQLLRQSVLTR-
UniRef90_UPI00030AB192_16_277 RGYSTTEVRNWFIGVWQLLRIFFLSR-
UniRef90_K7W9H0_9_269         RGNSTSQVRSWFTNIWQMLRVSCLNR-
UniRef90_UPI00034A5E0C_10_263 RGDSTSEVRNWFMAVWQMLR-------
UniRef90_A0A0A0E4Q3_4_265     FANTTQVIERIFSEIHRHISHEWFQ--
UniRef90_UPI000717263C_4_266  LANSTQVIERILNNCHKLISEKWLNR-
UniRef90_K9QPZ4_10_269        RGASTSEVRHWFTSVWQMLRVDYFKR-
UniRef90_UPI0002D28249_14_276 RGSSTAEVRNWFISVWQLLRVSFLSRV
UniRef90_F9DU20_4_266         MANYTQVIERIISACHHVISDEWYQK-
UniRef90_UPI00047933CA_3_266  LANSTQVIEKMFSEFHQIISQEWFNK-
UniRef90_W7RFE8_4_266         LANQTQIVERLFSEFHHILNQEWFNK-
UniRef90_UPI0007441C40_4_265  LGNSTQLIERIFNQCHKIISEKWNN--
UniRef90_A0A0T7BRT9_7_268     RGNSTSEVRNWFISVWQLLRQSIGNR-
UniRef90_UPI0006A78357_4_266  LAKNTQTIESMFFDFHHIISQEWFNK-
UniRef90_UPI0007C7BB46_4_266  LAHSTQKIEELMAACSSFLRKEWYDR-
UniRef90_UPI000472611B_4_265  LANTTQVIEALFAKCHMIVNKEWYD--
UniRef90_UPI0002ACBA0D_11_270 RGSSSTEVRNWFTEIWQLLRSPYLNR-
UniRef90_A0A0C2KR59_11_269    RGSTTSEVRNWFIGVWQLLRQSFL---
UniRef90_K9WE93_4_277         RGSSTSEVRNWFTEVWQLLRLSFLGR-
UniRef90_A0A0M0SP30_16_277    RGSSTAEVRNWFIGVWQLLRMSFLSR-
UniRef90_K9Q9K1_10_268        RGASTSEVRNWFTAVWQILRTSFLSR-
UniRef90_A0A0S3TTX4_17_278    RGSSTAEVRNWFIGVWQLLRMSFLNR-
UniRef90_K9VQF3_56_323        RGTSTTAARDWFVNIWQLLRLSFSQR-
UniRef90_W1SM28_4_265         LANTTQVIEKLFSEIHRVISQEWFN--
UniRef90_A0A0C1N935_15_285    RGSGTTEVRNWFTSVWQLLRQSVLAR-
UniRef90_A0A139SK67_11_275    LGPACEPGRAWFARLWAVLRPALSGRA
UniRef90_UPI00041C7CDC_4_265  LANSTQIIERIFADCHQMICEEWFG--
UniRef90_K9XHH8_15_275        RGDSTTEVRQWFTEVWHLLRLSSLGK-
UniRef90_UPI0002E7D901_4_265  LADKTQLIESIVSACHLAVSEKWHQ--
UniRef90_A0ZB05_10_269        RGASTSEVRNWFTVVWELLRGDFLHR-
UniRef90_D4ZSS4_5_270         RGPSSQEAIAWFTQIWQLLRPNLSGK-
UniRef90_A0A0F5YFY1_6_271     RGNTTQEVINWFTTVWQLIRQNHQGRV
UniRef90_K9T9Y9_59_322        RGSSTADVRNWFTGVWQLLRLSLMKR-
UniRef90_K9VAD2_10_273        RGSGSSEVRNWFTDVWQLLRMSYLHRV
UniRef90_W7Z4J7_4_265         LANSTQVIEKIHTKCHHIIYQQFFN--
UniRef90_Q47G52_13_278        LGNSSEAARLWFAELWTILRPACCGR-
UniRef90_UPI0004024FB0_3_269  RGHSISEVRNWFTNIWQSLRINYQNR-
```

```
UniRef90_K8GMD3_12_265        RGHSTSEARRWFIQVWQ----------
UniRef90_A0A0S3UB12_16_269    RGHSSTEARRWLIAVWSMVR-------
UniRef90_A0A0F7D4R9_4_265     IADKTQVIERVTAACHAAISEEWQQ--
UniRef90_A0YQS4_5_271         RGNTTQEVINWFTNIWQFIRQNYQGRV
UniRef90_A0A0Q8RCL2_14_276    LGHSSQTARRLMMHTWQRLRPALTGR-
UniRef90_UPI0005625009_5_271  LGQRTESILRLFTAIWRRVRPEI----
UniRef90_UPI000305E365_16_284 RGSSTTEVRNWFIGVWQLLRMSFLSR-
UniRef90_UPI0002E38242_11_269 LAGQALHARAWLIELWRLLRPALLGR-
UniRef90_G8Q511_12_270        LASEALLARGWLIELWRLLRPALLGR-
UniRef90_A0A168IUL4_4_266     LAYSTQEIEKIFRSVQQLIREQWFGK-
UniRef90_UPI0004799FB1_14_281 LGDNSEAARLWFAELWKILRPACCGR-
UniRef90_Q3KIS7_11_269        LASEALLARAWLIDLWRLLRPALLGR-
UniRef90_K9XZ09_12_271        RGYSTTEAKEWLSQVWQILRSEI----
UniRef90_A0A090RU86_25_288    LGNWSEVILDCFQQVWQVAREHWTG--
UniRef90_A0A0M5LWB1_11_269    LAVEALQARAWLIELWRLLRPALLGR-
UniRef90_A0A0C2I8A1_11_269    LAGEALHARAWLIELWRLLRPALLGR-
UniRef90_H3SNI3_3_265         LGNSTQAIETLFGRIANAVRESWFG--
UniRef90_A0A0Q0XPT8_11_269    LATEALQARAWLIELWRLLRPALLGR-
UniRef90_K4ZJ70_4_265         LANSTQTIEAIFACCHRFINQSIFN--
UniRef90_A0A085V951_11_269    LADEALHARAWMIELWKLLRPAVLGR-
UniRef90_A0A098SV87_11_271    LADEALHARAWLIELWKLLRPALLGR-
UniRef90_A0A0A1GEL6_9_267     LAGEALHARAWLIDLWRLLRPELLDR-
UniRef90_I4CP19_9_267         LADEALHARAWLIDLWRLLRPELLGR-
UniRef90_A0A0W0P2S5_11_269    LASEALHARAWLIDLWHLLRPALLGR-
UniRef90_C3K5A6_11_269        LASEALLARGWLIDLWKLLRPVMFGR-
UniRef90_A0A078LV36_9_267     LANEALHARSWLIALWQLLRPELLER-
UniRef90_J3GGT8_11_269        LASEALLARGWLIDVWRLLRPALLGR-
UniRef90_A5L5M9_32_304        LGNWSEDILQAFGQIWQATRSHLYG--
UniRef90_A0A0J6GPD8_11_270    LAAEALQARGWLIDLWHLLRPALLGRA
UniRef90_Q4KJ05_11_269        LASEALQARAWLIELWRLLRPALLGR-
UniRef90_A0A0V7ZQT1_2_281     RGNSTAEVRNWFINVWLLLRTSFLSR-
UniRef90_A0A075PF54_11_269    LAGETLQARAWLIALWRLLRPALLGR-
UniRef90_A0A0D5Y774_11_269    LASETLLARAWLIDLWRLLRPALLGR-
UniRef90_A0A0Q5EC30_11_271    LADEALHARAWLIELWRHLRPALLGR-
UniRef90_I4N4V9_11_269        LAHEGLHARDWLIGLWKLLRPEALGK-
UniRef90_D4TUH6_7_266         RGNSTSWAKKWFTNVWQDLRQSLLNR-
UniRef90_UPI00048AE6EC_1_266  LSHDTQTIEEVFNYIHQTIRKELLQK-
UniRef90_UPI0007398261_11_285 RGTSTQEARRWFTTVWHLLRWELLNR-
UniRef90_B0C790_12_273        RGPSSQAARQWFIQVWNLLRSTHLGR-
UniRef90_A0A066UMR7_26_298    LGHWSEDILQAFGQIWQATRSHLCG--
UniRef90_A0A0Q9XWI2_4_265     MANSTQLIERIFTACHHTISMEWFQ--
UniRef90_A0A0S7ZTJ9_14_276    LGHHGEQAKKVFTSVWSAIRPACVNR-
UniRef90_A6SZ04_26_286        LGHSSETARLWMTRAWQRIRPELMQR-
UniRef90_U3H3U5_9_267         LASEALHARAWLIDLWRLLRPALLGR-
UniRef90_A0A073CY52_6_272     RGNSTQEVINWFTDVWQLLRQNYTGK-
UniRef90_U6ZYX1_11_270        LASEALQARAWLMDLWRLLRPALLGRA
UniRef90_UPI00067CF5D4_26_284 LGHSSESARHWMTRAWQRIRPELM---
UniRef90_A0A0K2BGW7_26_288    LGNSSQAARQWLTHAWQRIRPELMQR-
UniRef90_A0A011QEK6_40_302    LAPGAEAARGWLREVWQQLRPLALGK-
UniRef90_G4T117_15_272        LDHRADKLRDFFNEVWTSIREATVGR-
UniRef90_Q87VP5_11_271        LADEALHARAWLIEIWKRLRPALLGR-
UniRef90_A0A089YS31_11_269    LADEALHARDWLIQLWTLLRPALLGR-
UniRef90_UPI0004174C9A_9_267  LASEALHARAWLIELWRLLRPALLGR-
UniRef90_UPI000379F3E4_11_286 RGNSTAEVRNWFIGVWQLLRQSFWGR-
UniRef90_A4VQU8_9_267         LADEALHTRAWLIDLWRLLRPALLGR-
UniRef90_A0A0M3V4J1_7_297     RGASTSEVRNWFTAVWQILRTSFLSR-
UniRef90_A0A098ESZ3_4_266     MAHSTQAIEGIFSEYHHIISRDWFNK-
UniRef90_UPI000345DD51_13_277 LGDDSEQARRVMLAVWQRLRPFLLERA
UniRef90_UPI00047D06E7_9_270  LGNSTEQAHHLFRKIWLAIRPLVNGRL
UniRef90_I3BUX5_1_262         LGDSTAQAHRLFRSLWQAIRPLLTGRA
UniRef90_UPI000780B9AB_11_271 LADEALHARAWLIELWKLLRPALLGR-
UniRef90_UPI000255752C_11_273 LAREALHARAWLIQLWQLLRPAVLGR-
UniRef90_UPI000484E5E1_11_269 LASEALLARGWLIALWRLLRPALLGR-
UniRef90_A0A0D6AS13_1_264     HGNSVSEAKTCLTAIWQLLRAKY----
UniRef90_UPI00034A425D_18_281 RGASSLEARTWLIAVWQMLRVSFMG--
UniRef90_UPI00046A7B93_11_269 LASEALLARAWLIDLWRLLRPALLGR-
UniRef90_I3YAM1_6_262         LAPFAEPARRLFAAIWGILRPRLL---
UniRef90_A0A0D9AIT7_9_267     LASEALHARAWLIDLWRLLRPALLGR-
UniRef90_UPI0003FDB5F5_4_265  LANYTQVIERIISVCHHVISDEWYQ--
UniRef90_K9SBL5_6_267         RGDSTEEARAWMLRVWDLLRQALIQR-
UniRef90_UPI00040AC544_17_274 LDCRSDRLRVFFERVWAVVRPDCVRR-
UniRef90_A6D6Q9_31_296        LGNWSEVILACFQKIWQLVRGHWTG--
UniRef90_B8HW54_12_274        RGFSTQSARNWFMQVWHLLRLHYRHQ-
UniRef90_A0A0J6H3B0_11_269    LAGEAFQARGWLIDLWRLLRPALLGR-
```

Table 2-SI. ConSurf analysis of the multiple sequence alignment obtained using PSI-BLAST.

*Hp*UreD sequence number	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	Most conserved residue (%)	ConSurf Grade
1								8		8	38		23					23			M 38.462	6
2	8		4					12				31		4	4	23	15				N 30.769	3
3	10		6			1	1		1	1	4	1	4	4	3	12	48	1			T 47.826	3
4	3				3	1	1		4			4	47	12		12	12			1	P 47.368	2
5	7		2	2		33	6	1				2	1	3	1	36	6				S 36.275	3
6														1					99		W 99.315	9
7		1	1	1			47		3	1		2		14			28	1		3	H 46.980	6
8	35					64										1					G 64.430	9
9	1		1	26				8	18	2		7		3	9	10	3	11		1	E 26.174	2
10								1		99											L 99.333	9
11	4	1	10	26			7		1			21		3	13	12	3				E 26.000	2
12								1		99	1										L 98.667	9
13	9		19	3		27		3	5	1		2		1		3	5	25			G 26.667	1
14	8				17			2		9	2						1	6		55	Y 55.333	6
15	52	1		30		7		1	1			1		1		2	2	3			A 52.000	6
16								100													I 100.000	6
17	1	4	22	7	1	1	7		6	1		13	1	6	24	2		1		2	R 24.000	1
18	2	3	7	1	15		3	2	5			1		2	53	2	1	1		1	R 52.667	3
19	7		5	3		26	3		7	5		12	1	23	4	2	1			1	G 26.000	1
20	3		20	5		51	1		1			16	1			2					G 51.333	1
21	8	3		2		3		3	35					3	11	17	13	1			K 35.333	2
22		1														5	94				T 94.000	9
23	4			1		1	1	4	3					25	27		2	33		1	V 32.667	6
24	23					1		5		37	3		24			1	1	5			L 37.333	6
25	3		1	1		1		26	23	1	1	1		1	5	3	11	22			I 26.000	3
26			5	9	4	1	15		4	13	1	15		8	11	5	1			9	HN 14.667	1
27	7	1	1	1				2	1	1		19		2	32	9	1	23			R 32.000	5
28	1			1	3		11		1	2				15	28	5	1			31	Y 31.333	4
29	1				25	3	39	1			3	9		7	1	3	1	5		2	H 39.333	5
30	1	1		1	4	1	1	1	9	9		1		59	5	5	3			1	Q 58.667	5
31	36					61						1	1			1	1				G 61.333	9
32	27												73								P 72.667	9
33					19					77									1	3	L 77.333	7
34	5								63			2			27		1	1			K 63.333	9
35								11		5	1							84			V 84.000	7
36	1									1	25			72			1				Q 72.000	9
37	1								29				1		69	1					R 69.333	9
38	3						21						71			3	1				P 71.333	8
39					42		1	17	1	31								8			F 42.000	7
40							3													97	Y 97.333	9
41	20				3		17			4			55				1				P 55.333	7
42			25	71		1			1			2	1	1							E 70.667	9
43	1		17	3		69	2		1			5		1	1						G 69.128	5
44	8		6	28		30			9			2	1	9	1	5	1				G 30.000	1
45							2	15	2	1			2	21	1		1	53		2	V 53.333	6
46	9	72											7					12			C 72.000	7
47		26		1			51							23							H 50.667	9
48	6	1			1	1	21	1		1		1				33	5	1		27	S 33.333	7
49	1				1			29		1	1							35		33	V 34.667	8
50	1							64		23	1							10			I 64.000	6
51								1		71	3							25			L 70.667	8
52	1						72					27				1					H 72.000	9
53					1								61				37	1			P 61.333	9
54	36					29							35			1					A 36.000	9
55	1					97							1			1					G 97.333	9
56						100															G 100.000	9
57								42		1	4							26		27	I 42.000	7
58	23									26	1							50			V 50.000	9
59	2		25	1		71						1			1						G 70.667	8
60						99										1					G 99.333	9
61			100																		D 100.000	9
62	1	1		2			3	1	2					3	78	4	5	1			R 78.000	6
63										71		1		1						27	L 70.667	9
64	7	1	15	1	1	1	3	1	7	2		5		9	11	32	4				S 32.000	4
65	1	4			2	1		35		13	21			7		7	2	1		7	I 35.333	5
66	1	1	13	11			5		14			18		7	5	18	5				NS 18.000	1
67	23				19			35		8								16			I 34.667	3
68	4		2	3		1	35		1	1		5		4	9	25	10				H 35.333	1
69	11							1		57						1		30			L 57.333	7
70	8	1	11	18		23	1	1	3	1		1		30	1	2					Q 30.000	2
71	10		5	25			1		6			1	35	3	4	6	3				P 35.333	1

72	2		15	5		25	7		4			25		15	3						N 25.333	1
73	67	1														20	12				A 67.333	7
74			1	2		1	17		18		1	2		27	10	1			21		Q 26.667	4
75	61									15	2					3	3	16			A 60.667	6
76					1			1		49				23	1		23	2			L 49.333	8
77								40		58	1							1			L 58.000	7
78																	100				T 100.000	9
79														1		23	76				T 76.000	9
80	33												37	28		1	1				P 37.333	9
81	35					36										25	4				G 36.000	8
82	96				1											3					A 96.000	9
83	23			1		18						1				23	33				T 33.333	7
84									99						1						K 99.333	9
85					7			40		1								25	27		I 40.000	7
86							1													99	Y 99.333	9
87						1			27			1			71						R 71.333	8
88	24															43	33				S 43.333	8
89	14		1	3		5		1	1	1		37	26	1	1	5	4	1			N 37.333	5
90	1		1	1		65	1		10			11		1	1	5	3	1		1	G 64.667	4
91	3		4	2	1	3	1		11	17	3	4	23	5	6	9	6		1	1	P 22.667	1
92	2	1		11	5	1	14	3	1	8		1	8	19	2	3	4	3		13	Q 19.492	1
93	88												1			5		7			A 88.000	8
94	1			1	7	1	3	1	14	1	1	1		7	17	8	2	1		36	Y 36.000	2
95							1			1		1		96	1	1					Q 96.000	9
96			5	28			1	6	5	1	1	9		8	3	7	23	1		1	E 28.000	2
97	3							25		19	3	1		7		3	19	20			I 24.667	5
98	3		9	36			4	3	3	1	1	13		9	4	5	9	1		1	E 36.000	1
99		1			9			54		30	3							3			I 54.000	7
100	1	1	2	3	3		11	2	15	1		7		19	11	10	11	2		1	Q 19.333	1
101	3							16		29	3							49			V 48.667	5
102	21		25	19		7			21			1		5	1	1	1				D 24.667	3
103	35		5	11		1			20			3	10	2	1	8	4				A 34.667	1
104	1		7	1		78	4					8		1	1	1					G 78.000	3
105	64	1				5										30	1				A 64.430	8
106	3	29			2		1	2		1	1	1			3	2	25	3	1	28	C 28.667	3
107		3								95								2			L 95.333	8
108			1	99																	E 99.333	9
109					3					1									62	33	W 62.000	6
110	1				3			11		79	4							2			L 79.333	6
111													98			1	1				P 98.000	9
112			27		1					2				67	3						Q 66.667	9
113	2		2	69		1							25			1					E 69.333	9
114	1							3		25		6				4	61	1			T 60.667	7
115								99		1											I 99.333	9
116	21				8	5		8		9	2		1					47			V 46.667	6
117					64															36	F 64.000	8
118	2	1	11	9					5			31	5	9	7	21					N 31.333	5
119	17		18	3		34	5		1			13		5		5					G 34.000	2
120	95	1				2										1	1				A 95.333	9
121	1	1	5	3			13	21	7	5		4		21	15	2		1			IQ 21.333	2
122	25	1			5	1				1			1					7		61	Y 60.667	7
123	1			21		1	1	1	20	2		1		5	41	2	1	2		1	R 41.333	6
124	1					1		1		28	3			65		1					Q 65.333	9
125	1		35	7			1		26			5		4	2	7	10	1		1	D 34.667	2
126							2	1		30	3	22		1		1	37	3			T 36.667	7
127	1	3	1	3			3	8	2		1	3		8	41	5	5	17		1	R 40.667	1
128					1			49										51			V 50.667	6
129	1		10	42			15		4			7		1	11	1	2	1	1	3	E 42.000	2
130								3		73	24							1			L 72.667	8
131	35		5	36		4	6	1				1		9	1	1	1				E 36.000	4
132	7	1	2	3		19			15			1	20	3	5	7	16	1			P 20.000	2
133	1		31	3		48			3			7		1		2	4				G 48.000	3
134	85	1				3										10	1				A 85.333	8
135	3	3		1			1	1	2	2		7		4	27	19	30	1	1		T 30.000	4
136					32			1		35	1							1	18	13	L 34.667	4
137		5			19			22		47	3						1	2	1		L 47.333	2
138	3				1	45		1		2						1				47	Y 46.667	6
139	1	1				3										16	11	1	68		W 68.000	8
140			50	50																	DE 50.000	9
141								77		2	9							11			I 77.333	8
142	1	1					4	14		21	1	3		1		1	31	23			T 31.333	5
143	22	11													37	1	27	1			R 37.333	9
144	1	1			37					33			26			1		1		1	F 36.667	6
145						100															G 100.000	9
146															73				27		R 72.667	9
147	1									1	1		26	1	1	52	17	1			S 52.000	7

148	71												27				1	1			A 70.667	9
149	5	1	9	7		1		1	1	1		5		1	37	30					R 37.333	6
150	1		4			88			1			5		1			1				G 88.000	6
151			3	87		1			5						1	1	1				E 87.333	8
152	1			1		3	3	2	33	3		2	1	8	35	3	3		1		R 35.333	2
153					98					1										1	F 98.000	9
154	6	1	16	9	1	1		1	1	19	1	2			2	25	6	7		3	S 24.667	1
155	7	1	1	5	1	1	7	1	1	6				27	3	7	3			27	QY 27.333	2
156			18	1		72	2		1			3		1		1	1				G 72.000	6
157	1		3	26	1		21	1	5	3	4	7		3	7	4	13			3	E 26.000	2
158	3				23			8		19	3							11	33		W 33.333	5
159	3		3	1		1	2		3	4				41	40	1	1	2			Q 41.333	6
160	11				3	3		1		27	1	2		4		46	1	1			S 46.000	6
161	1	3	1	1	1		37	9	13	3	2	3		9	9	1		5		3	H 36.667	2
162		1			1			1		23		16				2	56		1		T 56.000	8
163	1		18	66			1		1		1	3		4	3	3					E 66.000	7
164								79		3								18			I 79.333	7
165				2	1		1			1		1		1	23	1	1	1	36	32	W 36.000	6
166		1					1	1	1	12	13			35	31			4	2	1	Q 34.667	7
167	3		50	8		3	1		1	1	1	9		22	1	1					D 50.000	5
168			13	4		71			1			9		1		1					G 70.667	3
169	1	1	3	19			1	3	25	1				17	15	1	2	11			K 25.333	1
170	1							1	1	47			47		2		1		1		P 47.333	6
171	5				1	7		7		63				1				17			L 63.333	6
172	2				1			1		7	2					1		21	64	2	W 64.000	6
173					19		21	39	1	5							3	1	1	11	I 38.667	6
174			67	33																	D 67.333	9
175						3	27					1	1	5	63						R 63.333	8
176	1				1	6		21		10				59			1	1			Q 58.667	8
177	1	1	1		1	1	3	3	23	3	1	2		3	28	1	1	1	23	3	R 28.000	3
178					3			18		67	1						1	11			L 67.333	5
179	5	1	5	4	1	1	4	3	2	1	1	7	31	5	5	7	5	13			P 30.667	1
180	9					61						1	29								G 61.333	7
181	15	1	8	7		25	1		3	1		13	1	2	4	19	1				G 25.170	1
182	3		28	32		1	1	1	4	1	1	3	2	6	2	7	5	3	1		E 31.757	1
183	10		7	21		21	1		6			3	2	22	1	3	3	1			Q 22.000	3
184	4		5			1	1	12	1	28	4	13	2		1	5	5	16	1		L 27.891	1
185					25			16		41	15							3	1		L 40.667	4
186	4		27	9		2	21	2	3			10		2	3	7	7			3	D 26.667	1
187	6		1	1		20			1	1		3				66	1				S 66.000	7
188	3				1		2	9		21	1	1	58		1	1		1		1	P 58.219	4
189	4	1				27	33	19	1	4	2	3	1		1	1		3	2		H 33.333	6
190	4		1		20	65	1			5	1				1	2				1	G 65.333	6
191					1					73	25								1		L 72.667	9
192	33		19	27	1	4	1		1	3		4		1	3	3					A 32.667	5
193			1			93						3			1	1					G 93.333	7
194	1	3	1	1	11		9		11	1		6		27	1				1	29	Y 29.333	1
195	5												53	1		15	26				P 52.667	6
196	1						28	19		3	3					1		45			V 45.333	8
197	5	5			23			7		24	4			1		1	3	27		1	V 26.667	4
198	36					63				1											G 63.333	6
199												3				53	44				S 52.667	8
200					15					61	24							1			L 60.667	6
201	10				4			35		23	3					1	3	17	1	3	I 35.333	4
202	18	2			2	1		3		5	1							37	22	9	V 37.333	3
203	1							27		6	1					6	20	37		2	V 36.667	3
204	3	1	2		1	73		1	1			1	4	1		11					G 73.333	4
205	1	1	10	42	1	3	3		9	1		3	3	15	1	5	1	1			E 41.667	2
206	7		2	11		1			9	2		3	23	23	2	5	6	5		2	PQ 23.148	1
207	8							27		3	6					3	15	36	1		V 36.301	3
208	2	1	26	7		2						16	3			33	10				S 33.333	2
209	19		4	7		3	3		7	1	1	7	24	5	3	13	3	1			P 24.000	1
210	11		21	54		1	1					1		4		5		1			E 54.000	3
211	2	1	2		11	1		17		56	4				1		2	3	1	1	L 56.000	2
212								18		57	4				1			19		1	L 56.667	5
213	7		28	49			3		2			3		5	1	1	2				E 48.667	5
214	6			13			3	2	24	5	1	1		13	24	2	5	1			KR 24.000	1
215	26	25						7		30						1	1	11			L 30.000	5
216	1	1				1	3	1	1			1		5	66	1				21	R 66.000	7
217	9		6	23		3	1		1	2	1	13		9		26	4	1		1	S 26.000	1
218	12	1		4	1	1	1	5	1	59	3			1		3	5	5		1	L 58.667	1
219	1	2	1		6	2		26		12	4		1		2	3	4	5	31	1	W 30.702	1
220	5	2	13	5		5	7	3	7	5	1	10	3	15	2	7	7	1		2	Q 15.315	1
221	9		13	15		5	5	1	5	3	1	7	11	7	6	8	3	2			E 15.333	1
222	5	3	1		11	25	2	1		4		1	4			7	2	24	1	11	G 24.667	1
223	3	1	3	21	1	1	1	2	17	7		1		9	27	3	3			1	R 26.667	1

224	12	1			14	23	1	15		1	3			1	1	3	7	17	1	1	G 22.667	1
225	3	1	21			75						1				1					G 74.667	8
226	10				1			11		40	1		1					37			L 40.000	6
227																37	63				T 62.667	8
228	4			3	1		1	1	4	11	3	1		35	33	2	2				Q 35.333	4
229					1			4		86	4		1				2	3			L 86.000	7
230	9		4	15				2	3	3	4	1	37	7		9	4	1			P 36.667	2
231	4			2		2		33								2	2	53			V 53.333	5
232	1	3	3	3		27	11	1	2	1	1	9	19	5	1	11	1				G 26.667	1
233						65	1	1		26	1							7			G 64.667	7
234					33			7		57	1							2			L 56.667	5
235		2				1		5		37	1					11	13	29			L 37.333	6
236	27	39				2		17		6								9			C 39.333	6
237									1						99						R 99.333	9
238	5	23						13			1							14	1	44	Y 44.000	7
239					1		1	1		57	2			2	37						L 57.333	8
240	48		1			50										1					G 50.000	8
241	5	1	11		1	3	11		1	1		32	3	1	1	23	2	1		3	N 32.000	2
242	2		1	21	1	4	1		3	1		2		1	2	53	5		3	1	S 52.667	6
243	21	1				1		1								15	61	1			T 60.667	7
244	6		1	11	1	1		1		22			1	33		17	7				Q 32.667	6
245	6		3	34			12	2	2	9	1		2	8	1	2	9	9	1		E 34.000	2
246	38					1		29		2							1	29			A 38.000	8
247				27			1	3	3	4					62						R 62.000	8
248	19		3	2		5	4		7	2		26		4	21	3	2	1			N 26.000	2
249	1	2			1			18		10	2							3	63		W 62.667	5
250					63		1	3		27	5				1		1	1			F 62.667	6
251	7		1	1	1	2		33		2	3	8		1	3	9	27	1			I 33.333	3
252	15	1	15	23		6	1		3			8		11	5	9	1	2		1	E 22.667	1
253	3	19			3			13		23							1	37		1	V 36.667	6
254	1						24							2		1			72		W 72.000	8
255	3	1	1	2		1	9		9	2	1	2		43	21	2	3				Q 43.333	4
256	5	1	1		2		1	18		52	6			1	5	1	1	4			L 52.027	2
257	1							30		61							1	6			L 61.486	7
258		1					1		1			2			75	19	1			1	R 75.000	8
259	1		1	14		2	1	2	3	7	5		33	15	2	4	2	7	1		P 32.877	1
260	25		3	21	1		3		8	2		4	1	2	1	25	2	1			AS 24.658	4
261	1	3			17		1	3		32	1					1	1	3	27	9	L 32.192	1
262		2			19	1	1	1		50	3	6		3	1	3	5	2	1	3	L 49.650	1
263	1		2	3		45	6		1			22		11	1	7	1				G 44.928	1
264									16					1	83						R 82.727	8
265	45									9								45			AV 45.455	7

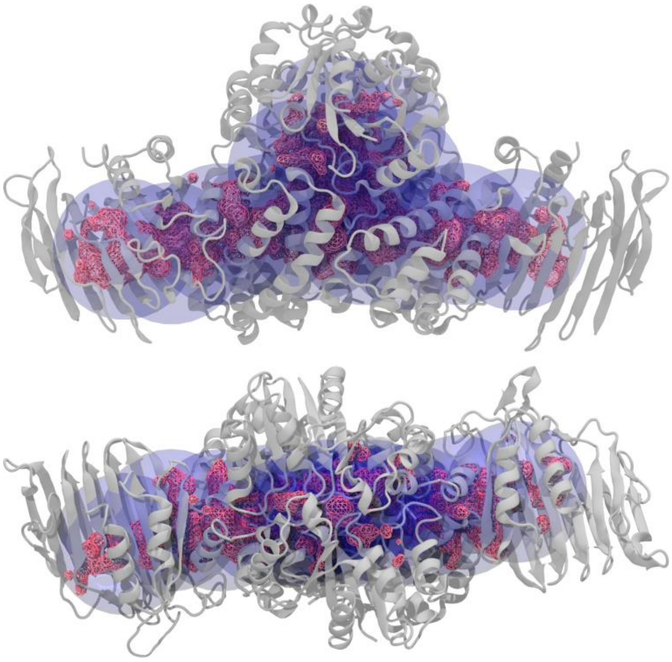


Figure 12-SI. Density isosurface of the water molecules found inside the *Hp*UreDFG tunnels contoured at 0.025 oxygen atom Å^{-3} (purple mesh) superimposed on the fourteen spheres (light blue) used to discriminate the time frames in which selected waters lied inside or outside the tunnels. The *Hp*UreDFG complex is reported as white ribbons. In the bottom panel, the *Hp*UreDFG complex is rotated by 90° around the horizontal axis with respect to the orientation in the top panel.

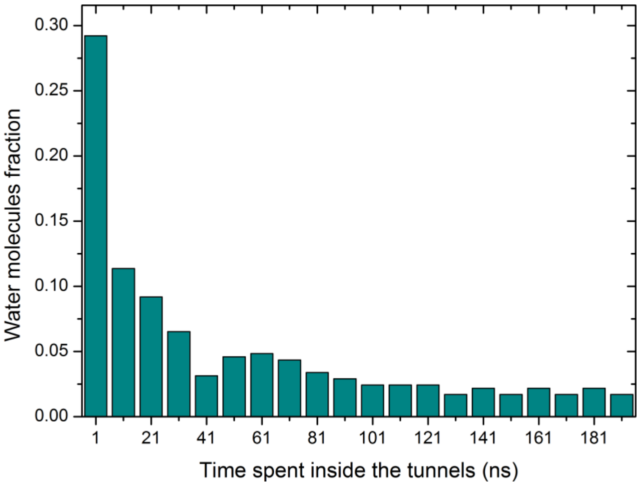


Figure 13-SI. Distribution of the time spent inside the tunnels of the 370 water molecules that are able to enter in the tunnels from the bulk of the solvent or that escape from the tunnels during the course of the simulation (see main text for details).

Figure 14-SI (part one). Trajectory of selected water molecules inside the *Hp*UreDFG tunnels (see main text for details). In each frame, the water molecule is reported as a sphere colored from red to green and finally to blue, accordingly to the simulation time. The starting frame is indicated by a red sphere of larger radius with respect to the others and a black arrow. The *Hp*UreDFG complex is reported as white ribbons and GDP is depicted as balls-and-sticks.

Figure 14-SI (part two).

Figure 14-SI (part three).

REFERENCES

1. Musiani, F., Ippoliti, E., Micheletti, C., Carloni, P., and Ciurli, S. (2013) Conformational fluctuations of UreG, an intrinsically disordered enzyme, *Biochemistry 52*, 2949-2954.
2. Celniker, G., Nimrod, G., Ashkenazy, H., Glaser, F., Martz, E., Mayrose, I., Pupko, T., and Ben-Tal, N. (2013) ConSurf: Using Evolutionary Data to Raise Testable Hypotheses about Protein Function, *Isr. J. Chem. 53*, 199-206.