

observations to be true, we see no reason to doubt the statements of these excellent "ethologists."

REFERENCES AND NOTES

1. E. Whittall, *Bokmakierie* **20**, 73 (1968).
2. H. Pager, *Bee World* **54**, 61 (1973).
3. G. W. B. Huntingford, *Anthropos* **50**, 602 (1955).
4. M. Ichikawa, *Afr. Study Monogr.* **1**, 55 (1981).
5. H. A. Isack, thesis, Oxford University, Oxford (1987).
6. H. Friedmann, *Bull. U.S. Natl. Mus.* **208**, 1 (1955).
7. L. L. Short and J. F. M. Horn, *Am. Mus. Novit.* **2825**, 1 (1985).
8. E. Bartschelet, in *Animal Migration, Navigation and Homing*, K. Schmidt-Koenig and W. T. Keeton, Eds. (Springer-Verlag, Berlin, 1978), p. 3.
9. The observation that stopping distances decrease as the bees' nest is approached was also reported by G. W. Stow [*The Native Races of South Africa* (Swan Sonnenschein, London, 1905)].
10. We measured distances by counting paces and later converting them into meters. Perching heights were estimated to the nearest 0.5 m. In cases of skewed distribution, original data were log-transformed.
11. The relation between Stopdist and Stopno was tested within guidings, yielding one regression coefficient (b) for each tour. The relations between Stopdist and Nestdist and Distflown, respectively, were tested across guidings. In the case of Nestdist, separate regressions were calculated for each Stopno, whereas, in the case of Distflown, separate regressions were calculated for each 50-m category of Nestdist. In all cases, the regression coefficients were then tested against the 0 hypothesis of no relation ($b = 0$) by means of a Wilcoxon matched-pairs, signed-ranks test.
12. The small sample size did not allow us to calculate separate regressions as in the case of Stopdist. Therefore, Perch data from all guidings were pooled.
13. Our reanalysis of Friedmann's data (6) indicates that even his birds showed directional guiding and probably had prior knowledge of the hive location (H. A. Isack and H.-U. Reyer, in preparation).
14. E. Curio, G. Klump, K. Regelman, *Oecologia (Berlin)* **60**, 83 (1983).
15. M. Milinski, *Nature* **325**, 433 (1987).
16. J. R. Krebs, in *Behavioural Ecology—An Evolutionary Approach*, J. R. Krebs and N. B. Davies, Eds. (Blackwell, Oxford, 1978), p. 23.
17. M. P. Hassel and R. M. May, *J. Anim. Ecol.* **43**, 567 (1974).
18. J. N. M. Smith, *Behaviour* **49**, 1 (1964).
19. R. Wehner and M. V. Srinivasan, *J. Comp. Physiol.* **142**, 315 (1981).
20. We thank those who supported the study, in particular R. E. Leakey, C. M. Perrins, I. Orto, S. Orto, and the honey collectors G. Dambi, D. Galgallo, and A. Mariiqo. Financial and logistic support was provided by the African Wildlife Foundation (Nairobi), East African Wildlife Society (Nairobi), Frank Chapman Memorial Fund (New York), Max-Planck-Institut für Verhaltensphysiologie (Seewiesen, FRG), National Museum of Kenya (Nairobi), Percy Sladen Memorial Fund (London), and the Christopher Welch Scholarship (Oxford). H. G. Wallraff and A. D. Barbour helped with the statistics. R. Diesel, J. Lamprecht, F. Trillmich, W. Wickler, and two anonymous referees made useful comments on an earlier draft of the paper. G. Louw improved the English. This report is dedicated to the late H. Friedmann, the pioneer of honeyguide research.

21 June 1988; accepted 14 December 1988

Structure of Recombinant Human Renin, a Target for Cardiovascular-Active Drugs, at 2.5 Å Resolution

ANITA R. SIELECKI, KOTO HAYAKAWA, MASAO FUJINAGA,* MICHAEL E. P. MURPHY, MARIE FRASER, ALASTAIR K. MUIR, CYNTHIA T. CARILLI, JOHN A. LEWICKI, JOHN D. BAXTER, MICHAEL N. G. JAMES

The x-ray crystal structure of recombinant human renin has been determined. Molecular dynamics techniques that included crystallographic data as a restraint were used to improve an initial model based on porcine pepsinogen. The present agreement factor for data from 8.0 to 2.5 angstroms (Å) is 0.236. Some of the surface loops are poorly determined, and these disordered regions border a 30 Å wide solvent channel. Comparison of renin with other aspartyl proteinases shows that, although the structural cores and active sites are highly conserved, surface residues, some of which are critical for specificity, vary greatly (up to 10 Å). Knowledge of the actual structure, as opposed to the use of models based on related enzymes, should facilitate the design of renin inhibitors.

RENIN (E.C. 3.4.23.15) IS A HIGHLY specific aspartyl proteinase with only one known substrate, angiotensinogen. In humans, a decapeptide, angiotensin I, is released from angiotensinogen by the catalytic hydrolysis of the Leu¹⁰-Val¹¹ bond. Angiotensin I is processed by the angiotensin-converting enzyme (ACE) to angiotensin II, a potent vasoconstrictor involved in regulating blood pressure and fluid balance. However, presently available therapeutic agents for reducing blood pressure target ACE (1) and not renin.

Because only very small quantities of kidney renin have been available, much of the biochemical characterization of renin has been done on the mouse submaxillary gland enzyme. Mouse renin has been protein- (2) and cDNA-sequenced (3), and crystals of it have been reported by several groups (4, 5). In the five crystal forms of mouse renin

obtained, the corresponding asymmetric units contain multiple copies of the molecules (5), which complicates the crystallographic problem enormously. No further progress in this analysis has been reported.

The primary structure of the human enzyme has been deduced solely from cDNA (6) and gene sequences (7). We converted secreted human prorenin (8) from transfected Chinese hamster ovary cells (8, 9) to active renin by cleavage with immobilized trypsin. The purified recombinant human (rh) renin (10) was treated with endoglycosidase F to remove attached carbohydrates without affecting the specific activity of the final product (11).

Sample homogeneity from batch to batch was difficult to regulate. Most preparations exhibited three to five bands on isoelectric focusing gels. From ~1500 different crystallization trials, the optimal conditions were 5

to 6% polyethylene glycol 600 buffered with 50 mM NaH₂PO₄-K₂HPO₄ to pH 4.7. The resulting crystals exhibited tetragonal symmetry, space group *I*4, with unit-cell dimensions, $a = b = 133.5$ Å, $c = 41.7$ Å, with one renin molecule per asymmetric unit, $V_M = 2.53$ Å³ per dalton.

Intensity data were collected on a twin multiwire detector system (12). A total of 60,512 measurements (13,343 unique data) were measured from a single renin crystal. The overall symmetry agreement factor [$= \sum(I_i - \langle I \rangle) / \sum \langle I \rangle$, where I are the net intensities] was 0.09. The structure was solved by the molecular replacement method (13). Several renin models based on the known structures of three fungal aspartic proteinases have been built (14). The more extensive homology among mammalian species led us to construct a search model for renin based on the molecular structure of porcine pepsinogen (15, 16). The correlation coefficient between observed and calculated structure factors based on the oriented and translated model was 0.39 (9.5 σ above the mean) for the data in the 6 to 4 Å resolution shell. The corresponding *R* factor [$= \sum ||F_o| - |F_c|| / \sum |F_o|$, where $|F_o|$ and $|F_c|$ are observed and calculated structure factor amplitudes]

A. R. Sielecki, K. Hayakawa, M. Fujinaga, M. E. P. Murphy, M. Fraser, A. K. Muir, M. N. G. James, Medical Research Council of Canada Group in Protein Structure and Function, Department of Biochemistry, University of Alberta, Edmonton, Alberta, Canada T6G 2H7.
C. T. Carilli and J. A. Lewicki, California Biotechnology Inc., Mountain View, CA 94043.
J. D. Baxter, Metabolic Research Unit, University of California, San Francisco, CA 94143.

*Present address: Laboratory of Physical Chemistry, University of Groningen, Nijenborgh 16, 9747 AG, Groningen, The Netherlands.

was 0.48 (0.55 for all data to 2.5 Å resolution).

An electron density map with calculated phases and coefficients designed to minimize model bias (17) was computed at this stage, and showed that the molecular replacement had provided a correct solution but that extensive modifications to the search model

were needed to trace the polypeptide chain for some regions (Fig. 1A). The relatively small radius of convergence of standard refinement procedures (18) prompted us to use the method of Brünger *et al.* (19) of performing molecular dynamics (MD) simulations with the crystallographic data as an extra energy term. Details of the implemen-

tation of this refinement method within the GROMOS MD program package (20) are described elsewhere (21). The MD simulations were carried out at 300 K with bond lengths restrained with the SHAKE algorithm (22); 0.002-ps time steps were used. Each sequence of MD simulations was preceded and followed by 50 to 100 steps of steepest descent energy minimization (EM) that also included the crystallographic terms (21). Following each such set, conventional restrained parameters least-squares refinement (18) was carried out until stereochemical parameters had values comparable to those found in accurate small molecule determinations. During these conventional refinement cycles, individual temperature factors were refined with tight restraints. The refined models and associated electron density maps were examined and manually updated at intermediate stages in an MMS-X graphics system (23) with the program M3 (24).

A total of 350 steps of EM and 12 ps of MD simulations with crystallographic restraints (21) were performed on the initial model based on the molecular replacement solution. The present crystallographic *R* factor is 0.236 for 12,769 data, 8 to 2.5 Å resolution, with $|F_o| > 0.5 \sigma(F_o)$ [$\sigma(F_o)$ as the estimated error in the measurement of the observed structure-factor amplitudes. Only 15 possible solvent sites were included in the last few cycles of conventional least-squares refinement. The current structure has good stereochemistry [root-mean-square (rms) deviations from bond distances, 0.016 Å; from "angle" distances, 0.044 Å; from planarity, 0.006 Å; and from planarity of peptide bonds, 1.9°; coordinates will be submitted to Brookhaven Data Bank].

The MD refinement had its most striking effect on the residues that constitute some of the loops on the periphery of rh-renin. Such loops can exhibit conformational variability among homologous proteins due, in part, to differences in molecular packing. Alternatively, they suffer from various forms of disorder in the crystalline arrays (25). Some loop regions moved up to 9 to 10 Å in going from the initial model to the present refined structure. For example, in the loop near the NH₂-terminus, residues Thr⁷ to Tyr¹⁵ [the numbering throughout is based on our alignment of the human renin sequence with that of porcine pepsin (26)], the C α atoms of residues Tyr⁹, Met¹⁰, Asp¹¹, and Thr¹² moved an average of 2.9 Å, whereas some of the side-chain atoms moved 9 to 11 Å (Fig. 1). The lack of interpretable electron density in the initial map of this loop dramatically improved after the restrained MD and least-squares refinement. However, several regions were not

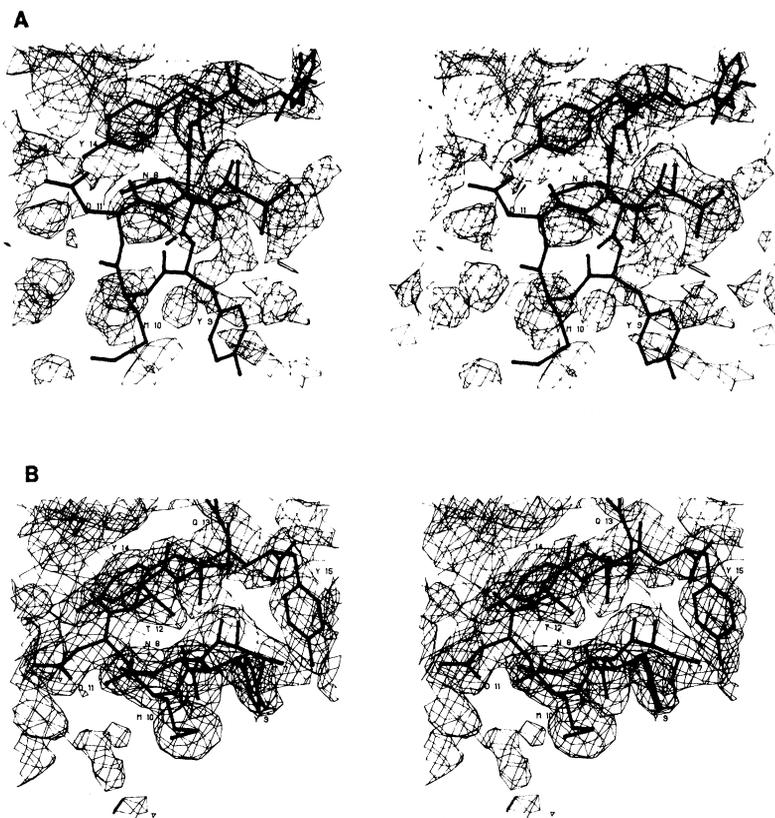


Fig. 1. (A) Electron density distribution and initial model in the region of the NH₂-terminal strand, Thr⁷ to Tyr¹⁵, as seen in the first map computed with the phases deduced from the molecular replacement solution (contour level = 0.3 eÅ⁻³). (B) Same region and orientation as in (A) but with the current electron density and the superimposed refined model. Tyr⁹ OH moved by 11.5 Å during the refinement process. Comparison of (A) and (B) shows that model bias has little influence on these electron density maps. Residues are denoted by the one-letter code (36).

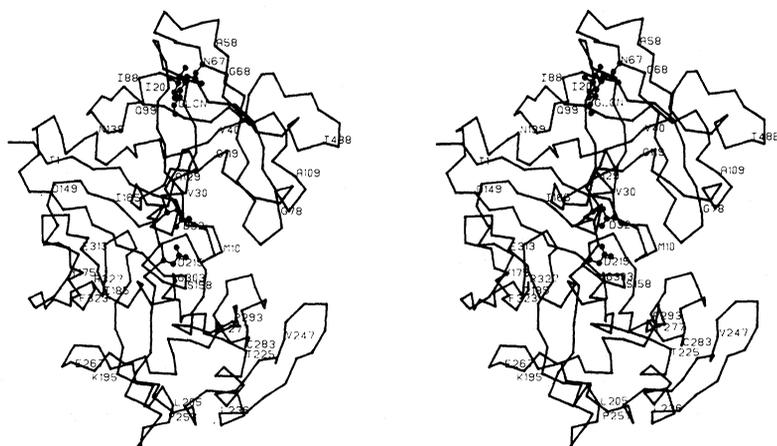


Fig. 2. The C α -atom representation of the rh-renin molecule (numbering derived from our alignment with pepsin). The side-chain atoms of Asp³², Asp²¹⁵, and Asn⁶⁷ (Asn⁷⁵ in renin numbering) with an attached carbohydrate moiety (GLCN) are indicated with filled circles. There is no electron density associated with the first four of the six residues preceding Thr1 (T1), the residue aligned to the NH₂-terminus of pepsin. The COOH-terminus of the molecule is labeled R327.

antiparallel β sheet located opposite the active site cleft. The interdomain dyad passes between the central two strands [i and q in the nomenclature in (30)]. Only at the periphery of the sheet do the strands depart from one another to accommodate differences in the sequences. Structural similarity of the core is also evident in the vicinity of the active site (Fig. 4B). Two highly conserved strands in renin (Val³⁰ to Ser³⁵ and Leu²¹³ to Ala²¹⁸) and the residues that flank them (Val¹²⁰ to Gly¹²⁴ and Ala³⁰¹ to Thr³⁰⁵), a total of 105 main-chain atoms, were used to compare the active sites of rh-renin and the fungal enzymes. The rms differences are ~ 0.45 Å, whereas the active site structures of the fungal enzymes agree more closely, with rms differences of ~ 0.24 Å for the same 22 residues.

The regions in renin that differ from the other enzymes provide its specificity, but are difficult to model accurately. However, such regions may be critical for designing inhibitors of the enzyme. Major differences in the aspartic proteinases are seen in the loops of the molecular periphery. In particular, the loops that border the entrance to the active site substrate binding cleft display differences in position of up to 8 to 10 Å (Fig. 5). Some of these conformational differences can be explained by insertions or deletions in the sequences, others by unshared structural features. For example, the disulfide bridge at Cys²⁰⁶-Cys²¹⁰ and the preceding bulge in the rh-renin and pepsinogen probably shift the position of the α helix at 224 to 235 relative to its position in penicillopepsin.

Other more subtle differences are found in the core of renin. The relatively high *pH* optimum of renin [5.5 to 7.5, compared with 2.0 to 4.0 for other aspartyl proteinases (33)] may be due to Ala³⁰⁴ (317 in renin) replacing Asp³⁰⁴ in all other family members (14, 34). However, a site-specific mutation

of Ala³⁰⁴ in rh-renin to Asp³⁰⁴ downshifts the *pH* profile by only 0.5 units (35). The carboxyl group of Asp³⁰⁴ in the fungal enzymes is partially buried, most likely protonated, and forms a hydrogen bond with Thr²¹⁶ O. In rh-renin, the analogous hydro-

gen bond to Thr²¹⁶ O comes from Tyr¹⁵⁵ OH (Fig. 6), a residue substituted by a hydrophobic amino acid in the other enzymes. Although this hydrogen-bonding pattern could be preserved, a mutation of Ala³⁰⁴ to an Asp would disrupt the surround-

Fig. 5. Conformational variability in the surface loops that line the entrance to the substrate binding clefts of four aspartic proteinases. Color coding and enzyme identification are given in the figure. The "flap" (residues 72 to 84) is shown in the top right; helix 222 to 232 (penicillopepsin numbers) is at the top left of the figure. Some of the residues in these loops in renin are poorly ordered (see Fig. 8).

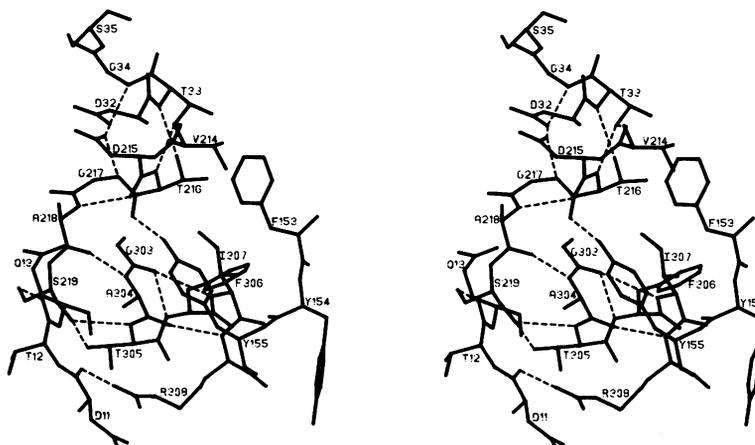
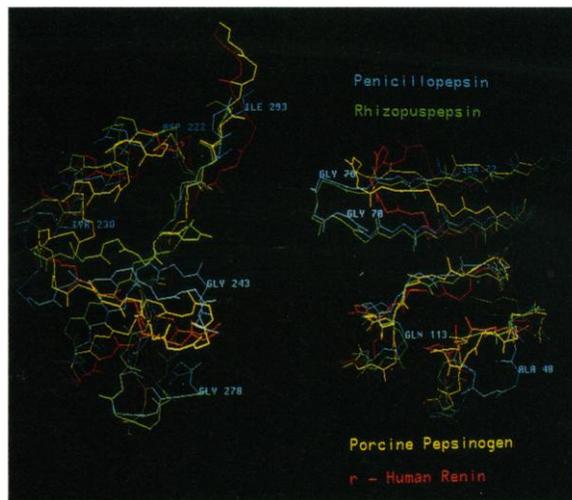


Fig. 6. Stereoscopic view of the environment of Ala³⁰⁴ (317 in the renin numbering) in rh-renin. Four residues along the short internal helix lies Arg³⁰⁸, a basic residue in all aspartic proteinases. The guanidinium group of Arg³⁰⁸ forms an ion pair with Asp¹¹. Lys³⁰⁸ in penicillopepsin forms an analogous hydrogen-bonded ion pair with Asp¹¹. Incorporating an Asp side chain in place of the methyl group of Ala³⁰⁴ in rh-renin would be disruptive.

Table 1. Residues (36) lining the substrate binding sites in human renin. The numbering is derived from our sequence alignment with pepsin.

Angiotensinogen substrate	Renin enzyme
P ₄ Pro	S ₄ : T12, S219, Y220
P ₃ Phe	S ₃ : T12, Q13, P111, F112, A115, F117, S219
P ₂ His	S ₂ : A218, Y220, S222, H288, M290, A301
P ₁ Leu	S ₁ : V30, D32, Y75, F112, F117, V120, D215, G217
P ₁ ' Val	S ₁ ': Y75, L213, D215, M290, T299, A301
P ₂ ' Ile	S ₂ ': G34, S35, L73, Y75, Q128, I130
P ₃ ' His	S ₃ ': Q128, V189, Q191, L213, D291, P293, P298, T299

Table 2. Structural comparison of four aspartic proteinases. A least-squares fitting procedure [(37); program of W. Bennett] was used to do the structural alignment of the enzymes based on C α -atoms only. The numbers in the upper triangular part of the matrix are the rms differences (Å); the numbers in parentheses following are the number of C α -atom pairs considered structurally equivalent and included in each calculation. The lower triangular part of the matrix has the number of identical residues between pairs of enzymes with these structural alignments. Percent values are given in parentheses. Abbreviations: PPEP, penicillopepsin (30); RPEP, rhizopus pepsin (31); EPEP, endothia pepsin (32); PGEN, porcine pepsinogen (15); RHRE, recombinant human renin. Coordinates for RPEP and EPEP were obtained from the Brookhaven Protein Data Bank (38).

Enzymes	PPEP	RPEP	EPEP	PGEN	RHRE
PPEP		0.92 (247)	0.92 (268)	1.63 (275)	1.59 (254)
RPEP	123 (38.1)		1.10 (264)	1.53 (278)	1.40 (280)
EPEP	141 (43.7)	122 (37.5)		1.91 (268)	1.96 (279)
PGEN	102 (31.6)	118 (36.5)	90 (27.9)		1.33 (279)
RHRE	78 (24.1)	86 (26.6)	70 (21.2)	131 (40.3)	

Fig. 7. The carbohydrate (GLCN) attachment site in rh-renin. The electron density is contoured at $0.30 \text{ e}\text{\AA}^{-3}$. The residue numbering follows that of pepsin so that Asn⁷⁵, in renin, is labeled Asn⁶⁷. The conformation of the polypeptide chain is that of an Asx-turn with the main-chain NH of Thr⁶⁹ forming a hydrogen bond to Asn⁶⁷ O δ 1. There is no density for an N-acetyl group on N2 of the sugar moiety.

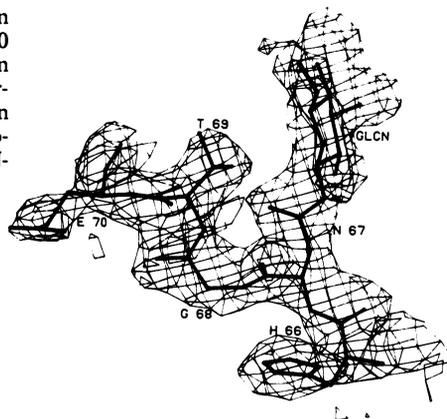
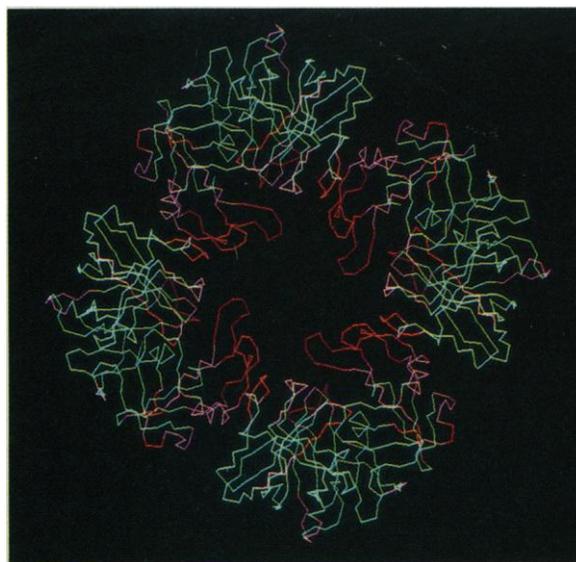


Fig. 8. Molecular packing of rh-renin near the fourfold axis. The large solvent channel in the center is $\sim 30 \text{ \AA}$ in diameter. Only C α -atoms of rh-renin are represented; blue, average B -factor $< 25 \text{ \AA}^2$ (the molecular average) magenta average B between 25 \AA^2 and 35 \AA^2 ; and red, average $B > 35 \text{ \AA}^2$ (poorly determined in the current structure).



ing region (Fig. 6), making it difficult to evaluate the effects of the substitution. Thus other amino acid differences may contribute to the increased pH optimum of human renin. The overall amino acid composition may be a factor. In rh-renin the ratio of negative to positive charges is almost 1:1 (33:31), whereas in pepsin this ratio is 8.5:1 (43:5). The pH optimum of human renin could also be affected by the substitution of Ala²¹⁸ for either Thr or Ser in other aspartyl proteinases (14). The side chains of these residues form a hydrogen bond to the catalytic Asp²¹⁵. Removal of this interaction should have far-reaching electrostatic consequences.

Carbohydrate attachment sites on proteins involve N-linked Asn residues in a canonical sequence Asn-X-Ser/Thr (X is any amino acid). There are two such sequences in rh-renin, at Asn⁵ and Asn⁷⁵ (in the renin numbering). Although the rh-renin was treated with endoglycosidase F (11), a sugar residue is covalently bonded to Asn⁷⁵ (Fig. 7). In our structure Asn⁵ is the NH₂-terminal, as there is no electron density for the tracing of the first four residues. An ordered

carbohydrate (if present) is also not visible either. The nature of the carbohydrate on Asn⁷⁵ is not known. We have incorporated a single β D-glucosamine residue into the structure, although other weak peaks in the electron density map could indicate a longer carbohydrate chain. As this carbohydrate is far from the active site ($\sim 25 \text{ \AA}$), carbohydrate residues probably have little influence on the catalytic function of rh-renin (11).

Stretches of polypeptide in the outermost regions of five of the surface loops of rh-renin remain poorly ordered and have weak associated electron density in the current maps (Arg⁴⁷ to Tyr^{48A}; Arg⁷⁴ to Thr⁷⁷; Glu^{158B} to Leu¹⁶¹; Arg²⁴² to Asn²⁵¹; and Glu²⁷⁹ to Cys²⁸³). Most of these residues face towards a solvent channel and have no intermolecular contacts. The most disordered parts of the structure converge into this area (Fig. 8). The channels are $\sim 30 \text{ \AA}$ in diameter, with no apparent order in their aqueous environment, and extend throughout the length of the crystals parallel to the crystallographic c -axis.

The rapid progress from the initial molecular replacement structure to the present re-

fined model could only have been achieved with the recently developed molecular dynamics refinement methods. On the basis of this renin structure the exploitation of the large database of structure-activity relations for the design of specific inhibitors should be possible. However, understanding the extreme substrate specificity of renin probably awaits the determination of an enzyme-inhibitor complex structure.

REFERENCES AND NOTES

1. M. A. Ondetti and D. W. Cushman, *Annu. Rev. Biochem.* **51**, 283 (1982).
2. K. S. Misono, J.-J. Chang, T. Inagami, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 4858 (1982).
3. J.-J. Panthier *et al.*, *Nature* **298**, 90 (1982).
4. S. Cohen, J. M. Taylor, K. Murakami, A. M. Michelakis, T. Inagami, *Biochemistry* **11**, 4286 (1972); J. P. Mornon, E. Surcouf, J. Berthou, P. Corvol, S. Foote, *J. Mol. Biol.* **155**, 539 (1982).
5. M. A. Navia, J. P. Springer, M. Poe, J. Boger, K. Hoogsteen, *J. Biol. Chem.* **259**, 12714 (1984).
6. T. Imai *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 7405 (1983).
7. J. A. Hardman *et al.*, *DNA* **3**, 457 (1984); P. M. Hobart, M. Fugliano, B. A. O'Connor, I. M. Schaefer, J. M. Chirgwin, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 5026 (1984); H. Miyazaki *et al.*, *ibid.*, p. 5999.
8. L. C. Fritz *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 4114 (1986).
9. R. A. Poorman *et al.*, *Proteins Struct. Funct. Genet.* **1**, 139 (1986).
10. C. T. Carilli, L. C. Wallace, L. M. Smith, M. A. Wong, J. A. Lewicki, *J. Chromatogr.* **444**, 203 (1988).
11. C. T. Carilli *et al.*, *Hypertension* **11**, 713 (1988).
12. R. Hamlin, *Methods Enzymol.* **114**, 416 (1985); A. J. Howard, C. Neilsen, Ng-H. Xuong, *ibid.*, p. 452.
13. M. G. Rossmann, Ed., *The Molecular Replacement Method*, vol. 13 of the *International Science Review* (Gordon and Breach, New York, 1972).
14. K. Akahane *et al.*, *Hypertension* **7**, 3 (1985); W. Carlson, M. Karplus, E. Haber, *ibid.*, p. 13; B. L. Sibanda *et al.*, *FEBS Lett.* **174**, 102 (1984).
15. M. N. G. James and A. R. Sielecki, *Nature* **319**, 33 (1986).
16. The model was oriented in the renin cell with the Crowther fast rotation algorithm [R. A. Crowther, in (13), pp. 173-178]. Its positioning was then searched by a "brute force" program BRUTE [M. Fujinaga and R. J. Read, *J. Appl. Cryst.* **20**, 517 (1987)]. The final orientation and position of the model was refined by a six-dimensional fine search using the same program.
17. R. J. Read, *Acta Cryst.* **A42**, 140 (1986).
18. W. A. Hendrickson, *Methods Enzymol.* **115**, 252 (1985).
19. A. T. Brünger, J. Kuriyan, M. Karplus, *Science* **235**, 458 (1987).
20. W. F. van Gunsteren and H. J. C. Berendsen, *GROMOS: Groningen Molecular Simulation Library* (BIOMOS B. V., Groningen, 1987).
21. M. Fujinaga, P. Gros, W. F. van Gunsteren, *J. Appl. Cryst.*, in press.
22. W. F. van Gunsteren and H. J. C. Berendsen, *Mol. Phys.* **34**, 1311 (1977).
23. C. D. Barry, C. E. Molnar, F. U. Rosenberger, *Tech. Memo. No. 229* (Computer Systems Laboratory, Washington University, St. Louis, 1976).
24. A. R. Sielecki, M. N. G. James, C. G. Broughton, in *Computational Crystallography*, D. Sayre, Ed., (Clarendon, Oxford, 1982), pp. 409-419.
25. W. S. Bennett and R. Huber, *CRC Crit. Rev. Biochem.* **15**, 291 (1984).
26. J. Tang *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **70**, 3437 (1973).
27. M. N. G. James and A. R. Sielecki, in *Biological Macromolecules and Assemblies*, vol. 3, *Active Sites of Enzymes*, F. A. Jurnak and A. McPherson, Eds. (Wiley, New York, 1987).
28. J. Tang, M. N. G. James, I.-N. Hsu, J. A. Jenkins, T.

- L. Blundell, *Nature* **271**, 618 (1978).
29. T. L. Blundell, B. T. Sewell, A. D. McLachlan, *Biochim. Biophys. Acta* **580**, 24 (1979).
30. M. N. G. James and A. R. Siclecki, *J. Mol. Biol.* **163**, 299 (1983).
31. K. Suguna *et al.*, *ibid.* **196**, 877 (1987).
32. L. Pearl and T. Blundell, *FEBS Lett.* **174**, 96 (1984).
33. T. Inagami, in *Biochemical Regulation of Blood Pressure*, R. L. Soffer, Ed. (Wiley, New York, 1981), pp. 39–73.
34. T. Blundell, B. L. Sibanda, L. Pearl, *Nature* **304**, 273 (1983).
35. T. Yamauchi, M. Nagahama, H. Hori, K. Murakami, *FEBS Lett.* **230**, 205 (1988).
36. Abbreviations for the amino acid residues are: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
37. M. G. Rossmann and P. Argos, *J. Biol. Chem.* **250**, 7525 (1975).
38. F. C. Bernstein *et al.*, *J. Mol. Biol.* **112**, 535 (1977).
39. Supported by Wyeth-Ayerst Laboratories, California Biotechnology, Inc., and the Medical Research Council of Canada.

30 June 1988; accepted 14 November 1988

A Dwarf Mutant of *Arabidopsis* Generated by T-DNA Insertion Mutagenesis

KENNETH A. FELDMANN, M. DAVID MARKS,
MICHAEL L. CHRISTIANSON, RALPH S. QUATRANO

Most plant genes that control complex traits of tissues, organs, and whole plants are uncharacterized. Plant height, structure of reproductive organs, seed development and germination, for example, are traits of great agronomic importance. However, in the absence of knowledge of the gene products, current molecular approaches to isolate these important genes are limited. Infection of germinating seeds of *Arabidopsis thaliana* with *Agrobacterium* results in transformed lines in which the integrated T-DNA from *Agrobacterium* and its associated kanamycin-resistance trait cosegregate with stable, phenotypic alterations. A survey of 136 transformed lines produced plants segregating in a manner consistent with Mendelian predictions for phenotypes altered in height, flower structure, trichomes, gametogenesis, embryogenesis, and seedling development. This report is the characterization of a dwarf mutant in which the phenotype is inherited as a single recessive nuclear mutation that cosegregates with both the kanamycin-resistance trait and the T-DNA insert.

TRANSPOSABLE ELEMENTS HAVE been successfully used to tag genes in a limited number of eukaryotic organisms. Transposons simultaneously mutagenize and tag genes when insertion of the transposable element into a DNA region modifies expression of the gene. A number of genes have been identified and isolated by this technique in *Caenorhabditis elegans* (1), *Drosophila* (2), *Antirrhinum* (3), *Zea mays* (4), and mouse (5). However, because most transposons in plants occur in multiple copies with low transposition frequencies, detection of the insert responsible for the altered phenotype is difficult. Also, relatively few plant species contain endogenous transposons. Recently, the potential of using a transposable element from maize to tag genetic loci in transgenic plants was demonstrated (6).

An alternative method of tagging genes in plants is the use of the T-DNA of *Agrobacterium tumefaciens* as the insertional mutagen. A large number of plant species can be transformed with *Agrobacterium*, which integrates a portion of its DNA (T-DNA) into the plant genome (7). However, insertional mutagenesis via *Agrobacterium*-mediated transformation is inefficient in most higher plant

species because of the abundance of noncoding DNA, which decreases the probability of insertion into a functional gene, and also because of the space required to grow hundreds of thousands of transformants and the plethora of variations introduced by the tissue culture process—that is, somaclonal variation (8). The use of *Arabidopsis thaliana* as an experimental organism to develop an *Agrobacterium*-mediated insertional mutagenesis system has several advantages. *Arabidopsis* has a haploid genome size of 70,000 kb consisting of only 10% to 14% highly repetitive DNA (9) and is sufficiently small that thousands of individually transformed plants can be grown separately to maturity in growth chambers or in a small greenhouse in a short time (4 to 6 weeks). However, somaclonal variation is a major obstacle in regenerating *Arabidopsis* (10), making it difficult to separate the effect of the T-DNA insert from the effect of the tissue culture process, even under the best transformation and regeneration conditions (11).

We previously reported a non-tissue culture method for transformation, which involves infecting germinating seeds of *Arabidopsis* with *Agrobacterium* and results in trans-

formants that do not exhibit somaclonal variation (12). An initial population of more than 1000 transformants containing an average of 1.4 functional inserts was generated by the successful application of this seed transformation technique in three separate infection experiments (12, 13). Of the 156 families scored for resistance to kanamycin, 20 lines contained inserts that failed to segregate in a Mendelian fashion on selective medium, produced families with no kanamycin-resistant (Kan^R) progeny, or were lost (12, 14). The remaining 136 lines segregated for at least 190 independent functional inserts when scored on kanamycin. Screening of the 136 transformed lines in growth chamber conditions (15) produced 36 lines that segregated in a Mendelian fashion for a variety of altered phenotypes. These phenotypes include a dwarf (Fig. 1A) and other height variants, floral structure variants (agamous, Fig. 1B), stems lacking trichomes (glabrous, Fig. 1C), and embryo lethals (Fig. 1D). Genetic and molecular analyses of the dwarf mutant described in this report are consistent with the hypothesis that the T-DNA was inserted into a gene that in the homozygous state confers dwarfism to the plant.

The dwarf, in almost all characteristics, is a miniature of the wild-type plant, 7 to 8 cm in height compared to 30 to 40 cm for the wild type. The hypocotyls of dwarf seedlings are reduced in length, and the dwarf cotyledons have a square shape in comparison to the round, wild-type cotyledons. The leaves are short, dark green, and have an epinastic growth habit (Fig. 1A). The number of rosette leaves (six to eight; refer to cover) is similar to that of the wild type. Wild-type plants have three to five primary inflorescences when grown separately in 3-inch pots. Although dwarf plants initially have only one inflorescence, additional inflorescences arise as the plants age, so that by the time the plants senesce there may be more than 40 inflorescences arising from the base of the plant (14). This phenotype indicates a reduction in apical dominance. Seed set is reduced, owing in part to a failure of the stamens to elongate, thus decreasing the amount of pollen shed on the stigma. Seeds are reduced in size, more concave than those of the wild type, and have decreased viability with increasing age (14). Wild-type plants complete senescence in 8 to 9 weeks under the described growth conditions, whereas aging of dwarf plants is delayed 1 month.

K. A. Feldmann and R. S. Quatrano, E. I. du Pont de Nemours & Co., Wilmington, DE 19880-0402.

M. D. Marks, School of Biological Sciences, University of Nebraska, Lincoln, NE 68588.

M. L. Christianson, Sandoz Crop Protection, Palo Alto, CA 94304.