Model for prediction of probability of viral entry

Note:- When all the parameters were considered the glm logistic regression converged with Interaction energy, entropy side chain and entropy main chain as independent variables.

```
> logit <- glm(Infection ~ Interaction_Energy
+entropy_mainchain+Entropy_sidechain,data=data,family="binomial"
)
> summary(logit)

Call:
glm(formula = Infection ~ Interaction_Energy + entropy_mainchain
+
    Entropy_sidechain, family = "binomial", data = data)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-1.46547  -0.00002   0.01061   0.02567   1.20598

Coefficients:
                    Estimate Std. Error z value Pr(>|z|)
(Intercept)        -89.65044   16.70336  -5.367 8.00e-08 ***
Interaction_Energy  -0.19841    0.08769  -2.263   0.0237 *
entropy_mainchain    9.04423    1.53148   5.906 3.51e-09 ***
Entropy_sidechain    1.06761    0.56215   1.899   0.0575 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1235.83  on 1019  degrees of freedom
Residual deviance:  237.73  on 1016  degrees of freedom
AIC: 245.73

Number of Fisher Scoring iterations: 9

> r2Log(logit)
[1] 0.8076323
```

- *Hosmer and Lemeshow goodness of fit (GOF) test*

```
> hoslem.test(data$Infection, fitted(logit))

data:  data$Infection, fitted(logit)
```

```
X-squared = 13.686, df = 8, p-value = 0.09034
```

- Test for significant reduction in null deviance on inclusion of the three parameters – Interaction_Energy + Entropy_sidechain + entropy_mainchain

```
> 1-pchisq(1235.83-237.73,1019-1016)
[1] 0
```