

plicated in Alzheimer's disease (Fig. 3B, cytoskeletal/neuronal cluster). Most of the other known genes in this cluster are implicated in neuronal pathfinding and cell adhesion, including *E-cadherin*, which encodes a protein associated with the presenilin complex (28), and *Notch*, which encodes a substrate of the presenilin complex (29, 30). The cluster of 21 genes is enriched for components and substrates of the presenilin complex.

These data (24) provide an overview of gene expression profiles during *Drosophila* development. An unusually high proportion of the genes are developmentally regulated, but of 4028 genes analyzed, only 903 are previously named *Drosophila* genes with a known mutant phenotype, biochemical function, or protein homology. Fifty-one percent of the genes fall into 50 clusters with correlation coefficients greater than 0.80 (for an annotated hierarchical cluster, see fig. S7, green bars). Virtually all the clusters contain genes with known or predicted roles in development or physiology, and genes to which a biochemical or cellular function has been assigned by the GO project (12) [all genes in these clusters are listed in the online database (24)]. A large number of the clusters contain genes that are used together in specific developmental or biochemical processes. On the basis of their developmental expression patterns, we have tentatively assigned 53% of the genes to a developmental or biological functional category (for example, male germ line, female germ line, eye, muscle, early zygotic, biochemical complex, or cell biology function).

In addition to providing functional annotation of the *Drosophila* genome, these studies are a step toward a complete description of the genetic networks that control development.

References and Notes

1. L. S. Levy, J. E. Manning, *Dev. Biol.* **85**, 141 (1981).
2. M. Grunstein, D. S. Hogness, *Proc. Natl. Acad. Sci. U.S.A.* **72**, 3961 (1975).
3. W. Bender, P. Spierer, D. S. Hogness, *J. Mol. Biol.* **168**, 17 (1983).
4. J. L. DeRisi, V. R. Iyer, P. O. Brown, *Science* **278**, 680 (1997).
5. M. Schena, D. Shalon, R. W. Davis, P. O. Brown, *Science* **270**, 467 (1995).
6. K. P. White, S. A. Rifkin, P. Hurban, D. S. Hogness, *Science* **286**, 2179 (1999).
7. M. D. Adams *et al.*, *Science* **287**, 2185 (2000).
8. Materials and methods are available as supporting material on Science Online.
9. M. B. Eisen, P. T. Spellman, P. O. Brown, D. Botstein, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 14863 (1998).
10. P. Tamayo *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 2907 (1999).
11. M. Akam, *Development* **101**, 1 (1987).
12. M. Ashburner *et al.*, *Nature Genet.* **25**, 25 (2000).
13. M. Bate, in *The Development of Drosophila melanogaster*, A. M. A. Michael Bate, Ed. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1993), vol. II, pp. 1013–1090.
14. B. L. Black, E. N. Olson, *Annu. Rev. Cell. Dev. Biol.* **14**, 167 (1998).
15. D. Schmucker *et al.*, *Cell* **101**, 671 (2000).
16. R. E. Boswell, A. P. Mahowald, *Cell* **43**, 97 (1985).
17. D. Lindsley, K. T. Tokuyasu, in *Genetics and Biology of*

- Drosophila*, M. Ashburner, T. R. Wright, Eds. (Academic Press, New York, 1980), pp. 225–294.
18. A. C. Spradling, in *The Development of Drosophila melanogaster*, A. M. A. Michael Bate, Ed. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1993), pp. 1–70.
19. T. Hazelrigg *et al.*, *Genetics* **126**, 607 (1990).
20. E. Gateff, *Prog. Clin. Biol. Res.* **85** (part B), 621 (1982).
21. T. Barnett, C. Pachel, J. P. Gergen, P. C. Wensink, *Cell* **21**, 729 (1980).
22. M. J. Bertram, D. M. Neubaum, M. F. Wolfner, *Insect Biochem. Mol. Biol.* **26**, 971 (1996).
23. C. S. Zuker, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 571 (1996).
24. The data are available at <http://flygenome.yale.edu/Lifecycle>.
25. G. M. Rubin *et al.*, *Science* **287**, 2204 (2000).
26. L. T. Reiter, L. Potocki, S. Chien, M. Gribskov, E. Bier, *Genome Res.* **11**, 1114 (2001).
27. A complete annotated listing of the transcript profiles of human disease gene homologs is available in the database supplement (24).
28. L. Baki *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 2381 (2001).
29. G. Struhl, I. Greenwald, *Nature* **398**, 522 (1999).
30. Y. Ye, N. Lukinova, M. E. Fortini, *Nature* **398**, 525 (1999).
31. T. Lecuit, R. Samanta, E. Wieschaus, *Dev. Cell* **2**, 425 (2002).

32. We thank Berkeley *Drosophila* Genome Project/Howard Hughes Medical Institute (HHMI) expressed sequence tag (EST) sequencing project and Research Genetics for providing the EST library, P. Lem and G. Gibson for resequencing the EST library, C. Fan for technical assistance, the Minx Fuller lab for testes RNA, T. Jones for assistance in EST data management and analysis, and I. SanGil for database support. M.P.S., E.E.M.F., and B.H.N. were supported by a Defense Advanced Research Projects grant. M.A.K. and M.P.S. are investigators of the HHMI. F.I. was supported by an NIH Medical Scientist Training Program fellowship. B.S.B. and M.A.K. acknowledge support from National Institute of General Medical Studies (NIGMS) and National Institute of Neurological Disorders and Stroke. M.N.A. acknowledges support from NIGMS. R.W.D. is supported by the National Human Genome Research Institute (NHGRI). Supported by a grant from the NHGRI to K.P.W.

Supporting Online Material

www.sciencemag.org/cgi/content/full/297/5590/2270/DC1

Materials and Methods

Figs. S1 to S7

Tables S1 to S30

29 April 2002; accepted 8 August 2002

Structural Basis for Gluten Intolerance in Celiac Sprue

Lu Shan,¹ Øyvind Molberg,⁵ Isabelle Parrot,¹ Felix Hausch,¹ Ferda Filiz,¹ Gary M. Gray,² Ludvig M. Sollid,⁵ Chaitan Khosla^{1,3,4*}

Celiac Sprue, a widely prevalent autoimmune disease of the small intestine, is induced in genetically susceptible individuals by exposure to dietary gluten. A 33-mer peptide was identified that has several characteristics suggesting it is the primary initiator of the inflammatory response to gluten in Celiac Sprue patients. In vitro and in vivo studies in rats and humans demonstrated that it is stable toward breakdown by all gastric, pancreatic, and intestinal brush-border membrane proteases. The peptide reacted with tissue transglutaminase, the major autoantigen in Celiac Sprue, with substantially greater selectivity than known natural substrates of this extracellular enzyme. It was a potent inducer of gut-derived human T cell lines from 14 of 14 Celiac Sprue patients. Homologs of this peptide were found in all food grains that are toxic to Celiac Sprue patients but are absent from all nontoxic food grains. The peptide could be detoxified in in vitro and in vivo assays by exposure to a bacterial prolyl endopeptidase, suggesting a strategy for oral peptidase supplement therapy for Celiac Sprue.

Celiac Sprue (also known as Celiac disease or gluten-sensitive enteropathy) is an autoimmune disease of the small intestine caused by the ingestion of gluten proteins from widely prevalent food sources such as wheat, rye, and barley. In many human leukocyte antigen (HLA) DQ2 (or DQ8)-positive individuals, exposure of the small intestine to gluten in-

duces an inflammatory response, leading to destruction of the villous structure of the intestine (1–3). It commonly appears in early childhood, with severe symptoms including chronic diarrhea, abdominal distension, and failure to thrive. In many patients, symptoms may not develop until later in life, when the disease symptoms include fatigue, diarrhea, and weight loss due to malabsorption, anemia, and neurological symptoms. Celiac Sprue is a life-long disease, and if untreated it is associated with increased morbidity and mortality (4, 5). Despite its high prevalence in most population groups (>1:200) and serious manifestations, the only effective therapy is strict dietary abstinence from these food grains.

¹Department of Chemical Engineering, ²Department of Medicine, ³Department of Chemistry, and ⁴Department of Biochemistry, Stanford University, Stanford, CA 94305–5025, USA. ⁵Institute of Immunology, Rikshospitalet, University of Oslo, N-0027 Oslo, Norway.

*To whom correspondence should be addressed. E-mail: ck@chemeng.stanford.edu

REPORTS

The principal toxic components of wheat gluten are a family of closely related Pro- and Gln-rich proteins called gliadins (6, 7). However, given the enormous biological diversity and unusual chemistry of gluten proteins, and the absence of satisfactory assays for gluten toxicity, the structural basis for gluten toxicity in Celiac Sprue remains unclear (2). Notwithstanding the heterogeneity of T cell epitopes in gluten, a few epitopes appear to account for most of the α -gliadin-specific recognition by CD4⁺ T cells from patients (8, 9). These peptides are also substrates of tissue transglutaminase (tTGase) (10, 11), and the products of this enzymatic reaction bind to the HLA DQ2 molecule (12, 13).

To identify the physiologically stable regions of gliadin, recombinant α 2-gliadin, a representative α -gliadin (14), was digested with gastric and pancreatic enzymes and analyzed by liquid chromatography coupled with tandem mass spectroscopy and ultraviolet spectroscopy (LC-MS/MS/UV) (15) (Fig. 1). The most noteworthy of the digestive products was a relatively large fragment, the 33-mer LQLQPFPPQQLPYPQPQLPYPQPQLPYPQPQP (residues 57 to 89) (16). This peptide was of particular interest for two reasons: (i) Whereas most other relatively stable proteolytic fragments were cleaved to smaller fragments when the reaction times were extended, the 33-mer peptide remained intact despite prolonged exposure to proteases. (ii) Three distinct patient-specific T cell epitopes identified previously in T cell proliferation assays (8, 17) are present in this peptide, namely, PFPQPQLPY, PQQQLPYPQ (three copies), and PYPQPQLPY (two copies). In addition to this Gln- and Pro-rich

33-mer, the peptide WQIPEQSR was also identified and was used as a control in many of the studies that followed. Under similar proteolytic conditions, myoglobin (another common dietary protein) is rapidly broken down into much smaller products (18). No long intermediate is observed to accumulate.

The small intestinal brush-border membrane (BBM) enzymes are vital for breaking down any remaining peptides from gastric or pancreatic digestion into amino acids, dipeptides, or tripeptides (19). BBM fractions were prepared from rat small intestinal mucosa (20). The specific activities of known BBM peptidases were verified to be within the previously reported range (21). Whereas the half-life of disappearance of WQIPEQSR was \sim 60 min in the presence of BBM protein (12 ng/ μ l), the half-life of LQLQPFPPQQLPYPQPQLPYPQPQLPYPQPQP digestion was $>$ 20 hours (18). Therefore, the latter peptide must remain intact throughout the digestive process in the stomach and upper small intestine and is poised to act as a potential antigen for T cell proliferation and intestinal toxicity in genetically susceptible individuals.

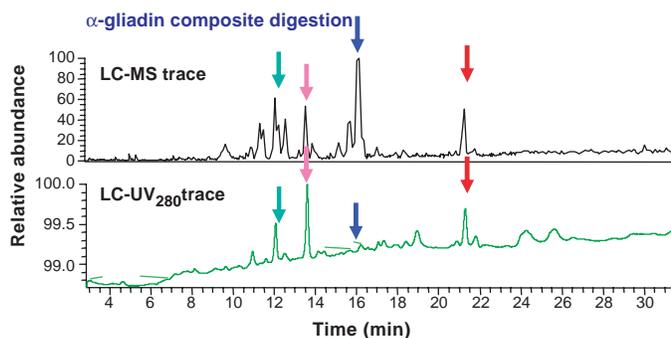
To validate the initial findings with rat BBM preparations in human small intestine, we prepared small intestinal biopsy material taken as part of the care of five individuals, one of whom carried the diagnosis of Celiac Sprue and was in remission. The other four patients proved to have normal intestinal histology. LQLQPFPPQQLPYPQPQLPYPQPQLPYPQPQP, QLQPFPPQQLPY (an internal sequence from the 33-mer used as a control), WQIPEQSR, and other control peptides (100 μ M) were incubated with BBM

prepared from each human biopsy (final aminopeptidase N activity \sim 13 μ U/ μ l, total protein \sim 1 μ g/ μ l) at 37°C for varying time periods. Although LQLQPFPPQQLPY, WQIPEQSR, and other control peptides were nearly completely proteolyzed within 1 to 5 hours, the 33-mer peptide remained largely intact for at least 15 hours (Table 1).

The proteolytic resistance of the 33-mer gliadin peptide, observed in vitro with BBM from rats and humans, was confirmed in vivo by a perfusion protocol in intact adult rats (22). Whereas $>$ 90% of LQLQPFPPQQLPY was proteolyzed in the perfusion experiment, the 33-mer gliadin peptide was highly resistant to digestion (Fig. 2). These results demonstrate that the 33-mer peptide is very stable when it is exposed to the BBM of the mammalian upper small intestine.

Regiospecific deamidation of immunogenic gliadin peptides by tTGase increases their affinity for HLA DQ2 as well as the potency with which they activate patient-de-

Fig. 1. Products of gastric plus pancreatic protease mediated digestion of α 2-gliadin under physiological conditions. Analysis was performed by reverse-phase liquid chromatography coupled with electrospray ionization mass spectroscopy (LC-ESIMS). The longest peptides are highlighted by arrows coded to indicate the sequence of α 2-gliadin (bottom right). Although pepsin-catalyzed cleavage of the NH₂-terminal Leu residue of the red-coded 33-mer was observed only in prolonged incubations, this residue was included in the later analysis because it stabilized the peptide from noncatalytic conversion into pyro-Gln peptide. (Pyroglutamination readily occurs during both synthesis and biochemical assays involving peptides with NH₂-terminal Gln residues.)



α -gliadin composite digestion

MVRVFPVQLQPNPSQQQVPLVQ
 QQQFPQQQFPQQPYPQPQFPSSQ
 PYLQLQPFPPQQLPYPQPQLPYPQPQL
 PYPQPQFPFRPQQPYPQSQPQYSQPQQP
 ISQQQQQQQQQQKQQQQQQQILQQ
 ILQQQLIPCRDVLVQHSIAYGSSQVL
 QQSTYQLVQQLCCQQLWQIPEQSRCA
 IHNVAHAIILHQQQQQQQQQPLSQ
 VSFQQPQQQYPSGQGSFQPSQQNPQAQ
 GSVQPQQLPQFEEIRNLALETLPAMCN
 VYIIPPYCTIAPVGFGTNYR

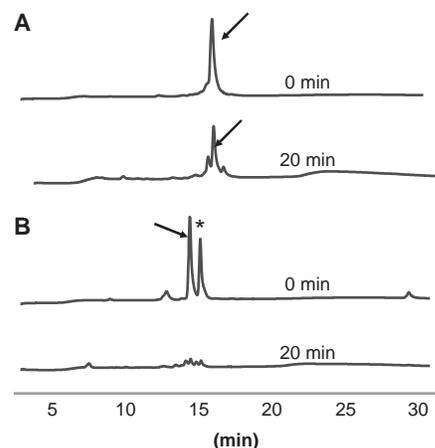


Fig. 2. In vivo BBM digestion of peptides. (A) Reverse-phase liquid chromatography at UV 215 nm (RPLC-UV₂₁₅) trace of 25 μ M of LQLQPFPPQQLPYPQPQLPYPQPQP (shown by arrow) before perfusion and after perfusion of a 20-cm segment of upper small intestine (residence time = 20 min). (B) RPLC-UV₂₁₅ trace of 50 μ M of QLQPFPPQQLPY (shown by arrow) before perfusion and after perfusion of the 20-cm segment (residence time = 20 min). *, pyroLQLQPFPPQQLPY.

Table 1. Human BBM catalyzed digestion (%) (over 15 hours) of LQLQPFPPQQLPYPQPQLPYPQPQP ("33-mer"), QLQPFPPQQLPY ("Control A"), and WQIPEQSR ("Control B") derived from a panel of adult biopsies. Asterisk indicates participant diagnosed with Celiac Sprue.

	33-mer	Control A	Control B
H1	<20	90	90
H2	<20	61	85
H3	<20	87	95
H4*	<20	96	95
H5	<20	96	95

