

1/17

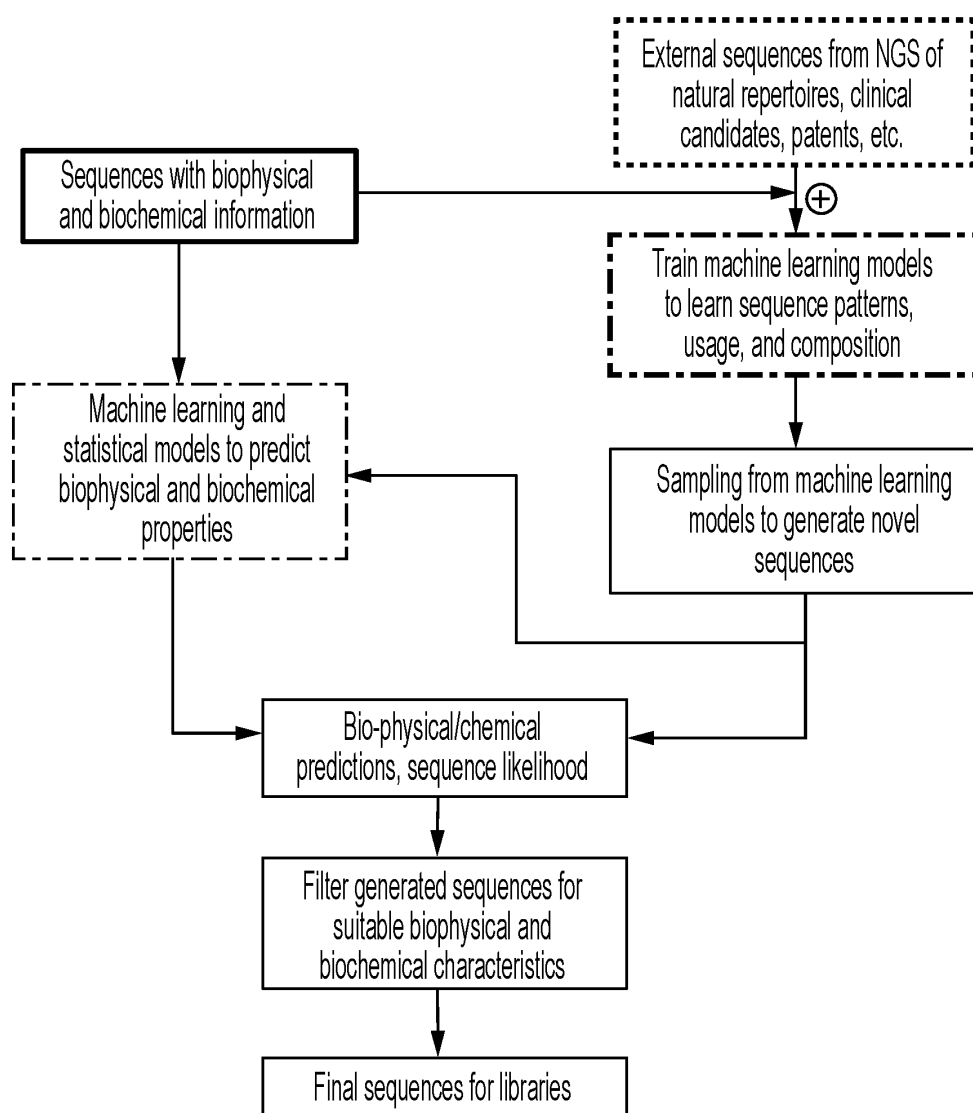


FIG. 1

2/17

Output layer, O
Hydrophobic Interaction
Chromatography Retention Time,
HIC RT

- Combine extracted features across HC and LC

- Convolution layers
- Recurrent layers
- Dense layers
- Attention layers

- Input layer, I
- One-hot encoding
 - Existing amino-acid property scales

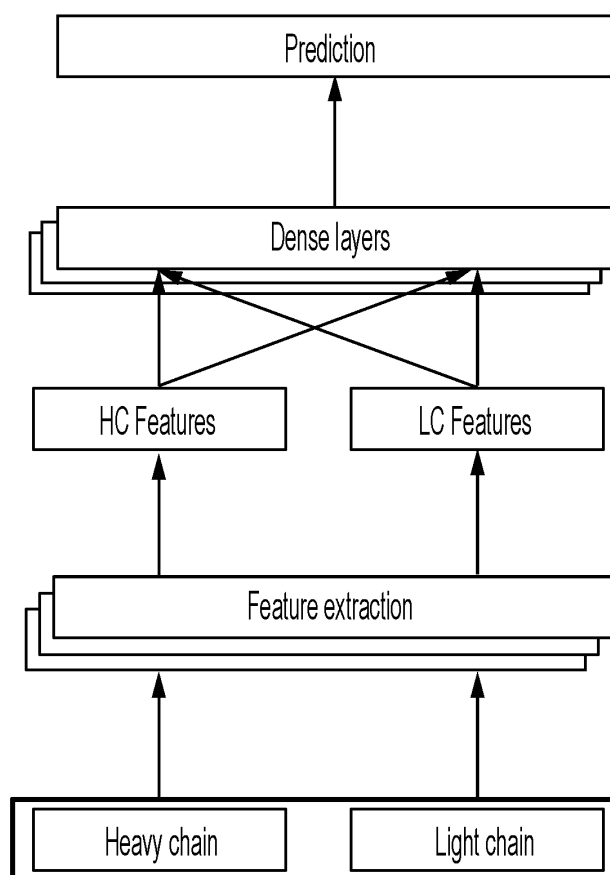


FIG. 2

Updating segment weights after matching for H3 sequence AAEPGGGSWPRYEYNF

Match 1: AAEP <u>PSGGGS</u> WPRYEYN <u>F</u> D-segment = <u>PSGGGS</u> J-segment = <u>EYNY</u> Len(D+J) = 9, Match(D+J) = 8 Candidate N1 = AAE Candidate N2 = WPRY S = 7.11	Match 2: AAEP <u>SGGS</u> WPRYEYN <u>F</u> D-segment = <u>SGGSW</u> J-segment = <u>EYNY</u> Len(D+J) = 9, Match(D+J) = 8 Candidate N1 = AAEP Candidate N2 = PRY S = 7.11	Match 3: AAEP <u>SGGS</u> WPRYEYN <u>F</u> D-segment = <u>SGGS</u> J-segment = <u>EYNY</u> Len(D+J) = 8, Match(D+J) = 7 Candidate N1 = AAEP Candidate N2 = WPRY S = 6.13
--	---	---

N=2, since there are two matches with identical maximal scores of 7.11

Ignore this match since 7.11 > 6.13

Increment usage of segments:

$$\begin{aligned} P(\text{PSGGGS}) &= P(\text{PSGGGS}) + 1/N \\ P(\text{EYNY}) &= P(\text{EYNY}) + 1/N + 1/N \\ P(\text{AAE}) &= P(\text{AAE}) + 1/N \\ P(\text{AAEP}) &= P(\text{AAEP}) + 1/N \\ P(\text{WPRY}) &= P(\text{WPRY}) + 1/N \\ P(\text{PRY}) &= P(\text{PRY}) + 1/N \end{aligned}$$

FIG. 3

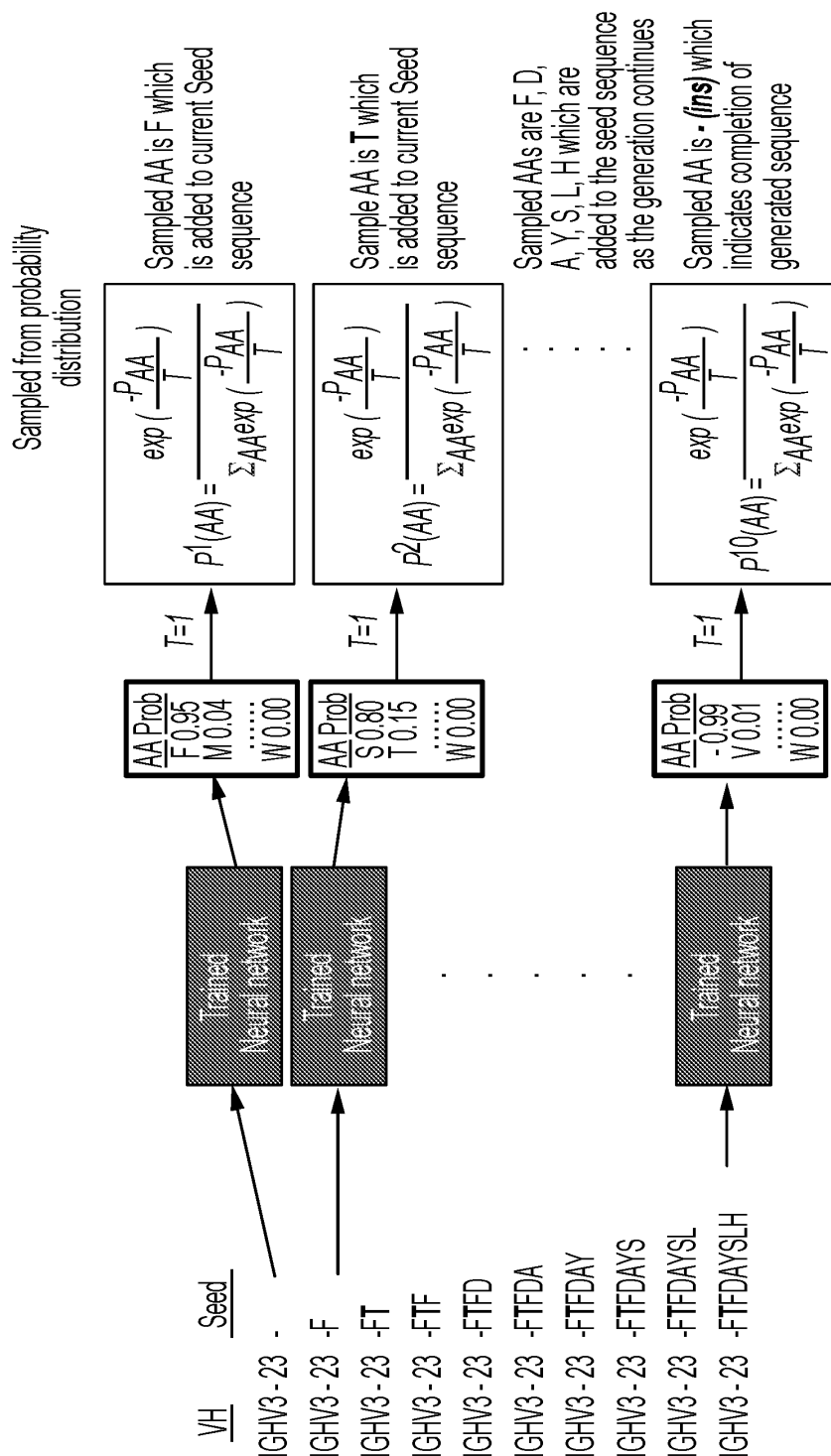
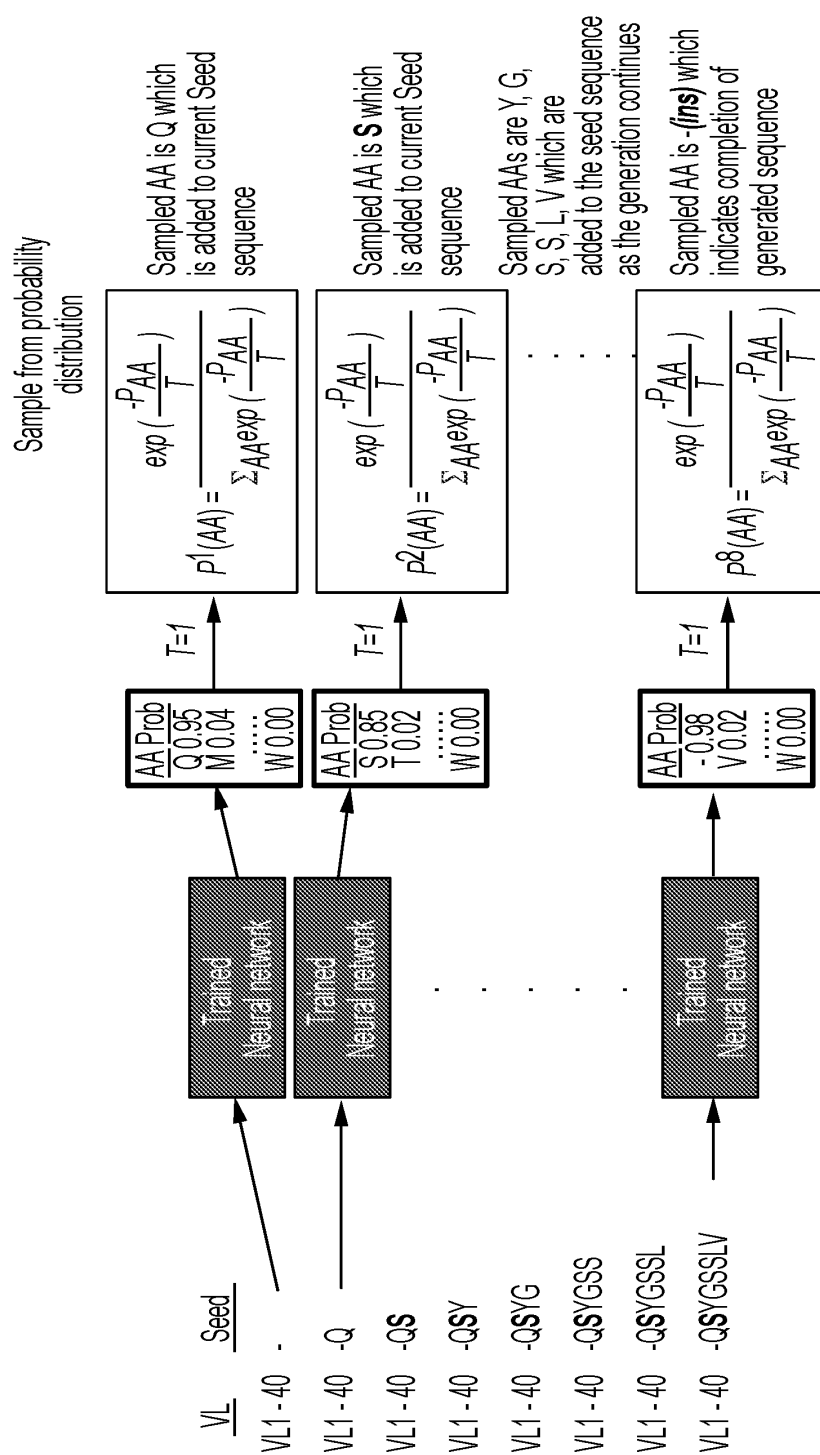


FIG. 4



	VH1.02	VH1.18	VH1.69.01	VH3.07	VH3.09	VH3.21	VH3.23	VH3.30	VH3.33	VH3.48	VH4.31	VH4.34	VH4.39	VH4.59	VH4.b	VH5.51
Target	•	•	●	•	•	•	●	•	•	•	•	•	●	•	•	•
G	•	•	●	•	•	•	●	•	•	•	•	•	●	•	•	•
A	•	•	●	•	•	•	●	•	•	•	•	•	●	•	•	•
V	•	•	●	•	•	•	●	•	•	•	•	•	●	•	•	•
S	•	•	●	•	•	•	●	•	•	•	•	•	●	•	•	•
T	•	•	●	•	•	•	●	●	•	•	•	•	●	•	•	•
P	•	•	●	•	•	•	●	•	•	•	•	•	●	•	•	•
L	•	•	●	•	•	•	●	•	•	•	•	•	●	•	•	•
E	•	•	●	•	•	•	•	•	•	•	•	•	●	•	•	•
R	•	•	●	•	•	•	•	•	•	•	•	•	●	•	•	•
Q	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
I	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
K	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
D	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
Y	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
F	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
H	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
N	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
X	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
C	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•

FIG. 5

7/17

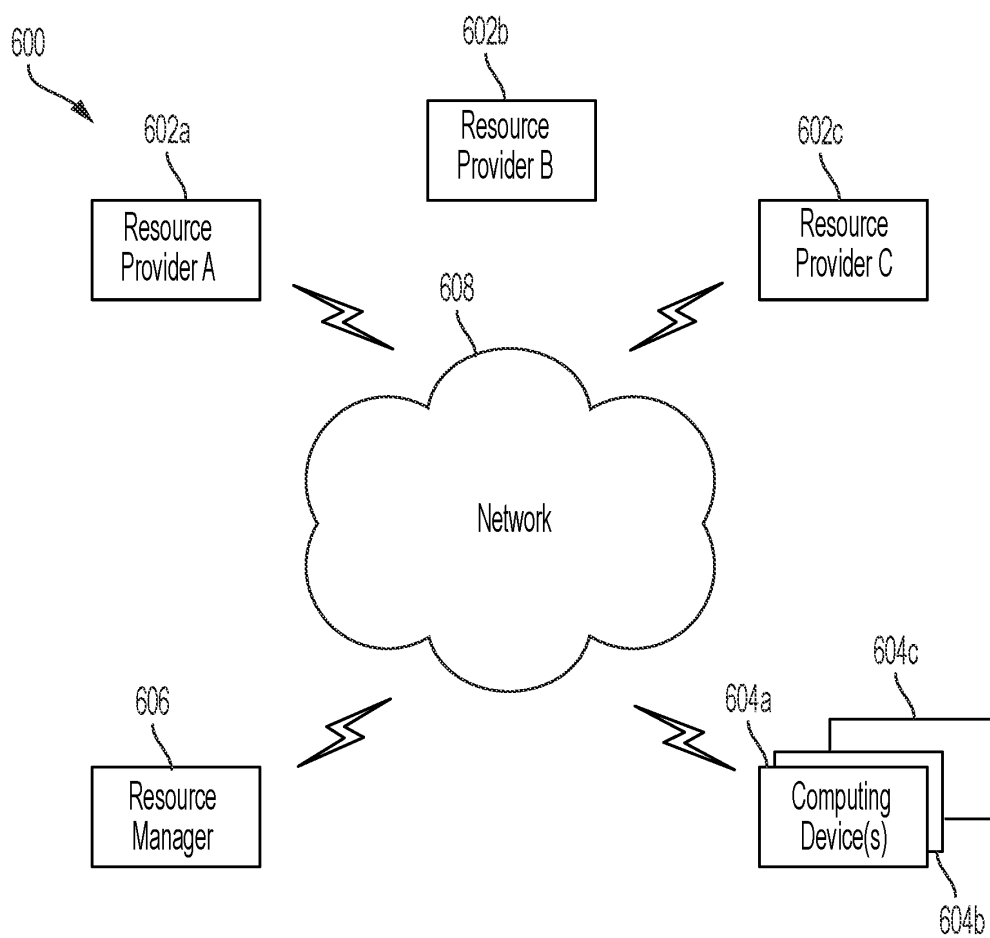
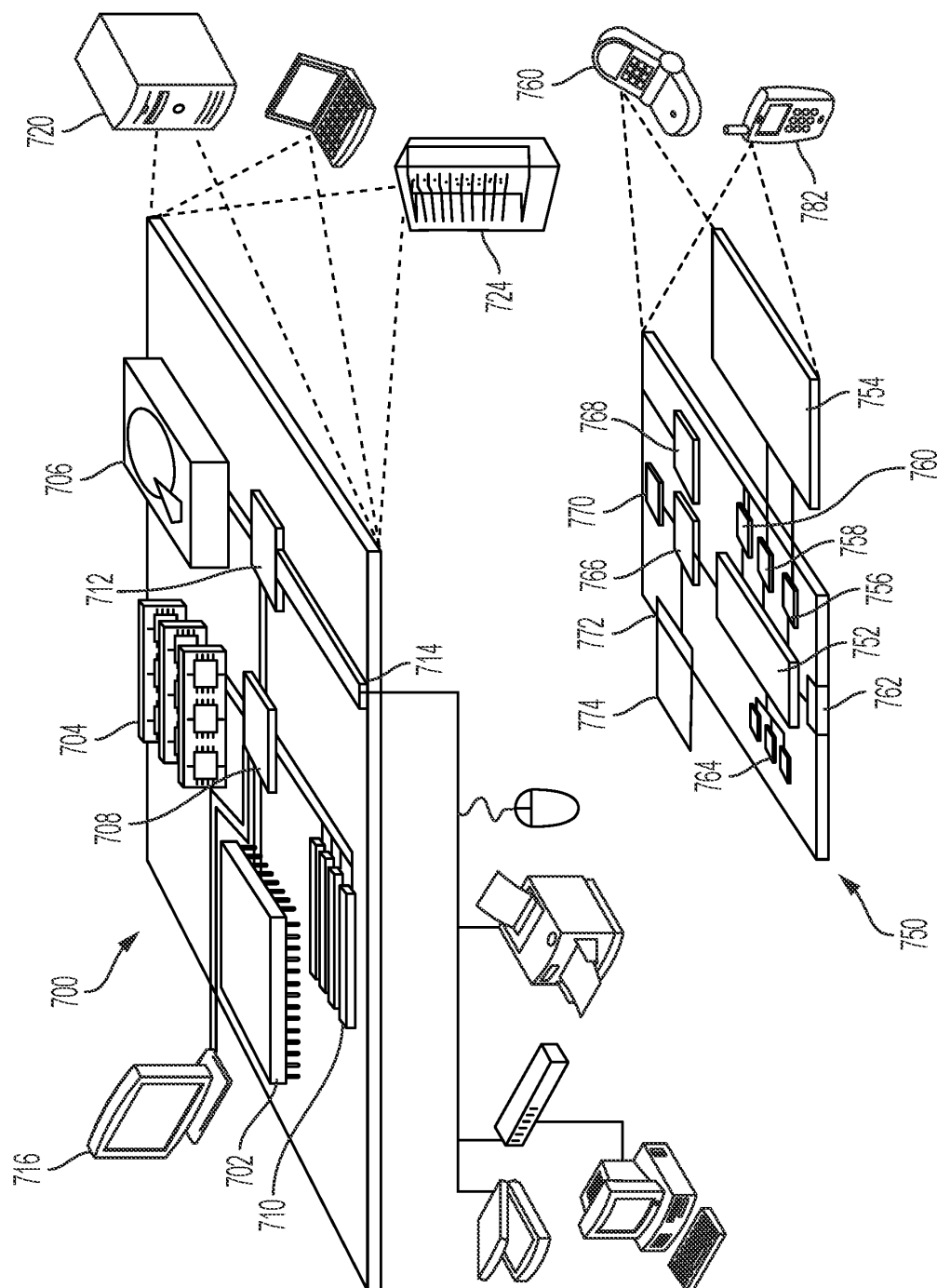


FIG. 6

8/17



$x_i' = W^G x_i$	Learn central residue features	L	Length of the sequence
$x_j' = W^H x_j$	Learn neighbor influencing features	l	Index of the layer of the model
		x	input features for a residue from prior layers or starting input to the model
		x'	intermediate features learned in the GCN layer
$x_i^{l+1} = f(x_i' + \sum_{j=1}^L C_{ij}^{lh} x_j')$	Residual connection with activation	W	learned weight matrices in the GCN layer to transform from input features to intermediate features
		C_{ij}^{lh}	learned neighbor attention or influence weights on the central residue to combine input and/or learned intermediate features. The superscript h indicates multiple independent sets of attention weights that can be learned
$x_i^{l+1} = f(W^d(x_i' \sum_{j=1}^L C_{ij}^{lh} x_j') + b)$	Feature concatenation followed by convolution with activation	x^{l+1}	learned output features from the model
		f	activation functions - can be linear or non-linear
		b	learned bias values

FIG. 8A

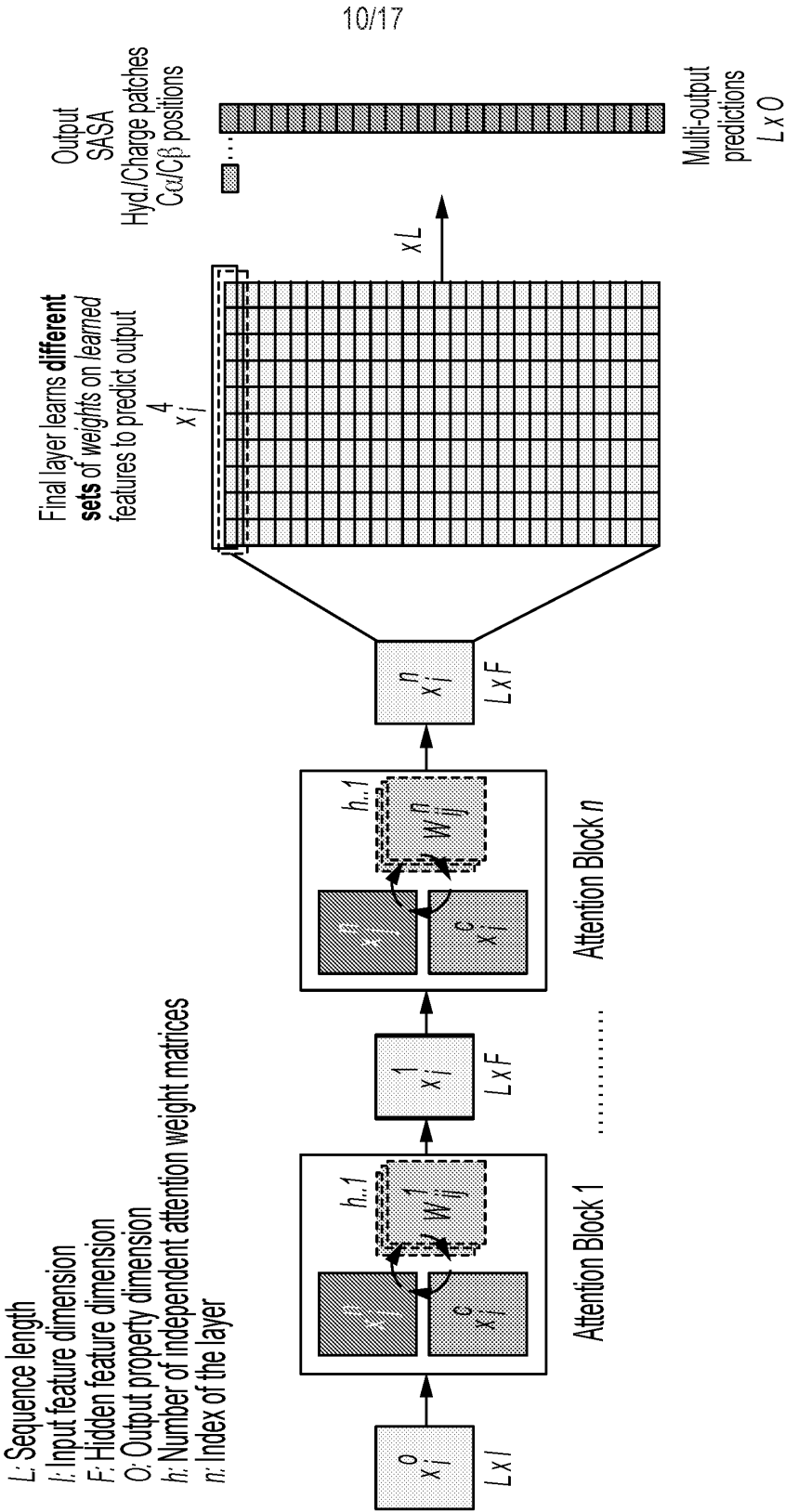
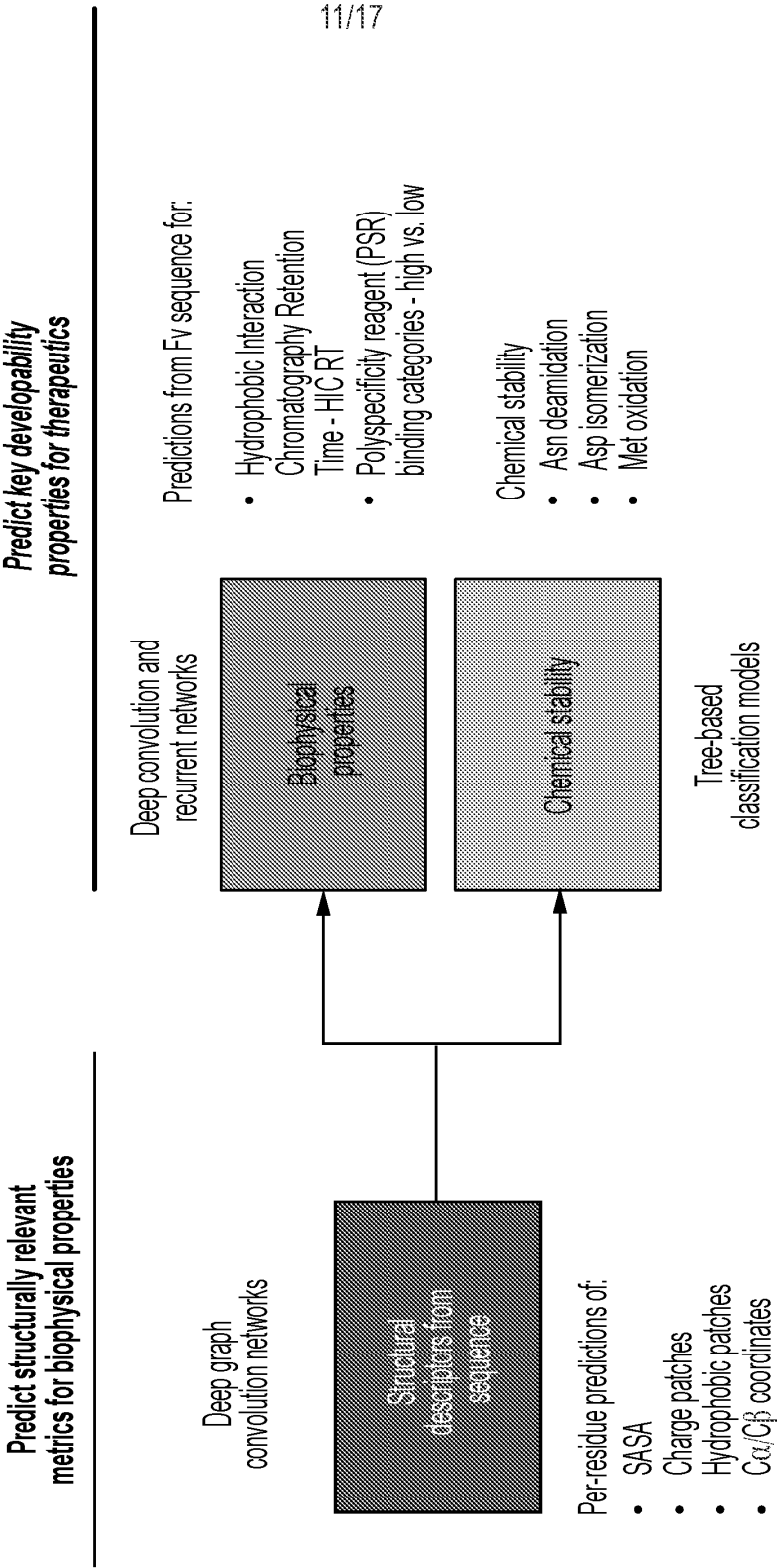


FIG. 8B



11/17

FIG. 9

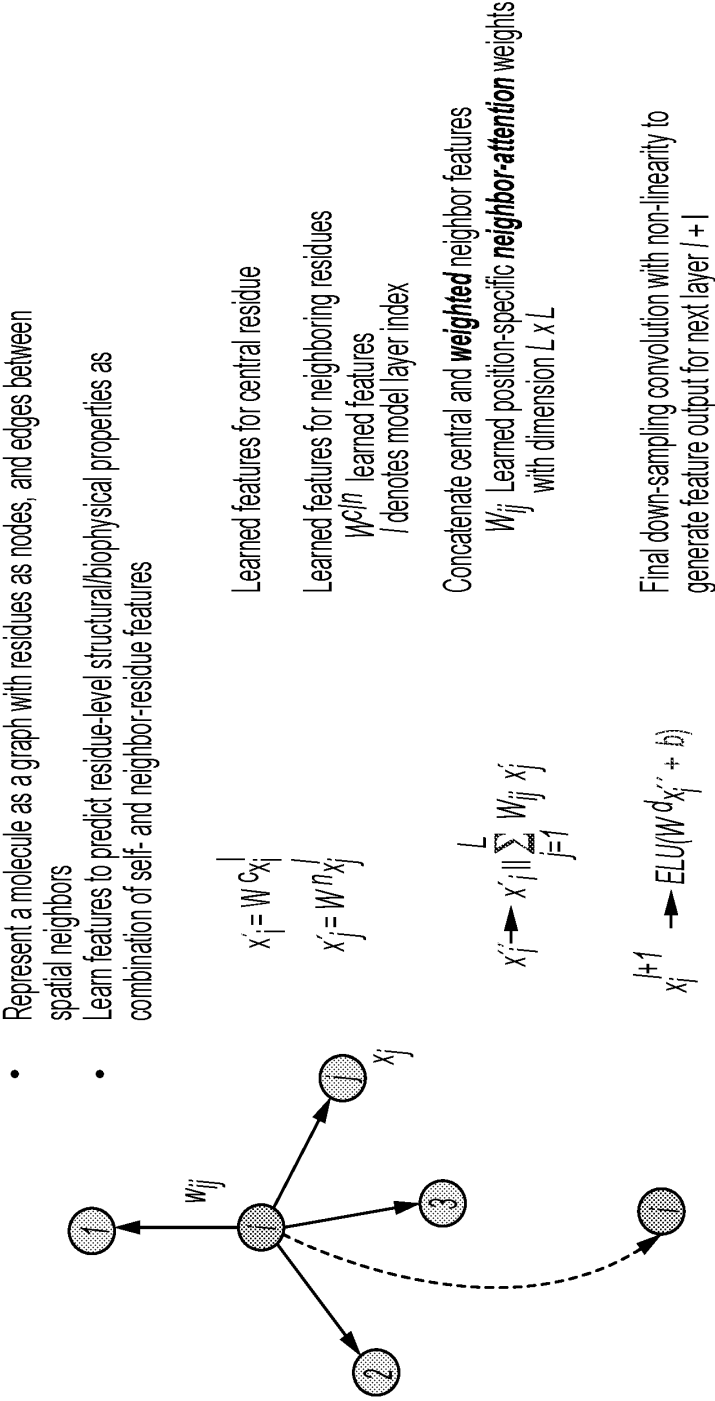


FIG. 10A

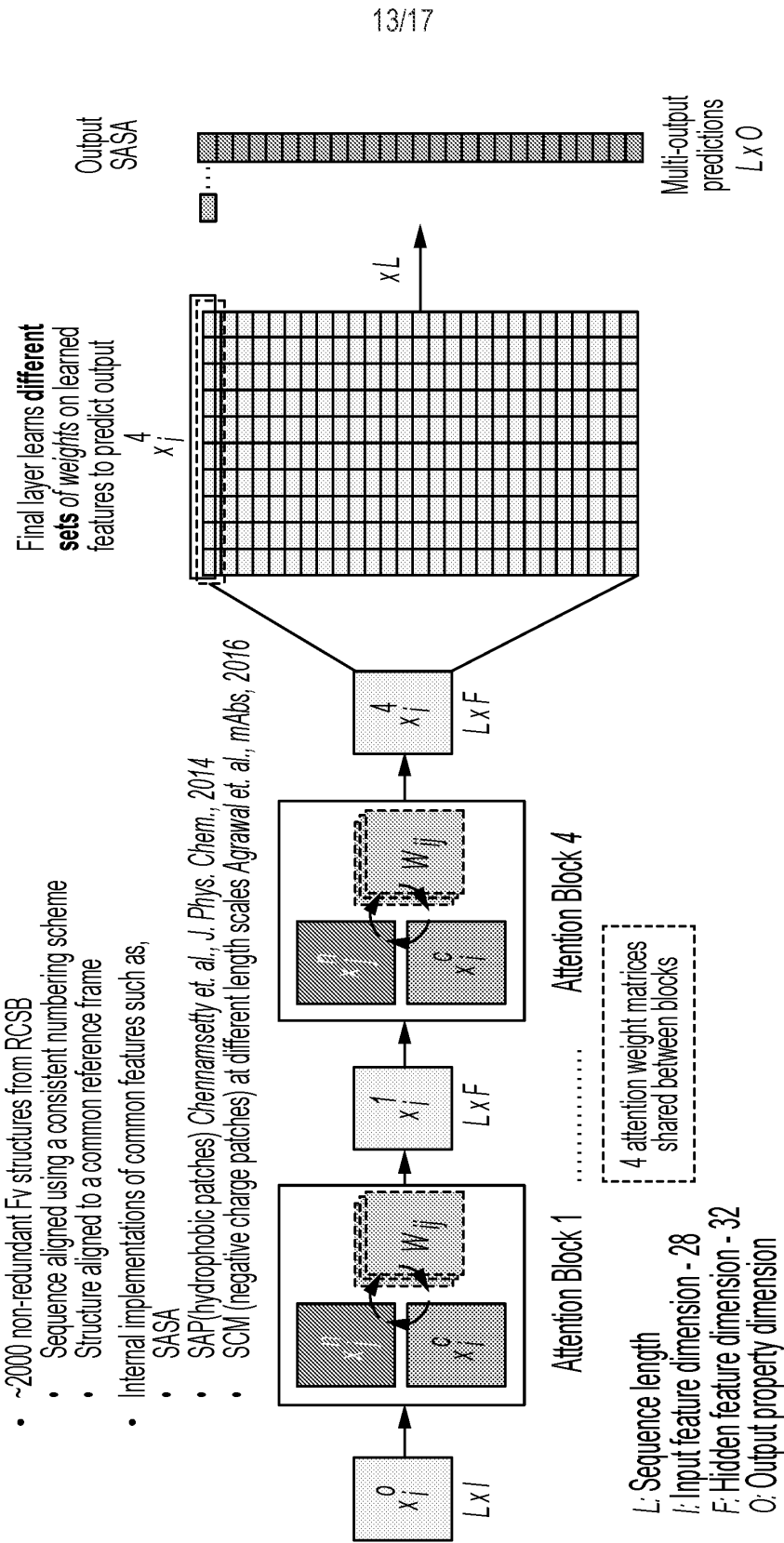


FIG. 10B

- ~2000 non-redundant Fv structures from RCSB
- Sequence aligned using a consistent numbering scheme
- Structure aligned to a common reference frame
- Internal implementations of common features such as,
 - SASA
 - SAP(hydrophobic patches) *Chenramsetty et. al., J. Phys. Chem., 2014*
 - SCM (negative charge patches) at different length scales *Agrawal et. al., mAbs, 2016*

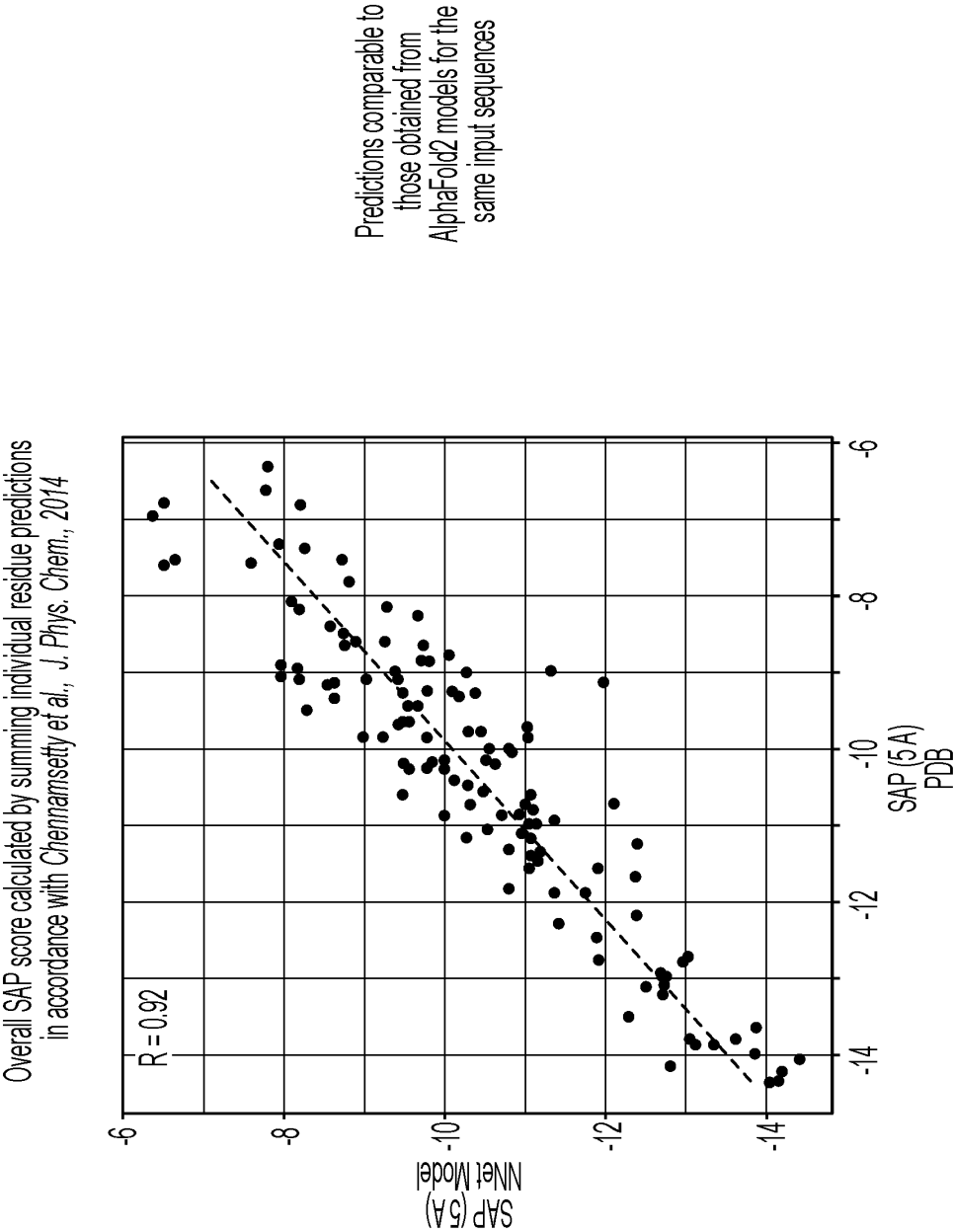


FIG. 11

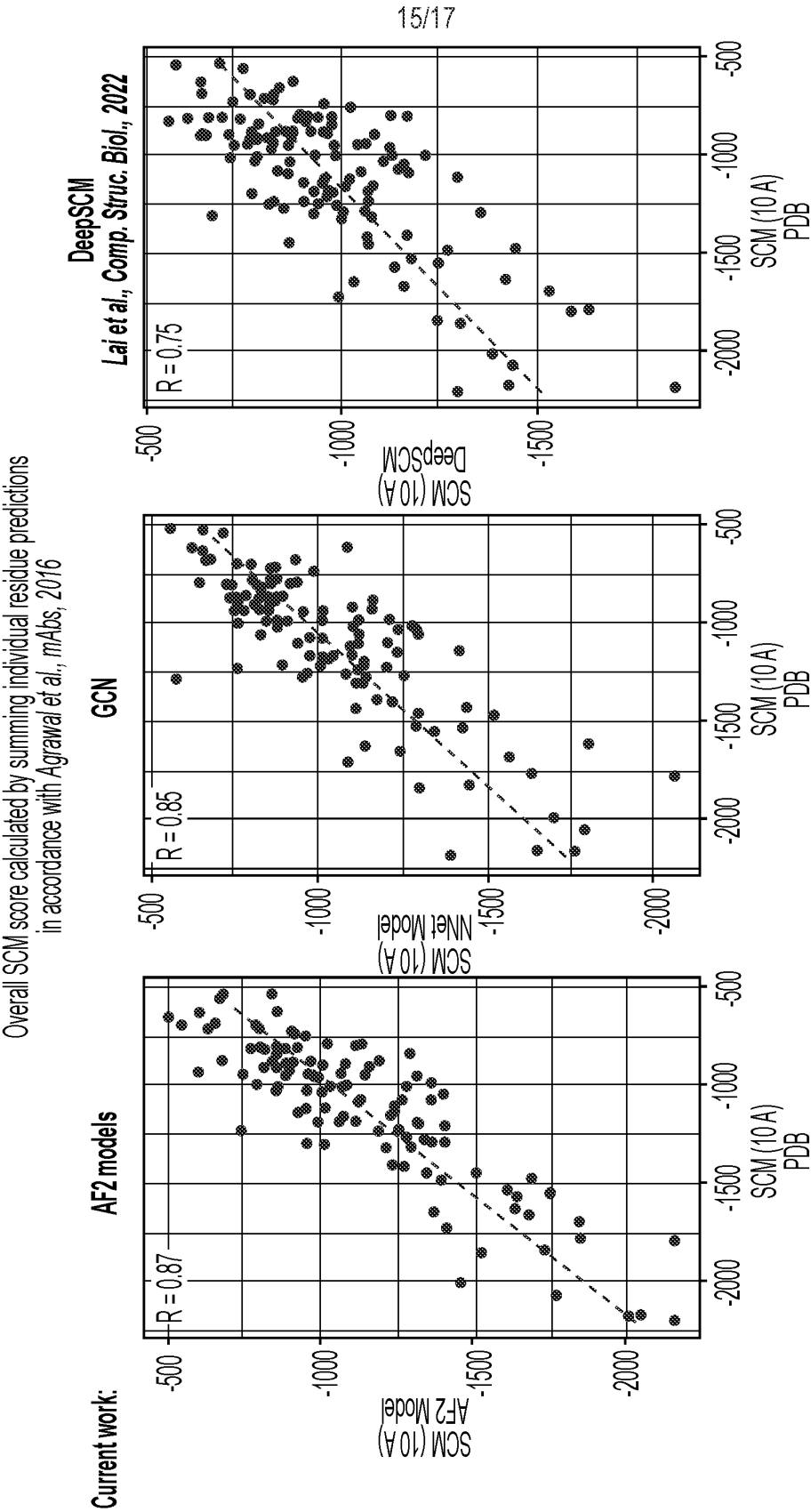


FIG. 12

