Synthesis, Biological Activity, and Hologram Quantitative Structure−Activity Relationships of Novel Allatostatin Analogues[†]

ZHEN-PENG KAI,[§,#,ǁ] JUAN HUANG,[§,ǁ] YONG XIE,[§] STEPHEN S. TOBE,*[,#] YUN LING,[§]
LI ZHANG,[§] YI-CHEN ZHAO,[§] AND XIN-LING YANG*[,§]

[§]Department of Applied Chemistry, College of Science, China Agricultural University, Beijing 100193,
People's Republic of China, and [#]Department of Cell and Systems Biology, University of Toronto,
25 Harbord Street, Toronto, Ontario, Canada M5S 3G5. [ǁ] Both authors contributed equally to this paper.

Cockroach-type allatostatins (FGLamides) (ASTs) can inhibit the production of juvenile hormone in vitro, and they therefore are regarded as possible insect growth regulator (IGR) candidates for pest control. However, several shortcomings, such as the absence of in vivo effects, rapid degradation, and high production costs, preclude their practical use in pest management. To discover new IGRs, 25 novel analogues of pentapeptide (Y/FXFGLa) were designed and synthesized with different aromatic acids, fatty acids, and dicarboxylic acids as the Y/FX region replacements on the basis of previous results. Their bioactivities in vitro were determined, and the results showed that eight analogues (**K14**, **K15**, **K17**, **K18**, **K19**, **K23**, **K24**, and **K25**) were more active than the lead, core region pentapeptide. The IC_{50} values of **K15** and **K24** (IC_{50} = 1.79 and 5.32 nM, respectively) were even lower than that of the natural AST, Dippu-AST 1(IC_{50} = 8 nM), which indicated both analogues have better activity than Dippu-AST 1; particularly, **K15** has better activity than most natural Dippu-ASTs. A predictable and statistically meaningful hologram quantitative structure−activity relationship (HQSAR) model of 32 AST analogues (28 as training sets and 4 as test sets) was obtained. The final model suggested that a potent AST analogue should contain an aromatic group, a linker of appropriate length, and the FGLa portion. These results will be useful in the design of new AST analogues that are structurally related to the training set compounds.

KEYWORDS: Allatostatin; AST analogues; HQSAR; juvenile hormone; insect growth regulators; cockroach

INTRODUCTION

The juvenile hormones (JH) play a significant role in insect development, reproduction, sex pheromone production, and mating (*1*) among other functions. In many insect species, a reduction in endogenous titer of JH is critical for metamorphosis from the nymph to the adult stage. Therefore, any substance with the ability to modulate the release of JH can be regarded as a potential candidate insect growth regulator (IGR). The cockroach allatostatins (ASTs) are a family of neuropeptides, 6−18 amino acids in length, originally isolated from the cockroach, *Diploptera punctata* (*2−5*). ASTs can inhibit the biosynthesis of JH by the corpora allata (CA) in vitro (*2, 6*). Subsequent studies show that these molecules also exhibit other biological properties, including modulation of myotropic activity in both gut (*7*) and dorsal vessel (*8*), inhibition of vitellogenin production in the fat body (*9*), and stimulation of carbohydrate enzyme activity in the midgut (*10*). The cockroach-type ASTs are regarded as the

potential leads for the discovery of new IGRs, because of their inhibitory action on JH biosynthesis, but some limitations, such as inactivation by peptidases, poor absorption and transportation in the insect, and high production cost (*11*), preclude their use in pest management. To overcome the above issues, structural modifications of natural ASTs and structure−activity relationships (SARs) have been carried out in recent years.

Earlier structure−activity relationships indicated that cockroach-type ASTs share the common C-terminus Y/FXFGL-NH₂. This pentapeptide sequence is the minimal sequence required for functional inhibition of JH in vitro biosynthesis. The C-terminus is generally regarded as the "active core" region responsible for direct receptor interaction (*12*). Hayes et al. (*12*) indicated that Leu8, Phe6, and Tyr4 in Dippu-AST 4 (DRLYSFGLa) are the most important side chains for inhibition of JH biosynthesis with the Ala scanning technique. Ketomethylene and methyleneamino pseudopeptide analogues of allatostatins Dippu-AST 5 were also designed by Piulachs et al., who altered the nature of the peptide bond with the aim of reducing the susceptibility of the bond to hydrolysis (*13*). Bioassay showed that both analogues were similarly active to the model peptides with respect to the inhibition of JH in vitro biosynthesis from CA of

---

*Authors to whom correspondence should be addressed (telephone +86-10-62732223; fax 86-10-62732223; e-mail yangxl@cau.edu.cn or stephen.tobe@utoronto.ca).

Published on Web 12/02/2009

virgin *Blatella germanica*. The first, third, or fourth amino acid residues of the C-terminal pentapeptide were substituted by sterically hindered amino acids or aromatic acids in several ASTs analogues (*11*). The bioassay results showed that those analogues retained significant bioactivity. To reduce susceptibility to metabolic inactivation, Dippu-AST 6 analogues were designed both to eliminate those amino acids that are not required for biological activity by replacement with nonpeptide moieties and to increase the rigidity of the molecule (*14*). Radiochemical assay for JH biosynthesis indicated that the activities of these analogues approached those of the native neuropeptides (*11*, *15*). The above efforts emphasize the bright future for the discovery of IGRs that are neuropeptide-like but not readily degradable by peptidases.

We first chose the core region pentapeptide (Y/FXFGLa) as the lead compound and designed three series of its analogues with the peptidomimetics strategy. The three types of pentapeptide analogues were designed respectively by way of using aromatic acids and dicarboxylic fatty acids as mimics of the Tyr/Phe-Xaa region in series I, replacing the Phe-Gly region with amino-Indane carboxylic acid, L-1,2,3,4-tetrahydroisoquinoline-3-carboxylic acid and L-proline in series II to validate the hypothesis of the turn conformation in the core region and reducing the C-terminal Leu-NH$_2$ region in series III to clarify its function. The bioassay revealed that the Y/FX mimics showed significant effects both in vitro and in vivo. The SARs of the analogues suggest that the FGLa sequence is more important for bioactivity than the Y/FX portion. Y/FX can be mimicked by substituents, whereas the FGLa sequence must be retained in the structure (*16*). To find more Y/FX mimics with higher activity, 25 novel pentapeptide analogues designed with different aromatic acids, fatty acids, and dicarboxylic acids to replace the Y/FX region are reported in the present work. For the further purpose of investigating theoretically the structure factors influencing activity, their HQSAR is also studied in this paper.

EXPERIMENTAL PROCEDURES

Synthesis. *Materials*. Rink Amide-AM resin (0.52 mmol/g substitution), HOBt (1-hydroxybenzotriazole anhydrate), HBTU (*O*-benzotriazole-*N,N,N',N'*-tetramethyluronium hexafluorophosphate), DIEA (*N, N*-diisopropylethylamine), TFA (trifluoroacetic acid), Fmoc-protected amino acids, and Z-Gly-OH were purchased from GL Biochem (Shanghai) Ltd. HPLC grade DMF (*N,N*-dimethylformamide), DCM (dichloromethane), and acetonitrile were purchased from DIMA Technology Inc. (USA). Benzoic acid, substituted benzoic acids, hydrocinnamic acid, 4-phenylbutanoic acid, 5-phenylpentanoic acid, 4-methylbenzene-1-sulfonyl chloride, hexanoic acid, and glutaric acid were purchased from Alfa Aesar, USA. (*E*)-Cinnamic acid and substituted (*E*)-cinnamic acids were prepared using the method of Sun et al. (*17*).

Synthesis of Analogues. (*a*) FGL with Resin and GFGL with Resin. These two key intermediates were synthesized from Rink Amide-AM resin (198 mg, 0.1 mmol) using standard Fmoc/tBu chemistry and an HBTU/HOBt protocol (*18*). Incoming amino acids (Leu, Gly, and Phe) were activated with HOBt (41 mg, 0.3 mmol), HBTU (114 mg, 0.3 mmol), and DIEA (105 μL, 0.6 mmol) in DMF (5 mL) for 5 min, and couplings were run for 2 h. Removal of the N-terminal Fmoc group from the residues was accomplished with 20% piperidine in DMF (5 mL) for 20 min.

(*b*) **K1**. Hydrocinnamic acid (45 mg, 0.3 mmol) was coupled to the FGL with resin with HOBt, HBTU, and DIEA in DMF for 3 h at room temperature. **K1** was cleaved from the resin with TFA containing 5% phenol, 2.5% thioanisole, and 5% water for 2 h.

(*c*) **K2−K10**, **K12**, and **K13** were obtained from the relevant acids using the same method as for **K1**.

(*d*) **K11** was synthesized from Rink Amide-AM resin (198 mg, 0.1 mmol) using the standard Fmoc/tBu chemistry and HBTU/HOBt protocol (*18*). Incoming amino acids (0.3 mmol) were activated with HOBt

(41 mg, 0.3 mmol), HBTU (114 mg, 0.3 mmol), and DIEA (105 μL, 0.6 mmol) in DMF (5 mL) for 5 min, and couplings were run for 2 h. Removal of the N-terminal Fmoc group from the residues was accomplished with 20% piperidine in DMF (5 mL) for 20 min. **K11** was cleaved from the resin with TFA (4 mL) containing 5% phenol, 2.5% thioanisole, and 5% water for 2 h.

(*e*) **K14**. (*E*)-Cinnamic acid (44 mg, 0.3 mmol) was coupled to the GFGL with resin with HOBt, HBTU, and DIEA in DMF for 3 h at room temperature. **K14** was cleaved from the resin with TFA containing 5% phenol, 2.5% thioanisole, and 5% water for 2 h.

(*f*) **K15−K24** were obtained from the relevant acids using the same method as for **K14**.

(*g*) **K25** Benzyloxycarbonylglycine (Z-Gly-OH) (627 mg, 0.1 mmol) was coupled to the FGL with resin (0.1 mmol) with HOBt, HBTU, and DIEA in DMF for 3 h at room temperature. **K25** was cleaved from the resin with 2% TFA/DCM (4 mL) for 2 h.

All of the crude compounds were purified on a C$_{18}$ reversed-phase preparative column (250 mm × 4.6 mm, 10 μm) with a flow rate of 10 mL/min using acetonitrile/water (50:50) containing 0.06% TFA as an ion-pairing reagent. UV detection was at 215 nm. The structures of the analogues were confirmed by the presence of the following molecular ions using an 1100 series LC/MSD Trap (VL) (Agilent Technologies, USA). The structures, purity, and MS data of all target compounds are shown in **Table 1**.

Bioassays. *Animals*. Newly emerged mated female *D. punctata* (day 0) were isolated from stock cultures. Mating was confirmed by the presence of a spermatophore. Stocks and isolated females, fed lab chow and water ad libitum, were kept at 27 ± 1 °C and 50 ± 5% relative humidity with a 12 h light/12 h dark cycle.

Bioassays in Vitro. All radiochemical assays for JH biosynthesis were performed using individual pairs of CA from day 7 mated females. Compounds were dissolved in tissue culture medium 199 (GIBCO, Grand Island, NY, with Hank's salts) for assay as described previously (*19*, *20*). Compounds were used in the bioassay on the same day that the samples were prepared. The solutions were discarded at the end of each day. Rates of JH release were determined using the modified in vitro radiochemical assay (*19*, *20*). This assay measures the incorporation of the radiolabeled *S*-methyl moiety of radiolabeled methionine into JH III in the final step of biosynthesis by CA maintained in vitro. CA were incubated for 3 h in 100 μL of medium 199 (GIBCO, 1.3 mM Ca^{2+}, 2% Ficoll, methionine-free) containing L-[^{14}C-*S*-methyl]methionine (40 μM, specific radioactivity = 1.48−2.03 GBq/mmol (Amersham). Samples were extracted, and JH release was determined. (JH biosynthesis is equal to the sum of JH release plus CA content of JH, which is usually only a small percentage of release; therefore, JH release is generally regarded as equivalent to JH biosynthesis and the terms are used interchangeably here). Each data point on the dose−response figure represents replicate incubations of 10−27 experimental CA compared to control CA (i.e., no analogue added).

HQSAR Modeling. Thirty-two analogues were used for the HQSAR analyses. Twenty-eight analogues (**K1**, **K2**, **K4**, **K6−K25**, and **I1−I5**) were used for the training set, whereas four other compounds (**K3**, **K5**, the core pentapeptide, and **H17**) were used for the test set. The chemical structures and biological properties of compounds are listed in **Table 1**. The IC$_{50}$ values ranged from 1.79 to 2060 nM and were measured under identical experimental conditions, a fundamental requirement for successful QSAR studies. The test set compounds **K3**, **K5**, and **H17** were selected because they had different aromatic acids as the replacements of Y/FX region. The core pentapeptide was used to test whether the QSAR model can predict the natural peptides. The IC$_{50}$ values of the test set compounds were from 12 to 1140 nM. The IC$_{50}$ values were converted to the corresponding pIC$_{50}$ (−log IC$_{50}$) and used as dependent variables in the QSAR investigations.

All molecular modeling studies were performed using the molecular modeling package SYBYL 7.3 (*21*) installed on a Silicon Graphics Fuel Workstation. The force field was MMFF94 (*22*) with an 8 Å cutoff for nonbonded interactions, and the atomic point charges were also calculated by MMFF94. The minimizations were implemented with the steepest descent method for the first 100 steps, followed by the Broyden−Fletcher−Goldfarb−Shanno (BFGS) method, which is a method to solve an unconstrained nonlinear optimization problem (*23*) until the rms of gradient was < 0.005 kcal/(mol·Å).

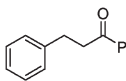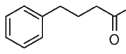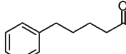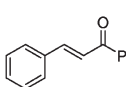**Table 1.** Structures, Mass Results, and Bioactivities of the Analogues[a]
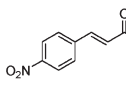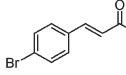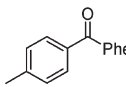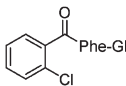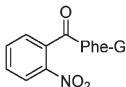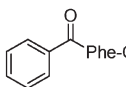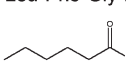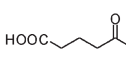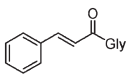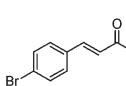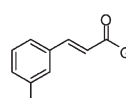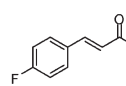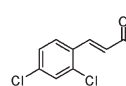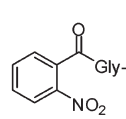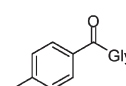
No.	structure	M. W.[a]	Mass result	Purity (%)	IC$_{50}$ (nM)	pIC$_{50}$	Pred.[b]	Res.[c]
K1	Phe-Gly-Leu-NH$_2$	466.6	MH+: 467.3	96	1120	5.95	6.14	-0.19
K2	Phe-Gly-Leu-NH$_2$	480.6	MNa+: 503.4	97	450	6.35	6.42	-0.07
K3[f]	Phe-Gly-Leu-NH$_2$	494.6	MNa+: 517.4	95	150	6.82	6.70	0.12
K4	Phe-Gly-Leu-NH$_2$	464.6	MH+: 465.4	95	850	6.07	6.20	-0.13
K5[f]	Phe-Gly-Leu-NH$_2$	509.6	MNa+: 532.3	96	1140	5.94	6.13	-0.19
K6	Phe-Gly-Leu-NH$_2$	542.2	MNa+: 565.2	98	1430	5.84	6.44	-0.60
K7	Phe-Gly-Leu-NH$_2$	452.6	MH+: 453.5	95	3450	5.46	5.25	0.21
K8	Phe-Gly-Leu-NH$_2$	473.2	MH+: 474.1	93	1270	5.90	5.80	0.10
K9	Phe-Gly-Leu-NH$_2$	483.5	MH+: 484.3	97	2040	5.69	5.56	0.13
K10	Phe-Gly-Leu-NH$_2$	438.1	MH+: 439.0	96	1730	5.76	5.64	0.12
K11	Leu-Phe-Gly-Leu-NH$_2$	447.3	MH+: 448.2	100	990	6.00	5.91	0.09
K12	Phe-Gly-Leu-NH$_2$	432.3	MH+:433.2	99	1270	5.90	6.28	-0.38
K13	HOOC Phe-Gly-Leu-NH$_2$	448.5	MH+: 449.1	98	870	6.06	5.97	0.09
K14	Gly-Phe-Gly-Leu-NH$_2$	521.6	MNa+: 544.3	98	39.6	7.40	7.66	-0.26
K15	Gly-Phe-Gly-Leu-NH$_2$	600.5	MNa+: 624.2	98	1.79	8.75	7.90	0.85
K16	Gly-Phe-Gly-Leu-NH$_2$	535.6	MNa+: 558.3	96	380	6.42	7.15	-0.73
K17	Gly-Phe-Gly-Leu-NH$_2$	539.6	MH+: 540.3	98	10.8	7.97	7.68	0.29
K18	Gly-Phe-Gly-Leu-NH$_2$	590.5	MH+:591.2	96	10.2	7.99	7.95	0.04
K19	Gly-Phe-Gly-Leu-NH$_2$	540.6	MH+: 541.5	97	100	7.01	7.02	-0.01
K20	Gly-Phe-Gly-Leu-NH$_2$	509.6	MH+: 510.5	97	130	6.88	6.71	0.17

Article

J. Agric. Food Chem., Vol. 58, No. 5, 2010   **2655**

Table 1. Continued

No.	structure	M. W.[a]	Mass result	Purity (%)	IC_{50} (nM)	pIC_{50}	Pred.[b]	Res.[c]
K21	Gly-Phe-Gly-Leu-NH$_2$	545.7	MH+: 546.5	95	400	6.40	6.36	0.04
K22	Gly-Phe-Gly-Leu-NH$_2$	495.6	MH+: 496.3	99	160	6.81	7.10	-0.29
K23	Gly-Phe-Gly-Leu-NH$_2$	530.0	MH+: 530.8	96	120	6.93	7.26	-0.33
K24	Gly-Phe-Gly-Leu-NH$_2$	551.7	MH+: 552.3	98	5.32	8.27	8.16	0.11
K25	Gly-Phe-Gly-Leu-NH$_2$	525.2	MH$^+$: 526.1	100	15.5	7.81	7.79	0.02
Pentapeptide[d, f]	Tyr-Asp-Phe-Gly-Leu-NH$_2$				130	6.90	6.97	-0.07
I1[d]	Asp-Phe-Gly-Leu-NH$_2$				140	6.85	6.89	-0.04
I2[d]	Asp-Phe-Gly-Leu-NH$_2$				990	6.00	5.95	0.05
I3[d]	Phe-Gly-Leu-NH$_2$				60	7.22	6.98	0.24
I4[d]	Phe-Gly-Leu-NH$_2$				870	6.06	5.69	0.37
I5[d]	Phe-Gly-Leu-NH$_2$				160	7.22	6.98	0.24
H17[e, f]	Gly-Phe-Gly-Leu-NH$_2$				12	7.92	7.59	0.33

[a] The data for the HQSAR study show experimental and predicted pIC_{50} values and residuals. [a] molecular weight; [b] predicted pIC_{50} by HQSAR model; [c] residual (experimental pIC_{50} − predicted pIC_{50}); [d] see ref *16*; [e] see ref *25*; [f] four compounds for test set.

RESULTS AND DISCUSSION

Effects of AST Analogues on JH Biosynthesis in Vitro. The ability of all AST analogues to inhibit JH biosynthesis was evaluated using bioassay methods in vitro. The IC_{50} value of each analogue shows that they all have some effect on JH biosynthesis by CA in vitro, compared with the core region (**Table 1**). Eight analogues (**K14**, **K15**, **K17**, **K18**, **K19**, **K23**, **K24**, and **K25**) are more active than the lead pentapeptide. On the basis of the IC_{50} values, **K15** and **K24** (IC_{50} = 1.79 and 5.32 nM, respectively) are more potent than Dippu-AST 1, Dippu-AST 3, and Dippu-AST

13. Particularly, the IC_{50} value of **K15** is lower than those of most natural ASTs, such as Dippu-AST 1, Dippu-AST 3, Dippu-AST 6, Dippu-AST 8, Dippu-AST 9, Dippu-AST 11, Dippu-AST 12, and Dippu-AST 13 (*24*), which indicates that analogue **K15** has better activity than most natural Dippu-ASTs. Both **K15** and **K24** have lower IC_{50} values than **H17**, a potential IGR reported previously (*25*). The highly active Y/FX region mimics are cinnamamido acetic acid, substituted cinnamamido acetic acids, benzamido acetic acid, substituted benzamido acetic acids, 5-phenylpentanamido acetic acid, and Z-Gly. This demonstrates that appropriate changes in AST structure can improve their activity. The IC_{50} values of the analogues also suggest that the Y/FX region does contribute weakly to the biological activity. Therefore, the Y/FX sequence could be reduced for new analogue design.

HQSAR Model. HQSAR is a new QSAR technique that avoids many of the problems associated with classical or 3D QSAR approaches (*26*). As part of our ongoing research program aimed at discovering new, potent AST analogues, we have employed the HQSAR method to generate predictive 2D QSAR models so as to investigate QSAR of a large series of AST analogues (**Table 1**).

In HQSAR, the molecular structures were broken down into all of the constituent fragments. Each unique fragment in a data set was then assigned by a cyclic redundancy check (CRC) algorithm. Each of these integers was then hashed into a bin (molecular hologram) in an integer array of fixed length. Molecular hologram fingerprints were generated for each molecule in the data set using lengths in the range from 53 to 401 and fragment sizes in the range from 2 to 10 atoms. These fragment counts are then related to biological data using partial least-squares (PLS) analysis. The model quality was determined such that a cross-validated r^2_{cv} (or q^2) > 0.5 normally indicates significant predictive ability.

Table 2. Results of HQSAR Analyses for Various Fragment Distinctions on the Key Statistical Parameters Using Default Fragment Size (4-7)[a]

model	fragment distinction	q^2	SEP	r^2	SEE	HL	N
1-1	A	0.513	0.647	0.789	0.426	353	3
1-2	A/B	0.408	0.729	0.768	0.457	61	4
1-3	A/C	0.498	0.686	0.885	0.329	353	5
1-4	A/Ch	0.495	0.673	0.833	0.387	401	4
1-5	A/H	0.523	0.654	0.845	0.373	83	4
1-6	A/DA	0.459	0.729	0.864	0.365	61	6
1-7	A/B/C	0.517	0.658	0.816	0.406	83	4
1-8	A/B/H	0.456	0.684	0.784	0.431	353	3
1-9	A/B/Ch	0.396	0.736	0.808	0.415	151	4
1-10	A/B/DA	0.283	0.802	0.714	0.507	61	4
1-11	A/C/H	0.546	0.668	0.911	0.296	151	6
1-12	A/C/Ch	0.456	0.714	0.877	0.339	353	5
1-13	A/C/DA	0.437	0.727	0.878	0.338	71	5
1-14	**A/H/Ch**	**0.591**	**0.619**	**0.874**	**0.343**	**257**	**5**
1-15	A/H/DA	0.450	0.735	0.899	0.314	307	6
1-16	A/Ch/ DA	0.410	0.761	0.866	0.363	307	6
1-17	A/B/C/H	0.501	0.700	0.906	0.304	59	6
1-18	A/B/C/Ch	0.512	0.662	0.829	0.391	59	4
1-19	A/B/C/DA	0.459	0.713	0.827	0.403	83	5
1-20	A/B/H/Ch	0.393	0.722	0.774	0.441	61	3
1-21	A/B/H/DA	0.522	0.685	0.918	0.284	53	6
1-22	A/B/Ch/DA	0.225	0.800	0.562	0.601	151	2
1-23	A/C/H/Ch	0.532	0.662	0.891	0.320	83	5
1-24	A/C/H/DA	0.369	0.769	0.892	0.318	83	5
1-25	A/H/Ch/DA	0.529	0.664	0.884	0.330	307	5
1-26	A/C/H/Ch/DA	0.435	0.728	0.874	0.343	71	5
1-27	A/B/H/Ch/DA	0.564	0.612	0.821	0.392	61	3
1-28	A/B/C/Ch/DA	0.468	0.723	0.915	0.289	53	6
1-29	A/B/C/H/DA	0.560	0.657	0.932	0.259	151	6
1-30	A/B/C/H/Ch	0.500	0.685	0.882	0.333	59	5
1-31	A/B/C/H/Ch/DA	0.531	0.649	0.859	0.356	257	4

[a] q^2, cross-validated correlation coefficient; SEP, cross-validated standard error; r^2, non-cross-validated correlation coefficient; SEE, non-cross-validated standard error; HL, hologram length; N, optimal number of components. Fragment distinction: A, atoms; B, bonds; C, connections; H, hydrogen atoms; Ch, chirality; DA, donor and acceptor.

There is still considerable debate regarding the 3D structures of ASTs and AST analogues (*12*, *14*, *27*, *28*), which precludes the building of 3D QSAR models. HQSAR is of particular interest because it eliminates the need for determining 3D structure, putative binding conformation, and molecular alignment.

A HQSAR run requires selecting values for parameters that specify the size of the hologram that is to be used, as well as the size and type of fragment substructures that are to be encoded. Final HQSAR models can be affected by a number of parameters concerning hologram generation: hologram length, fragment size, and fragment distinction. In our studies, holograms were generated using the standard parameters implemented in SYBYL 7.3. The generation of molecular fragments was carried out using the following fragment distinctions: atoms (A), bonds (B), connections (C), hydrogen atoms (H), chirality (Ch), and donor and acceptor (DA). To assess the process of hologram generation, several combinations of these parameters were considered using the default fragment size (4-7) shown in **Table 2**. The HQSAR analysis was performed by screening the 12 default series of hologram length values ranging from 53 to 401 bins. The patterns of fragment counts from the training set compounds were then related to the measured biological activity. All models generated in our studies were investigated using a fully cross-validated r^2 (q^2) partial least-squares (PLS) leave-one-out (LOO) method. The predictive ability of the models was assessed by their q^2 values. The statistical results from the PLS analyses for the

Table 3. HQSAR Analysis Showing the Influence of Various Fragment Sizes on the Key Statistical Parameters Using the Best Fragment Distinction (Atoms, Bonds, Connections, and Hydrogen Atoms)[a]

model	fragment size	q^2	SEP	r^2	SEE	HL	N
2-1	**2-5**	**0.638**	**0.583**	**0.878**	**0.338**	**61**	**5**
2-2	3-6	0.585	0.624	0.879	0.337	53	5
2-3	4-7	0.591	0.619	0.874	0.343	257	5
2-4	5-8	0.522	0.655	0.803	0.421	61	4
2-5	6-9	0.591	0.605	0.844	0.374	151	4
2-6	7-10	0.575	0.617	0.822	0.399	151	4

[a] Abbreviations as in **Table 2**.
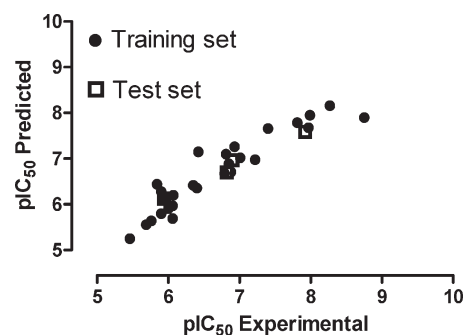
Figure 1. Plot of experimental versus predicted pIC$_{50}$ values of the training and test set molecules. The training set and test set molecules are shown in round and square symbols, respectively.

28 training set compounds using several fragment distinction combinations are presented in **Table 2**.

In **Table 2**, the best statistical result among all models was obtained for model 1-14 ($q^2 = 0.591$ and $r^2 = 0.874$). This model was derived using three fragment distinctions, with five being the optimum number of PLS components. The influence of different fragment sizes in the statistical parameters was further investigated for the best HQSAR model generated previously (model 1-14, **Table 2**). Fragment size parameters control the minimum and maximum lengths of fragments to be included in the hologram fingerprint. The HQSAR results for the different fragment sizes used are summarized in **Table 3**. The increment in q^2 was significant ($\Delta q^2 = 0.047$), when the fragment size changed from four to seven atoms to two to five atoms. Further increase or decrease in the atom count reduced the overall statistical power of the model. To avoid overinterpretation, model 2-1 was chosen as our HQSAR model with the fragment size 2-5, using the fragment distinction (A/H/Ch). This model was obtained using a hologram length of 61, 5 being the optimum number of components (onc) as the fragment type ($q^2 = 0.638$ and $r^2 = 0.878$). Model 2-1 was chosen as the final HQSAR model because of the high q^2. The detailed experimental pIC$_{50}$ and predicted pIC$_{50}$ values of the training set analogues for the best HQSAR model are given in **Table 1** and plotted in **Figure 1**. The low residual values show that the HQSAR model obtained is highly reliable and can be used to explain the structure–activity relationship of our AST analogues and to predict the biological activity of novel analogues. In the present study, all four molecules of the test set were predicted well by the model. The plots of actual versus predicted pIC$_{50}$ of the training and test set molecules are shown in **Figure 1** with a notable predictive r^2 of 0.883. The residual value (experimental pIC$_{50}$ − predicted pIC$_{50}$) of pentapeptide was −0.07, which indicated that the HQSAR model can not only predict the AST analogues (pentapeptide mimics) but also predict the natural core region reliably.

HQSAR Contribution Map. The HQSAR module in SYBYL uses color coding to show the atomic contributions to the activity.

Article

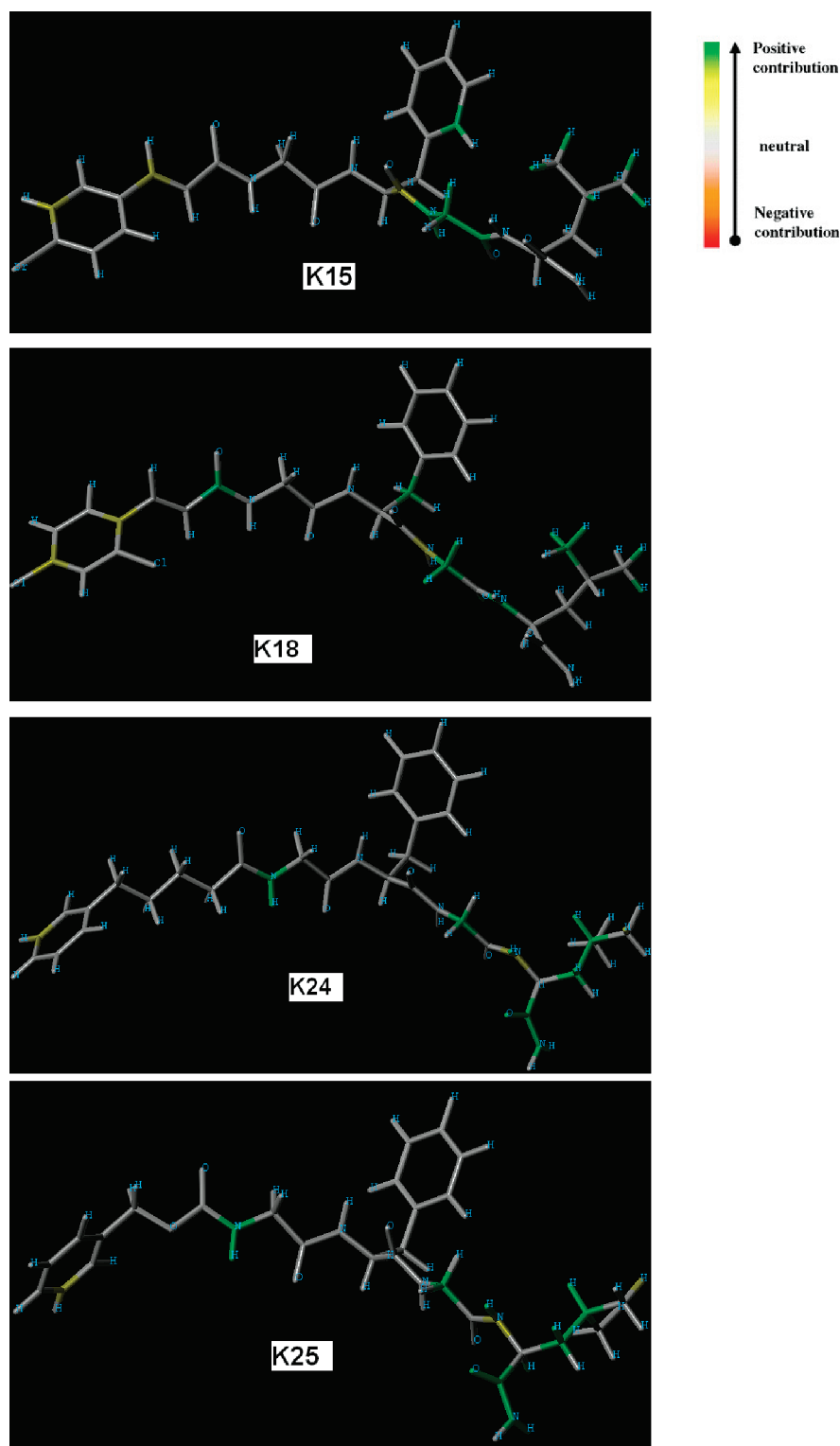*J. Agric. Food Chem.,* Vol. 58, No. 5, 2010 **2657**

Figure 2. Contribution map obtained for **K15**, **K18**, **K24**, and **K25** by HQSAR analysis.

Whereas the color codes red, red-orange, and orange show the unfavorable or negative contribution to the activity, the color codes yellow, green-blue, and green denote favorable or positive contribution to the activity. The white color shows intermediate contribution to the activity. It is clear that the region of FGLa is very important to the bioactivity, because this region was colored green and yellow in **Figure 2**. The result of our HQSAR study was in accordance with the previous results (*16*). The yellow color mapped some phenyl groups, which show that the contributions of some special aromatic acids such as phenylpropanoic acid, phenylacrylic

acid, and phenylpentanoic acid are better than the fatty acids and dicarboxylic acids (**Figure 2**). The aromatic acids can be divided into two parts: phenyl and linker. Five analogues, **K10**, **K1**, **K2**, **K3**, and **I3** (the only difference between those compounds is the carbon numbers of the linkers, the numbers are from 0 to 5), were selected to calculate this relationship between the length of linker and the pIC_{50} (**Figure 3**). **Figure 3** shows that longer linkers have more contribution to activity. The best carbon number was five in our study. Some parts of the linker are green and yellow in **Figure 2**, which demonstrates that the small steric hindrance of the linker is
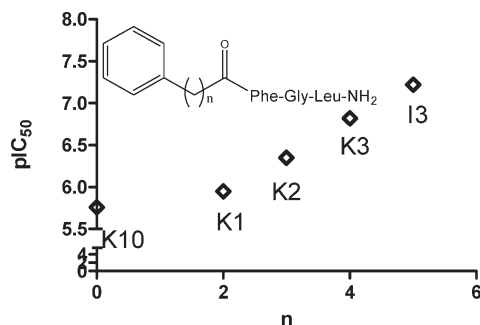
Figure 3. Contribution of the length of the linker. Longer linkers provide a positive contribution to the activity.

favorable to the activity. Substituted groups in the phenyl with the longest linker did not make significant contributions to the activity, which can also be seen in **Figure 2**.

In this paper, a predictable and statistically meaningful HQSAR model of 32 AST analogues was obtained. The final model suggested that a bioactive AST analogue must contain an aromatic group, an appropriate linker in the Y/FX region, and the FGLa portion. These results will be useful for the design of new AST analogues that are structurally related to the training set compounds.

LITERATURE CITED

(1) Schooley, D. A.; Judy, K. J.; Bergot, B. J.; Hall, M. S.; Jennings, R. C. Determination of the physiological levels of juvenile hormones in several insects and biosynthesis of the carbon skeletons of the juvenile hormones. In *The Juvenile Hormones*; Gilbert, L. I., Ed.; Plenum Press: New York, 1976; pp 107−117.

(2) Woodhead, A. P.; Stay, B.; Seidel, S. L.; Khan, M. A.; Tobe, S. S. Primary structure of four allatostatins: neuropeptide inhibitors of juvenile hormone synthesis. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 5977–6001.

(3) Woodhead, A. P.; Khan, M. A.; Stay, B.; Tobe, S.S. Two new allatostatins from the brains of *Diploptera punctata. Insect Biochem. Mol. Biol.* **1994**, *24*, 257–263.

(4) Pratt, G. E.; Farnsworth, D. E.; Siegel, N. R.; Fok, K. F.; Feyereisen, R. Identification of an allatostatin from adult *Diploptera punctata. Biochem. Biophys. Res. Commun.* **1989**, *163*, 1243–1247.

(5) Pratt, G. E.; Farnsworth, D. E.; Fok, K. F.; Siegel, N. R.; McCormack, A. L.; Shabanowitz, J.; Hunt, D. F.; Feyereisen, R. Identity of a second type of allatostatin from cockroach brains: an octadecapeptide amide with tyrosine-rich address sequence. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 2412–2416.

(6) Tobe, S. S.; Bendena, W. G. Allatostatins in the insects. In *Handbook of Biologically Active Peptides*; Kastin, A. J., Ed.; Academic Press: San Diego, CA, 2006; pp 201−206.

(7) Lange, A. B.; Bendena, W. G.; Tobe, S. S. The effect of the thirteen Dip-allatostatins on myogenic and induced contractions of the cockroach (*Diploptera punctata*) hindgut. *J. Insect Physiol.* **1995**, *41*, 581–588.

(8) Vilaplana, L.; Maestro, J. L.; Piulachs, M. D.; Bellés, X. Modulation of cardiac rhythm by allatostatins in the cockroach *Blattella germanica* (L.) (Dictyoptera Blattellidae). *J. Insect Physiol.* **1999**, *45*, 1057–1064.

(9) Martin, D.; Piulachs, M. D.; Bellés, X. Inhibition of vitellogenin production by allatostatin in the German cockroach. *Mol. Cell. Endocrinol.* **1996**, *121*, 191–196.

(10) Fuse, M.; Zhang, J. R.; Partridge, E.; Nachman, R. J.; Orchard, I; Bendena, W. G.; Tobe, S. S. Effects of an allatostatin and a myosuppressin on midgut carbohydrate enzyme activity in the cockroach *Diploptera punctata. Peptides* **1999**, *20*, 1285–1293.

(11) Nachman, R. J.; Garside, C. S.; Tobe, S. S. Hemolymph and tissue-bound peptidase-resistant analogs of the insect allatostatins. *Peptides* **1999**, *20*, 23–29.

(12) Hayes, T. K.; Guan, X. C.; Johnson, V.; Strey, A.; Tobe, S. S. Structure−activity studies of allatostatin 4 on the inhibition of juvenile hormone biosynthesis by corpora allata: the importance of individual side chains and stereochemistry. *Peptides* **1994**, *15*, 1165–1171.

(13) Piulachs, M. D.; Vilaplana, L.; Bartolomé, J. M.; Carreño, C.; Martín, D.; González-Muñiz, R.; Herranz, R.; García-López, M. T.; Andreu, D.; Bellés, X. Ketomethylene and methyleneamino pseudopeptide analogs of insect allatostatins inhibit juvenile hormone and vitellogenin production in the cockroach *Blattella germanica. Insect Biochem. Mol. Biol.* **1997**, *27*, 851–858.

(14) Nachman, R. J.; Moyna, G.; Williams, H. J.; Tobe, S. S.; Scott, A. I. Synthesis, biological activity, and conformational studies of insect allatostatin neuropeptide analogues incorporating turn-promoting moieties. *Bioorg. Med. Chem.* **1998**, *6*, 1379–1388.

(15) Nachman, R. J.; Moyna, G.; Williams, H. J.; Garside, C. S.; Tobe, S. S. Active conformation and peptidase resistance of conformationally restricted analogues of the insect allatostatin neuropeptide family. In *Advances in Comparative Endocrinology*; Kawashima, S., Kikuyama, S., Eds.; Monduzzi Editore: Yokohama, Japan, 1997; pp 1353−1359.

(16) Kai, Z. P.; Xie, Y.; Huang, J.; Tobe, S. S.; Zhang, J. R.; Ling, Y.; Zhang, L.; Zhao, Y. C.; Yang, X. L. Peptidomimetics in the discovery of new insect growth regulators: structure−activity relationships study of allatostatin neuropeptides. *J. Agric. Food Chem.* **2009**, submitted for publication.

(17) Sun, D. Q.; Lai, P. X.; Xie, W. D.; Deng, J. G.; Jiang, Y. Z. Concise synthesis of pentenyl phenyl acrylic acid. *Synth. Commun.* **2007**, *37*, 2989–2994.

(18) Chan, W. G.; White, P. D. *Fmoc Solid Phase Peptide Synthesis: A Practical Approach*; Oxford University Press: New York, 2000; pp 9−74.

(19) Tobe, S. S.; Clarke, N. The effect of L-methionine concentration on juvenile hormone biosynthesis by corpora allata of the cockroach *Diploptera punctata. Insect Biochem.* **1985**, *15*, 175–179.

(20) Tobe, S. S.; Pratt, G. E. The influence of substrate concentrations on the rate of insect juvenile hormone biosynthesis by corpora allata of the desert locust *in vitro. Biochem. J.* **1974**, *144*, 107–113.

(21) *SYBYL7.3*; Tripos Associate Inc.: St. Louis, MO.

(22) Halgren, T. MMFF VI. MMFF94s option for energy minimization studies. *J. Comput. Chem.* **1999**, *20*, 720.

(23) Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; Vetterling, W. T. *Numerical Recipes in C: The Art of Scientific Computing*; Cambridge University Press: Cambridge, U.K., 1988.

(24) Tobe, S. S.; Zhang, J. R.; Bowser, P. R. F.; Donly, B. C.; Bendena, W. G. Biological activities of the allatostatin family of peptides in the cockroach, *Diploptera punctata*, and potential interactions with receptors. *J. Insect Physiol.* **2000**, *46*, 231–242.

(25) Kai, Z. P.; Huang, J.; Tobe, S. S.; Yang, X. L. A potential insect growth regulator: synthesis and bioactivity of an allatostatin mimic. *Peptides* **2009**, *30*, 1249–1253.

(26) Hurst, T.; Heritage, T. HQSAR − A Highly Predictive QSAR Technique Based on Molecular Holograms, 213th ACS National Meeting of the American Chemical Society, San Francisco, CA; American Chemical Society: Washington, DC, 1997; CINF 019.

(27) Kai, Z.; Ling, Y.; Liu, W.; Zhao, F.; Yang, X. The study of solution conformation of allatostatins by 2-D NMR and molecular modeling. *BBA−Proteins Proteomics* **2006**, *1764*, 70–75.

(28) Banerjee, M.; Meyerowitz, E.; Huang, C.; Mohanty, S. Probing the conformation and Dynamics of allatostatin neuropeptides: a structural model for functional differences. *Peptides* **2008**, *29*, 375–385.