

Letter to the Editor

Are dust mite allergens more abundant and/or more stable than other *Dermatophagoides pteronyssinus* proteins?

To the Editor:

Understanding the characteristics that differentiate allergens from nonallergens remains a perplexing question in allergy research. General features such as protein stability and abundance from source are frequently cited as likely determinative factors. However, rigorous statistical comparisons of allergens versus nonallergens on genomic and proteomic scales are lacking. In this study, the house dust mite *Dermatophagoides pteronyssinus* (DP) proteome was evaluated using RNA-seq methods as a proxy to assess the abundance of all proteins in this source. In addition, the thermodynamic stabilities of 656 nonallergens and 19 allergens were evaluated using a combined chemical denaturation and mass spectrometry approach. The results showed that when expression and stability are considered in combination, the allergens are a statistically different population from other DP proteins. The allergens are more stable *and* more highly expressed. The combination of high levels of transcription and stability for the major allergens Der p 1, Der p 2, and Der p 23 is rare (<2% of all DP proteins) but not unique.

Although it has been hypothesized that allergens are the abundant and/or stable proteins from an allergen's source, experimental evidence to support or refute such hypotheses is lacking. To address the question of allergen abundance, the transcriptome of the house dust mite DP was analyzed with RNA-seq technologies (see this article's Online Repository at www.jacionline.org). All 19 DP allergens in the official Allergen Nomenclature database of the World Health Organization and the

International Union of Immunological Societies (www.allergen.org) were identified in the transcriptome. Fig 1 shows the relative abundance (fragments per kilobase per million reads [FPKM]) of the DP allergen transcripts compared with DP as a whole. Clearly, the allergens as a group are more highly expressed than the norm. In fact, many are outliers in the quartile analysis of the DP transcriptome. However, there are still hundreds of other transcripts with similar abundance. There is a positive correlation of 0.4 for protein abundance versus RNA transcription levels in a well-studied murine system.¹ This correlation indicates that using RNA-seq to infer protein levels is not rigorously accurate, but the general conclusions for the levels of allergens and nonallergens are likely valid. Relative transcription levels for the allergens were similar to homologues in *D. farinae* ($r = 0.59$).²

The stability of allergens is frequently addressed in the context of food allergens and generally refers to resistance to low pH and digestive enzymes. Recent studies have suggested that the stability of Bet v 1 to endolysosomal digestion in dendritic cells is important for sensitization.³ Here, a mass spectrometry-based proteomics technique was used to evaluate the chemical denaturant-induced equilibrium-unfolding properties of 656 DP proteins, and the 19 DP allergens all derived from mite extract, except for a recombinant Der p 23.⁴ Connecting the protein identifications with RNA-seq data shows that the nonallergen proteins span a wide range of RNA expression values, and trend toward the more highly expressed proteins (Fig 1, A-C). The stabilities of the DP proteins are reported as the concentration of guanidinium chloride at the transition midpoint of the chemical denaturation curve ($GND_{1/2}$), which is directly related to thermodynamic stability (Fig 1, D and E). More stable proteins have higher $GND_{1/2}$ values. Based on testing with purified Der p 1, the $GND_{1/2}$ values of the protease allergens (groups 1, 3, 6,

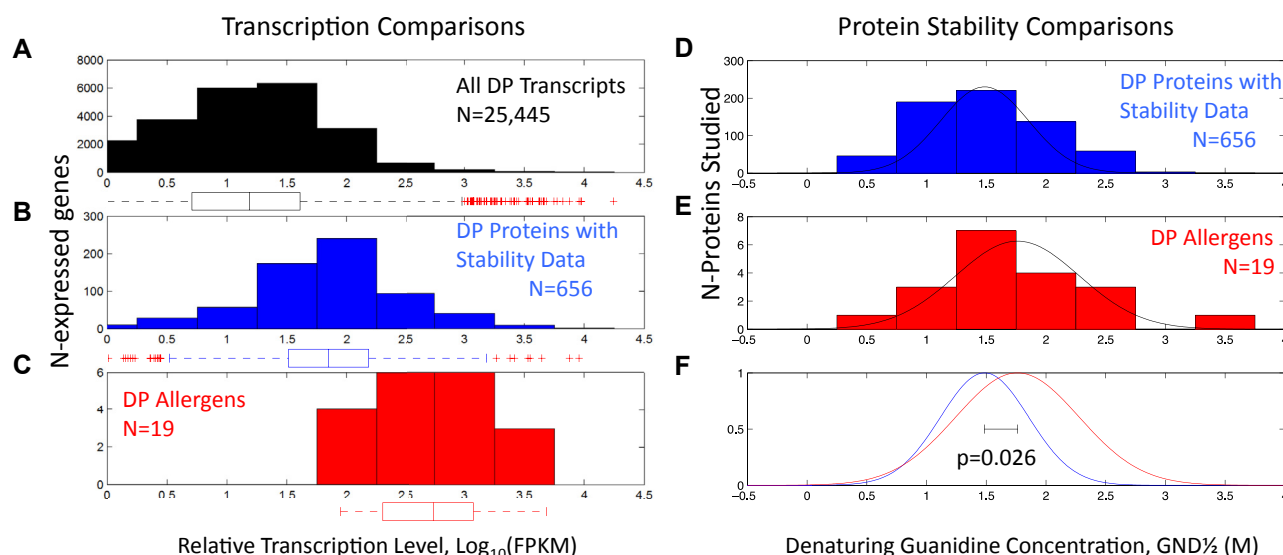


FIG 1. Analysis of DP proteome and protein stability measurements. Histograms and box plots of relative expression (FPKM) of all DP transcripts (A), DP proteins studied for stability (B), and DP allergens (C). Panels D and E show histograms of the stability data for nonallergens and allergens, respectively, overlaid with Gaussian fits to the distribution. Panel F compares the Gaussian fit of stability data for the nonallergens (blue) with the allergens (red).

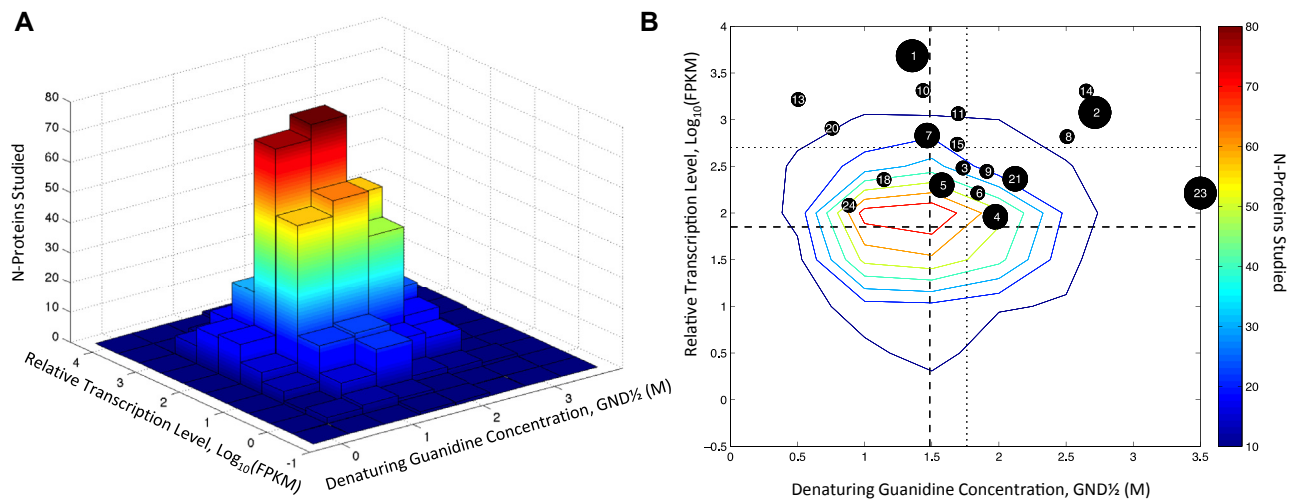


FIG 2. Two-dimensional histograms of stability and relative expression for DP. Data are presented as a bar chart (A) or as a contour plot (B). Numerical labels indicate allergen groups. Three black circle sizes (large, medium, and small) correspond to the allergen importance rankings of major, mid-tier, and minor/"needs further study," respectively.⁵ Dashed and dotted lines are the means for the nonallergen and allergen groups, respectively.

and 9) were adjusted lower by 0.4 to account for additional stability from added protease inhibitors (see this article's Online Repository at www.jacionline.org). A comparison of the set of allergens ($GND_{1/2} = 1.8 \pm 0.7$) to the nonallergens ($GND_{1/2} = 1.5 \pm 0.5$) shows that the allergens trend toward greater stability. These distributions exhibit significant overlap, but a *t* test reveals that they are statistically different ($P = .026$; Fig 1, F).

The 2-dimensional histogram in Fig 2, A, shows that there is a Gaussian-like distribution of abundance and stability for the 656 DP proteins. Fig 2, B, shows the same data as a contour plot and adds the abundance and stability data for the set of allergens for comparison. The allergens cluster primarily in the quadrant corresponding to high stability and high abundance. Hotelling's T^2 statistic was used to test whether the allergen group differed from the nonallergen group with respect to $GND_{1/2}$ and $\log$ (FPKM) simultaneously.⁶ The difference is significant ($P < .0001$), where the value of T^2 is 34.8, which is equivalent to an *F* statistic of 17.7 with 2 and 672 degrees of freedom.

Considering the major allergens, Der p 1, Der p 2, and Der p 23, the stability of Der p 1 is slightly less than the mean for DP proteins, but it is exceptionally highly transcribed. In contrast, Der p 23 is not as highly expressed but extremely stable, whereas the measurements on Der p 2 appear in between the other 2. The major allergens were analyzed to see whether they were substantially different from the rest of the allergens. They are all reasonably close to the group of allergens, falling within the 85th percentile of the distribution of allergens. However, they are significantly distinct from the group of nonallergens. The probability of obtaining the Der p 1 $GND_{1/2}$ and $\log$ (FPKM) values for proteins in the nonallergen population is 0.015; for Der p 2, the probability is 0.0002; and for Der p 23, the probability is less than 0.0001. To reiterate, the properties of the allergens are rare, but considering the size of the DP transcriptome they are not unique.

The results of this study have important implications for understanding the factors that lead to sensitization. Other influences on sensitization include additional environmental

exposures, adjuvants (natural⁷ and anthropogenic⁸), genetic susceptibility,⁹ and a modern hygienic lifestyle.¹⁰ A limitation of this protocol is that these studies were performed with whole mite extract to sample the maximum number of RNA transcripts and proteins for stability assays. Because allergens are found in both whole bodies and fecal pellets,¹¹ human exposures are at least partially represented in the whole body extracts used here. Future studies will attempt to quantify the amount of DP proteins in dust mite fecal particles and samples of house dust to better compare the allergens and nonallergens to which humans are exposed. Understanding the biophysical properties of the mite proteins involved in human exposure will lead to a better understanding of why the human immune system consistently develops an allergic response to less than 0.1% of the DP proteome.

We thank the National Institutes of Health Intramural Sequencing Center, Comparative Sequence Group, for expert technical assistance with high-throughput sequencing. We thank Marjorie S. Morgan, PhD, for preparation of the extracts and DiAnn L. Vyszynski-Moher, MS, for assistance with culturing the mites.

Ryenne N. Ogburn, BS^a
 Thomas A. Randall, PhD^b
 Yingrong Xu, BS^a
 Julia H. Roberts, BS^a
 Betelihem Mebrahtu^a
 Jaret M. Karnuta^a
 S. Dean Rider, PhD^c
 Grace E. Kissling, PhD^b
 Robert E. London, PhD^b
 Anna Pomés, PhD^d
 Larry Arlian, PhD^c
 Michael C. Fitzgerald, PhD^d
 Geoffrey A. Mueller, PhD^b

From ^athe Chemistry Department, Duke University, Durham, NC; ^bthe Intramural Program, National Institute of Environmental Health Sciences, Research Triangle Park, NC; ^cWright State University, Fairborne, Ohio; and ^dIndoor Biotechnologies, Charlottesville, Va. E-mail: Geoffrey.Mueller@nih.gov.

This study was supported in part by the Intramural Research Program of the National Institute of Environmental Health Sciences (grant no. Z01-ES102906-01 to R.E.L.), the Extramural Research Program of the National Institute of General

Medical Sciences (grant no. 2R01-GM08174 to M.C.F.), the National Institute of Allergy and Infectious Diseases of the National Institutes of Health (grant no. R01AI077653 to A.P.), and the National Science Foundation (grant no. CHE-1308093 to M.C.F.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Disclosure of potential conflict of interest: R. N. Ogburn, Y. Xu, and M. C. Fitzgerald have received support from the National Science Foundation and the National Institutes of Health. J. H. Roberts has received a grant from the National Institutes of Health. B. Mehbratu has received grants from the National Science Foundation and North Carolina American Chemical Society Project Science Educational Experience for the Disadvantaged. A. Pomés has received a grant from the National Institutes of Health/National Institute of Allergy and Infectious Diseases (R01 Award) and is employed by Indoor Biotechnologies. The rest of the authors declare that they have no relevant conflicts of interest.

REFERENCES

1. Vogel C, Marcotte EM. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat Rev Genet* 2012;13:227-32.
2. Chan TF, Ji KM, Yim AK, Liu XY, Zhou JW, Li RQ, et al. The draft genome, transcriptome, and microbiome of *Dermatophagoides farinae* reveal a broad spectrum of dust mite allergens. *J Allergy Clin Immunol* 2015;135:539-48.
3. Machado Y, Freier R, Scheiblhofer S, Thalhamer T, Mayr M, Briza P, et al. Fold stability during endolysosomal acidification is a key factor for allergenicity and immunogenicity of the major birch pollen allergen. *J Allergy Clin Immunol* 2016;137:1525-34.
4. Xu Y, Strickland EC, Fitzgerald MC. Thermodynamic analysis of protein folding and stability using a tryptophan modification protocol. *Anal Chem* 2014;86:7041-8.
5. Thomas WR. Hierarchy and molecular properties of house dust mite allergens. *Allergol Int* 2015;64:304-11.
6. Morrison DF. *Multivariate statistical methods*, 2nd ed. New York: McGraw-Hill; 1976.
7. Gilles-Stein S, Traidl-Hoffmann C. Pollen are more than allergen carriers. *Allergologie* 2016;39:69-76.
8. Baldacci S, Maio S, Cerrai S, Sarno G, Baiz N, Simoni M, et al. Allergy and asthma: effects of the exposure to particulate matter and biological allergens. *Resp Med* 2015;109:1089-104.
9. Bonnelykke K, Sparks R, Waage J, Milner JD. Genetics of allergy and allergic sensitization: common variants, rare mutations. *Curr Opin Immunol* 2015;36:115-26.
10. Platts-Mills TA. The allergy epidemics: 1870-2010. *J Allergy Clin Immunol* 2015;136:3-13.
11. Batard T, Baron-Bodo V, Martelet A, Le Mignon M, Lemoine P, Jain K, et al. Patterns of IgE sensitization in house dust mite-allergic patients: implications for allergen immunotherapy. *Allergy* 2016;71:220-9.

<http://dx.doi.org/10.1016/j.jaci.2016.08.016>

METHODS

DP mite material

DP was cultured at room temperature on diet A as described.^{E1} As food became depleted, mites of all active life stages that migrated onto the lids were collected by aspiration onto a 38-mm stainless steel mesh and were killed by freezing. Once food was completely consumed, the spent culture was frozen to kill the mites. Both mites and spent culture were lyophilized before use.

DP aqueous extract

Aqueous mite extract was prepared from the mite material described above using methods adapted from Arlian and Morgan.^{E2} The extract was prepared in sterile Dulbecco's PBS without protease inhibitors and frozen before the analysis below.

DP nucleic acids

Mites were collected as indicated above from culture lids and surface sterilized as described for scabies mites.^{E3} Genomic DNA was isolated from living mites of all active stages using the Wizard SV genomic DNA purification system. To enable proper digestion of mite tissues using this kit, mites were ground in digestion buffer using a Dounce homogenizer.

RNA was isolated from living surface-sterilized mites using the Direct-zol RNA MiniPrep w/TRI-reagent. The sample prep was similar to that for DNA with approximately 50 mg of live mites being ground in 500 μ L of TRI Reagent on ice.

RNA-seq analysis

A de novo reconstruction of DP RNA-seq data as described^{E4} was used for the prediction of open reading frames with Transdecoder 2.0.1.^{E5} A predicted proteome of 26,692 proteins was generated, and of these, 25,446 had an FPKM of 1.0 or more; these were used in our analysis. This compares to 16,376 genes reported for the draft of the *D. farinae* genome.^{E6}

Supplement to denaturation analysis

The mass spectrometry-based proteomics analyses performed here identified a total of 1579 proteins from the predicted DP proteome. Based on the predicted expression levels, the proteins span almost 4 orders of magnitude of FPKM values. In addition, all 19 DP allergens officially recognized by the World Health Organization and the International Union of Immunological Societies were observed. (Including all species of mites, there are currently 36 groups of mite allergens in the allergen nomenclature database of the World Health Organization/the International Union of Immunological Societies.) Eighteen of the DP allergens were studied for their denaturation behavior from extract including the major allergens Der p 1 and Der p 2. The other major allergen Der p 23 was not detected in the proteomic analysis of the extract but the chemical denaturation behavior of a purified recombinant protein construct of Der p 23^{E4} was characterized using a tryptophan-modification protocol similar to that used in the proteomic analysis. One of the 2 tryptophan residues in Der p 23 remained unreacted even at high denaturant concentrations (eg, 7 mol/L GND), suggesting that the protein is remarkably resistant to chemical denaturation. Thus, for the statistical analyses performed in this work, Der p 23 was assigned the highest level of GND $\frac{1}{2}$, 3.5 mol/L, assayed in the proteomewide analysis. Recombinant Der p 1 expressed in *Pichia pastoris* and Der p 5 expressed in *Escherichia coli*^{E7} were compared with the natural allergen in the lysate to verify the methodology. The GND $\frac{1}{2}$ value determined for recombinant Der p 5 using a methionine-modification protocol^{E8} was similar to that measured in the lysate, but Der p 1 was more stable in the lysate analysis. As a known cysteine protease, we reasoned that the increased stability of Der p 1 in the

lysate analysis was due to the presence of protease inhibitors that were added to the lysate sample before the proteomewide analysis (see below). Indeed, an analysis of the recombinant Der p 1 in the presence of the protease cocktail described below revealed a 0.4 mol/L stabilization.

The chemical denaturant-induced equilibrium-unfolding properties of the proteins in the mite extract (prepared in sterile Dulbecco's PBS) were evaluated using the so-called hybrid SPROX protocol, which has been previously described.^{E9} Aliquots (20 μ L) of 5.7 mg/mL mite extract containing 1 mmol/L 4-(2-aminoethyl)benzenesulfonyl fluoride (AEBSF), 0.01 mmol/L pepstatin A, 0.02 mmol/L leupeptin, 0.015 mmol/L E-64, and 0.05 mmol/L bestatin (ie, the so-called protease inhibitor cocktail) were incubated in increasing concentrations of guanidinium chloride (0, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, and 3.5 mol/L) for 2 hours before initiating the chemical modification reactions in the hybrid SPROX protocol, which included the hydrogen peroxide-mediated oxidation of methionine residues and the dimethyl (2-hydrogen-5-nitrobenzyl)sulfonium bromide modification of tryptophan residues. The chemical modification reactions and subsequent quantitative bottom-up proteomics analyses were performed as previously described, except that the chemically modified methionine- and tryptophan-containing peptides were not used in the analysis.^{E9} GND $\frac{1}{2}$ values were determined only for the wild-type methionine- and wild-type tryptophan-containing peptide probes assayed in the hybrid SPROX protocol. This helped eliminate potential artifactual stability changes that the chemical modification might introduce. The GND $\frac{1}{2}$ values were determined using a JAVA-based program, developed in-house, to fit the chemical denaturation data sets to a 4-parameter sigmoidal equation using a nonlinear least squares analysis. In some cases, multiple GND $\frac{1}{2}$ values were determined for a given DP protein because multiple methionine- and/or tryptophan-containing peptide probes from different regions of the protein were assayed. The distributions of GND $\frac{1}{2}$ values shown in Fig 1, D-F, were generated using only 1 GND $\frac{1}{2}$ value per protein. In cases in which multiple GND $\frac{1}{2}$ values were generated for a given protein (eg, the stabilities of multiple folding domains within the protein were determined), the largest GND $\frac{1}{2}$ value detected for the protein was used in statistical analyses. The lysate-determined GND $\frac{1}{2}$ values for Der p 1, 3, 6, and 9 (all proteases) were also adjusted to account for the artifactual stability increase due to the presence of the protease inhibitor cocktail. Based on results obtained for Der p 1 (see above), the measured GND $\frac{1}{2}$ values were reduced by 0.4 mol/L.

REFERENCES

- Avula-Poola S, Morgan MS, Arlian LG. Diet influences growth rates and allergen and endotoxin contents of cultured *Dermatophagoides farinae* and *Dermatophagoides pteronyssinus* house dust mites. *Int Arch Allergy Immunol* 2012;159:226-34.
- Arlian LG, Morgan MS. Immunomodulation of skin cytokine secretion by house dust mite extracts. *Int Arch Allergy Immunol* 2011;156:171-8.
- Morgan MS, Arlian LG, Markey MP. *Sarcoptes scabiei* mites modulate gene expression in human skin equivalents. *Plos One* 2013;8:e71143.
- Mueller GA, Randall TA, Glesner J, Pedersen LC, Perera L, Edwards LL, et al. Serological, genomic and structural analyses of the major mite allergen Der p 23. *Clin Exp Allergy* 2016;46:365-76.
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 2011;29:644-52.
- Chan TF, Ji KM, Yim AK, Liu XY, Zhou JW, Li RQ, et al. The draft genome, transcriptome, and microbiome of *Dermatophagoides farinae* reveal a broad spectrum of dust mite allergens. *J Allergy Clin Immunol* 2015;135:539-48.
- Mueller GA, Gosavi RA, Krahn JM, Edwards LL, Cuneo MJ, Glesner J, et al. Der p 5 crystal structure provides insight into the group 5 dust mite allergens. *J Biol Chem* 2010;285:25394-401.
- West GM, Tang L, Fitzgerald MC. Thermodynamic analysis of protein stability and ligand binding using a chemical modification- and mass-spectrometry based strategy. *Anal Chem* 2008;80:4175-85.
- Xu Y, Strickland EC, Fitzgerald MC. Thermodynamic analysis of protein folding and stability using a tryptophan modification protocol. *Anal Chem* 2014;86:7041-8.