Index to Supporting Information

A. Supplemental Materials and Methods

and Derivation of Equations **Page 2**

B. Supplemental Figure 1 **Page 8**

B. Supplemental Figure 2 **Page 9**

Supporting Information

Supplementary Materials and Methods

I. General Methods

A. Cell culture

Studies were done in cultured kidney epithelial cells, originally produced by Vandewalle and colleagues（1）and later re-cloned by Yu et al（2）to maximize aquaporin-2 expression and response to vasopressin (mpkCCD-clone11).

B. Dynamic SILAC for protein half life and translation rate profiling

Media. Cells were initially labeled in presence of 'light' SILAC medium and then switched to 'heavy' SILAC medium at time 'zero'; or vice versa. 'Light' medium was Advanced DMEM／F-12 (Dulbecco's Modified Eagle Medium／Ham's F-12, 12634-010, Invitrogen, Calsbad, CA) plus light ('natural') amino acids [91mg/L of (^{12}C) lysine (Supplied with MS10033, Invitrogen Corp.) and 147 mg/L of (^{12}C6, ^{14}N) arginine (Supplied with MS10033, Invitrogen Corp.). The 'heavy' medium was the same DMEM/F12 medium plus heavy amino acids [91mg/L of [^{13}C at 6 sites] lysine (Supplied with MS10033, Invitrogen Corp.) and 147 mg/L of [^{13}C at 6 sites, ^{15}N at 4 sites] arginine (MS10009, Invitrogen Corp.). Peptides labeled with heavy isotopes have an extra mass of either +10.008269 kDa (arginine) or +6.020129 kDa (lysine). The media also contained 2% dialyzed fetal bovine serum except for the 5 day period after the cells reached confluency (*vide infra*).

Labeling and sampling procedure. Previous studies demonstrated that labeling is at least 99% complete in these cells after seven days in SILAC medium（3）. Thus, all cells were pre-labeled with the initial medium for at least 12 days including initial growth in T25 flasks, followed by seeding of ~20,000 of these cells per cm^2 on membrane supports (Transwell, Corning Inc., Lowell, MA). Growth on permeable supports is required for full differentiation of epithelial cells. Both apical and basolateral media were changed daily. We monitored cell confluence by measuring transepithelial electrical resistance (EVOMTM, Epithelial Voltohmmeter, WPI). After cells became confluent (greater than 5 kOhm/cm^2), cells were exposed to the V2R-selective vasopressin analog dDAVP (10^{-10} M) or vehicle added to the basolateral medium for 5 days prior to the medium switch from light to heavy or *vice versa*. Cells were collected at various time points after the medium switch for MS analysis (2, 6, 16, or 24 hours). To collect the cells, they were washed twice with PBS and harvested by lysis with 150µL of 8M urea lysis buffer and scaping. Sample protein concentration was measured by the BCA method (Pierce Biotechnology, Rockford, IL).

C. Standard SILAC

Cells were grown to steady state, with either the 'light' or 'heavy' media described above. For different experiments, the light medium was used for dDAVP-treated cells or vehicle-treated cells, while the heavy medium was used for the opposite condition. Cells were harvested as described above (with no medium switch) and mixed 1:1 light:heavy and processed for MS as described（3）

D. In-gel Trypsinization

550 micrograms of protein sample were stratified on a one-dimensional SDS-PAGE gel (4-15% polyacrylamide Ready Gel, BioRad, Hercules, CA) divided among several adjacent lanes. Gels were stained with 25 mL of Imperial ™ protein stain (Pierce, # 24615) at room temperature with agitation (15 mins). Gels were destained for 1 hr in water at room temperature with agitation. Protein bands were excised with a razor blade and each gel slice was further cut into small pieces (1-2 mm^3) and placed into 1.5 mL Eppendorf tubes. Gel pieces were further destained with 25 mM ammonium bicarbonate in 50% Acetonitrile, vortexed, briefly spun and incubated for 10 min. The supernatant was removed, and gel pieces were dried in a SpeedVac. Proteins were reduced using 10 mM DTT followed by alkylation with iodoacetamide (40 mM). Proteins in gel pieces were digested with 12.5 ng/uL trypsin for 12 hour at 37°C. After extraction of tryptic peptides with of 50% Acetonitrile/ 0.5% Formic Acid, samples were desalted with C18 Ziptips (Millipore Corp., Bedford, MA) and then dried and reconstituted with 0.1% Formic Acid before analysis by nanospray LC-MS/MS.

E. LC-MS/MS Analysis

All samples were analyzed on an LTQ-Orbitrap XL (Thermo-Fisher Scientific) interfaced with a nano-LC 1D plus system (Eksigent Technologies). Fragmentation was conducted via collision-induced dissociation. Briefly, samples were loaded onto an Agilent Zorbax 300SB-C18 trap column (0.3 mm i.d. × 5mm length, 5-µm particle size) at a flow rate of 10 µl/min for 15 min. Reversed-phase C18 chromatographic separation of trapped peptides was carried out on a prepacked BetaBasic C18 PicoFrit column (75 µm i.d. × 10 cm length; New Objective) at 300 nL/min using the following gradient: 2–5% solvent B for 2 min; 5–45%solvent B for 45 min; 45–50% solvent B for 5 min; 50–95% solvent B for 5 min (solvent A: 0.1% formic acid in 98% water, 2% acetonitrile; solvent B: 0.1% formic acid in 100% acetonitrile). Eluted peptides were introduced into the LTQ-Orbitrap mass spectrometer via a nano-spray ion source. LTQ-Orbitrap settings were as follows: spray voltage: 1.5 kV; temperature of ion transfer tube: 180 °C; and full MS mass range: 300–2000 m/z. The LTQ-Orbitrap was operated in a data-dependent mode such that a single survey MS1 scan for precursor ions was followed by six data-dependent MS2 scans for precursor ions above a threshold ion count of 2,000 with collision energy of 35%. Survey MS scans were acquired in the Orbitrap component with a resolution of 30,000, and MS2 scanning were performed in the linear ion trap.

F. Peptide Identification

MS spectra were matched to peptide sequences using three search engines: SEQUEST（4）, InsPecT（5）and OMSSA（6）. Peak masses were searched

against the most current version of the Mouse Refseq Database (National Center for Biotechnology Information, NCBI). For isotopic labeling, variable modifications of +10.008269 Da for arginine and +6.020129 Da for lysine were included. Target-decoy analysis was used to limit peptide false discovery rate to <1%（7）. Peptides that matched to more than one protein were labeled as "ambiguous" using ProMatch software（8）and were not used in subsequent analyses.

G. Peptide Quantitation

Relative quantification of isotopically-labeled peptides was performed using QUIL software（9）. Protein half-lives were calculated only for proteins identified by at least 3 unique peptides. K-means clustering（10）of peak intensities was used throughout the MS run to define a background noise level filter. Based on this, peak intensities below 3500 were not considered further. Transcription factors were clustered according to the DNA-binding domain sequences (Figure 3C) using UPGMA (Unweighted Pair Group Method with Arithmetic Mean). Dendrogram was produced produced using iTOL（11）.

H. Identification of N-terminal peptides by mass spectrometry (acetylation and initiator methionine cleavage)

Separate MS experiments were done to identify N-terminal peptides in mpk-CCD (clone 11) cells. To increase the number of identifications, unlabeled cells were pre-fractionated to obtain cytoplasmic and nuclear fractions using the (NE-PER® kit; Pierce) and further separated by 1-dimensional electrophoresis (0.3 mg protein for nuclear extract and 0.750 mg protein each for nuclear pellet and cytoplasmic fractions, each loaded among multiple lanes on 10% polyacrylamide Ready Gels, BioRad, Hercules, CA). After in gel trypsinization, samples were analyzed on an LTQ-Orbitrap XL (Thermo-Fisher Scientific) interfaced with a nano-LC 1D plus system (Eksigent Technologies) as described above. Variable protein N-terminal acetylation (+42.010565 Da) searches were performed using Mascot, InsPecT, and OMSSA (*vide supra*). Peak masses were searched against the most current version of the Mouse Refseq Database (National Center for Biotechnology Information, NCBI). These searches identify N-terminal peptides with or without removal of initiator methionine, and/or acetylation of the N-terminus.

I. Generation of N-terminal sequence logos for short and long half life proteins

45-amino acid sequences aligned from the N-terminus were used. The calculations can be done for individual amino acids or for defined amino acid groups (i.e. defining hydrophobic amino acids as one group). For each individual amino acid (or group), the information content I(a,i) of the given amino acid a (or group) at the position i compares the observed probability P(a,i) to the background probability of finding that amino acid (or group) in the mouse proteome $P_{ref}(a)$（12）

$$I(a,i) \equiv P(a,i) * \log_2(P(a,i)/ P_{ref}(a)).$$

The calculation for specific amino acid groups such as shown in Figure 6 of the main paper is analogous to that for individual amino acids. The logo displayed consists of columns of characters in each position relative to the N-terminus. Each column consists of characters representing only the overrepresented residues (positive information content values). Each residue's size is proportional to its I(a,i) value, and the total column height is proportional to the sum of the favored residues' I(a,i) values. The characters are ordered from top to bottom in order of decreasing I(a,i) values. The calculations and visualization were coded in Java.

J. Identification of PEST domains in mouse proteome

The algorithm of Rogers et al.（6）for identifying PEST domains was re-coded in Java to allow batch analysis, using mouse gene symbols as input. In brief, the program searched for a region of at least 12 amino acids that contained at least one P, D or E, and S or T; and absented of positive charge amino acids (K, R, or H). Enrichment of D, E, P, S, and T [DEPST, in mass percent (w/w)] and the Kyte-Doolittle hydrophathy index (KDHI) were determined on the selected region. PEST score was then calculated from the DEPST and KDHI values as follow: PEST score = 0.55 * DEPST – 0.5 * KDHI. We used a threshold of +5.0.

II. <u>Derivation of Equations for Protein Half Life and Translation Rate Determination using Dynamic SILAC</u>

Note on Terminology: Throughout the following mathematical analysis, we use the terms "total protein" or "protein" to refer to any specific protein quantified by mass spectrometry. Thus the analysis would be assumed to apply to any specific protein successfully studied by dynamic SILAC labeling.

Initial Conditions, Assumptions, and Experimental Design: At steady state, the rate of change in total protein (X) in a cell or ensemble of cells is zero. This is mathematically stated as:

$$\frac{d\,X}{dt} = 0 \qquad \text{(Equation 1)}$$

Under the steady state condition, the rate of protein formation in the cell by translation (R_p) must be equal to the rate of protein loss which is the sum of the rate of protein degradation (R_d), and the rate of total protein lost into the medium (R_e). (Note that the latter term was found in experiments to be negligible for all but very long half life proteins, but we carry it along in the equations to maintain generality.) Equation 1 becomes:

$$\frac{dX}{dt} = R_p - R_d - R_e = 0 \qquad \text{(Equation 2)}$$

The dynamic SILAC experiments involve letting cells grow in one medium (e.g. light isotope-labeled R and K) until a steady state is reached, and switching at time zero to the second medium. (In this derivation a light-to-heavy switch will be assumed although the change was made in both directions in different experiments.) Other than the switch from light to heavy amino acids, all conditions are assumed to remain unchanged at time zero. That is, the steady state is assumed to be maintained with respect to total protein. When the medium is switched from a light to heavy amino acids, newly synthesized proteins are assumed to incorporate the newly added heavy amino acids immediately (See below for a discussion of this assumption.) The rate of change in a protein labeled with the heavy amino

acids (^hX) or light amino acids (^lX) is a function on the rates of production, degradation, and loss to the medium:

$$\frac{d^hX}{dt} = {}^hR_p - {}^hR_d - {}^hR_e \quad \text{(Equation 3)}$$

$$\frac{d^lX}{dt} = {}^lR_p - {}^lR_d - {}^lR_e \quad \text{(Equation 4)}$$

Unlike the total protein levels, however, the labeled proteins are not at steady state. It also can be noted that the rates for total protein are functions of corresponding rates for the light and heavy isotopic proteins:

$$R_p = {}^hR_p + {}^lR_p \quad \text{(Equation 5)}$$

$$R_d = {}^hR_d + {}^lR_d \quad \text{(Equation 6)}$$

$$R_e = {}^hR_e + {}^lR_e \quad \text{(Equation 7)}$$

Since the total of any given protein remains constant at steady state, and the total protein levels are a function of the heavy and light isotopic proteins, the experiment can be normalized by setting the total amount of protein to be 1 at steady state:

$$X = {}^hX + {}^lX = 1 \quad \text{(Equation 8)}$$

At the time of the medium change (t=0), the following initial conditions exist:

$$\begin{aligned} {}^hX_o &= 0 \\ {}^lX_o &= 1 \end{aligned} \quad \text{(Equation 9)}$$

There are two major assumptions in this experiment. The first assumption is that degradation of all proteins follow first-order kinetics, i.e. the rate of a degradation is proportional to its molar amount (implicitly per unit volume) in the cell and a rate constant (k_d) specific to that protein. This assumption was verified in main text of this paper. Second, it is assumed that transport of amino acids is rapid compared to the rate of incorporation into new proteins ("no recycling" assumption, see experimental verification below.) Therefore, the rate of light isotopic protein production is assumed to be zero after t=0. Incorporating these assumptions and the steady-state condition into Equation 5, the rate of total protein production is described in terms of heavy and light isotopic proteins:

$$\begin{aligned} R_p &= {}^hR_p + {}^lR_p = {}^tR_d + {}^tR_e \\ R_p &= {}^hR_p = (k_d + k_e)({}^hX + {}^lX) \end{aligned} \quad \text{(Equation 10)}$$

Formulating and Solving Balance Equations for Light and Heavy-Labeled Proteins

Equation 10 is substituted into Equation 3 to form the basic differential equation:

$$\begin{aligned} \frac{d^hX}{dt} &= {}^hR_p - {}^hR_d - {}^hR_e \\ \frac{d^hX}{dt} &= (k_d + k_e)({}^hX + {}^lX) - k_d{}^hX - k_e{}^hX \\ \frac{d^hX}{dt} &= k_d{}^hX - k_d{}^hX + k_d{}^lX + k_e{}^hX - k_e{}^hX + k_e{}^lX \\ \frac{d^hX}{dt} &= k_d{}^lX + k_e{}^lX = {}^lX(k_d + k_e) \end{aligned}$$

(Equation 11)

Equation 11 describes the formation of heavy peptide over time in terms of light peptide. However, it is more useful to have this equation in terms of heavy peptide; as shown by substituting Equation 8 into Equation 11:

$$\frac{d^hX}{dt} = {}^lX(k_d + k_e) = (1 - {}^hX)(k_d + k_e) \quad \text{(Equation 12)}$$

For simplification, allow the rate constant term $(k_d + k_e)$ to be substituted by a single term κ. Rearranging:

$$\begin{aligned} \frac{d^hX}{dt} &= (1 - {}^hX)\kappa \\ \frac{d^hX}{1 - {}^hX} &= \kappa dt \end{aligned}$$

(Equation 13)

and taking the integral with $^hX(0) = {}^hX_0$ gives:

$$1 - {}^hX = (1 - {}^hX_o)e^{-\kappa t} \quad \text{(Equation 14)}$$

Since the initial value of heavy isotopic protein is zero, this equation simplifies to:

$${}^hX = 1 - e^{-\kappa t} \quad \text{(Equation 15)}$$

From the normalization condition ($^hX + {}^lX = 1$), the amount of light isotopic protein is known:

$$^{l}X = e^{-\kappa t} \qquad \text{(Equation 16)}$$

To ensure that Equations 15 and 16 are conceptually valid, the derivative of the rate of change in heavy isotopic protein should equal the negative of the rate of change of light isotopic protein to ensure that the rate of change in total protein levels is zero:

$$\frac{d^{h}X}{dt} = e^{-\kappa t} \quad \frac{d^{l}X}{dt} = -e^{-\kappa t} \qquad \text{(Equation 17)}$$

Thus, the requirement is satisfied.

Applying Equations to the Experimental Question

We wish to determine the half-lives of all the proteins in mpkCCD$_{clone11}$ cells. The half-life of a protein is the time it takes for half of the amount of protein to disappear from the system. Mathematically, this is stated as:

At $t=t_{1/2}$, ^{h}X has the value 0.5, and

$$\frac{1}{2} = e^{-\kappa t_{1/2}} \qquad \text{(Equation 18)}$$

Rearranging for -κ:

$$-\kappa = \frac{\ln\left(\frac{1}{2}\right)}{t_{1/2}} \qquad \text{(Equation 19)}$$

The value of -κ is substituted into Equation 16 and evaluated at $t=t_s$ (the time of sampling):

$$^{l}X = e^{\frac{t_s}{t_{1/2}}\ln\left(\frac{1}{2}\right)} \qquad \text{(Equation 20)}$$

From the normalization condition, the amount of heavy isotopic protein is also known:

$$^{h}X = 1 - e^{\frac{t_s}{t_{1/2}}\ln\left(\frac{1}{2}\right)} \qquad \text{(Equation 21)}$$

Recognizing that with the SILAC method, the mass spectrometer gives values of $^{h}X/\,^{l}X$, we can generate this ratio by dividing equation 21 by equation 20. This ratio is symbolized by Ψ and is the quantity measured in the dynamic SILAC experiments. Mathematically, this gives the relationship among the half-life of a protein $t_{1/2}$, the sampling time t_s and Ψ:

$$\frac{^{h}X}{^{l}X} = \Psi = \frac{1 - e^{\frac{t_s}{t_{1/2}}\ln\left(\frac{1}{2}\right)}}{e^{\frac{t_s}{t_{1/2}}\ln\left(\frac{1}{2}\right)}}$$

$$e^{\frac{t_s}{t_{1/2}}\ln\left(\frac{1}{2}\right)} = \frac{1}{(\Psi+1)} \qquad \text{(Equation 22)}$$

$$t_{1/2} = t_s \frac{\ln\left(\frac{1}{2}\right)}{\ln\left(\frac{1}{\Psi+1}\right)}$$

To confirm the validity of Equation 22, we examine some special cases. If Ψ has a value of 0 for all $t_s>0$, indicating that the light form is never degraded and replaced by heavy, then $t_{1/2}$ goes to infinity as expected. If Ψ has a value of infinity for all $t_s>0$, indicating that all of the light form is immediately degraded and replaced by heavy, equation 22 gives a value for $t_{1/2}$ of zero. Finally, when Ψ=1 (heavy and light forms equal) at the sampling time t_s, $t_{1/2}= t_s$. That is, the appearance curve for the heavy form crosses the disappearance curve for the light form at $t_{1/2}$.

Effect of Vasopressin on Rate of Protein Production

The next objective is to formulate an equation giving the rate of protein production in terms of the degradation rate constants and Ψ. Using equations 19 and 22, we have:

$$-\kappa = \frac{1}{t_s}\ln\left(\frac{1}{\Psi+1}\right)$$

$$\kappa = \frac{1}{t_s}\ln(\Psi+1) \qquad \text{(Equation 23)}$$

Substituting in the definition of κ, the equation becomes:

$$k_d + k_e = \frac{1}{t_s}\ln(\Psi+1) \qquad \text{(Equation 24)}$$

However, at this point there are two unknowns: k_d and k_e. Re-writing in terms of k_d:

$$k_d = \frac{1}{t_s}\ln(\Psi+1) - k_e \qquad \text{(Equation 25)}$$

(Note that k_e was found experimentally to be small relative to k_d for most proteins, but we retain this term for generality.)

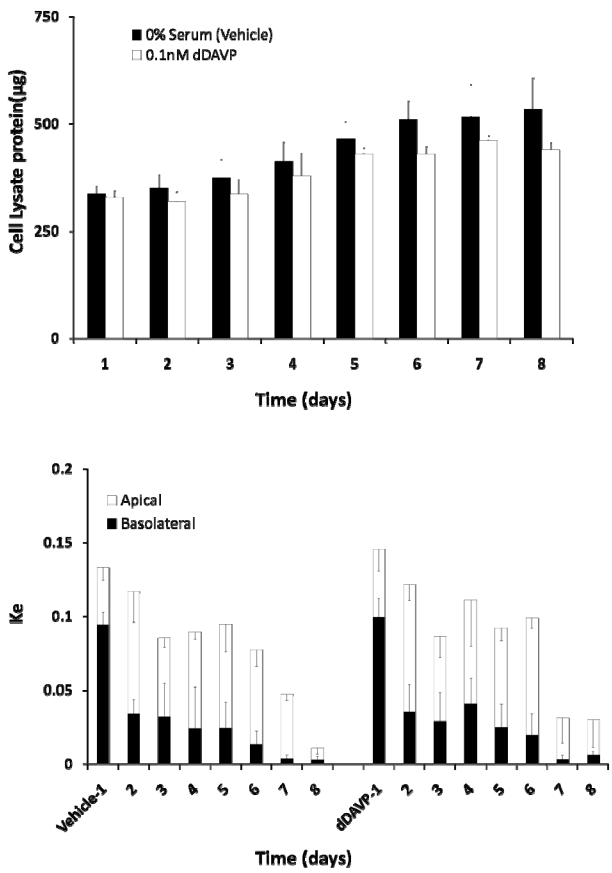
Recall that at steady state, the translation rate is

$$\mathbf{R_p = R_d + R_e = (k_d + k_e)\,X} \qquad \text{(Equation 26)}$$

for any given protein. If we consider two conditions, here dDAVP treatment and vehicle and form a ratio of the translation rates for the two, $R_{p\text{-}dDAVP}/R_{p\text{-}Vehicle}$, we obtain

$$\frac{R_{p-dDAVP}}{R_{p-Vehicle}} = \frac{(k_d + k_e)_{dDAVP}}{(k_d + k_e)_{Vehicle}} \times \frac{X_{dDAVP}}{X_{Vehicle}} \quad \text{(Equation 27)}$$

we can calculate this ratio from the measured values for k_d and k_e if we add one more piece of data for each protein, viz. $X_{dDAVP}/X_{Vehicle}$, which can be measured in additional SILAC experiments in non-dynamic mode, i.e. comparing labeled cells that are in the steady state with regard to dDAVP concentration with labeled cells that are at steady-state in the absence of dDAVP.



Legend. *Upper panel*, amounts of protein in cultured mpkCCD cells per well. *Lower panel*, amount of protein retreieved from medium after consecutive 24-hour periods normalized by amount of protein in cell lysate.

Estimate of k_e

Loss of proteins into the medium was measured daily during serum-free portion of a set of dynamic SILAC experiments extended beyond the 5-day protocol. Proteins were measured using the BCA assay as decribed above. Loss was compared to the total amount of protein in the confluent cells to calculate k_e. On day 5, the loss of protein into the medium is slightly less than 10 percent of the total protein into the medium per day. That is, it would take >10 days to lose the equivalent of the total cell mass of protein into the medium. This presumably occurs by exosome secretion, apoptosis and protein secretion. Based on this value, the average k_e for all proteins in the cells, can be estimated as ~0.1 day^{-1}. This would correspond to an apparent half life of ln(0.5)/-k_e or ~7 days if loss into the medium were the only mechanism of protein removal from cells. Since most of the proteins in mpkCCD cells were found to have half lives of <2 days, we conclude that loss of protein into the medium is unlikely to make a significant contribution to the cell turnover of these proteins. However, for proteins estimated to have very long half lifes, we assume that loss into the medium does contribute to the overall turnover.

Thus, equation 27 can be simplified as

$$\frac{R_{p-dDAVP}}{R_{p-Vehicle}} = \frac{(k_d)_{dDAVP}}{(k_d)_{Vehicle}} \times \frac{X_{dDAVP}}{X_{Vehicle}} \quad \text{(Equation 28)}$$

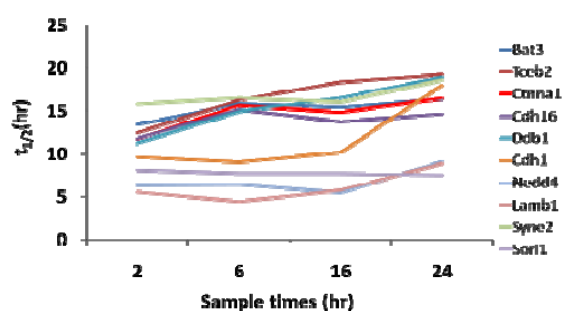
Recognizing that $t_{1/2}$ varies inversely as k_d, we can rewrite this equation as

$$\mathbf{R_{dDAVP}/R_{Veh} = (t_{1/2,Veh}/t_{1/2,dDAVP}) \cdot (X_{dDAVP}/X_{Veh})} \quad \text{(Equation 29)}$$

This equation is used for all estimates of relative translation rate given in this paper.

Justification of 'no-recycling' assumption

The model assumes that transport of lysine and arginine occurs much more rapidly than utilization for protein synthesis, maintaining equilibration with the medium in the time frame of the half-life measurements ('no recycling assumption'). Absence of equilibration would slow incorporation of the new label in a dynamic SILAC experiment and this would be expected to artificially increase apparent protein half life for measurements made at short sampling times. To test this, we plotted apparent protein half lives for several relatively short half life proteins, calculated according to the single-point method equations derived above:



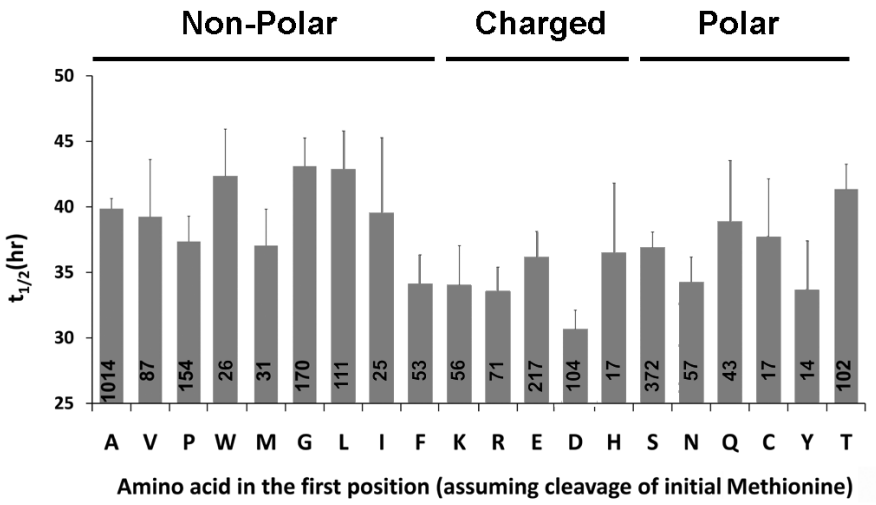
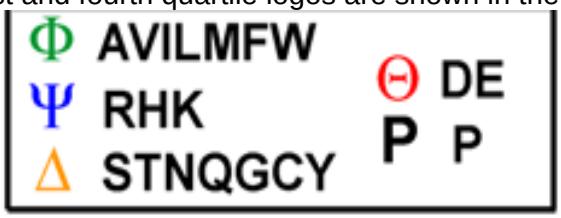
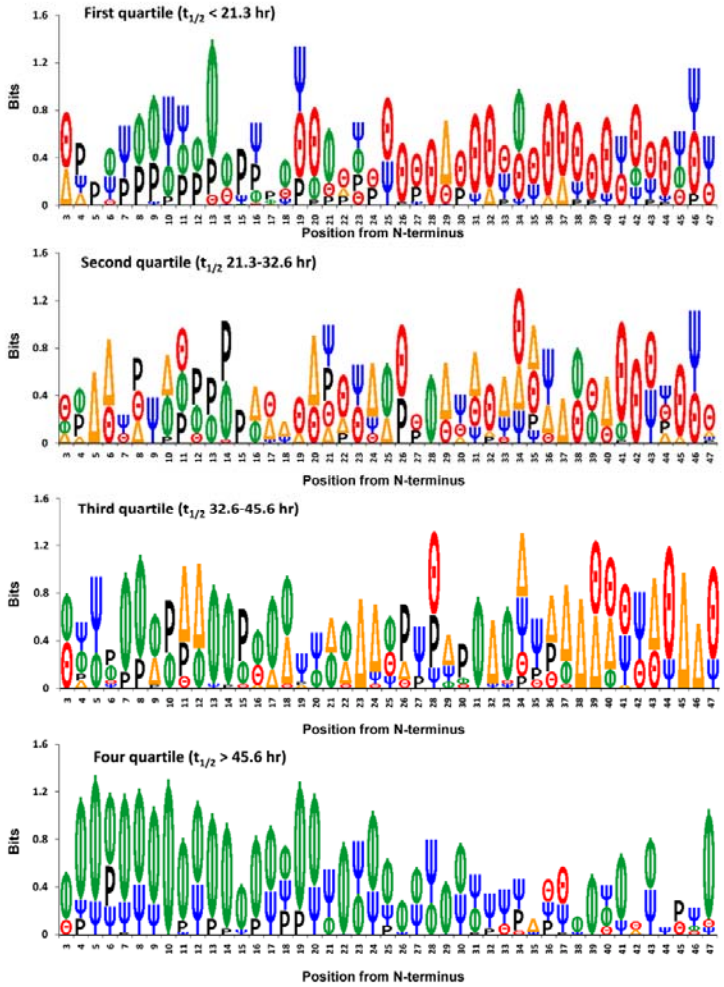
Legend. Estimates of protein half-life using the single point method for several time points inr selected proteins with short hal-life.

As seen, apparent half life is not greater at the 2-hour sampling time than for longer sampling times, providing evidence supporting the 'no recycling' assumption for 2 or more hour sampling times. A general analysis including all proteins is also consistent with this conclusion.

References for Supplement

1. Duong Van Huyen JP, Bens M, & Vandewalle A (1998) Differential Effects of Aldosterone and Vasopressin on Chloride Fluxes in Transimmortalized Mouse Cortical Collecting Duct Cells. *Journal of Membrane Biology* 164:79-90.
2. Yu MJ et al (2009) Systems-level analysis of cell-specific AQP2 gene expression in renal collecting duct. *Proc.Natl.Acad.Sci.U.S.A* 106:2441-2446.
3. Khositseth S et al (2010) Quantitative protein and mRNA profiling shows selective post-transcriptional control of protein expression by vasopressin in kidney cells. *Mol.Cell Proteomics.* 10:M110.004036
4. Eng JK, McCormack AL, & Yates JR, III (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J.Am.Soc.Mass.Spectrom.* 5:976-989.
5. Tanner S et al. (2005) InsPecT: identification of posttranslationally modified peptides from tandem mass spectra. *Anal.Chem.* 77:4626-4639.
6. Geer LY et al. (2004) Open mass spectrometry search algorithm. *J.Proteome.Res.* 3:958-964.
7. Elias JE & Gygi SP (2007) Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry 2. *Nat.Methods* 4:207-214.
8. Tchapyjnikov D et al (2010) Proteomic profiling of nuclei from native renal inner medullary collecting duct cells using LC-MS/MS. *Physiol Genomics* 40:167-183.
9. Wang G et al. (2006) Automated quantification tool for high-throughput proteomics using stable isotope labeling and LC-MSn. *Anal.Chem.* 78:5752-5761.
10. Hartigan JA & Wong MA (1979) A K-Means Clustering Algorithm. *Journal of the Royal Statistical Society* 28(:100-108.
11. Letunic I & Bork P (2011) Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Res.* 39 (Web Server issue):W475-W478.
12. Schneider TD & Stephens RM (1990) Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.* 18:6097-6100.
13. Rogers S, Wells R, & Rechsteiner M (1986) Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis. *Science* 234:364-368.

Suppl. Figure 1. Proteome-wide analysis of relationship between initial amino acid and protein half-life in renal mpkCCD (clone 11) cells. In contrast to the analysis done in the main text which used experimentally determined N-terminal sequences, here we calculated mean half-life for proteins for which valid measurements of protein half life were obtained, with the assumption that the initiator methionine had been cleaved. Based on the data in the text, this assumption should be true for about 80% of proteins. As seen, the protein half-lives tended to be greater when the first amino acid was a non-polar residue than for proteins that had a charged or polar residue in position 1.



Suppl. Fig. 2. Sequence preference logos for all four quartiles of protein half-life data. The first and fourth quartile logos are shown in the main text in Figure 6. Key: