Protein Feature Extractor

Jesse Thomas, Alyssa Capeheart, Harica BNL

Mar 17, 2021

PROTEIN FEATURE EXTRACTOR:

CHAPTER

ONE

PROTEIN FEATURE EXTRACTOR REPORT

1.1 Understanding Proteins:

Proteins are polymers of alpha amino acids, arranged in a linear sequence and connected by covalent bonds. Amino acids are the major building blocks of a protein. Each amino acid contains a central C atom, an amino group (NH2), a carboxyl group (COOH), and a specific R group. The R group determines the characteristics (size, polarity, and pH) for each type of amino acid.

1.2 Feature Selection:

Polarizability usually refers to the response of a molecule to an electric field. As molecules are made up of elementary particles with electric charge, namely protons and electrons, when subject to an electric field, the negatively charged electrons and positively charged atomic nuclei are subject to opposite forces and undergo charge separation.

In our assignment we have chosen Polarizability as one of the features of proteins to be extracted.

1.3 Normalization:

the goal of normalization is to change the values of numeric columns in the dataset to a common scale, without distorting differences in the ranges of values. For machine learning, every dataset does not require normalization. So we normalize the data to bring all the variables to the same range

The normalization we chose for this project is, for each sequence, to divided by the length of the sum of the char of the sequence

1.4 Data Available:

List of protein sequences which belong to 4 groups (4 places in the Gram-Negative Bacterial protein cells). The proteins, their class labels, and a list of 55 physicochemical properties for amino acids (20 alphabets that build the protein sequence) are available.

1.5 Logic Implemented:

- Step 1: Read in protein data.
- Step 2: Read in protein attribute data.
- Step 3: Extract attribute headers.
- Step 4: Convert protein sequences to given properties.
- Step 5: Normalize count vector.
- Step 6: Pick attribute to extract and apply it.

For a more detail read *Main.py*

To see a basic summary of the data navigate to *Polarizability Data*. To see the full data extraction open *Polarizability.csv*

CHAPTER

TWO

GETTING STARTED

2.1 Install requirements

From inside the project folder run:

```
> python3 pip install -r requirements.txt
```

2.2 Run program

To run the code as is use:

```
> python3 main.py
```

2.3 Extracting a specific attribute

To extract a specific attribute, in the file named main, change the *ATTRIBUTE_TO_EXTRACT* to the index of the attribute wanted and then run the code. The extracted attribute file will appear in the *features* folder

To extract all given attributes into separate files set *ATTRIBUTE_TO_EXTRACT* to 0.

CHAPTER

THREE

MAIN.PY

The file named main is the application driver.

3.1 Logic Implementation

3.1.1 Step 1: Read in protein data.

We read in protein data held in a csv file using the python module Pandas and a custom wrapper class called *Protein Data Handler*.

3.1.2 Step 2: Read in protein property data.

The protein attribute data, a list of vectorized properties that a protein has, is read in from a csv file using the python module Pandas and a custom wrapper class classed *Protein Attribute Data Handler*.

3.1.3 Step 3: Extract attribute headers.

To keep the code module, we extract the headers associated with the protein property attributes, this is just in case we are presented with an attribute list that doesn't want to take into account certain attribute values.

3.1.4 Step 4: Convert protein sequences to given attributes.

Here we count how many times an attribute appears in a sequence and place the result under it's associated name and attribute.

3.1.5 Step 5: Normalize count vector.

We divide each attribute count by the length of the sequence.

3.1.6 Step 6: Pick properties to extract and apply it.

We take the chosen index from the protein properties csv and grab it's associated position in the dataframe. We then multiplied each attribute with it's associated count. We also summed the result under the column with the properties name.

CHAPTER

FOUR

DATA HANDLER

4.1 Data Handler Class

class DataHandler(*filename: str*)
I'm used to read csv data in.

Parameters **filename** (*str*) – The filename, including the path to read from, used to pull desired data.

__slots__ = ['__data']
Reserve space for writable attributes and limits addition attribute creation.

__init__(*filename: str*) → None
Constructor Method.

4.2 Data Handler Properties

data = <property object>
I'm used to access the data that was read in.

Returns Dataframe

Return type pandas.DataFrame

4.3 Data Handler Module Global Variables

logger = <CustomLogger.CustomLogger object>
Local logger

CHAPTER

FIVE

PROTEIN DATA HANDLER

5.1 Protein Data Handler Class

class ProteinDataHandler (*filename: str*)

I'm used to read in csv data pertaining to protein data.

Inherits from *Data Handler*

5.2 Protein Data Handler Public Methods

covert_to_count_vector (*attributes_values: numpy.ndarray*) → *Handlers.Processor.Processor*

I convert all sequences to a count vector based on attributes given.

Parameters **attributes_values** (*numpy.ndarray*) – List of attributes to use to convert file to.

Returns A processor object containing processed data.

Return type *Processor.Processor*

5.3 Protein Data Handler Module Global Variables

logger = <CustomLogger.CustomLogger object>

Local logger

5.4 Protein Data Handler Module Global Functions

count_sequence (*sequence: str*) → dict

I count the number of times a character is seen in a sequence.

Parameters **sequence** (*str*) – Protein sequence.

Returns Dictionary with the keys being a character found in the given sequence and the value being the number of time it was seen.

Return type dict

CHAPTER

SIX

PROTEIN ATTRIBUTE DATA HANDLER

6.1 Protein Attribute Data Handler Class

class ProteinAttributeDataHandler (*filename: str*)

I'm used to read in csv data pertaining to protein attribute data.

Inherits from *Data Handler*.

6.2 Protein Attribute Data Handler Public Methods

get_attribute_headers () → numpy.ndarray

I get the attribute headers from the given data.

Returns An array of attributes.

Return type numpy.ndarry

get_attribute_values (*value: int*) → tuple

I get the attribute name and list of associated values associated with the given value - 1.

Parameters **value** (*int*) – Row number associated with desired attribute.

Returns Attribute name and attributes

Return type tuple

6.3 Protein Attribute Data Handler Module Global Variables

logger = <CustomLogger.CustomLogger object>

Local logger

CHAPTER

SEVEN

PROCESSOR

7.1 Processor Class

class Processor (*file: pandas.core.frame.DataFrame*)
I'm used to process a given dataframe.

Parameters **file** (*pandas.DataFrame*) – Dataframe to work on.

__slots__ = ['__data']
Reserve space for writable attributes and limits addition attribute creation.

__init__ (*file: pandas.core.frame.DataFrame*) → None
Constructor Method.

7.2 Processor Properties

data = <property object>
I'm used to access the data that was read in.

Returns Dataframe

Return type pandas.DataFrame

7.3 Processor Public Methods

apply_attribute (*attribute_name: str*, *attribute: pandas.core.series.Series*) → pandas.core.frame.DataFrame
I apply a given attribute series to the data currently held, sum the result, and adds it as a new column to the data.

Parameters

- **attribute_name** (*str*) – Name of the attribute being applied.
- **attribute** (*pandas.Series*) – Series of attribute to be applied.

Returns Nothing

Return type NoneType

normalize_via_length () → None
I normalize the data held by dividing each attribute by the length of the sequence.

Returns Nothing

Return type NoneType

save_processed_data(*path: str*) → None
I save the data to a given path name.

Parameters **path** (*str*) – The path name to save the data held.

Returns Nothing

Return type NoneType

7.4 Processor Private Methods

__get_attribute_header() → numpy.ndarray
I get the header from the held data associated with the attributes.

Returns Nothing

Return type numpy.ndarray

7.5 Processor Module Global Variables

any_uint
Typing that's a union of all the unsigned integers

alias of Union[numpy.uint8, numpy.uint16, numpy.uint32, numpy.uint64]

logger = <CustomLogger.CustomLogger object>
Local logger

CHAPTER

EIGHT

CUSTOM LOGGER

8.1 Custom Logger Class

class CustomLogger (*filename: str*, *level: numpy.uint8 = 4*)
I am used to log information to specific files associated with the module that called me.

Parameters

- **filename** (*str*) – The name of the module or file used for this logger.
- **level** (*numpy.uint8*) – The level of logging wanted

Note:

- 0: NOTSET - Doesn't log anything.
- 1: CRITICAL - logs only critical log calls.
- 2: ERROR - logs error log calls and everything before it.
- 3: INFO - logs info log calls and everything before it.
- 4: DEBUG - logs debug log calls and everything before it.

__slots__ = ['__level', '__logger']
Reserve space for writable attributes and limits addition attribute creation.

__init__ (*filename: str*, *level: numpy.uint8 = 4*) → None
Constructor Method.

8.2 Custom Logger Properties

level = <property object>
I'm used to access the current logging level.

Returns Unsigned 8 bit Integer.

Return type numpy.uint8

logger = <property object>
I'm used to access the logger object.

Returns Logger object.

Return type logging.Logger

8.3 Custom Logger Methods

flow(*message: str*) → None
I'm used to capture the flow of your application. Eg. 'Starting connection'.

Parameters **message** (*str*) – The string wanting to be logged.

Returns Nothing

Return type NoneType

sanity_check(*message: str*) → None
I'm used to capture debugging information from your application. Eg. the current value of a specific variable at a specific point in the application state.

Parameters **message** (*str*) – The string wanting to be logged.

Returns Nothing

Return type NoneType

CHAPTER

NINE

POLARIZABILITY DATA

To get the break of the individual attributes open the file: *Polarizability.csv*

The below

Table 1: Polarizability

	Group	Fold	Name	Polarizability
0	1	Fold1	>P23453	0.15671686746987956
1	1	Fold1	>P07373	0.15866120218579235
2	1	Fold1	>P12921	0.16700507614213195
3	1	Fold1	>P19579	0.15875912408759124
4	1	Fold1	>A2RMA8	0.15875251509054325
5	1	Fold1	>A6QIG2	0.15633561643835614
6	1	Fold1	>P35594	0.15206117021276594
7	1	Fold1	>P0A3G5	0.14909282700421939
8	1	Fold1	>P39755	0.15912871287128716
9	1	Fold1	>P33379	0.14572769953051645
10	1	Fold1	>P0C0N4	0.17809278350515462
11	1	Fold1	>Q2FV54	0.1692703150912106
12	1	Fold1	>P0A4D8	0.15608778625954198
13	1	Fold1	>P35865	0.14872255489021954
14	1	Fold1	>Q9CD58	0.14501905972045742
15	1	Fold1	>P46922	0.14972696245733788
16	1	Fold1	>P46105	0.13780276816609
17	1	Fold1	>O31652	0.16259375
18	1	Fold1	>Q07833	0.1476520994001714
19	1	Fold1	>P33251	0.16417508417508417
20	1	Fold1	>O06493	0.15384146341463417
21	1	Fold1	>O34894	0.16227758007117435
22	1	Fold1	>P53663	0.15919621749408983
23	1	Fold1	>P35165	0.17432989690721645
24	1	Fold1	>Q08429	0.1621875
25	1	Fold1	>P12667	0.14829931972789115
26	1	Fold1	>P13485	0.17301608579088473
27	1	Fold1	>P33116	0.1697394136807818
28	1	Fold1	>Q07428	0.15112068965517242
29	1	Fold1	>Q5HKQ0	0.1704611650485437
30	1	Fold1	>P33113	0.1629193899782135
31	1	Fold1	>P42177	0.16838565022421526
32	1	Fold1	>P34956	0.16848998459167952

continues on next page

Table 1 – continued from previous page

	Group	Fold	Name	Polarizability
33	1	Fold1	>P0A4Y0	0.14507900677200902
34	1	Fold1	>P37948	0.16427927927927927
35	1	Fold1	>Q79VE9	0.15877551020408162
36	1	Fold1	>P39843	0.15457500000000002
37	1	Fold1	>P20166	0.14911301859799714
38	1	Fold1	>P14677	0.1475733333333333
39	1	Fold1	>P22094	0.14833910034602077
40	1	Fold1	>P25051	0.14810495626822157
41	1	Fold1	>P26829	0.1415799614643545
42	1	Fold1	>P0A4N3	0.16532142857142856
43	1	Fold1	>P0A5I8	0.13355932203389828
44	1	Fold1	>Q9ZEP5	0.13022222222222224
45	1	Fold1	>P0C0H0	0.16606205250596662
46	1	Fold1	>O07002	0.15803846153846152
47	1	Fold1	>Q05523	0.1595353982300885
48	1	Fold1	>P54178	0.15564766839378236
49	1	Fold1	>P68782	0.15931464174454826
50	1	Fold1	>P40411	0.15563451776649745
51	1	Fold1	>P24011	0.15941011235955058
52	1	Fold1	>O30417	0.15928429423459245
53	1	Fold1	>P0A3S3	0.14558394160583943
54	1	Fold1	>O52351	0.15722891566265063
55	1	Fold1	>O66043	0.16807157057654076
56	1	Fold1	>P65563	0.15515625
57	1	Fold1	>P0A609	0.11979166666666669
58	1	Fold1	>P43439	0.16381024096385546
59	1	Fold1	>Q93QY7	0.17178359096313905
60	1	Fold1	>P35166	0.15459239130434785
61	1	Fold1	>Q45479	0.16616883116883116
62	1	Fold1	>Q99027	0.17137841352405722
63	1	Fold1	>P12669	0.1587121212121212
64	1	Fold1	>P82599	0.10945652173913042
65	1	Fold1	>Q8GFE2	0.14642201834862387
66	1	Fold1	>P37105	0.1531614349775785
67	1	Fold1	>Q01467	0.1752906976744186
68	1	Fold1	>P39774	0.16849514563106796
69	1	Fold1	>P0A670	0.1231818181818182
70	1	Fold1	>Q08430	0.15916083916083917
71	1	Fold1	>Q04664	0.17229398663697104
72	1	Fold1	>P14319	0.1558878504672897
73	1	Fold1	>P0A4G4	0.15183870967741936
74	1	Fold1	>P36497	0.16102209944751383
75	1	Fold1	>P54544	0.16159999999999997
76	1	Fold1	>Q8NNK5	0.1375971731448763
77	1	Fold1	>P30345	0.13949999999999999
78	1	Fold1	>P45829	0.15657370517928285
79	1	Fold1	>Q8NNK2	0.1513649025069638
80	1	Fold1	>P42175	0.1589657980456026
81	1	Fold1	>P46339	0.14906148867313918

continues on next page

Table 1 – continued from previous page

	Group	Fold	Name	Polarizability
82	1	Fold1	>Q45068	0.15217204301075266
83	1	Fold1	>P39215	0.1430513595166163
84	1	Fold1	>P45706	0.16276595744680855
85	1	Fold1	>P34959	0.15596774193548388
86	1	Fold1	>P39141	0.15279898218829518
87	1	Fold1	>P42086	0.14499999999999996
88	1	Fold1	>A5U127	0.1419205298013245
89	1	Fold1	>O34742	0.15021834061135372
90	1	Fold1	>Q5HKP9	0.180990099009901
91	1	Fold1	>P31141	0.1285969387755102
92	1	Fold1	>Q01464	0.14858208955223878
93	1	Fold1	>O07628	0.17154639175257733
94	1	Fold1	>Q02113	0.1463971631205674
95	1	Fold1	>P50847	0.14036363636363636
96	1	Fold1	>P08656	0.15882812500000001
97	1	Fold1	>P16449	0.1755050505050505
98	1	Fold1	>P35160	0.15687150837988828
99	1	Fold1	>P32959	0.15999999999999998
100	1	Fold1	>P42432	0.156104513064133
101	1	Fold1	>P54493	0.15528599605522678
102	1	Fold1	>P72058	0.15175738724727836
103	1	Fold1	>Q9L523	0.15761578044596913
104	1	Fold1	>O33654	0.15054852320675102
105	1	Fold1	>P46921	0.15109929078014186
106	1	Fold1	>P75409	0.16389170896785107
107	1	Fold1	>P63713	0.14969325153374233
108	1	Fold1	>P37608	0.16814761215629523
109	1	Fold1	>Q9X3P3	0.1632951289398281
110	1	Fold1	>O69282	0.14940000000000003
111	1	Fold1	>Q02251	0.13983893889152058
112	1	Fold1	>P26235	0.15409921671018273
113	1	Fold1	>P24943	0.15156769596199526
114	1	Fold1	>P54104	0.15210762331838562
115	1	Fold1	>Q10900	0.13293538461538462
116	1	Fold1	>Q79VE8	0.15276960784313723
117	1	Fold1	>P22821	0.15222891566265062
118	1	Fold1	>P81594	0.15332584269662924
119	1	Fold1	>P28611	0.14914814814814814
120	1	Fold1	>Q50367	0.135414364640884
121	1	Fold1	>P15363	0.16148883374689826
122	1	Fold1	>P23648	0.15031446540880505
123	1	Fold1	>P31306	0.14879336349924585
124	1	Fold1	>P39128	0.17204402515723272
125	1	Fold1	>P0A334	0.1574375
126	1	Fold1	>P37386	0.1464676616915423
127	1	Fold1	>P30195	0.17302231237322516
128	1	Fold1	>P28628	0.1633695652173913
129	1	Fold1	>O08362	0.14855
130	1	Fold1	>P96710	0.1538362068965517

continues on next page

Table 1 – continued from previous page

	Group	Fold	Name	Polarizability
131	1	Fold1	>P49022	0.15143174250832406
132	1	Fold1	>P50012	0.16348754448398575
133	1	Fold1	>P25959	0.17112903225806453
134	1	Fold1	>P28612	0.15643678160919539
135	1	Fold1	>P54571	0.150491452991453
136	1	Fold1	>P42237	0.15230769230769234
137	1	Fold1	>P24012	0.16362318840579712
138	1	Fold1	>P35596	0.14731907894736843
139	1	Fold1	>O34360	0.1631736526946108
140	1	Fold1	>P21608	0.1461764705882353
141	1	Fold1	>P50738	0.16784403669724768
142	1	Fold1	>P54582	0.14899159663865547
143	1	Fold1	>P29761	0.15381766381766382
144	1	Fold1	>Q9ZFB5	0.17177464788732394
145	1	Fold1	>P21625	0.13603305785123967
146	1	Fold1	>P39603	0.17482758620689653
147	1	Fold1	>A6QH29	0.13775720164609054
148	1	Fold1	>P33247	0.16141045958795563
149	1	Fold1	>P20487	0.17884353741496598
150	1	Fold1	>P0A4N1	0.16119540229885054
151	1	Fold1	>P25953	0.16058988764044946
152	1	Fold1	>P36948	0.14416149068322978
153	1	Fold1	>O52866	0.15846153846153846
154	1	Fold1	>P39129	0.16183050847457628
155	1	Fold1	>P02983	0.16043572984749452
156	1	Fold1	>P39694	0.1360975609756098
157	1	Fold1	>P50002	0.14826180257510727
158	1	Fold1	>Q9ZBM2	0.15679999999999997
159	1	Fold1	>P19580	0.16213362068965517
160	1	Fold1	>P39793	0.14133479212253827
161	1	Fold1	>P12012	0.1426785714285714
162	1	Fold1	>P39570	0.1701358695652174
163	1	Fold1	>Q04959	0.12177852348993289
164	1	Fold1	>P65565	0.15606060606060607
165	1	Fold1	>P65561	0.15531707317073173
166	1	Fold1	>P42252	0.17326530612244895
167	1	Fold1	>P25958	0.15598425196850393
168	1	Fold1	>P54559	0.15495121951219515
169	1	Fold1	>P95246	0.14829174664107486
170	1	Fold1	>P46349	0.15479744136460555
171	1	Fold1	>P43440	0.15547671840354765
172	1	Fold1	>O34853	0.17141129032258065
173	1	Fold1	>P54146	0.14382743362831854
174	2	Fold2	>Q02114	0.13903225806451616
175	2	Fold2	>P13692	0.13089147286821706
176	2	Fold2	>P38536	0.14843095110155827
177	2	Fold2	>P09333	0.135796812749004
178	2	Fold2	>P35825	0.13877035830618895
179	2	Fold2	>P0C0P9	0.15839080459770113

continues on next page

Table 1 – continued from previous page

	Group	Fold	Name	Polarizability
180	2	Fold2	>Q05044	0.12944086021505377
181	2	Fold2	>Q59634	0.145670356703567
182	2	Fold2	>Q06852	0.13587980977086034
183	2	Fold2	>P35829	0.13324324324324324
184	2	Fold2	>P06546	0.14903133903133903
185	2	Fold2	>P55111	0.127992700729927
186	2	Fold2	>O32083	0.1579078014184397
187	2	Fold2	>Q55242	0.13807017543859648
188	2	Fold2	>P0A5B7	0.15500000000000003
189	2	Fold2	>P54423	0.14286353467561524
190	2	Fold2	>P36921	0.16022222222222224
191	2	Fold2	>Q8GFE2	0.14642201834862387
192	3	Fold3	>P47437	0.16338709677419355
193	3	Fold3	>Q8G7W9	0.160265306122449
194	3	Fold3	>P39802	0.15384615384615385
195	3	Fold3	>Q828D2	0.15112033195020746
196	3	Fold3	>Q88Z97	0.1569985358711567
197	3	Fold3	>P42450	0.1519350473612991
198	3	Fold3	>P59934	0.14201257861635222
199	3	Fold3	>O53079	0.14275390625
200	3	Fold3	>Q82K53	0.12823135755258128
201	3	Fold3	>Q3AB78	0.15460829493087552
202	3	Fold3	>A0JSV8	0.1475359342915811
203	3	Fold3	>Q8XMP3	0.16434782608695656
204	3	Fold3	>P0A0A5	0.1691428571428571
205	3	Fold3	>Q53778	0.1634109816971714
206	3	Fold3	>P07515	0.1359090909090909
207	3	Fold3	>Q08291	0.14696969696969697
208	3	Fold3	>Q2SS99	0.1645088161209068
209	3	Fold3	>Q55055	0.15886666666666668
210	3	Fold3	>Q8EUR6	0.1643655913978495
211	3	Fold3	>Q8G4V2	0.14400000000000002
212	3	Fold3	>Q02VB0	0.16006379585326952
213	3	Fold3	>P28618	0.1526046511627907
214	3	Fold3	>P50589	0.15116788321167882
215	3	Fold3	>Q56198	0.1410975609756098
216	3	Fold3	>P27623	0.17341085271317824
217	3	Fold3	>P59959	0.13708520179372197
218	3	Fold3	>O53441	0.13992957746478873
219	3	Fold3	>Q8G864	0.1446137339055794
220	3	Fold3	>Q6YP15	0.16110526315789475
221	3	Fold3	>P81102	0.15089108910891094
222	3	Fold3	>O05724	0.16157575757575757
223	3	Fold3	>Q46338	0.13384236453201973
224	3	Fold3	>Q59280	0.12610824742268043
225	3	Fold3	>P77949	0.14577142857142858
226	3	Fold3	>O32130	0.16131805157593124
227	3	Fold3	>Q6YPH3	0.16561264822134386
228	3	Fold3	>Q8DS12	0.14373563218390806

continues on next page

Table 1 – continued from previous page

	Group	Fold	Name	Polarizability
229	3	Fold3	>P56968	0.14843511450381675
230	3	Fold3	>Q9AGJ6	0.15193388429752067
231	3	Fold3	>P26379	0.15253424657534245
232	3	Fold3	>O34777	0.16517006802721085
233	3	Fold3	>P39071	0.1367816091954023
234	3	Fold3	>Q2RKV6	0.16225705329153603
235	3	Fold3	>Q2G1N7	0.17764000000000005
236	3	Fold3	>P50846	0.14183673469387756
237	3	Fold3	>Q7NAT8	0.1618857142857143
238	3	Fold3	>P07343	0.14725108225108227
239	3	Fold3	>Q9Z4Q7	0.14178423236514526
240	3	Fold3	>P26380	0.1557668711656442
241	3	Fold3	>P46712	0.13437500000000002
242	3	Fold3	>P68575	0.16480769230769232
243	3	Fold3	>P20805	0.149375
244	3	Fold3	>Q9EYW6	0.1706422018348624
245	3	Fold3	>Q10765	0.15846301633045148
246	3	Fold3	>P0A0C1	0.17102296450939455
247	3	Fold3	>Q2FZE2	0.16869158878504678
248	3	Fold3	>Q46337	0.13907962771458118
249	3	Fold3	>P45855	0.13661577608142497
250	3	Fold3	>P16400	0.16831615120274915
251	3	Fold3	>P56220	0.1471897810218978
252	3	Fold3	>Q06539	0.137032967032967
253	3	Fold3	>P06632	0.14532374100719422
254	3	Fold3	>Q9S4K9	0.15701388888888895
255	3	Fold3	>P47722	0.15240909090909088
256	3	Fold3	>P38037	0.17099999999999999
257	3	Fold3	>Q9ZN78	0.13463768115942026
258	3	Fold3	>P0A4V0	0.16377723970944308
259	3	Fold3	>P54322	0.1410289389067524
260	3	Fold3	>P11931	0.16721590909090908
261	3	Fold3	>P19080	0.16141732283464566
262	3	Fold3	>P13267	0.15808629088378562
263	3	Fold3	>Q03377	0.16964028776978418
264	3	Fold3	>P80734	0.15778625954198475
265	3	Fold3	>Q9F0R1	0.1736521739130435
266	3	Fold3	>P35155	0.15401015228426396
267	3	Fold3	>Q04942	0.15408888888888891
268	3	Fold3	>P11959	0.13704255319148936
269	3	Fold3	>O86963	0.1485878489326765
270	3	Fold3	>Q98R33	0.17583892617449665
271	3	Fold3	>O31749	0.14533333333333337
272	3	Fold3	>Q8Y960	0.16150326797385625
273	3	Fold3	>Q2NJZ2	0.16282608695652173
274	3	Fold3	>P28598	0.13595588235294118
275	3	Fold3	>A5U654	0.13332075471698113
276	3	Fold3	>P96614	0.1544331983805668
277	3	Fold3	>P45870	0.1596195652173913

continues on next page

Table 1 – continued from previous page

	Group	Fold	Name	Polarizability
278	3	Fold3	>P40859	0.1514102564102564
279	3	Fold3	>P32397	0.14770212765957444
280	3	Fold3	>P0A3Q0	0.16314851485148518
281	3	Fold3	>A5U697	0.14259259259259263
282	3	Fold3	>Q2RKH8	0.13110849056603774
283	3	Fold3	>O34767	0.1535204081632653
284	3	Fold3	>P0A672	0.14873913043478262
285	3	Fold3	>O68575	0.1638783269961977
286	3	Fold3	>P55179	0.1499756097560976
287	3	Fold3	>Q8EWP5	0.15707446808510636
288	3	Fold3	>P46807	0.1406958762886598
289	3	Fold3	>Q83GH3	0.1614666666666667
290	3	Fold3	>Q08788	0.16272727272727272
291	3	Fold3	>Q9X895	0.14863793103448275
292	3	Fold3	>P0A2X4	0.15654545454545452
293	3	Fold3	>Q9KWG2	0.14110192837465566
294	3	Fold3	>O32271	0.14607375271149675
295	3	Fold3	>Q83FF5	0.16058685446009385
296	3	Fold3	>Q5FIJ4	0.1482722513089005
297	3	Fold3	>P38645	0.15340380549682875
298	3	Fold3	>O53078	0.14990066225165563
299	3	Fold3	>P00512	0.14188087774294675
300	3	Fold3	>Q47N75	0.15045698924731182
301	3	Fold3	>P81101	0.15308108108108112
302	3	Fold3	>P32081	0.14656716417910445
303	3	Fold3	>Q7TWW7	0.1487878787878788
304	3	Fold3	>O50515	0.12516129032258067
305	3	Fold3	>Q4A6A2	0.1634573002754821
306	3	Fold3	>Q10744	0.16111358574610243
307	3	Fold3	>Q8G4D8	0.15392097264437687
308	3	Fold3	>Q4JVH7	0.1422384428223844
309	3	Fold3	>P46713	0.14073746312684365
310	3	Fold3	>P39804	0.16815126050420168
311	3	Fold3	>Q59112	0.14939849624060147
312	3	Fold3	>Q8RP81	0.14292237442922373
313	3	Fold3	>Q9Z4P6	0.1417142857142857
314	3	Fold3	>P0A5T0	0.1495544554455446
315	3	Fold3	>Q9AGY7	0.1461111111111111
316	3	Fold3	>Q59118	0.14909356725146197
317	3	Fold3	>P42014	0.1502249134948097
318	3	Fold3	>Q03VA1	0.15483146067415732
319	3	Fold3	>P19368	0.1685304659498208
320	3	Fold3	>O32242	0.15999999999999998
321	3	Fold3	>Q1AU26	0.1517062937062937
322	3	Fold3	>P46831	0.15498349834983496
323	3	Fold3	>Q59111	0.15512499999999999
324	3	Fold3	>P0A5L8	0.15021201413427562
325	3	Fold3	>P24301	0.1431
326	3	Fold3	>P47481	0.1584654088050314

continues on next page

Table 1 – continued from previous page

	Group	Fold	Name	Polarizability
327	3	Fold3	>P71756	0.14514666666666665
328	3	Fold3	>Q11010	0.14990665110851809
329	3	Fold3	>Q04FF6	0.15689045936395757
330	3	Fold3	>Q93PS3	0.15070205479452053
331	3	Fold3	>O32885	0.15229946524064172
332	3	Fold3	>Q45066	0.15849875930521093
333	3	Fold3	>P00469	0.16765822784810125
334	3	Fold3	>P54264	0.16292899408284023
335	3	Fold3	>P16304	0.1505069124423963
336	3	Fold3	>Q7A338	0.15237654320987654
337	3	Fold3	>Q45588	0.15634615384615388
338	3	Fold3	>P41019	0.14572207084468666
339	3	Fold3	>P37487	0.1479935275080906
340	3	Fold3	>P0A5P0	0.1621854304635762
341	3	Fold3	>P46815	0.14056390977443606
342	3	Fold3	>P11018	0.13695924764890277
343	3	Fold3	>P59076	0.15293075684380028
344	3	Fold3	>P13375	0.15601336302895322
345	3	Fold3	>P75361	0.15682281059063133
346	3	Fold3	>P34024	0.16432176656151415
347	3	Fold3	>P0A5S6	0.14355048859934855
348	3	Fold3	>O53077	0.14422680412371136
349	3	Fold3	>A1R7K6	0.15106719367588933
350	3	Fold3	>P46033	0.16033755274261605
351	3	Fold3	>P53627	0.14901812688821753
352	3	Fold3	>Q831U3	0.15811239193083576
353	3	Fold3	>P37552	0.14768
354	3	Fold3	>P22346	0.16349934469200522
355	3	Fold3	>Q4A5X2	0.17060240963855416
356	3	Fold3	>P0A512	0.15809312638580936
357	3	Fold3	>O34559	0.16458445040214476
358	3	Fold3	>Q97EB7	0.16421359223300971
359	3	Fold3	>P45618	0.1468831168831169
360	3	Fold3	>P21938	0.15643652561247218
361	3	Fold3	>Q5FKI4	0.16045602605863193
362	3	Fold3	>P39127	0.16727272727272727
363	3	Fold3	>P34020	0.14160493827160492
364	3	Fold3	>O31760	0.1517837837837838
365	3	Fold3	>Q5HJF4	0.1571562082777036
366	3	Fold3	>P26900	0.15072948328267474
367	3	Fold3	>P81100	0.14773869346733667
368	3	Fold3	>Q9RC92	0.1588063660477454
369	3	Fold3	>P00343	0.14736196319018408
370	3	Fold3	>P32396	0.15912903225806452
371	3	Fold3	>P17893	0.1568456375838926
372	3	Fold3	>Q9CJ45	0.15236641221374045
373	3	Fold3	>Q9L4Q8	0.16245322245322244
374	3	Fold3	>Q08352	0.13714285714285712
375	3	Fold3	>Q83GS8	0.1462993762993763

continues on next page

Table 1 – continued from previous page

	Group	Fold	Name	Polarizability
376	3	Fold3	>P0A574	0.13420895522388063
377	3	Fold3	>Q97II1	0.14617977528089887
378	3	Fold3	>Q53062	0.14822695035460992
379	3	Fold3	>P12876	0.1462135922330097
380	3	Fold3	>P06535	0.1684895833333333
381	3	Fold3	>P29422	0.15626242544731614
382	3	Fold3	>O32221	0.13333333333333336
383	3	Fold3	>P11540	0.1621111111111111
384	3	Fold3	>Q933K8	0.12919565217391304
385	3	Fold3	>P0A0I7	0.1684873949579832
386	3	Fold3	>P50849	0.1483971631205674
387	3	Fold3	>P27580	0.14647642679900746
388	3	Fold3	>Q97R46	0.14139433551198255
389	3	Fold3	>Q46336	0.16285714285714287
390	3	Fold3	>P22326	0.15720379146919433
391	3	Fold3	>O07151	0.15433048433048432
392	3	Fold3	>P40875	0.1475061728395062
393	3	Fold3	>P23532	0.14571428571428569
394	3	Fold3	>Q59331	0.148625
395	3	Fold3	>P35154	0.170398406374502
396	3	Fold3	>P60611	0.1624390243902439
397	3	Fold3	>Q03Q83	0.15691428571428573
398	3	Fold3	>P29094	0.17160142348754448
399	3	Fold3	>P39043	0.15900369003690037
400	4	Fold4	>O33702	0.1350344827586207
401	4	Fold4	>Q09T02	0.15029411764705883
402	4	Fold4	>P43163	0.12882096069868995
403	4	Fold4	>P29958	0.14254452926208652
404	4	Fold4	>P45702	0.15852482269503548
405	4	Fold4	>P0C1U8	0.1394940476190476
406	4	Fold4	>P39597	0.1489423076923077
407	4	Fold4	>P19571	0.16023166023166022
408	4	Fold4	>P11001	0.14877269881026928
409	4	Fold4	>Q03424	0.12204481792717085
410	4	Fold4	>Q8RP81	0.14292237442922373
411	4	Fold4	>P82594	0.15692073170731713
412	4	Fold4	>P0A312	0.1404
413	4	Fold4	>O34344	0.14876847290640396
414	4	Fold4	>P80696	0.12846153846153846
415	4	Fold4	>O87237	0.1410769230769231
416	4	Fold4	>P81715	0.1308447488584475
417	4	Fold4	>P09401	0.13151447661469934
418	4	Fold4	>P82593	0.1462060606060606
419	4	Fold4	>P15555	0.13682266009852215
420	4	Fold4	>Q6E3K9	0.15109803921568624
421	4	Fold4	>O31097	0.14530026109660574
422	4	Fold4	>P10547	0.13941176470588235
423	4	Fold4	>Q05622	0.15087500000000004
424	4	Fold4	>P15698	0.1485891089108911

continues on next page

Table 1 – continued from previous page

	Group	Fold	Name	Polarizability
425	4	Fold4	>Q04707	0.14908205841446454
426	4	Fold4	>Q45882	0.16464926590538342
427	4	Fold4	>Q8GJ44	0.1414900153609831
428	4	Fold4	>Q93M42	0.16158102766798416
429	4	Fold4	>P38578	0.11971830985915492
430	4	Fold4	>P0A4L0	0.15084063047285465
431	4	Fold4	>P00691	0.14822727272727274
432	4	Fold4	>P22222	0.13897810218978102
433	4	Fold4	>Q46134	0.15492000000000003
434	4	Fold4	>P30234	0.13592991913746633
435	4	Fold4	>P26831	0.15133292383292388
436	4	Fold4	>P0A564	0.12989473684210526
437	4	Fold4	>P80172	0.1508
438	4	Fold4	>Q08002	0.1563013698630137
439	4	Fold4	>P0C046	0.1547422680412371
440	4	Fold4	>Q5YLG1	0.15975720789074355
441	4	Fold4	>P0A5Q2	0.13528301886792454
442	4	Fold4	>P11701	0.1459245283018868
443	4	Fold4	>Q2FWV6	0.1556034482758621
444	4	Fold4	>P22629	0.12857923497267762
445	4	Fold4	>P07883	0.16042071197411004
446	4	Fold4	>P40943	0.16461916461916462
447	4	Fold4	>P09879	0.1529203539823009
448	4	Fold4	>P39790	0.1418849840255591
449	4	Fold4	>P0C0J0	0.14459798994974876
450	4	Fold4	>P68802	0.16018404907975461
451	4	Fold4	>P08954	0.16510638297872335
452	4	Fold4	>A5HZZ9	0.16402777777777777
453	4	Fold4	>P20910	0.1401272727272727
454	4	Fold4	>A6QG59	0.16109090909090906
455	4	Fold4	>P96501	0.14389261744966445
456	4	Fold4	>P04957	0.1574793388429752
457	4	Fold4	>Q8VL79	0.16706896551724137
458	4	Fold4	>O33635	0.14963295880149813
459	4	Fold4	>Q44052	0.14694227769110765
460	4	Fold4	>P02968	0.1377302631578947
461	4	Fold4	>Q45070	0.15635071090047392
462	4	Fold4	>P0A5P8	0.12113095238095237
463	4	Fold4	>P29767	0.1437179487179487
464	4	Fold4	>P0C0I6	0.16610169491525426
465	4	Fold4	>P22266	0.13889855072463766
466	4	Fold4	>P60158	0.12935622317596565
467	4	Fold4	>P25959	0.17112903225806453
468	4	Fold4	>A6QIG7	0.16221476510067112
469	4	Fold4	>P17670	0.15570048309178744
470	4	Fold4	>P22637	0.1434057971014493
471	4	Fold4	>P37957	0.14500000000000002
472	4	Fold4	>Q6TYB1	0.16816608996539797
473	4	Fold4	>P09616	0.15467084639498432

continues on next page

Table 1 – continued from previous page

	Group	Fold	Name	Polarizability
474	4	Fold4	>P34020	0.14160493827160492
475	4	Fold4	>Q9KWM4	0.15105263157894736
476	4	Fold4	>P00644	0.15073593073593072
477	4	Fold4	>P10424	0.14885620915032682
478	4	Fold4	>P0A668	0.11854922279792747
479	4	Fold4	>P0A618	0.1430057803468208
480	4	Fold4	>Q03091	0.1502076124567474
481	4	Fold4	>P85147	0.12258064516129032
482	4	Fold4	>P29141	0.1377543424317618
483	4	Fold4	>P0A566	0.1376
484	4	Fold4	>Q9KJT6	0.13957198443579766
485	4	Fold4	>Q48919	0.12446194225721784
486	4	Fold4	>O34310	0.14868778280542985
487	4	Fold4	>P39046	0.13635135135135135
488	4	Fold4	>P42983	0.1497058823529412
489	4	Fold4	>P0C2S1	0.15620111731843575
490	4	Fold4	>A5H1G9	0.13407894736842105
491	4	Fold4	>P01552	0.17139097744360907
492	4	Fold4	>Q53728	0.13832733812949638
493	4	Fold4	>P35804	0.13393034825870648
494	4	Fold4	>P39045	0.12611524163568774
495	4	Fold4	>P21543	0.1461371237458194
496	4	Fold4	>A9Q0M7	0.14957142857142858
497	4	Fold4	>Q53591	0.15646341463414634
498	4	Fold4	>P16397	0.1412142358688067
499	4	Fold4	>Q9K6W0	0.1502443609022556
500	4	Fold4	>Q2FZL3	0.15587786259541986
501	4	Fold4	>P16169	0.163125
502	4	Fold4	>Q0TV31	0.15894472361809045
503	4	Fold4	>P45741	0.15234718826405869
504	4	Fold4	>P39652	0.15140624999999996
505	4	Fold4	>P43131	0.14239263803680982
506	4	Fold4	>Q9RLV9	0.16367164179104474
507	4	Fold4	>P10335	0.14849275362318842
508	4	Fold4	>Q2QBT0	0.13854545454545458
509	4	Fold4	>O87236	0.15728813559322036
510	4	Fold4	>Q44856	0.15872881355932203
511	4	Fold4	>P33673	0.15192691029900332
512	4	Fold4	>P00649	0.15314814814814812
513	4	Fold4	>P39116	0.14190476190476192
514	4	Fold4	>P0A5P2	0.15410714285714286
515	4	Fold4	>P0C047	0.1495192307692308
516	4	Fold4	>P0C1D7	0.1430593607305936
517	4	Fold4	>Q9LAB5	0.15245714285714285
518	4	Fold4	>P0A5Q4	0.14714912280701756
519	4	Fold4	>P06886	0.15547008547008545
520	4	Fold4	>P29148	0.13957627118644067
521	4	Fold4	>P15917	0.16411619283065515
522	4	Fold4	>P54422	0.14775127768313456

CHAPTER

TEN

INDICES AND TABLES

- genindex
- modindex
- search

INDEX

Symbols

A

C

D

F

G

L

N

P

S