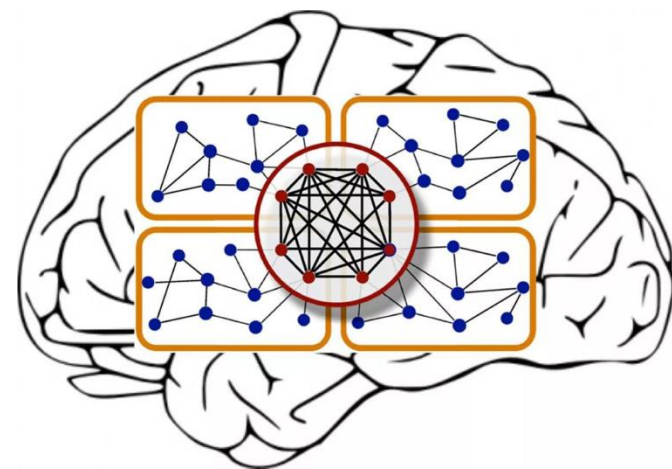


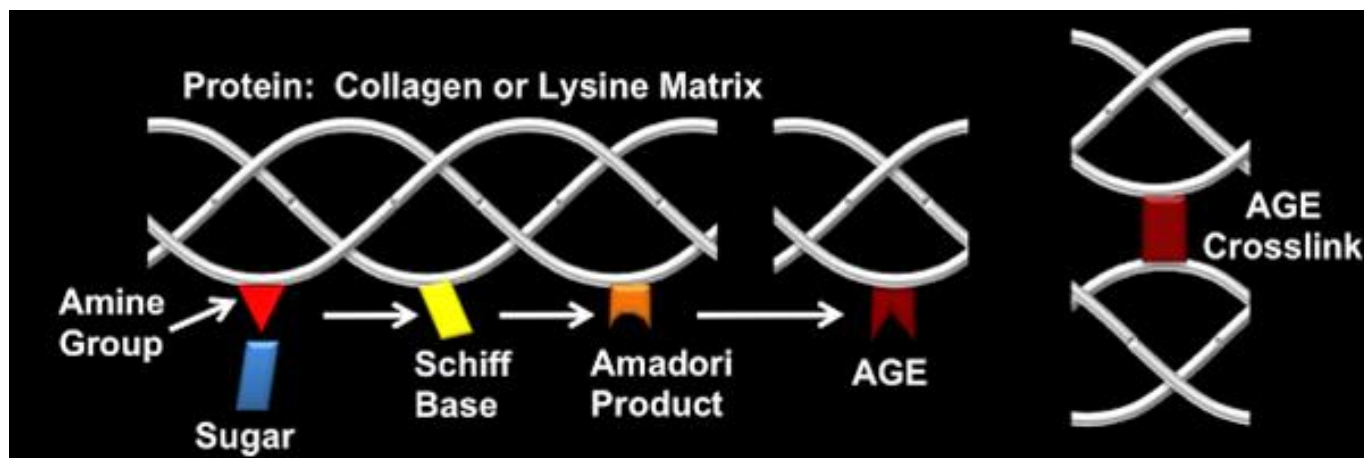
МЕТОДОЛОГИЯ ПОСТРОЕНИЯ НЕЙРОСЕТЕВЫХ КВАНТОВО- ХИМИЧЕСКИХ QSAR-МОДЕЛЕЙ НА ПРИМЕРЕ ИНГИБИТОРОВ РЕАКЦИИ МАЙЯРА

Змиевская Екатерина Анатольевна,
студентка 4 курса медико-
биологического факультета ВолгГМУ
Ворфоломеева В.В.
Васильев П.М.
Кузнецова В.А.
Соколова Е.В.
Волгоград, Россия



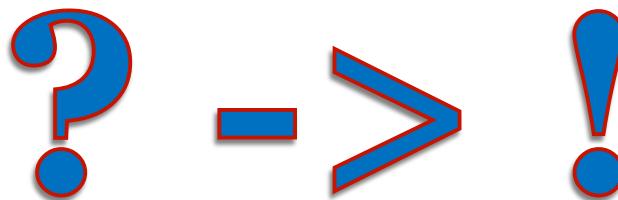
Актуальность

- Реакция Майяра – начальный процесс неферментативного гликирования белков, приводящий в конечном итоге к тяжелым осложнениям сахарного диабета.
- Ингибиторы этой реакции являются новым перспективным классом противодиабетических средств.



Цель

Разработать методологию построения нейросетевых квантово-химических QSAR-моделей общего назначения.



Материалы и методы

Методом PM7 в программе MORAS2012 рассчитывали следующие квантово-химические параметры:

- общая энергия соединения E
- энтальпия образования H
- энергии граничных орбиталей E_{HOMO} и E_{LUMO}
- их разность $E_{\text{HOMO}} - E_{\text{LUMO}}$.



Материалы и методы

данные по антигликирующей активности 303 структурно-разнородных соединений (в тестовой концентрации 10^{-3}M и 10^{-4}M) кластеризовали методом k-средних на три класса: высокой, умеренной и низкой активности.



Материалы и методы

- Границы классов активности

	A	B	C
65	S-0040a	60.70	h
66	GEMA-0011	60.09	h
67	03590	59.75	h
68	KHFA-0588	59.41	h
69	RU-1250	59.31	h
70	S-0032a	58.34	m
71	HC-0028	57.50	m
72	HC-0052	56.52	m
73	RUS-0219	56.35	m
74	06317	55.74	m

	A	B	C
130	SHQU-0022	27.73	m
131	BUM-0005	27.26	m
132	TONS-0226	27.08	m
133	DAB-0008	26.99	m
134	RU-1270	26.79	m
135	SHQU-0007	25.93	l
136	TONS-0031	25.01	l
137	KHFA-0147	24.57	l
138	FM-0003	24.55	l
139	KHFA-0453	24.42	l

Материалы и методы

- Нейросетевое моделирование проводили в программе Statistika 8.0, архитектура сети - многослойный перцептрон MLP. При моделировании задавалось пять вариантов сэмплинга, с объемами обучающей-тестовой-валидационной выборки 60-20-20%.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Code	d10-3	level 0	h	hm	E	H	HOMO	LUMO	H-L	Train1	Train2	Train3	Train4	Train5
2	HC-0054	97.36	h	h	hm	-5011.410	84.07	-10.41	-2.46	-7.95	train	val	test	train	train
3	KC-GKC-G	97.20	h	h	hm	-4189.710	36.32	-10.08	-2.02	-8.06	train	train	val	test	train
4	HC-0009	94.51	h	h	hm	-2650.070	155.68	-9.23	-1.72	-7.51	train	train	train	val	test
5	BUIMZ-0010	93.25	h	h	hm	-2064.291	41.32	-9.27	-0.81	-8.46	test	train	train	train	val
6	AB-0052	93.01	h	h	hm	-3179.868	-22.84	-10.45	-2.27	-8.18	val	test	train	train	train
7	RUS-0217	92.90	h	h	hm	-1711.189	77.22	-8.56	-0.19	-8.37	train	val	test	train	train
8	IS-0201	91.68	h	h	hm	-2500.606	116.28	-11.40	-2.24	-9.15	train	train	val	test	train
9	AB-0053	91.57	h	h	hm	-3236.221	-29.27	-9.42	-1.68	-7.74	train	train	train	val	test
10	AB-0054	91.10	h	h	hm	-3535.851	-37.80	-9.31	-1.61	-7.70	test	train	train	train	val

Материалы и методы

- Для каждого уровня активности в заданной концентрации, на каждом варианте сэмплинга программа рассчитывала 500 сетей, автоматически отбирая из них 100 лучших.
- Из этих 100 сетей по совокупности точности обучения, тестирования и валидации выбирались 5 лучших

10-3 class h-in train1* - Summary of active networks (MRI for network d10-3)

10-3 class h-in train1*

SANN (MRI for network d10-3)

STATISTICA NN Results I

Summary of active n

Network weights (MF)

Data statistics (MRI f

h (Confusion matrix)

h (Classification sumr

Sensitivity analysis fc

Confidence levels (M

Samples: Train, Tes

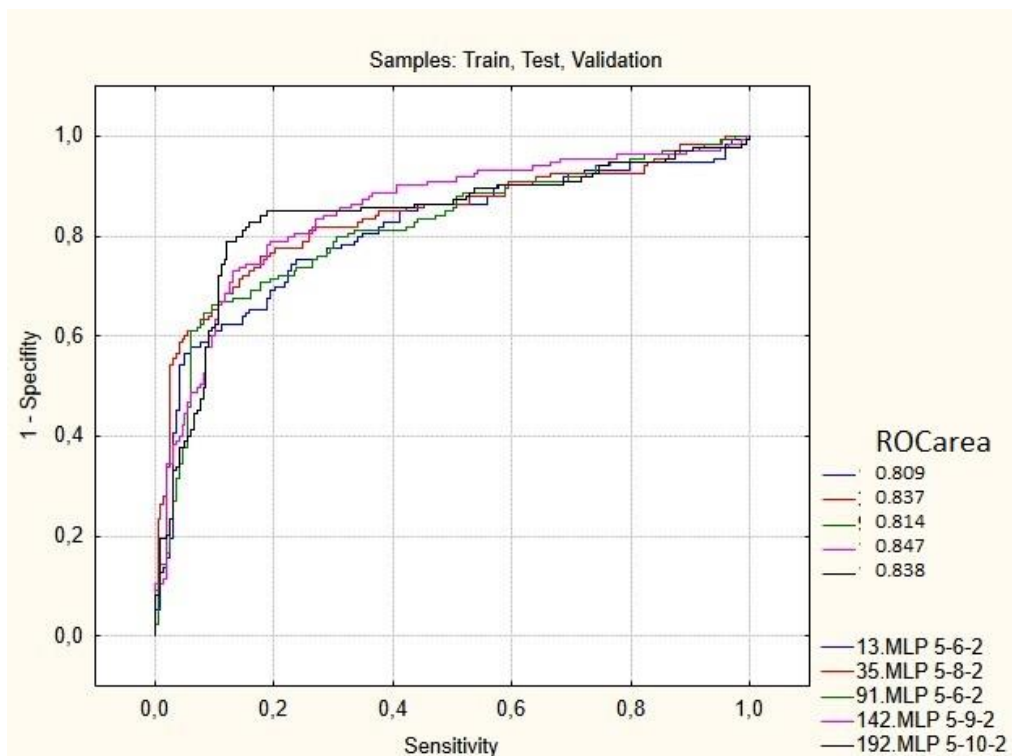
ROC (MRI for netwoi

ROC areas and thres

Index	Net. name	Training perf.	Test perf.	Validation perf.	Training algorithm	Error function	Hidden activation	Output activation
13	MLP 5-5-2	89,39394	89,39394	84,84848	BFGS 33	SOS	Logistic	Logistic
35	MLP 5-10-2	88,88889	89,39394	78,78788	BFGS 32	SOS	Tanh	Logistic
40	MLP 5-8-2	91,41414	89,39394	81,81818	BFGS 51	SOS	Tanh	Logistic
64	MLP 5-6-2	89,89899	89,39394	84,84848	BFGS 32	SOS	Tanh	Logistic
88	MLP 5-6-2	89,89899	89,39394	81,81818	BFGS 62	SOS	Tanh	Logistic

Материалы и методы

- Для них проводили оценку точности на объединенной выборке и ROC-анализ
- По полученным параметрам выбирали лучшую модель



Результаты

- Лучшая модель была получена для градации «наличие антигликирующей активности» при концентрации веществ 10^{-3}M , для которой показатели точности прогноза составили $F_o=84\%$, $F_a=78\%$, $F_n=87\%$, $\text{ROC}_{\text{AREA}}=0.838$
- Архитектура сети MLP 5-10-2 (logistic, logistic)

Практическое применение - виртуальный скрининг

- Отобрано по прогнозу и испытано в концентрации 10^{-3}M - 7 веществ
- Показали выраженную антигликирующую активность - 6
- Из них 4 высокоактивных:
- сопоставимы по активности с аминогуанидином - 2
- превышают по активности аминогуанидин - 2
- Точность поиска $6/7=86\%$

Выводы

- Разработана методология построения нейросетевой QSAR-модели на примере ингибиторов реакции Майяра
- Показана возможность использования для нейросетевого моделирования общих квантово-химических параметров молекул, что допускает применение такого подхода для других видов активности
- Одна из полученных моделей показала высокую эффективность при поиске новых антигликирующих веществ

Спонсирование

- Исследование выполнено в ВолгГМУ за счет гранта РФФИ (проект № 14-25-00139)

Спасибо за внимание!

