

Дополнительная информация

Структурные особенности каротиноид-связывающих белков

М.М. Сурков¹, А.Ю. Литовец¹, А.А. Мамчур¹, Т.Б. Станишневa-Коновалова²,

И.А. Ярошевич*¹

¹Кафедра биофизики, биологический факультет, Московский государственный университет имени М.В. Ломоносова, Россия, 119234, г. Москва, Ленинские горы, д. 1, стр.

24;

²Кафедра биоинженерии, биологический факультет, Московский государственный университет имени М.В. Ломоносова, Россия, 119234, г. Москва, Ленинские горы, д. 1, стр.

73

Таблица SI1. Полный список лигандов, использованных в работе.

Код лиганда в базе PDB	Код лиганда в базе PDB
0IE	IWJ
32N	K3I
45D	KGD
45H	LUT
5X6	LYC
7OT	NEX
8CT	NS0
A1EFU	NS1
A1L0S	NS5
A1LXP	O1U
A86	PID
AXT	PQ9
BCR	Q6L

C7Z
CRT
DD6
ECH
EQ3
ET4
H4X
HEQ
I7D
IHT
II0
II3
IRM

QDL
RAW
RRX
SP2
SPN
SPO
U4Z
UIX
V7N
WVN
XAT
ZE0
ZEX

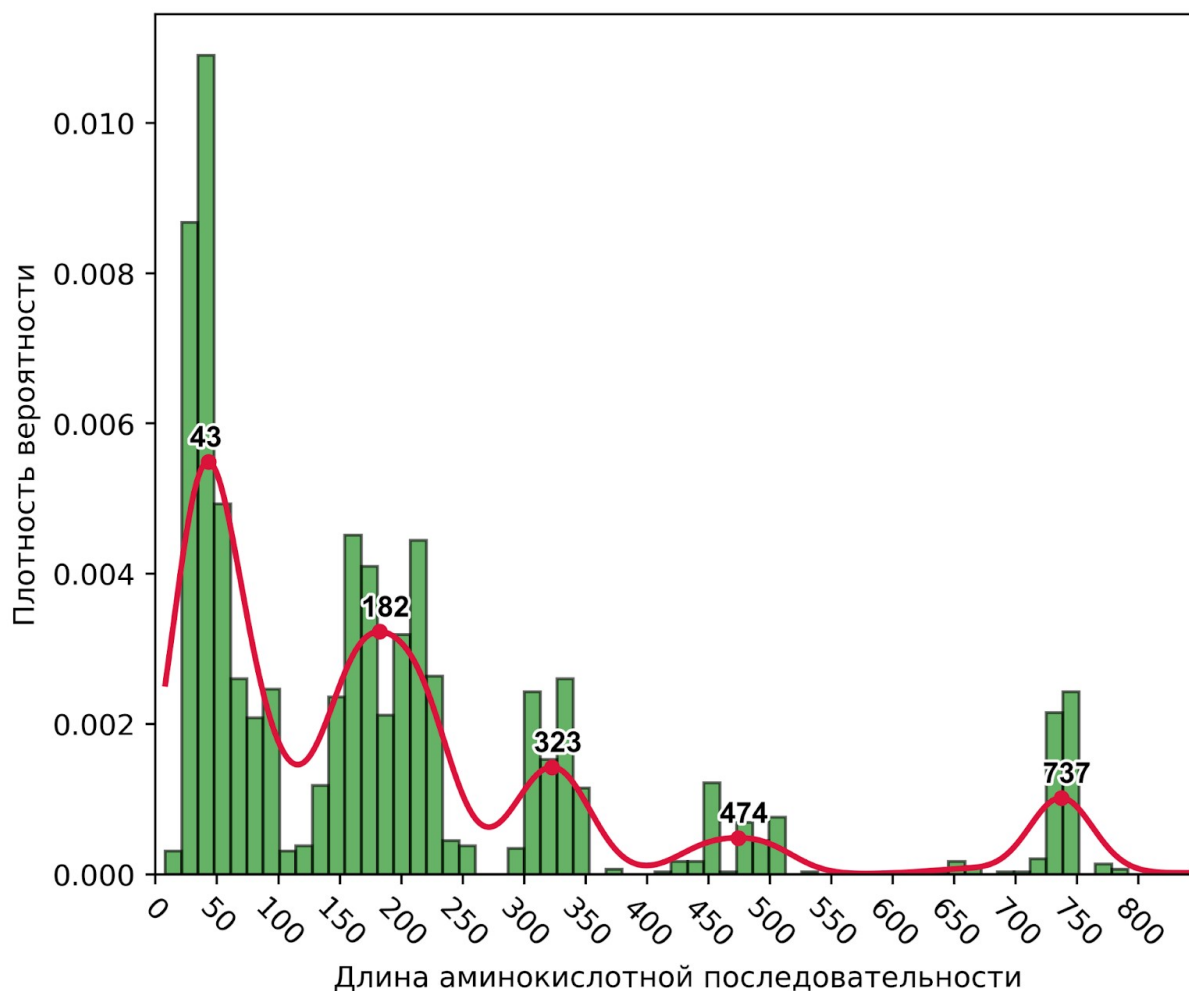


Рисунок SI1. Столбчатая диаграмма с распределением длин аминокислотных последовательностей уникальных связывающих каротиноиды цепей.

Распределение длин аминокислотных последовательностей было визуализировано с помощью столбчатой диаграммы, отображающей плотность вероятности (Рис SI1).

Анализ последовательностей, соответствующих основным пикам на графике распределения длин, выявил характерные классы белков, ассоциированные с каждым пиком:

- До 48 а.к.: единичные альфа-спирали, содержащие в основном гидрофобные аминокислоты.
- От 140 до 183 а.к.: в основном последовательности, содержащие несколько гидрофобных альфа-спиралей. Среди них, в основном, LHC-like белки. Встречаются последовательности из краustiанинов, красного каротиноидного белка (RCP), спирального каротиноидного белка (Helical carotenoid protein).

- От 184 до 200 а.к.: в основном, различные хлорофилл-связывающие белки: хлорофилл а-в связывающие белки, хлорофилл а-с связывающие белки и так далее.
- От 300 до 350 а.к.: в основном, белки реакционных центров фотосистем: белки D2, D1; М-субъединица реакционного центра у пурпурных бактерий; последовательности оранжевого каротиноидного белка (OCP).
- От 440 до 500 а.к.: в основном, белки фотосистемы II: коровые светособирающие белки; попадают белки фотосистемы I, такие как белок, кодируемый геном *psaL*.
- От 700 до 760 а.к.: в основном, белки реакционного фотосистемы I, кодируемые генами *psaA*, *psaB* (белки Photosystem I P700 chlorophyll a apoprotein A1 и Photosystem I P700 chlorophyll a apoprotein A2)

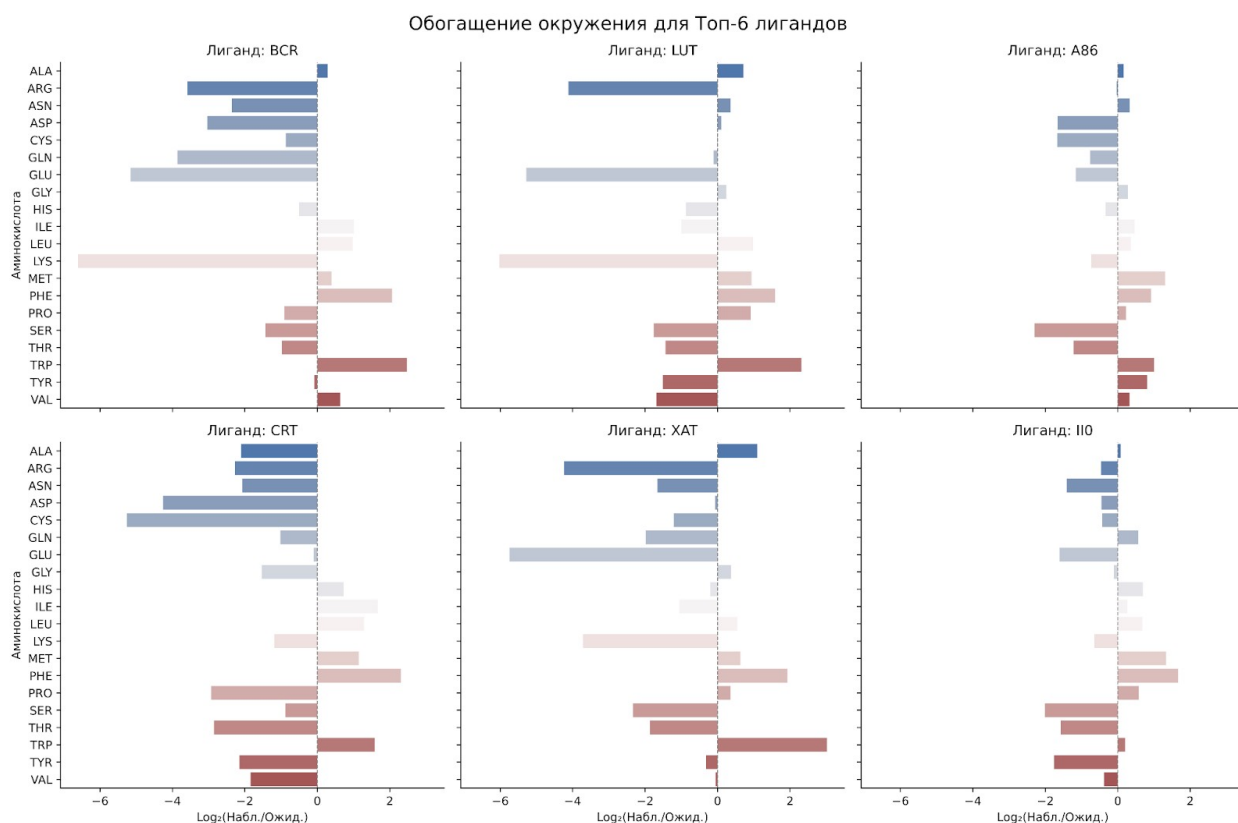


Рисунок SI2. Специфика аминокислотного окружения для шести наиболее представленных в выборке каротиноидов: β -каротин (BCR), лютеин (LUT), астаксантин (A86), кантаксантин (CRT), зеаксантин (XAT) и сфероиден (110). Обозначения аналогичны Рисунку 2.

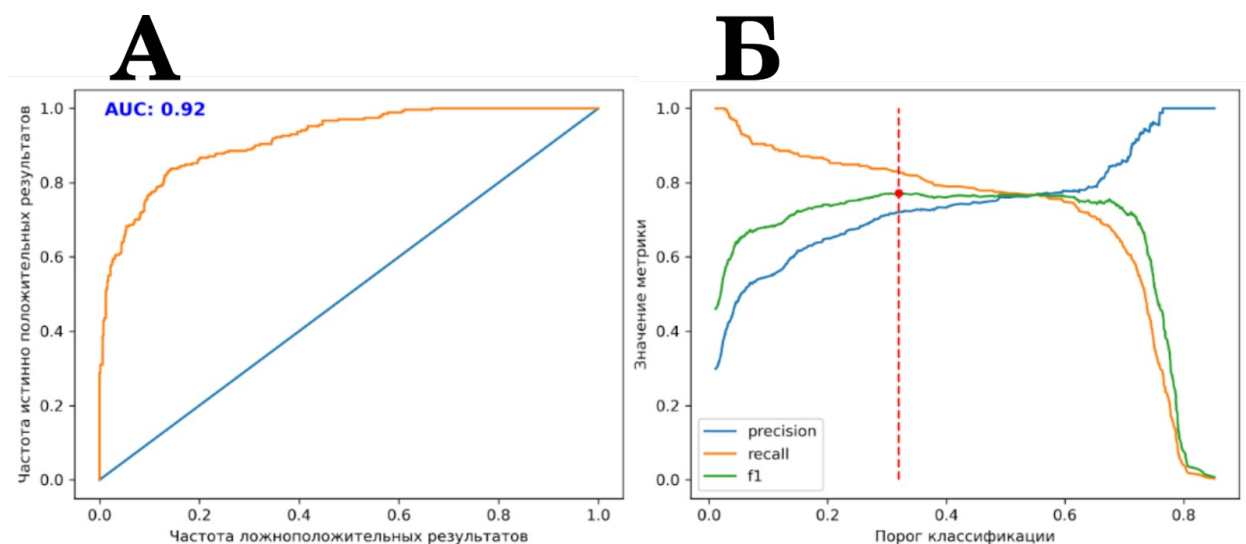


Рисунок SI3. Кривые производительности классификатора. А – ROC-кривая, Б – Precision-Recall-F1 кривая; красная точка соответствует максимальному значению F1.

Таблица SI2. Метрики Precision@k, рассчитанные на тестовом наборе данных

Precision@50	1,00
Precision@100	0,96
Precision@150	0,93