B-C stacking sequence like that of the (111) planes in fcc lattices. This sequence is the origin of the tripling of the periodicity along the growth direction. For a perfect fcc lattice, the D/L ratio between the vertical separation of the dot planes and the in-plane dot distance should be equal to 0.866. However, for our sample, this ratio equals 0.704. Thus, the trigonal dot lattice can be thought of as a fcc lattice that is compressed by 18% along the body diagonal. This yields a trigonal angle α of 39.5° between the a_i and the [111] directions instead of 35.26° for the cubic case.

To obtain such a stacking sequence solely by epitaxial growth, the nucleation of dots must occur along the $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$ directions with respect to the dots in the previous layer, in contrast to the preferred nucleation along the growth direction observed for other material systems (*12, 14, 15*). This behavior can be induced only by a nonuniform in-plane strain distribution on the surface that is produced by the lattice distortions around the buried dots, with an enhanced dot nucleation probability at surface positions of lower elastic energy density Σ. For elastically isotropic materials, however, the absolute minimum of Σ occurs exactly above the previous dots (*11, 12*); that is, the dots in a multilayer should be aligned vertically. To resolve this issue, we performed elastic strain field calculations that accounted for the elastic anisotropy of the lattice. To simplify these calculations, we replaced the actual dot with an infinitely small "force nucleus" and neglected the surface relaxation of the internal stresses. The in-plane strain on the surface was then calculated with a Fourier method to solve the Navier equation (*16*). As shown in Fig. 4, the calculated elastic energy density Σ on a (111) $Pb_{1-x}Eu_xTe$ surface with a PbSe dot at 47 nm below the surface exhibits a $3m$ point symmetry with three pronounced energy minima that are displaced laterally from the subsurface force nucleus. These minima coincide almost exactly with the experimentally observed dot positions (Fig. 4). In addition, a saddle point of Σ occurs in the center, above the buried dot, which indicates that this point is not favorable for the growth of a subsequent dot.

The directions of the Σ minima are determined only by the elastic anisotropy of the matrix material. Therefore, the symmetry of the dot arrangement should be essentially independent of the spacer thickness. To verify this conclusion, we fabricated a series of PbSe/ $Pb_{1-x}Eu_xTe$ dot superlattices with spacer thicknesses that varied from 35 to 60 nm. The lateral and vertical separation L and D, respectively, of the dots in each sample was determined from RSMs. We found that, for D_{SL} varying almost by a factor of 2, the trigonal angle α of the dot lattice indeed remains nearly constant (Fig. 5). Moreover, the lateral dot distance L and the dot crystal lattice constant a change linearly with spacer thickness (Fig. 5). Thus, the PbSe dot samples represent macroscopic crystals with a continuously tunable lattice constant. Furthermore, for a constant-deposited PbSe thickness, with varying L (varying dot density), the average dot volume changes with D_{SL} as well.

Because of their atomic-like density of states, semiconductor quantum dots offer many advantages for optoelectronic devices. For such applications, however, a high density of dots and a narrow size distribution are of crucial importance (*1, 6, 10*). The excellent control of absolute size, as well as of lateral distance that can be achieved by the growth of quantum-dot superlattices, thus allows great improvements for the precise tuning of the optical and electronic properties of self-assembled quantum dots. This opens promising perspectives for their applications in semiconductor diode lasers (*1, 6, 10*), intersubband infrared detectors (*17*), and microcavity devices (*18*).

References and Notes

1. A. P. Alivisatos, *Science* **271**, 933 (1996), and references therein.
2. C. B. Murray, C. R. Kagan, M. G. Bawendi, *ibid.* **270**, 1335 (1995).
3. D. Leonhard *et al.*, *Appl. Phys. Lett.* **63**, 3203 (1993); J. M. Moison *et al.*, *ibid.* **64**, 196 (1994).
4. R. Nötzel, J. Temmyo, T. Tamamura, *Nature* **369**, 131 (1994).
5. R. Leon P. M. Petroff, D. Leonard, S. Fafard, *Science* **267**, 1966 (1995).
6. S. Fafard *et al.*, *ibid.* **274**, 1350 (1996).
7. R. J. Asaro and W. A. Tiller, *Metall. Trans.* **3**, 1789 (1972); D. J. Srolovitz, *Acta Metall.* **37**, 621 (1989); C. Ratsch and A. Zangwill, *Surf. Sci.* **293**, 123 (1993).
8. D. J. Eaglesham and M. Cerullo, *Phys. Rev. Lett.* **64**, 1943 (1990); S. Guha, A. Madhukar, K. C. Rajkumar, *Appl. Phys. Lett.* **57**, 2110 (1990).
9. J. Y. Marzin *et al.*, *Phys. Rev. Lett.* **73**, 716 (1994); M. Grundman *et al.*, *ibid.* **74**, 4043 (1995); L. Landin, M. S. Miller, M.-E. Pistol, C. E. Pryor, L. Samuelson, *Science* **280**, 262 (1998).
10. N. Kirstaedter *et al.*, *Electron. Lett.* **30**, 1416 (1994); V. Y. Shchukin *et al.*, *Phys. Rev. B* **57**, 12262 (1998).
11. J. Tersoff, C. Teichert, M. G. Lagally, *Phys. Rev. Lett.* **76**, 1675 (1996).
12. Q. Xie, A. Madhukar, P. Chen, N. Kobayashi, *ibid.* **75**, 2542 (1995).
13. M. Pinczolits, G. Springholz, G. Bauer, *Appl. Phys. Lett.* **73**, 250 (1998).
14. A. Darhuber *et al.*, *Phys. Rev. B* **55**, 15652 (1997).
15. G. S. Solomon, J. A. Trezza, A. F. Marshall, J. S. Harris, *Phys. Rev. Lett.* **76**, 952 (1996).
16. R. A. Masumura and G. Sines, *J. Appl. Phys.* **41**, 3930 (1970); J. W. Flocken and J. R. Hardy, *Phys. Rev. B* **1**, 2447 (1970).
17. S. Sauvage *et al.*, *Appl. Phys. Lett.* **71**, 2785 (1997).
18. J. M. Gérard *et al.*, *Phys. Rev. Lett.* **81**, 1110 (1998).

19. This work was supported by the Austrian Science Fund (grant 11557), the Jubiläumfonds of the Austrian National Bank (grant 6690), the Gesellschaft für Mikroelektronik Austria, the grant agency of the Czech Republic (grant 202/97/0003), and the Ministry of Education of the Czech Republic (grant VS96102).

27 July 1998; accepted 15 September 1998

Rapid Identification of Subtype-Selective Agonists of the Somatostatin Receptor Through Combinatorial Chemistry

Susan P. Rohrer,* Elizabeth T. Birzin, Ralph T. Mosley, Scott C. Berk, Steven M. Hutchins, Dong-Ming Shen, Yusheng Xiong, Edward C. Hayes, Rupa M. Parmar, Forrest Foor, Sudha W. Mitra, Sylvia J. Degrado, Min Shu, John M. Klopp, Sheng-Jian Cai, Allan Blake, Wanda W. S. Chan, Alex Pasternak, Lihu Yang, Arthur A. Patchett, Roy G. Smith, Kevin T. Chapman, James M. Schaeffer

Nonpeptide agonists of each of the five somatostatin receptors were identified in combinatorial libraries constructed on the basis of molecular modeling of known peptide agonists. In vitro experiments using these selective compounds demonstrated the role of the somatostatin subtype-2 receptor in inhibition of glucagon release from mouse pancreatic alpha cells and the somatostatin subtype-5 receptor as a mediator of insulin secretion from pancreatic beta cells. Both receptors regulated growth hormone release from the rat anterior pituitary gland. The availability of high-affinity, subtype-selective agonists for each of the somatostatin receptors provides a direct approach to defining their physiological functions.

Somatostatin is distributed throughout the endocrine system and has multiple physiological functions including inhibition of secretion of growth hormone (*1*), glucagon (*2*), insulin (*3*), gastrin, and other hormones secreted by the pituitary and gastrointestinal tract (*4*). It also acts as a neuromodulatory peptide in the central nervous system (*5*) and has been im-

plicated as an inhibitor of tumor cell growth acting through somatostatin receptors coupled to tyrosine and serine-threonine phosphatases (*6*).

Somatostatin occurs naturally in two major forms: a tetradecapeptide (ss-14) and a 28–amino acid form (ss-28), both of which bind to somatostatin receptors that are coupled to heterotrimeric guanine nucleotide–binding proteins. Five distinct somatostatin receptors (sst1 through sst5) have been cloned from human tissues (*7*). Antibody probes have been developed for the sst2 and sst5 receptor subtypes and have been useful for establishing their potential functions (*8–10*). The sst2 receptor is localized with glucagon in rat pancreatic α cells (*8*), and the sst5 receptor localizes with insulin in the pancreatic β cells (*9*). In the pituitary gland, somatotrophs have sst2 and sst5 receptors, indicating that both receptor subtypes may have a role in regulation of growth hormone secretion (*10*).

In receptor-ligand binding assays, short peptide analogs of somatostatin, including MK-678 and octreotide, display selectivity for the sst2 receptor (*11*). These probes have been useful in assigning specific functions such as inhibition of secretion of growth hormone and glucagon to this particular receptor subtype (*12, 13*). However, the lack of somatostatin analogs with selectivity for the sst1, sst3, sst4, and sst5 receptors has impeded progress in understanding functions associated with these receptors.

Our objective was to identify nonpeptide agonists selective for each sst receptor subtype. We used an integrated approach of combinatorial chemistry and high-throughput receptor-binding techniques to rapidly identify subtype-selective compounds. A cyclic hexapeptide somatostatin agonist (L-363,377) (Fig. 1) was used as a probe in a search of the Merck chemical collection, which consists of approximately 200,000 randomly assembled compounds. The side chains of residues Tyr-Trp-Lys in the cyclic peptide were given priority in the search on the basis of their similarity to the pharmacophore created by residues Trp^8-Lys^9 in the somatostatin-14 peptide (Ala^1-Gly^2-Cys^3-Lys^4-Asn^5-Phe^6-Phe^7-Trp^8-Lys^9-Thr^{10}-Phe^{11}-Thr^{12}-Ser^{13}-Cys^{14}). Seventy-five compounds were selected based on their similarity to the pharmacophore for evaluation in binding assays. The most potent of these compounds (L-264,930) had an apparent inhibition constant of 100 nM for the human sst2 receptor. L-264,930 is tripartite in structure with an aromatic moiety, a tryptophan moiety, and a diamine moiety, making it quite amenable to a combinatorial chemistry approach.

Combinatorial library #1 was synthesized with 79 different substituents representing the aromatic moiety. Twenty different substituents for the Trp-amino acid module and 20 different substitutions for the diamine moiety yielded a library with $79 \times 20 \times 20$ possible compounds, each occurring as multiple stereoisomers for a total of approximately 130,000 compounds. This library was screened against all five receptors in ligand-binding assays adapted to a 96-well format. One of the most active and sst2-selective groups of compounds contained a family of 1330 spiroindane analogs with the original lead structure (L-264,930) among them. However, this group of compounds was not analyzed further, because traditional medicinal chemistry efforts already identified potent sst2-selective leads from this class (*14*). Another potent and sst2-selective group that consisted of 1330 compounds, related by virtue of a benzimidazolone substitution at the aromatic position rather than a spiroindole substitution, was chosen for further analysis.

In the second round of screening, 20 samples were examined, each of which contained a family of compounds with the same amino acid substitution linked to the benzimidazolone and to all possible diamine substitutions. The family of compounds bearing β-methyl tryptophan had greater potency than those with other substitutions for the Trp.

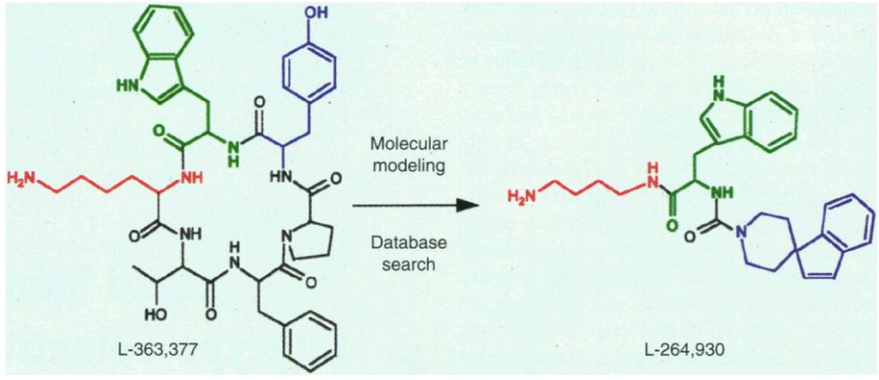

Fig. 1. Molecular modeling of the somatostatin pharmacophore. The three-dimensional model of L-363,377 was used as a probe in a search of the Merck sample collection database. The coloration of the molecules depicted in two dimensions schematically highlights the superposition between the residue side chains with the correspondingly colored moieties of L-264,930. Substitutions of the aromatic group are shown in blue, of the tryptophan in green, and of the diamine in red. The search strategy is described in greater detail elsewhere (*11*).

Table 1. Ligand binding assays. Receptor-ligand binding assays were done with membranes isolated from CHO-K1 cells expressing each of the cloned human sst receptors. Results shown are expressed as K_i values (in nanomoles). The ligand 3-[^{125}I]iodotyrosyl^{25}–somatostatin-28(Leu^8, D-Trp^{22}, Tyr^{25}) was obtained from Amersham and was used at a final concentration of 0.1 nM. Assays were done in 96-well polypropylene plates in a final volume of 200 μl. The assay buffer consisted of 50 mM tris-HCl (pH 7.8), 1 mM EGTA, 5 mM $MgCl_2$, leupeptin (10 μg/ml), pepstatin (10 μg/ml), bacitracin (200 μg/ml), and aprotinin (0.5 μg/ml). Test compounds were examined over a range of concentrations from 0.01 to 10,000 nM. CHO-K1 cell membranes, ligand, and test compound were incubated at room temperature for 45 min and then collected on Packard 96-well glass fiber filter plates treated with 0.1% polyethyleneimine. Plates were washed with cold 50 mM tris-HCl (pH 7.8), then dried before counting. K_i values were calculated with the Cheng-Prussof equation (*17*). K_d values for 3-[^{125}I]iodotyrosyl^{25}–somatostatin-28(Leu^8, D-Trp^{22}, Tyr^{25}) with each of the five receptors were determined from Scatchard plots of saturation binding curves. The apparent dissociation constants were 1.5, 0.1, 0.4, 4.2, and 0.7 nM for receptors sst1 through sst5, respectively.

Compound	hsstr1	hsstr2	hsstr3	hsstr4	hsstr5
ss-14	0.4	0.04	0.7	1.7	2.3
L-363,377	5664	0.5	3072	>10,000	2009
L-797,591	1.4	1875	2240	170	3600
L-779,976	2760	0.05	729	310	4260
L-796,778	1255	>10,000	24	8650	1200
L-803,087	199	4720	1280	0.7	3880
L-817,818	3.3	52	64	82	0.4

S. P. Rohrer, E. T. Birzin, E. C. Hayes, R. M. Parmar, F. Foor, S. W. Mitra, S.-J. Cai, A. Blake, W. W. S. Chan, R. G. Smith, J. M. Schaeffer, Department of Cell Biochemistry and Physiology, Merck Research Laboratories, Post Office Box 2000, Rahway, NJ 07065, USA. R. T. Mosley, S. C. Berk, S. M. Hutchins, D.-M. Shen, Y. Xiong, S. J. Degrado, M. Shu, J. M. Klopp, K. T. Chapman, Department of Molecular Design and Diversity, Merck Research Laboratories, Post Office Box 2000, Rahway, NJ 07065, USA. A. Pasternak, L. Yang, A. A. Patchett, Department of Medicinal Chemistry, Merck Research Laboratories, Post Office Box 2000, Rahway, NJ 07065, USA.

*To whom correspondence should be addressed. E-mail: susan_rohrer@merck.com

In the final round of screening, a restricted diamine moiety was identified as the most favorable for sst2 potency and selectivity. The compound isolated from this series of experiments has an inhibition constant (K_i) of 0.05 nM and greater than 6000-fold selectivity for the hsst2 receptor (Table 1). This compound (L-779,976) (Fig. 2) is a potent agonist that inhibited forskolin-stimulated accumulation of cyclic adenosine 3′,5′-monophosphate (cAMP) in Chinese hamster ovary cells (CHO-K1) expressing the sst2 receptor (Table 2). L-779,976 is also a potent inhibitor of growth hormone secretion from rat pituitary cells with a median inhibitory concentration (IC_{50}) of 0.025 nM and comparable efficacy to that of somatostatin-14 (Table 2). Arginine-stimulated secretion of glucagon from mouse pancreatic islet cells was inhibited with an IC_{50} of 0.1 nM, similar to the binding affinity of L-779,976 for the sst2 receptor. Inhibition of insulin secretion from mouse pancreatic islets required a 1000-fold higher concentration of L-779,976 (Table 2).

Combinatorial library #2 consisted of 350,000 different compounds. It was constructed according to the same principle as combinatorial library #1, but was expanded in all three dimensions (147 × 22 × 21). Of the 147 pools of compounds screened in the receptor-binding assays, two were chosen for further analysis, which resulted in identification of compounds selective for the sst1 and sst3 receptors (Fig. 2). In CHO-K1 cells expressing the hsst3 receptor, L-796,778 was a partial agonist with an IC_{50} value of 18 nM for inhibition of forskolin-stimulated production of cAMP. Efficient coupling of the hsst1 receptor to adenylate cyclase was not achieved in the CHO-K1 cell line (*15*). Therefore, an alternative cell line was chosen for evaluation of the sst1-selective compound. In the mouse L-cell line (*16*) transfected with the hsst1 receptor and a cAMP response element (CRE)–β-galactosidase reporter, L-797,591 displayed agonist activity with an IC_{50} of 3 nM. Neither of these compounds inhibited release of growth hormone, glucagon, or insulin (Table 2).

An sst4 receptor–selective compound was isolated from combinatorial library #3, an aryl indole library of limited complexity. The compound, L-803,087, has a diamine moiety that maps to lysine on the pharmacophore (Fig. 2), but relation of this molecule to the aromatic and the Trp substituents of the pharmacophore is not obvious. In binding and functional assays, L-803,087 was an hsst4 receptor agonist. L-803,087 did not inhibit secretion of growth hormone, insulin, or glucagon (Table 2).

An sst5-selective ligand was identified in combinatorial library #4, another extension of library #1. L-817,818 was 130-fold more selective for sst5 as compared to sst2 and eightfold more selective for sst1 (Table 1). The compound inhibited growth hormone secretion from rat primary pituitary cell cultures, with an IC_{50} of 3 nM, comparable to its binding affinity for sst5. Because a role for sst1 in growth hormone secretion was ruled out with the sst1-selective compound L-797,591, sst5 apparently participates in control of growth hormone release. L-817,818 also inhibited insulin secretion from pancreatic islets but did not inhibit glucagon secretion (Table 2).

Combinatorial chemistry proved useful for rapid refinement of new lead compounds. By adapting the sst receptor ligand binding assays to a 96-well format and screening complex mixtures, we were able to examine hundreds of thousands of new chemical entities in five different somatostatin receptor assays in a short period of time and to identify sst receptor–selective compounds that may be useful for dissecting functions of the individual receptors. These nonpeptide, small-molecule analogs of somatostatin may be useful in development of orally active chemotherapeutic agents capable of crossing the blood-brain barrier.

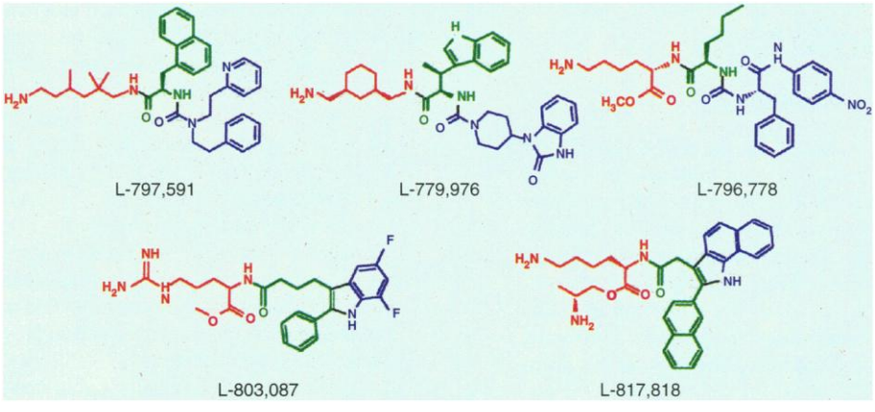

Fig. 2. Subtype-selective nonpeptide agonists of the somatostatin receptor. Color coding of sst receptor–selective compounds illustrates the relationship of various parts of each molecule to the original lead structure L-264,930. Colors are as in Fig. 1. L-797,591, L-779,976, L-796,778, L-803,087, and L-817,818 are selective for the sst1, sst2, sst3, sst4, and sst5 receptors, respectively.

Table 2. Functional activity of the sst receptor–selective somatostatin analogs. cAMP accumulation in cells expressing the sst receptors (*18*): Each receptor subtype was evaluated in at least three separate experiments. CRE–β galactosidase assay: The assay protocol used for measuring sst1-mediated inhibition of cAMP accumulation has been previously described (*11*). Somatostatin-14 had an IC_{50} of 0.2 nM in this assay. The sst1-selective compound was evaluated in duplicate in two separate experiments. Growth hormone (GH) release from primary cultures of rat pituitary cells: Primary cultures of rat anterior pituitary cells were maintained for 3 to 4 days at 37°C in 5% CO_2 and 95% air. Cells were treated with test compounds for two hours at 37°C. Secreted growth hormone was detemined by radioimmunoassay (*11*). Compounds were evaluated in two separate experiments. Mouse pancreatic islet preparations: The protocol for pancreatic islet preparation and measurement of secreted glucagon and insulin has been described previously (*11*). The sst1-selective compound was examined once in triplicate. All other compounds were evaluated in triplicate in two separate experiments. CHO K1, CHO K1 cells; RPC, rat pituitary cells; MI, mouse islets; ss-14, somatostatin-14.

Compound	SSTR selectivity	cAMP accumulation (CHO K1)*	GH release assay (RPC)*	Glucagon release (MI)*	Insulin release (MI)*
ss-14	–	–	0.04	1	1
L-797,591	sstr1	3.0†	>1000	>100	>100
L-779,976	sstr2	0.05	0.025	0.1	100
L-796,778	sstr3	18	>1000	>100	>100
L-803,087	sstr4	0.2	>1000	>100	>100
L-817,818	sstr5/1	1.3	3.1	>30	0.3

*Values given are IC_{50}s (in M). †Determined in mouse L cells transfected with hsst1 and CRE–β gal reporter.

References and Notes

1. P. Brazeau *et al.*, *Science* **179**, 77 (1973).
2. S. Reichlin, *N. Engl. J. Med.* **309**, 1495 (1983); M. Brown, J. Rivier, W. Vale, *Science* **196**, 1467 (1977).
3. N. Vaysse *et al.*, *Endocrinology* **108**, 1843 (1981); L. Mandarino *et al.*, *Nature* **291**, 76 (1981).
4. S. Bloom *et al.*, *Gut* **16**, 834 (1975); S. Bloom *et al.*, *Lancet* **ii**, 1106 (1974); W. Vale *et al.*, *Endocrinology* **95**, 968 (1974).
5. J. Epelbaum, *Prog. Neurobiol.* **27**, 63 (1986).
6. C. Liebow *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 2003 (1989); M. G. Pan, T. Florio, P. J. S. Stork, *Science* **256**, 1215 (1992); L. Buscail *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 2315 (1994); M. Le Romancer *et al.*, *J. Biol. Chem.* **269**, 17464 (1994); R. E. White, A. Schonbrunn, D. L. Armstrong, *Nature* **351**, 570 (1991).
7. Y. Yamada *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 251 (1992); F.-W. Kluxen, C. Bruns, H. Lubbert, *ibid.*, p. 4618;

M. Vanetti *et al.*, *FEBS Lett.* **311**, 290 (1992); K. Yasuda *et al.*, *J. Biol. Chem.* **267**, 20422 (1992); W. Meyerhof *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 10267 (1992); J. F. Bruno, Y. Xu, J. Song, M. Berelowitz, *ibid.*, p. 11151; L. Rohrer *et al.*, *ibid.* **90**, 4196 (1993); A. M. O'Carroll *et al.*, *Mol. Pharmacol.* **42**, 939 (1992).
8. B. Hunyady *et al.*, *Endocrinology* **138**, 2632 (1997).
9. S. W. Mitra *et al.*, *ibid.*, in press.
10. E. Mezey *et al.*, *ibid.* **139**, 414 (1998).
11. L. Yang *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 10836 (1998).
12. K. Raynor *et al.*, *Mol. Pharmacol.* **43**, 838 (1993).
13. W. J. Rossowski and D. H. Coy, *Biochem. Biophys. Res. Commun.* **205**, 341 (1994).
14. L. Yang *et al.*, *J. Med. Chem.* **41**, 2175 (1998).
15. Coupling of sst1 to adenylate cyclase was not observed with our hsst1 CHO-K1 cell line. Our result is consistent with results obtained by S. Rens-Domiano *et al.* [*Mol. Pharmacol.* **42**, 28 (1992)] and K. Raynor *et al.* (*12*). In contrast, Y. C. Patel, M. Greenwood, A. Warszynska, R. Panetta, and C. B. Srikant [*Biochem. Biophys. Res. Comm.* **198**, 605 (1994)] have observed coupling of the sst1 receptor to adenylate cyclase in CHO cells.
16. L cells are tissue culture cells originally derived from murine connective tissue. They are fibroblast-like in their morphology. The cell line was developed in the mid-1940s. The letter L is a strain designation assigned by the developers of the line.
17. Y. C. Cheng and W. H. Prusoff, *Biochem. Pharmacol.* **22**, 3099 (1973).
18. CHO-K1 cells stably expressing the sst2, sst3, sst4, and sst5 receptors were subcultured in 12-well plates and treated at 37°C for 30 min with growth medium containing 0.5 mM isobutylmethylxanthine. The medium was replaced with fresh growth medium, with or without 10 μM forskolin and test agents. After incubation for 5 min at 37°C, the medium was removed, and the cells were lysed by freeze-thaw in 0.1 M HCl. The cAMP content of duplicate wells was determined with a radioimmunoassay kit (Amersham). Data obtained from the dose-response curves were analyzed by nonlinear regression with GraphPad Prism, version 2.01 (GraphPad Software, San Diego, CA).
19. We thank K. Cheng, L.-Y. Pai, and T.-J. Wu for growth hormone secretion assays.

19 June 1998; accepted 22 September 1998

Pathways to a Protein Folding Intermediate Observed in a 1-Microsecond Simulation in Aqueous Solution

Yong Duan and Peter A. Kollman*

An implementation of classical molecular dynamics on parallel computers of increased efficiency has enabled a simulation of protein folding with explicit representation of water for 1 microsecond, about two orders of magnitude longer than the longest simulation of a protein in water reported to date. Starting with an unfolded state of villin headpiece subdomain, hydrophobic collapse and helix formation occur in an initial phase, followed by conformational readjustments. A marginally stable state, which has a lifetime of about 150 nanoseconds, a favorable solvation free energy, and shows significant resemblance to the native structure, is observed; two pathways to this state have been found.

Elucidation of the mechanism of protein folding is an important step in understanding the relation between sequence and structure of proteins. Understanding of the mechanism should allow more accurate prediction of protein structures, with wide-ranging implications in biochemistry, genetics, and pharmaceutical chemistry. The recent hypothesis of folding-related diseases is another example of the significance of folding (*1*). Yet despite great progress made by a variety of experimental and theoretical studies after decades of extensive research, it has been difficult to establish detailed descriptions of the folding process and mechanism (*2*).

Computer simulation of molecular systems can provide rich information at various levels of resolution, and this approach has been important in attempts to understand protein folding mechanisms. A simplified representation might treat the protein residues as (one or two) linked beads (*3*). Higher resolution models represent most or all of the atoms of the protein explicitly, with an implicit representation of the solvent (*4*, *5*). At an even higher level of detail are molecular dynamics (MD) simulations with full atomic representation of both protein and solvent. Such calculations are uniquely suited to the study of protein folding because of their resolution and accuracy. The simulation parameters (that is, the force field) are derived from experiments and from gas-phase quantum mechanical calculations and have been tested in smaller systems in many critical comparisons with experimental results (*6*). Because of the complexity of the representation, the large number of atoms (often exceeding 10,000), and the need to take time steps of 1 to 2 fs, such simulations have, to date, been limited to a few nanoseconds (*7*) (or a few million integration steps). This has precluded the simulation of even the early stages of protein folding. Nevertheless, insights have been gained from unfolding simulations (*8*) of the denaturation process. Attempts have also been made to construct the folding free-energy landscape from unfolding simulations (*9*). Direct folding simulations with this approach, however, have been limited to small peptide fragments and have been carried out for as long as 50 ns (*10*). Direct simulation of the protein folding process with such an approach has not been considered possible (*11*) "either now or in the foreseeable future" (*12*, p. 29). By using a Cray T3E, a massively parallel supercomputer consisting of hundreds of central processing units (CPUs) connected by low-latency, high-speed, and high-availability networks, with an efficiently parallelized program that scales well to the 256-CPU level for small protein–solvent systems and is six times faster than a typical current state-of-the-art program (*13*), we have conducted a 1-μs simulation at 300 K on the villin headpiece subdomain, a 36-residue peptide (HP-36) (*14*, *15*), starting from a fully unfolded extended state (Fig. 1A), including ~3000 water molecules (*16*–*18*). The simulation time scale is close to that required to fold small proteins. The simulation shows a mechanism for the protein to find a folding intermediate, an important step in proceeding to its fully folded state.

Proteins can have marginally stable nonnative states that are difficult to observe experimentally (*19*). Computer simulation can play an important role in identifying these structures because of its extremely high time resolution and detailed atomic level representation. Recent experimental studies suggest that the time for (small) proteins to reach their marginally stable states with partially formed secondary structures is on the order of 10 μs (*20*). It also has been shown that a small protein can fold within 20 μs (*21*), and it has been estimated that the lower limit of the folding time is 1 μs (*22*).

HP-36 is one of the smallest proteins that can fold autonomously. It contains only naturally occurring amino acids and does not require disulfide bonds, oligomerization, or ligand binding for stabilization; its melting temperature is above 70°C in aqueous solution (*14*). The estimated folding time of the protein is between 10 and 100 μs (*23*), which would make it one of the fastest folding proteins. Nuclear magnetic resonance (NMR) studies of the 36-residue subdomain revealed three short helices (Fig. 1C) (*15*). We refer to them as helices 1, 2, and 3, for residues 4 to 8, 15 to 18, and 23 to 30, respectively, as found in the NMR structure. They are held together by a loop (residues 9 to 14), a turn (residues 19 to 22), and

Department of Pharmaceutical Chemistry, University of California, San Francisco, CA 94143, USA.

*To whom correspondence should be addressed. E-mail: pak@cgl.ucsf.edu