

30. T. E. Cravens, *J. Geophys. Res.* **92**, 11083 (1987).
 31. V. R. Eshleman *et al.*, *Science* **204**, 976 (1979);
 ———, *ibid.* **206**, 959 (1979); J. H. Waite *et al.*,
 paper presented at 31st COSPAR Scientific Assembly,
 Birmingham, UK, 14 to 21 July 1996.
 32. Th. Encenaz *et al.*, *Astron. Astrophys.* **315**, L397
 (1996).
 33. A. Seiff *et al.*, *Science* **272**, 844 (1996); A. Seiff *et al.*,
ibid. **276**, 102 (1997).
 34. R. E. Young, M. A. Smith, C. K. Sobeck, *ibid.* **272**,
 837 (1996).
 35. Because low-latitude ion precipitation appears to
 be modulated by System III longitude, measure-
 ments made at other longitudes or local times
 would likely have revealed slightly cooler thermo-
 spheric temperatures.
 36. J. E. P. Connerney, *J. Geophys. Res.* **98**, 18659
 (1993).
 37. The Rayleigh (R) is a unit of surface brightness equiv-
 alent to an intensity of 10^6 photons $\text{cm}^{-2} \text{s}^{-1}$ per 4π

steradians. The flux received at a detector from a
 source with brightness B (in Rayleighs) is $F = 10^6$
 $B\omega/4\pi$ (in photons per square centimeter per sec-
 ond), where ω is the solid angle of the source (as
 seen from the detector) in steradians.

38. We gratefully acknowledge the work of the ROSAT
 team. ROSAT is supported by the Bundesministe-
 rium für Forschung und Technologie. We thank A.
 Seiff for furnishing the latest temperature profile de-
 rived from the ASI data. J.H.W. thanks the Centre
 National de la Recherche Scientifique and the staffs
 of the Observatoire de Paris-Meudon and the Obser-
 vatoire de Midi-Pyrénées, where most of the model-
 ing reported in this paper was performed. J.H.W.,
 G.R.G., and W.S.L. acknowledge support from the
 ROSAT Guest Observer Program (NAG5-2617) and
 the NASA Planetary Atmospheres Program (NAGW
 3624).

15 October 1996; accepted 30 January 1997

Gravity Waves in Jupiter's Thermosphere

Leslie A. Young,* Roger V. Yelle, Richard Young,
 Alvin Seiff, Donn B. Kirk

The Atmosphere Structure Instrument on the Galileo probe detected wavelike temperature fluctuations superimposed on a 700-kelvin temperature increase in Jupiter's thermosphere. These fluctuations are consistent with gravity waves that are viscously damped in the thermosphere. Moreover, heating by these waves can explain the temperature increase measured by the probe. This heating mechanism should be applicable to the thermospheres of the other giant planets and may help solve the long-standing question of the source of their high thermospheric temperatures.

The atmospheres of the giant planets (Jupiter, Saturn, Uranus, and Neptune) and Earth reach high temperatures in the low-pressure region of the upper atmosphere known as the thermosphere (1). On Earth, the energy to heat the thermosphere is supplied by absorption of solar extreme ultraviolet (EUV) radiation. However, for the giant planets, the solar heating rate is more than one order of magnitude too small to explain the high temperatures. Suggestions for Jupiter's thermospheric heat source include charged particle impacts, Joule heating, or the dissipation of gravity waves (2, 3, 4). It has been difficult to evaluate these energy sources because, until recently, the shape of the thermal profile above the 1- μ bar level was poorly constrained. Without knowing the first and second derivatives of the temperature profile, it is difficult to estimate the amount of energy required or the altitude at which the energy must be deposited by

mechanisms which seek to explain the thermal profile. The Atmosphere Structure Instrument (ASI) directly measured the temperature of Jupiter's upper atmosphere on 7 December 1995 (5). The ASI experiment detected wave-like fluctuations superimposed on a 700 K temperature increase in the thermosphere. Here we describe how these temperature fluctu-

ations are consistent with gravity waves, and show that viscous dissipation of these waves can be responsible for the elevated thermospheric temperatures.

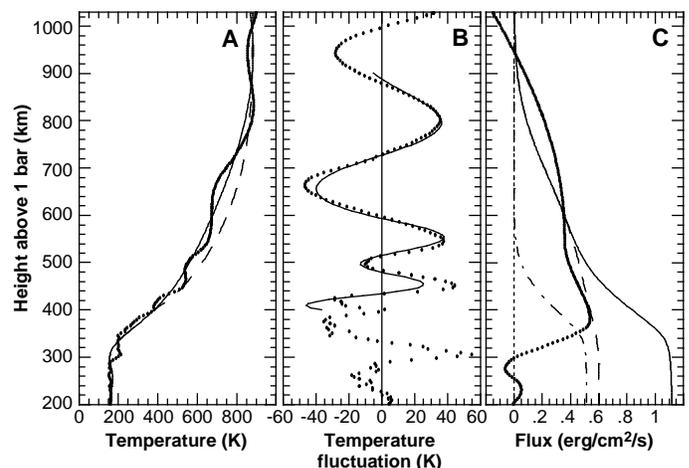
We decomposed the ASI temperature profile into a mean profile (Fig. 1A) and perturbations (6). Subtracting the mean thermal profile from the observed thermal profile leaves temperature fluctuations, peak-to-peak, of 50 to 80 K. These thermospheric waves have extrema at 940, 800, 665, 550, 495, 447, 425, and 395 km. Taking the distance between extrema as the half period, we estimated an observed vertical wavelength of ~ 260 km above 550 km, ~ 100 km between 450 and 550 km, and ~ 50 km between 395 and 450 km. The mesopause, which is the boundary between the middle atmosphere and the thermosphere, is located at the temperature minimum near 290 km (3.3 μ bar). The temperature gradient rises quickly above the mesopause, and reaches a maximum of 2.9 K/km at an altitude of 357 km. The positive thermal gradient implies that an energy of 0.53 $\text{erg}/\text{cm}^2/\text{s}$ is transported downward by thermal conduction. This conductive flux must be balanced by heating in steady state; therefore, the column-integrated heating rate in the thermosphere is also 0.53 $\text{erg}/\text{cm}^2/\text{s}$. The inferred upward flux at the top of the ASI observations is between 0 and 0.1 $\text{erg}/\text{cm}^2/\text{s}$, and depends on the method of defining the mean state (7).

We modeled how gravity waves could produce the estimated temperature perturbations under the following assumptions: (i) the frequency of the gravity wave is much larger than the Coriolis parameter, (ii) the mean state varies slowly compared to a wavelength, (iii) the period of the gravity

L. A. Young and R. V. Yelle, Center for Space Physics, Boston University, Boston, MA 02215, USA.
 R. Young, Space Science Division, NASA Ames Research Center, Moffett Field, CA 94035, USA.
 A. Seiff, Department of Meteorology, San Jose State University Foundation, CA 95192, USA.
 D. B. Kirk, University of Oregon, Eugene, OR 97403, USA.

*To whom correspondence should be addressed. E-mail: layoung@bu.edu

Fig. 1. (A) The observed ASI temperatures (thick dotted curve), an estimate of the mean thermal profile (solid curve), and the thermal profile which we ultimately derive by integrating the heating by the observed gravity waves (dashed curve). The dots represent individual measurements by the ASI. (B) The temperature fluctuations that remain after subtracting the mean state. The solid curve is the calculated sum of the two gravity waves summarized in Table 1. The dotted curve is the difference between the observed ASI temperatures and the estimated mean thermal profile. (C) The dots represent the rate at which energy is conducted downward, derived from the mean thermal profile. The solid curve is the total flux carried upward by the two waves, assuming $\epsilon F(z_0) = 0.2 \text{ erg}/\text{cm}^2/\text{s}$ for both waves.



wave is much shorter than the diffusion time over a vertical wavelength, (iv) the phase speed is much less than the sound speed, and (v) the zonal wind is negligible. We do not assume that the waves are hydrostatic, or that the mean state is isothermal (8).

The temperature fluctuation, δT , written as the product of an amplitude, $A(z)$, and a wave with wavenumbers k_x , k_y , and k_z and a frequency ω , is

$$\delta T = A(z) \exp[i(k_x x + k_y y - \omega t + \int k_z dz)] \quad (1)$$

The vertical wavenumber is given by

$$k_z^2 = \frac{N^2 - \omega^2}{\omega^2} k_h^2 - \frac{1}{4H^2} \quad (2)$$

where N is the Brunt-Väisälä frequency, H is the density scale height and $k_h = (k_x^2 + k_y^2)^{1/2}$ is the horizontal wavenumber. A negative k_z/ω corresponds to upward energy propagation. If ω is much smaller than N and k_z is much larger than $1/2H$, the dispersion relation simplifies to $k_z = -N/c$, where $c = \omega/k_h$ is the horizontal phase speed. Because of the probe's horizontal motion, k_z may differ from the observed vertical wavenumber, $k_{z,obs}$.

The amplitude of the gravity wave will grow with altitude until viscous dissipation becomes important. Under our assumptions, the amplitude is

$$A(z) \propto \sqrt{\frac{k_z T^2 N^4}{(N^2 - \omega^2) \rho}} \exp\left(\int_{z_{low}}^z \frac{\nu K^4}{\omega k_z} dz\right) \quad (3)$$

where $K = (k_h^2 + k_z^2)^{1/2}$ is the total wavenumber, and ρ , T , and ν are the density, temperature, and molecular (kinematic) viscosity of the mean state respectively. z_{low}

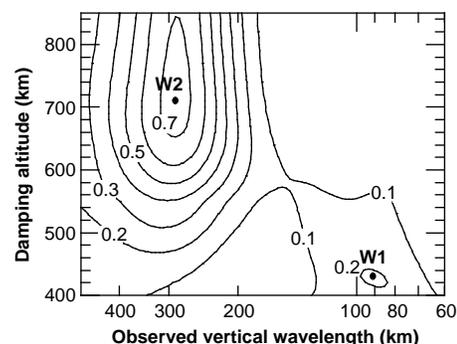


Fig. 2. The power spectral density is plotted as a function of the observed vertical wavelength ($k_{z,obs}$) and the viscous damping altitude (z_0). The dots indicate the local maxima, which correspond to two observed gravity waves (Table 1). Wave 1 is labeled W1 and wave 2 is labeled W2. The periodogram has been normalized so that a solitary, pure gravity wave will have unit power.

is arbitrary, because it simply introduces a multiplicative constant. The square-root term represents the growth of the wave in the absence of damping. The exponential term represents the decrease in amplitude due to damping by molecular viscosity. Since ν is inversely proportional to ρ , the damping term increases with increasing altitude. The maximum amplitude is reached at some damping altitude z_0 , at which $\partial A/\partial z = 0$. If $k_z \approx -N/c$, then at $z = z_0$

$$\frac{\omega^4}{k_h^3} = \omega c^3 = k_h c^4 \approx 2\tilde{H}_0 \nu_0 N_0^3 \quad (4)$$

where $\tilde{H} = [\partial \ln(T^2 N^3 / \rho) / \partial z]^{-1}$ is approximately the scale height, and N_0 , ν_0 , and \tilde{H}_0 are evaluated at z_0 . This simplifies the equation for the wave amplitude for a specified z_0 , which we write as $A(z, z_0)$,

$$A(z, z_0) \approx A_0 \sqrt{\frac{T^2 N^3 / \rho}{T_0^2 N_0^3 / \rho_0}} \exp\left(-\frac{1}{2\tilde{H}_0} \int_{z_0}^z \frac{\nu N^3}{\nu_0 N_0^3} dz\right) \quad (5)$$

where A_0 is the maximum temperature fluctuation amplitude. Notice that $A(z, z_0)$ no longer depends explicitly on the frequency or wavenumber of the wave.

The observed temperature perturbations cannot be fit by a single gravity wave. To determine the wave properties and spectrum from the observed perturbations, we calculate the periodogram, tailoring the analysis to be especially sensitive to waves with an amplitude modulation given by Eq. 5 (although other waves will be detectable as well). We vary z_0 from 400 to 900 km, and use $A(z, z_0)$ as the data window at each z_0 . We use the Lomb-Scargle periodogram, modified to include a data window (9), to determine both z_0 and $k_{z,obs}$ of the waves. The sliding periodogram has two local maxima, with observed vertical wavelengths of 288 and 91 km, and damping amplitudes of 710 and 430 km, respectively (Fig. 2 and Table 1). If other waves are present, they are of smaller amplitudes. The superposition

Table 1. Properties of modeled gravity waves.

	Wave 1	Wave 2
z_0 (km)	430	710
$2\pi/k_{z,obs}$ (km)	91	288
A_0 (K)	38	41
c (km/s)	0.186–0.205	0.18–0.43
ω (s ⁻¹)	9.9×10^{-5}	1.6×10^{-3}
$2\pi/k_h$ (km)	12,693	773
$2\pi/k_z$ (km)	92	149
$\varepsilon F(z_0)$ (erg/cm ² /s)	0.19	0.21

*For $c = 0.2$ km/s and $\varepsilon = 0.65$; For other values of c , use $\omega \propto c^{-3}$, $k_h \propto c^{-4}$, $k_z \propto c^{-1}$, and $F_0 \propto c^{-2}$.

of these two waves reproduce the temperature perturbations between 400 and 900 km (Fig. 1B). The relatively poor fit near and below 400 km may be due to gravity waves with smaller vertical wavelengths which are damped to insignificance at altitudes greater than 400 km.

If the probe trajectory were vertical, then the observed k_z would provide a direct measure of c , and all the wave properties could be inferred (assuming that these are gravity waves). However, the probe is moving horizontally as well as vertically, and the observed vertical wavenumber ($k_{z,obs}$) may be shifted from the actual vertical wavenumber (k_z). The effect of the probe's trajectory on the observed vertical wavenumber depends on the horizontal wavenumber, the vertical and horizontal velocities of the probe (V_z and V_h), and the angle (θ) between the horizontal component of the probe velocity and the horizontal phase velocity, such that

$$k_{z,obs} = k_z = k_h \cos\theta \frac{V_h}{V_z} - \frac{\omega}{V_z} \quad (6)$$

At the beginning of the observations, the probe was moving with $V_h = 46.3$ km/s and $V_z = 8.7$ km/s. The range of c for wave 1 is severely constrained by $k_{z,obs}$, so $0.186 < c_1 < 0.205$ km/s. The range of c for wave 2 is constrained to be less than 0.43 km/s by the observed wavelength, and is unconstrained on the short end. However, for $c_2 < 0.18$ km/s the horizontal and vertical wavelengths are similar. This violates the assumption of $k_z \approx -N/c$, used to generate the periodogram and find the wave amplitudes, so for a self-consistent analysis, we do not consider this case.

Finally, we show that heating by dissipating gravity waves can account for the observed thermal profile. The equation for the conservation of total energy can be written

$$\kappa \frac{dT}{dz} = \varepsilon [F_1(z) + F_2(z)] \quad (7)$$

where κ is the conductivity, F_1 and F_2 are the energy fluxes carried vertically by waves 1 and 2 (including internal, kinetic, and potential energy), and ε is an efficiency factor. In terms of the temperature perturbation, the wave flux is

$$F(z) = \frac{\rho}{2} \frac{\gamma}{\gamma - 1} \frac{g^2}{N^2} \left| \frac{\omega}{k_z} \right| \frac{N^2 - \omega^2}{N^2} \frac{A^2}{T^2} \quad (8)$$

where $\gamma = c_p/c_v$, the ratio of specific heats (3, 10). The derivative of Eq. 8 with respect to z represents the maximum possible heating rate, because some of the wave energy could be deposited as kinetic energy rather than heat, decreasing the heating efficiency of the waves.

We calculate the energy fluxes carried by both waves in the upper atmosphere, and compare them with $\kappa dT/dz$ calculated from the mean thermal profile. We find that we need $\epsilon F_2(z_0) = 0.2 \text{ erg/cm}^2/\text{s}$ to match the observed increase in temperature. We know the phase speed of wave 1 is about 0.2 km/s. If the phase speeds of the two waves are similar, than wave 2 must have a 65% efficiency. Alternatively, wave 2 could have $\epsilon F_2(z_0) = 0.2 \text{ erg/cm}^2/\text{s}$ if it has an efficiency of 100% and a phase speed of 0.25 km/s. We assumed $c = 0.2 \text{ km/s}$ and $\epsilon = 0.65$ for wave 1 and wave 2 (Table 1). Near the base of the thermosphere the magnitude of the total wave energy is greater than the thermally conducted flux (Fig. 1C). This could be due to some of the wave energy at altitudes below 440 km being radiated away by CH_4 , or the wave heating being less efficient near the mesopause. The thermally conducted flux varies with altitude similarly to the flux due to the two gravity waves in that there is a swift drop in the flux near 450 km (which we attribute mainly to wave 1) and a gradual decrease above 500 to 600 km (which we attribute mainly to wave 2). If we solve the thermal conduction equation at altitudes above 400 km the calculated thermal profile (Fig. 1A) is similar to the observed mean profile. Thus, it is possible that we have a self-consistent explanation above 400 km altitude in which the gravity waves that are manifest as temperature perturbations imply a heating profile which is consistent with the observed mean temperature profile.

Gravity waves are also present in Jupiter's mesosphere. Stellar occultations reveal temperature perturbations in the 1 to 10 μbar region with vertical scales of 2 to 20 km and amplitudes of $\sim 5 \text{ K}$ (3). The ASI data also show evidence for wave activity in the 1- to $10^5\text{-}\mu\text{bar}$ region. Although these data have yet to be analyzed in detail the temperature perturbations are on the order of 20 K with apparent vertical wavelengths of 50 km (5). Both observations show a complex spectral distribution that is consistent with breaking gravity waves (that is, waves that are critically damped by eddy viscosity, and whose thermal gradients approach the adiabatic lapse rate). As discussed by Seiff *et al.* (5), the short-wavelength waves likely damp at low altitudes. Previous discussions of gravity-wave heating on Jupiter have focused on the waves observed in the stellar occultation, which deposit their energy within one or two scale heights of the mesopause (3, 4). Breaking gravity waves have also been observed in the mesospheres of Earth (11) and Neptune (12).

In contrast with the mesospheric waves, the waves observed in Jupiter's thermosphere are damped by molecular, not eddy, viscosity. The exponential increase of mo-

lecular viscosity with altitude causes shorter waves to be damped at lower altitudes, and can lead to the dominance at any particular altitude of a single wavelength. Thus, the waves observed in the thermosphere are the surviving members of an ensemble of upwardly propagating waves and they tell us about the energy contained in the long wavelength end of the gravity wave spectrum. If we assume that the waves seen in the thermosphere were generated by convective motions in the troposphere and have propagated upward from the tropopause according to Eq. 5, then we infer amplitudes at 200 mbar of only 0.07 and 0.02 K for waves 1 and 2. These are the sizes of the temperature perturbations in the tropopause region required to generate the waves. The forcing perturbation in the troposphere must also have a horizontal scale comparable to that of the observed waves. Other suggestions for the heat source responsible for the high temperatures in Jupiter's thermosphere include heating by charged particle precipitation, transport of auroral energy from high to low latitudes, and Joule heating by winds in the ionosphere. These suggestions have been difficult to evaluate without observational constraints on the charged particle fluxes and winds in the upper atmosphere. Recently, Waite *et al.* (13) reported observations of x-rays from the jovian equatorial regions which may be caused by precipitation of energetic particles from the magnetosphere. The heating rates associated with the particle precipitation could contribute significantly to the upper atmospheric heating rate (13). Thus, there may be several contributors to the energy balance in the jovian upper atmosphere. The energetic particle precipitation (13) is a heating mechanism that is specific to Jupiter since it relies upon the magnetosphere as a source of energy, and Jupiter has the most powerful magnetosphere of the giant planets. Dissipation of gravity waves is not unique to Jupiter; it could occur in the thermospheres of all the giant planets and therefore may be a general cause for the hot thermospheres on the giant planets.

REFERENCES AND NOTES

1. Thermospheric temperatures were measured by the Voyager ultraviolet spectrometer (UVS) occultations for Jupiter [S. Atreya *et al.*, *Geophys. Res. Lett.* **6**, 795 (1979)], Saturn [Smith *et al.*, *J. Geophys. Res.* **88**, 8667 (1983)], Uranus [F. Herbert *et al.*, *J. Geophys. Res.* **92**, 15093 (1987)], and Neptune [A. L. Broadfoot *et al.*, *Science* **246**, 1459 (1989)]. The temperatures range from 420 to 1000 K, with no trend in size or distance from the sun.
2. For heating by charged particle impacts, see D. M. Hunten and A. J. Dessler [*Planet. Space Sci.* **25**, 817 (1977)]. For Joule heating, see J. T. Clarke *et al.* [*J. Geophys. Res.* **92**, 15139 (1987)] and Nishida and Watanabe [*ibid.* **86**, 9945 (1991)].

3. French and P. Gierasch, *J. Atmos. Sci.* **31**, 1707 (1974), who interpreted as gravity waves thermal oscillations detected from a stellar occultation by Jupiter's atmosphere [J. Veverka *et al.*, *Astron. J.* **79**, 73 (1974)].
4. R. V. Yelle *et al.*, *J. Geophys. Res.* **101**, 2149 (1996).
5. A. Seiff *et al.*, *Science* **272**, 844 (1996); A. Seiff *et al.*, *ibid.* **276**, 402 (1997).
6. We parameterize the mean profile as a cubic spline, with spline points every two scale heights, where a scale height is the e-folding distance of the pressure. The spline points are widely spaced so that the derived mean profile does not simply reproduce the long-wavelength fluctuations above 600 km. We considered other methods of defining the mean thermospheric profile, including a first- or second-order polynomial, a profile with constant energy flux, and a running average. Of the methods considered, we prefer the cubic spline because it best matched the profile between 400 and 500 km, where the curvature is large. The location of the maxima and minima in the thermal profiles does not depend on the method used to define the mean profile. The amplitudes are more sensitive, especially below 500 km. For example, the amplitudes of the negative peak at 665 km range from 34 (for the running average) to 55 K (for the constant-flux case), and the amplitudes of the negative peak at 425 km range from 2 (for the running average) to 40 K (for the first-order polynomial).
7. The inferred flux at 1000 km also depends on the estimate of the temperature at the first (highest) datum during the reduction of the ASI data (5). The upper-boundary temperature must be greater than 500 K, since cold upper-boundary temperatures lead to super-adiabatic gradients. An upper-boundary temperature of 1300 K is not ruled out by the ASI data, but is difficult to reconcile with H_3^+ observations [Marten *et al.*, *Planet. Space Sci.* **42**, 391 (1994)]. We prefer 900 K because of the small gradient at the upper boundary and the agreement with the H_3^+ observations.
8. The analysis used here follows Holton and Zhu [*J. Atmos. Sci.* **41**, 2653 (1993)] and Gill [*Atmosphere-Ocean Dynamics* (Academic Press, San Diego, 1982), pp. 294-296], but is extended to allow treatment of a non-isothermal atmosphere. We first find the wave equations assuming no viscosity, then integrate the wave equation with the WKBJ approximation. Finally, we introduce viscosity and thermal conduction as a perturbation to the non-viscous equations, by assuming $k_z = k_{z,\text{novisc}} + \nu k_{z,\text{novisc}}^2 / \partial \omega$. The WKBJ approximation says that $w = \sqrt{1/m} \exp(\pm i \int m dz)$ is a solution to the differential equation $d^2 w / dz^2 + m^2 w = 0$ if m varies slowly compared with a wavelength.
9. Because the data are sampled unevenly in altitude (because they are sampled evenly in time), we use the Lomb-Scargle periodogram analysis. Following the procedure outlined in appendix C of J. D. Scargle [*Astrophys. J.* **263**, 835 (1982)], we define the periodogram with a window function $A_i = A(z_i, z_0)$, normalized to a maximum value of 1 K. The power spectral density as a function of both z_0 and k_z is

$$P(z_0, k_z) = \frac{[\sum (\delta T)_i A_i \cos \phi_i]^2}{\sum A_i^2 \cos^2 \phi_i} + \frac{[\sum (\delta T)_i A_i \sin \phi_i]^2}{\sum A_i^2 \sin^2 \phi_i}$$

where $\phi_i = k_z(z_i - z_1)$, and z_1 is an offset defined by

$$\tan(2k_z z_1) = \frac{\sum A_i^2 \sin(2k_z z_i)}{\sum A_i^2 \cos(2k_z z_i)}$$

Only the 107 points between 400 and 900 km were used to calculate the periodogram, because the temperature fluctuation above 900 km is sensitive to the choice of the upper-boundary temperature and the fluctuation below 400 km is sensitive to the method used for determining the mean thermal profile.

10. The factor of $\gamma/(\gamma - 1)$ arises by including the internal energy in the total energy-balance equation. If there is no radiative heating, no chemical or phase changes, and the local time derivative of the total energy is

zero, then the total energy equation is

$$\rho w \left(E + \Phi + \frac{1}{2} u^2 \right) + \rho w - \kappa \frac{\partial(T + \delta T)}{\partial z} - \frac{v}{\rho} \frac{\partial u^2}{\partial z} = 0$$

where $u = (u, v, w)$ is the perturbation velocity, E , Φ , and $u^2/2$ are the internal, potential, and kinetic energies, ρ is the total pressure (mean plus perturbation), and κ is the thermal conductivity [Gill (8), eq. 4.7.3 and 4.7.4; French and Gierasch, (9)]. The first term represents the advection of energy, the second term is the mechanical-energy work term, the third term is the transport of energy by thermal conduction, and the fourth term represents the (reversible) diffusion of kinetic energy, and

is generally small. We drop terms that are higher than second order in perturbation variables, and take the average over a period (denoted by $\langle \rangle$), eliminating $\rho w \Phi$ and $\rho w(1/2)u^2$. Finally, we write ρE as $c_v \rho R$. Since $1 + c_v/R = \gamma/(\gamma - 1)$, this becomes

$$\kappa \frac{\partial T}{\partial z} = \frac{\gamma}{\gamma - 1} \langle (\delta \rho) w \rangle$$

where $\delta \rho$ is the perturbation pressure. Evaluation of $\langle (\delta \rho) w \rangle$ with the WKB wave solution leads to (8).

11. For example, J. A. Whiteway and A. I. Carswell, *J. Geophys. Res. (Atmos.)* **100**, 14,113 (1995).
12. F. Roques *et al.*, *Astron. Astrophys.* **288**, 985 (1994).
13. J. H. Waite Jr. *et al.*, *Science* **276**, 104 (1997).

21 January 1997; accepted 14 March 1997

The Receptor for the Cytotoxic Ligand TRAIL

Guohua Pan, Karen O'Rourke, Arul M. Chinnaiyan, Reiner Gentz, Reinhard Ebner, Jian Ni,* Vishva M. Dixit*†

TRAIL (also known as Apo-2L) is a member of the tumor necrosis factor (TNF) ligand family that rapidly induces apoptosis in a variety of transformed cell lines. The human receptor for TRAIL was found to be an undescribed member of the TNF-receptor family (designated death receptor-4, DR4) that contains a cytoplasmic "death domain" capable of engaging the cell suicide apparatus but not the nuclear factor kappa B pathway in the system studied. Unlike Fas, TNFR-1, and DR3, DR4 could not use FADD to transmit the death signal, suggesting the use of distinct proximal signaling machinery. Thus, the DR4-TRAIL axis defines another receptor-ligand pair involved in regulating cell suicide and tissue homeostasis.

Apoptosis, or programmed cell death, is a process fundamental to the normal development and homeostasis of multicellular organisms. Deregulation of programmed cell death leads to a number of human diseases, including cancer, neurodegenerative disorders, and acquired immunodeficiency syndrome (1, 2). The cell death machinery comprises effectors, activators, and negative regulators (3). Certain cytokines of the TNF ligand family and their cognate receptors, including TNFR-1 and Fas (also known as Apo-1 or CD95), are classic triggers of the suicide response (1, 3). TNF and Fas ligand (also known as Apo-1L or CD-95L) induce apoptosis by binding to their respective death domain-containing receptors, TNFR-1 and Fas. The death domain is a protein-protein interaction motif that orchestrates the assembly of a signaling complex leading to the recruitment of pro-apoptotic proteases (caspases) related to the product of the *Caenorhabditis elegans* death gene *ced-3* (3-7).

Another member of the TNF ligand family, named TRAIL (Apo-2L), has been iden-

tified (8), and like Fas ligand (FasL), it induces rapid apoptosis in transformed cell lines of diverse origin (8). Unlike FasL, whose transcripts are predominantly restricted to stimulated T cells and sites of immune privilege (for example, corneal epithelium), TRAIL expression is detected in many normal human tissues, suggesting that TRAIL must not be cytotoxic to most tissues in vivo (8, 9). Given that TRAIL is a member of the TNF ligand family and has marked proapoptotic potential for transformed cells, it was reasonable to assume that it bound a death domain-containing member of the TNF-receptor family. However, the inability of TRAIL to bind TNFR-1, Fas, or the recently identified death domain-containing receptor DR3 (also called Wsl-1, Apo-3, and TRAMP) (8, 10) suggested that TRAIL may interact with a yet unknown member of the TNF-receptor family.

To identify such a receptor, we searched an expressed sequence tag (EST) database with the death domain of TNFR-1 (Fig. 1B). A human EST clone was initially found that encoded a previously unidentified death domain. Subsequent cDNA library screening and DNA sequence analysis led to the isolation of a clone encoding an open reading frame of 468 amino acids with features characteristic of a cell surface receptor (Fig. 1A) (11). Database searches, sequence alignment, and hydropathy analysis revealed the encoded

protein (DR4) to be a member of the TNF-receptor family (Fig. 1). A putative signal peptide is present at the beginning of the molecule (amino acids -23 to -1), with the mature protein predicted to start at amino acid 24 (Ala). Residues 108 to 206 contain two cysteine-rich pseudorepeats that resemble corresponding regions in TNFR-1 (four repeats), DR3 (four repeats), Fas (three repeats), and CAR1 (two repeats) (12). Following the transmembrane domain (amino acids 227 to 245) is an intracellular region containing a 70-amino acid stretch with significant similarity to the death domains of TNFR-1, DR3, Fas, and CAR1 (Fig. 1B). The death domain of DR4 is 30, 29, and 28% identical to the corresponding domains in TNFR-1, CAR1, and DR3, respectively, but only 19% identical to the death domain of Fas. Four out of six residues in the death domain of TNFR-1 (Arg³³⁶, Leu³⁴⁰, Trp³⁶⁷, and Ile³⁹⁷) (13) that are essential for its signaling ability are conserved in the DR4 death domain (Arg³⁶², Leu³⁶⁶, Trp³⁹², and Ile⁴²¹). Given the presence of a death domain and its ability to engage the death pathway (discussed below), we termed this receptor DR4 for death receptor-4.

Tissue distribution of the DR4 transcript was examined by Northern (RNA) blot analysis. Three major transcripts of 2.6, 4.6, and 7.2 kb were detected in most human tissues, including spleen, peripheral blood leukocytes, small intestine, and thymus (Fig. 1C). DR4 expression was also found in K562 erythroleukemia cells, MCF7 breast carcinoma cells, and activated T cells (14).

Upon overexpression, TNFR-1, Fas, and DR3 are activated in a ligand-independent manner and induce apoptosis (7, 10). On the basis of this observation, we examined whether overexpression of DR4 could similarly trigger apoptosis. MCF7 human breast carcinoma cells and 293 human embryonic kidney cells were transiently transfected with a DR4-expressing construct. Most transfected cells underwent the morphological changes characteristic of apoptosis (Fig. 2, A to C). As anticipated, deletion of the death domain abolished the ability of DR4 to engage the death pathway. DR4-induced apoptosis was efficiently blocked by the caspase inhibitors z-VAD-fmk and CrmA, implicating the involvement of apical pro-apoptotic proteases (7) (Fig. 2D). Because apoptosis induced by TNFR-1, Fas, and DR3 is also attenuated by these same inhibitors (7, 10), it is likely that the downstream death effector molecules are similar in nature.

FADD (also called MORT1) is the common intracellular adaptor molecule recruited by Fas, TNFR-1, and DR3 to engage the downstream death machinery (6, 10, 15). We tested if DR4 could also use FADD to transmit the death signal. We cotransfected 293

G. Pan, K. O'Rourke, A. M. Chinnaiyan, V. M. Dixit, Department of Pathology, University of Michigan Medical School, Ann Arbor, MI 48109, USA.

R. Gentz, R. Ebner, J. Ni, Human Genome Sciences, 9410 Key West Avenue, Rockville, MD 20850-3338, USA.

*These authors share senior authorship.

†To whom correspondence should be addressed. E-mail: vmdixit@umich.edu