1. Introduction

* Hiện trạng ung thư, bệnh sinh ung thư, vai trò của HDAC và HAT trong cơ thể, HDAC hiện nay trở thành một trong các đích tiềm năng của ung thư với các chất ức chế đã được approved bởi FDA gồm có
* Giới thiệu chung về họ enzyme HDAC, cách phân nhóm và sự khác biệt của các isoform giữa các nhóm
* Isoform nhóm IIb có vai trò gì trong cơ thể, sự chọn lọc cơ chất của nhóm IIb khác gì với các nhóm còn lại. Tại sao phải tìm kiếm các chất ức chế chọn lọc HDAC6
* Các nghiên cứu gần đây về virtual screening các chất ức chế HDAC (hoặc HDAC6) và một số phương pháp sàng lọc đã được áp dụng => điểm mới của nghiên cứu này là sử dụng mô hình pharmacophore kết hợp với mô hình proteochemometrics để sàng lọc ra chất ức chế chọn lọc trên HDAC6.

Script 1:

In 1964, Allfrey collaborating with Faulkner and Mirsky firstly discovered the reversible acetylation of histone proteins and suggested that gene expression could be regulated by key transcriptional modification[[1]](#endnote-1)[[2]](#endnote-2)[[3]](#endnote-3). The balance of protein acetylation levels is regulated by two enzymes having opposite effect: histone acetyltranferases (HATs) and histone deacetylase (HDACs)[[4]](#endnote-4). HATs and HDACs regulate post-translational modifications by the acetylation and the deacetylation of the ε-amino group of lysine residues in histone tails and some non-histone proteins[[5]](#endnote-5). Acetylation on lysine could affected various aspects of protein function, therefore it is implicated in many biological processes such as epigenetics[[6]](#endnote-6)[[7]](#endnote-7), cell signaling[[8]](#endnote-8), and metabolism[[9]](#endnote-9)[[10]](#endnote-10). The balance of acetylation and deacetylation of histone maintains body homeostasis. However, the disruption of balanced acetylation-deacetylation can contribute to human diseases, such as cancer and neurodegenerative disease[[11]](#endnote-11)[[12]](#endnote-12)[[13]](#endnote-13). Relevant research also indicated that histone acetylation and deacetylation is recognized as hallmarks of cancer[[14]](#endnote-14). In recent studies, inhibition of HDACs has been proved to become promising strategies for cancer therapy[[15]](#endnote-15).

The histone deacetylase family in human consists 18 members that divided into four classes (I-IV) according to their homology to yeast HDACs[[16]](#endnote-16).

Consequently, histone deacetylase inhibitors (HDACi) have been discovered intensively. It makes HDACi has become one of the most studied antitumor drugs, especially the area of epigenetics is coming closer than ever.

2. Methods and Materials

2.1. Dataset

Data của 11 isoform HDAC được thu thập từ NCBI

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Protein | Locatization | PDB ID | Sequence | NCBI entry | Chiều dài  (a.a) | Class |
| HDAC1 | Nucleus | 4BKX | B |  | 482 | I |
| HDAC2 | Nucleus | 4LXZ | A |  | 488 | I |
| HDAC3 | Nucleus | 4A69 | A |  | 428 | I |
| HDAC4 | Nuc/Cyt | 2VQM | A |  | 972 | IIa |
| HDAC5 | Nuc/Cyt | Nd |  |  | 1122 | IIa |
| HDAC6 | Mainly Cyt | 5EDU | A |  | 1215 | IIb |
| HDAC7 | Nuc/Cyt | 3ZNR | A |  | 954 | IIa |
| HDAC8 | Nucleus | 5VI6 | A |  | 377 | I |
| HDAC9 | Nuc/Cyt | Nd |  |  | 1011 | IIa |
| HDAC10 | Mainly Cyt | 6VNQ | A |  | 669 | IIb |
| HDAC11 | Nuc/Cyt | Nd |  |  | 347 | IV |

Nd: no data published

Sequence:

>CAG46518.1 HDAC1 [Homo sapiens]

MAQTQGTRRKVCYYYDGDVGNYYYGQGHPMKPHRIRMTHNLLLNYGLYRKMEIYRPHKANAEEMTKYHSDDYIKFLRSIRPDNMSEYSKQMQRFNVGEDCPVFDGLFEFCQLSTGGSVASAVKLNKQQTDIAVNWAGGLHHAKKSEASGFCYVNDIVLAILELLKYHQRVLYIDIDIHHGDGVEEAFYTTDRVMTVSFHKYGEYFPGTGDLRDIGAGKGKYYAVNYPLRDGIDDESYEAIFKPVMSKVMEMFQPSAVVLQCGSDSLSGDRLGCFNLTIKGHAKCVEFVKSFNLPMLMLGGGGYTIRNVARCWTYETAVALDTEIPNELPYNDYFEYFGPDFKLHISPSNMTNQNTNEYLEKIKQRLFENLRMLPHAPGVQMQAIPEDAIPEESGDEDEDDPDKRISICSSDKRIACEEEFSDSEEEGEGGRKNSSNFKKAKRVKTEDEKEKDPEEKKEVTEEEKTKEEKPEAKGVKEEVKLA

>NP\_001518.3 histone deacetylase 2 [Homo sapiens]

MAYSQGGGKKKVCYYYDGDIGNYYYGQGHPMKPHRIRMTHNLLLNYGLYRKMEIYRPHKATAEEMTKYHSDEYIKFLRSIRPDNMSEYSKQMQRFNVGEDCPVFDGLFEFCQLSTGGSVAGAVKLNRQQTDMAVNWAGGLHHAKKSEASGFCYVNDIVLAILELLKYHQRVLYIDIDIHHGDGVEEAFYTTDRVMTVSFHKYGEYFPGTGDLRDIGAGKGKYYAVNFPMRDGIDDESYGQIFKPIISKVMEMYQPSAVVLQCGADSLSGDRLGCFNLTVKGHAKCVEVVKTFNLPLLMLGGGGYTIRNVARCWTYETAVALDCEIPNELPYNDYFEYFGPDFKLHISPSNMTNQNTPEYMEKIKQRLFENLRMLPHAPGVQMQAIPEDAVHEDSGDEDGEDPDKRISIRASDKRIACDEEFSDSEDEGEGGRRNVADHKKGAKKARIEEDKKETEDKKTDVKEEDKSKDNSGEKTDTKGTKSEQLSNP

>AAC52038.1 histone deacetylase 3 [Homo sapiens]

MAKTVAYFYDPDVGNFHYGAGHPMKPHRLALTHSLVLHYGLYKKMIVFKPYQASQHDMCRFHSEDYIDFLQRVSPTNMQGFTKSLNAFNVGDDCPVFPGLFEFCSRYTGASLQGATQLNNKICDIAINWAGGLHHAKKFEASGFCYVNDIVIGILELLKYHPRVLYIDIDIHHGDGVQEAFYLTDRVMTVSFHKYGNYFFPGTGDMYEVGAESGRYYCLNVPLRDGIDDQSYKHLFQPVINQVVDFYQPTCIVLQCGADSLGCDRLGCFNLSIRGHGECVEYVKSFNIPLLVLGGGGYTVRNVARCWTYETSLLVEEAISEELPYSEYFEYFAPDFTLHPDVSTRIENQNSRQYLDQILQTIFENLKMLNHAPSVQIHDVPADLLTYDRTDEADAEERGPEENYSRPEAPNEFYDGDHDNDKESDVEI

>AAH39904.1 HDAC4 protein [Homo sapiens]

MLAMKHQQELLEHQRKLERHRQEQELEKQHREQKLQQLKNKEKGKESAVASTEVKMKLQEFVLNKKKALAHRNLNHCISSDPRYWYGKTQHSSLDQSSPPQSGVSTSYNHPVLGMYDAKDDFPLRKTASEPNLKLRSRLKQKVAERRSSPLLRRKDGPVVTALKKRPLDVTDSACSSAPGSGPSSPNNSSGSVSAENGIAPAVPSIPAETSLAHRLVAREGSAAPLPLYTSPSLPNITLGLPATGPSAGTAGQQDAERLTLPALQQRLSLFPGTHLTPYLSTSPLERDGGAAHSPLLQHMVLLEQPPAQAPLVTDWYLSGLGALPLHAQSLVGADRVSPSIHKLRQHRPLGRTQSAPLPQNAQALQHLVIQQQHQQFLEKHKQQFQQQQLQMNKIIPKPSEPARQPESHPEETEEELREHQALLDEPYLDRLPGQKEAHAQAGVQVKQEPIESDEEEAEPPREVEPGQRQPSEQELLFRQQALLLEQQRIHQLRNYQASMEAAGIPVSFGGHRPLSRAQSSPASATFPVSVQEPPTKPRFTTGLVYDTLMLKHQCTCGSSSSHPEHAGRIQSIWSRLQETGLRGKCECIRGRKATLEELQTVHSEAHTLLYGTNPLNRQKLDSKKLLGSLASVFVRLPCGGVGVDSDTIWNEVHSAGAARLAVGCVVELVFKVATGELKNGFAVVRPPGHHAEESTPMGFCYFNSVAVAAKLLQQRLSVSKILIVDWDVHHGNGTQQAFYSDPSVLYMSLHRYDDGNFFPGSGAPDEVGTGPGVGFNVNMAFTGGLDPPMGDAEYLAAFRTVVMPIASEFAPDVVLVSSGFDAVEGHPTPLGGYNLSARCFGYLTKQLMGLAGGRIVLALEGGHDLTAICDASEACVSALLGNELDPLPEKVLQQRPNANAVRSMEKVMEIHSKYWRCLQRTTSTAGRSLIEAQTCENEEAETVTAMASLSVGVKPAEKRPDEEPMEEEPPL

>AAH51824.1 Histone deacetylase 5 [Homo sapiens]

MNSPNESDGMSGREPSLEILPRTSLHSIPVTVEVKPVLPRAMPSSMGGGGGGSPSPVELRGALVGSVDPTLREQQLQQELLALKQQQQLQKQLLFAEFQKQHDHLTRQHEVQLQKHLKQQQEMLAAKQQQEMLAAKRQQELEQQRQREQQRQEELEKQRLEQQLLILRNKEKSKESAIASTEVKLRLQEFLLSKSKEPTPGGLNHSLPQHPKCWGAHHASLDQSSPPQSGPPGTPPSYKLPLPGPYDSRDDFPLRKTASEPNLKVRSRLKQKVAERRSSPLLRRKDGTVISTFKKRAVEITGAGPGASSVCNSAPGSGPSSPNSSHSTIAENGFTGSVPNIPTEMLPQHRALPLDSSPNQFSLYTSPSLPNISLGLQATVTVTNSHLTASPKLSTQQEAERQALQSLRQGGTLTGKFMSTSSIPGCLLGVALEGDGSPHGHASLLQHVLLLEQARQQSTLIAVPLHGQSPLVTGERVATSMRTVGKLPRHRPLSRTQSSPLPQSPQALQQLVMQQQHQQFLEKQKQQQLQLGKILTKTGELPRQPTTHPEETEEELTEQQEVLLGEGALTMPREGSTESESTQEDLEEEDEEEDGEEEEDCIQVKDEEGESGAEEGPDLEEPGAGYKKLFSDAQPLQPLQVYQAPLSLATVPHQALGRTQSSPAAPGGMKSPPDQPVKHLFTTGVVYDTFMLKHQCMCGNTHVHPEHAGRIQSIWSRLQETGLLSKCERIRGRKATLDEIQTVHSEYHTLLYGTSPLNRQKLDSKKLLGPISQKMYAVLPCGGIGVDSDTVWNEMHSSSAVRMAVGCLLELAFKVAAGELKNGFAIIRPPGHHAEESTAMGFCFFNSVAITAKLLQQKLNVGKVLIVDWDIHHGNGTQQAFYNDPSVLYISLHRYDNGNFFPGSGAPEEVGGGPGVGYNVNVAWTGGVDPPIGDVEYLTAFRTVVMPIAHEFSPDVVLVSAGFDAVEGHLSPLGGYSVTARCFGHLTRQLMTLAGGRVVLALEGGHDLTAICDASEACVSALLSVELQPLDEAVLQQKPNINAVATLEKVIEIQSKHWSCVQKFAAGLGRSLREAQAGETEEAETVSAMALLSVGAEQAQAAAAREHSPRPAEEPMEQEPAL

>AAH69243.1 HDAC6 protein [Homo sapiens]

MTSTGQDSTTTRQRRSRQNPQSPPQDSSVTSKRNIKKGAVPRSIPNLAEVKKKGKMKKLGQAMEEDLIVGLQGMDLNLEAEALAGTGLVLDEQLNEFHCLWDDSFPEGPERLHAIKEQLIQEGLLDRCVSFQARFAEKEELMLVHSLEYIDLMETTQYMNEGELRVLADTYDSVYLHPNSYSCACLASGSVLRLVDAVLGAEIRNGMAIIRPPGHHAQHSLMDGYCMFNHVAVAARYAQQKHRIRRVLIVDWDVHHGQGTQFTFDQDPSVLYFSIHRYEQGRFWPHLKASNWSTTGFGQGQGYTINVPWNQVGMRDADYIAAFLHVLLPVALEFQPQLVLVAAGFDALQGDPKGEMAATPAGFAQLTHLLMGLAGGKLILSLEGGYNLRALAEGVSASLHTLLGDPCPMLESPGAPCRSAQASVSCALEALEPFWEVLVRSTETVERDNMEEDNVEESEEEGPWEPPVLPILTWPVLQSRTGLVYDQNMMNHCNLWDSHHPEVPQRILRIMCRLEELGLAGRCLTLTPRPATEAELLTCHSAEYVGHLRATEKMKTRELHRESSNFDSIYICPSTFACAQLATGAACRLVEAVLSGEVLNGAAVVRPPGHHAEQDAACGFCFFNSVAVAARHAQTISGHALRILIVDWDVHHGNGTQHMFEDDPSVLYVSLHRYDHGTFFPMGDEGASSQIGRAAGTGFTVNVAWNGPRMGDADYLAAWHRLVLPIAYEFNPELVLVSAGFDAARGDPLGGCQVSPEGYAHLTHLLMGLASGRIILILEGGYNLTSISESMAACTRSLLGDPPPLLTLPRPPLSGALASITETIQVHRRYWRSLRVMKVEDREGPSSSKLVTKKAPQPAKPRLAERMTTREKKVLEAGMGKVTSASFGEESTPGQTNSETAVVALTQDQPSEAATGGATLAQTISEAAIGGAMLGQTTSEEAVGGATPDQTTSEETVGGAILDQTTSEDAVGGATLGQTTSEEAVGGATLAQTTSEAAMEGATLDQTTSEEAPGGTELIQTPLASSTDHQTPPTSPVQGTTPQISPSTLIGSLRTLELGSESQGASESQAPGEENLLGEAAGGQDMADSMLMQGSRGLTDQAIFYAVTPLPWCPHLVAVCPIPAAGLDVTQPCGDCGTIQENWVCLSCYQVYCGRYINGHMLQHHGNSGHPLVLSYIDLSAWCYYCQAYVHHQALLDVKDIAHQNKFGEDMPHPH

>AAH64840.1 HDAC7 protein [Homo sapiens]

MHSPGADGTQVSPGAHYCSPTGAGCPRPCADTPGPQPQPMDLRVGQRPPVEPPPEPTLLALQRPQRLHHHLFLAGLQQQRSVEPMRLSMDTPMPELQVGPQEQELRQLLHKDKSKRSAVASSVVKQKLAEVILKKQQAALERTVHPNSPGIPYRTLEPLETEGATRSMLSSFLPPVPSLPSDPPEHFPLRKTVSEPNLKLRYKPKKSLERRKNPLLRKESAPPSLRRRPAETLGDSSPSSSSTPASGCSSPNDSEHGPNPILGSEADSDRRTHPTLGPRGPILGSPHTPLFLPHGLEPEAGGTLPSRLQPILLLDPSGSHAPLLTVPGLGPLPFHFAQSLMTTERLSGSGLHWPLSRTRSEPLPPSATAPPPPGPMQPRLEQLKTHVQVIKRSAKPSEKPRLRQIPSAEDLETDGGGPGQVVDDGLEHRELGHGQPEARGPAPLQQHPQVLLWEQQRLAGRLPRGSTGDTVLLPLAQGGHRPLSRAQSSPAAPASLSAPEPASQARVLSSSETPARTLPFTTGLIYDSVMLKHQCSCGDNSRHPEHAGRIQSIWSRLQERGLRSQCECLRGRKASLEELQSVHSERHVLLYGTNPLSRLKLDNGKLAGLLAQRMFVMLPCGGVGVDTDTIWNELHSSNAARWAAGSVTDLAFKVASRELKNGFAVVRPPGHHADHSTAMGFCFFNSVAIACRQLQQQSKASKILIVDWDVHHGNGTQQTFYQDPSVLYISLHRHDDGNFFPGSGAVDEVGAGSGEGFNVNVAWAGGLDPPMGDPEYLAAFRIVVMPIAREFSPDLVLVSAGFDAAEGHPAPLGGYHVSAKCFGYMTQQLMNLAGGAVVLALEGGHDLTAICDASEACVAALLGNRVDPLSEEGWKQKPNLNAIRSLEAVIRVHSKYWGCMQRLASCPDSWVPRVPGADKEEVEAVTALASLSVGILAEDRPSEQLVEEEEPMNL

>AAF73428.1 histone deacetylase 8 [Homo sapiens]

MEEPEEPADSGQSLVPVYIYSPEYVSMCDSLAKIPKRASMVHSLIEAYALHKQMRIVKPKVASMEEMATFHTDAYLQHLQKVSQEGDDDHPDSIEYGLGYDCPATEGIFDYAAAIGGATITAAQCLIDGMCKVAINWSGGWHHAKKDEASGFCYLNDAVLGILRLRRKFERILYVDLDLHHGDGVEDAFSFTSKVMTVSLHKFSPGFFPGTGDVSDVGLGKGWYYSVNVPIQDGIQDEKYYQICESVLKEVYQAFNPKAVVLQLGADTIAGDPMCSFNMTPVGIGKCLKYILQWQLATLILGGGGYNLANTARCWTYLTGVILGKTLSSEIPDHEFFTAYGPDYVLEITPSCRPDRNEPHRIQQILNYIKGNLKHVV

>NP\_478056.1 histone deacetylase 9 isoform 1 [Homo sapiens]

MHSMISSVDVKSEVPVGLEPISPLDLRTDLRMMMPVVDPVVREKQLQQELLLIQQQQQIQKQLLIAEFQKQHENLTRQHQAQLQEHIKELLAIKQQQELLEKEQKLEQQRQEQEVERHRREQQLPPLRGKDRGRERAVASTEVKQKLQEFLLSKSATKDTPTNGKNHSVSRHPKLWYTAAHHTSLDQSSPPLSGTSPSYKYTLPGAQDAKDDFPLRKTASEPNLKVRSRLKQKVAERRSSPLLRRKDGNVVTSFKKRMFEVTESSVSSSSPGSGPSSPNNGPTGSVTENETSVLPPTPHAEQMVSQQRILIHEDSMNLLSLYTSPSLPNITLGLPAVPSQLNASNSLKEKQKCETQTLRQGVPLPGQYGGSIPASSSHPHVTLEGKPPNSSHQALLQHLLLKEQMRQQKLLVAGGVPLHPQSPLATKERISPGIRGTHKLPRHRPLNRTQSAPLPQSTLAQLVIQQQHQQFLEKQKQYQQQIHMNKLLSKSIEQLKQPGSHLEEAEEELQGDQAMQEDRAPSSGNSTRSDSSACVDDTLGQVGAVKVKEEPVDSDEDAQIQEMESGEQAAFMQQPFLEPTHTRALSVRQAPLAAVGMDGLEKHRLVSRTHSSPAASVLPHPAMDRPLQPGSATGIAYDPLMLKHQCVCGNSTTHPEHAGRIQSIWSRLQETGLLNKCERIQGRKASLEEIQLVHSEHHSLLYGTNPLDGQKLDPRILLGDDSQKFFSSLPCGGLGVDSDTIWNELHSSGAARMAVGCVIELASKVASGELKNGFAVVRPPGHHAEESTAMGFCFFNSVAITAKYLRDQLNISKILIVDLDVHHGNGTQQAFYADPSILYISLHRYDEGNFFPGSGAPNEVGTGLGEGYNINIAWTGGLDPPMGDVEYLEAFRTIVKPVAKEFDPDMVLVSAGFDALEGHTPPLGGYKVTAKCFGHLTKQLMTLADGRVVLALEGGHDLTAICDASEACVNALLGNELEPLAEDILHQSPNMNAVISLQKIIEIQSMSLKFS

>NP\_114408.3 polyamine deacetylase HDAC10 isoform 1 [Homo sapiens]

MGTALVYHEDMTATRLLWDDPECEIERPERLTAALDRLRQRGLEQRCLRLSAREASEEELGLVHSPEYVSLVRETQVLGKEELQALSGQFDAIYFHPSTFHCARLAAGAGLQLVDAVLTGAVQNGLALVRPPGHHGQRAAANGFCVFNNVAIAAAHAKQKHGLHRILVVDWDVHHGQGIQYLFEDDPSVLYFSWHRYEHGRFWPFLRESDADAVGRGQGLGFTVNLPWNQVGMGNADYVAAFLHLLLPLAFEFDPELVLVSAGFDSAIGDPEGQMQATPECFAHLTQLLQVLAGGRVCAVLEGGYHLESLAESVCMTVQTLLGDPAPPLSGPMAPCQSALESIQSARAAQAPHWKSLQQQDVTAVPMSPSSHSPEGRPPPLLPGGPVCKAAASAPSSLLDQPCLCPAPSVRTAVALTTPDITLVLPPDVIQQEASALREETEAWARPHESLAREEALTALGKLLYLLDGMLDGQVNSGIAATPASAAAATLDVAVRRGLSHGAQRLLCVALGQLDRPPDLAHDGRSLWLNIRGKEAAALSMFHVSTPLPVMTGGFLSCILGLVLPLAYGFQPDLVLVALGPGHGLQGPHAALLAAMLRGLAGGRVLALLEENSTPQLAGILARVLNGEAPPSLGPSSVASPEDVQALMYLRGQLEPQWKMLQCHPHLVA

>NP\_079103.2 histone deacetylase 11 isoform 1 [Homo sapiens]

MLHTTQLYQHVPETRWPIVYSPRYNITFMGLEKLHPFDAGKWGKVINFLKEEKLLSDSMLVEAREASEEDLLVVHTRRYLNELKWSFAVATITEIPPVIFLPNFLVQRKVLRPLRTQTGGTIMAGKLAVERGWAINVGGGFHHCSSDRGGGFCAYADITLAIKFLFERVEGISRATIIDLDAHQGNGHERDFMDDKRVYIMDVYNRHIYPGDRFAKQAIRRKVELEWGTEDDEYLDKVERNIKKSLQEHLPDVVVYNAGTDILEGDRLGGLSISPAGIVKRDELVFRMVRGRRVPILMVTSGGYQKRTARIIADSILNLFGLGLIGPESPSVSAQNSDTPLLPPAVP

23430 chất ức chế HDAC được thu thập từ Binding Database (BindingDB, http:/www.bindingdb.org).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Protein | Raw Data | Filter hydroxamic | Refine duplicate | Refine cpd has IC50 | Training set | Test set |
| HDAC1 | 6302 | 2487 | 1555 | 1533 | 1073 |  |
| HDAC2 | 2472 | 940 | 305 | 288 | 202 |  |
| HDAC3 | 2596 | 948 | 443 | 427 | 299 |  |
| HDAC4 | 1724 | 851 | 358 | 343 | 240 |  |
| HDAC5 | 847 | 411 | 89 | 76 | 53 |  |
| HDAC6 | 3887 | 1744 | 705 | 685 | 480 |  |
| HDAC7 | 844 | 408 | 88 | 79 | 55 |  |
| HDAC8 | 2248 | 1092 | 552 | 536 | 375 |  |
| HDAC9 | 875 | 464 | 155 | 142 | 100 |  |
| HDAC10 | 811 | 390 | 111 | 98 | 68 |  |
| HDAC11 | 824 | 366 | 69 | 56 | 39 |  |
| Tổng | 23430 | 10101 | 4430 | 4263 | 2984 | 1279 |

2.2. Pharmacophore model

Mô hình Pharmacophore là gì? Hiện nay có bao nhiêu phương pháp tạo mô hình pharmacophore? Một số nghiên cứu về pharmacophore của các isoform HDAC hiện nay thực hiện theo phương pháp nào? Để thực hiện mô hình pharmacophore trong nghiên cứu này sử dụng phương pháp energy-based, là phương pháp như thế nào? Quy trình thực hiện trong nghiên cứu này là sử dụng công cụ nào?

Pharmacophore features được lựa chọn dựa trên phương pháp energy-based. Phương pháp sử dụng probe docking để lựa chọn các features tại ví trí mà probes have high interaction energies.

Alignment tất cả các protein có sẵn cấu trúc 3D => dùng FTmap để thực hiện mô phỏng probe docking

2.3. Proteochemometric model

Giới thiệu về mô hình proteochemometric: lịch sử, các thành phần, ứng dụng của mô hình trong một số nghiên cứu gần đây.

2.3.1. Ligand descriptors

MOE\_2D

MACCS fingerprints

2.3.2. Protein descriptors

Mô tả tương đồng chuỗi protein (sequence similarity): sequence-based

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| HDAC | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| 1 | 100.0 | 93.5 | 69.6 | 18.4 | 15.6 | 16.0 | 17.2 | 47.6 | 14.8 | 26.3 | 29.1 |
| 2 | 93.5 | 100.0 | 68.4 | 17.9 | 15.5 | 15.0 | 17.0 | 47.3 | 14.9 | 24.3 | 28.2 |
| 3 | 69.6 | 68.4 | 100.0 | 17.6 | 14.8 | 13.5 | 16.7 | 51.7 | 15.0 | 26.9 | 29.0 |
| 4 | 18.4 | 17.9 | 17.6 | 100.0 | 63.6 | 30.9 | 56.7 | 17.7 | 62.0 | 17.5 | 15.7 |
| 5 | 15.6 | 15.5 | 14.8 | 63.6 | 100.0 | 25.1 | 51.3 | 15.0 | 66.6 | 17.2 | 12.9 |
| 6 | 16.0 | 15.0 | 13.5 | 30.9 | 25.1 | 100.0 | 28.2 | 13.6 | 23.1 | 31.4 | 12.2 |
| 7 | 17.2 | 17.0 | 16.7 | 56.7 | 51.3 | 28.2 | 100.0 | 16.9 | 49.3 | 18.7 | 16.8 |
| 8 | 47.6 | 47.3 | 51.7 | 17.7 | 15.0 | 13.6 | 16.9 | 100.0 | 18.1 | 22.6 | 33.6 |
| 9 | 14.8 | 14.9 | 15.0 | 62.0 | 66.6 | 23.1 | 49.3 | 18.1 | 100.0 | 15.4 | 15.7 |
| 10 | 26.3 | 24.3 | 26.9 | 17.5 | 17.2 | 31.4 | 18.7 | 22.6 | 15.4 | 100.0 | 21.1 |
| 11 | 29.1 | 28.2 | 29.0 | 15.7 | 12.9 | 12.2 | 16.8 | 33.6 | 15.7 | 21.1 | 100.0 |

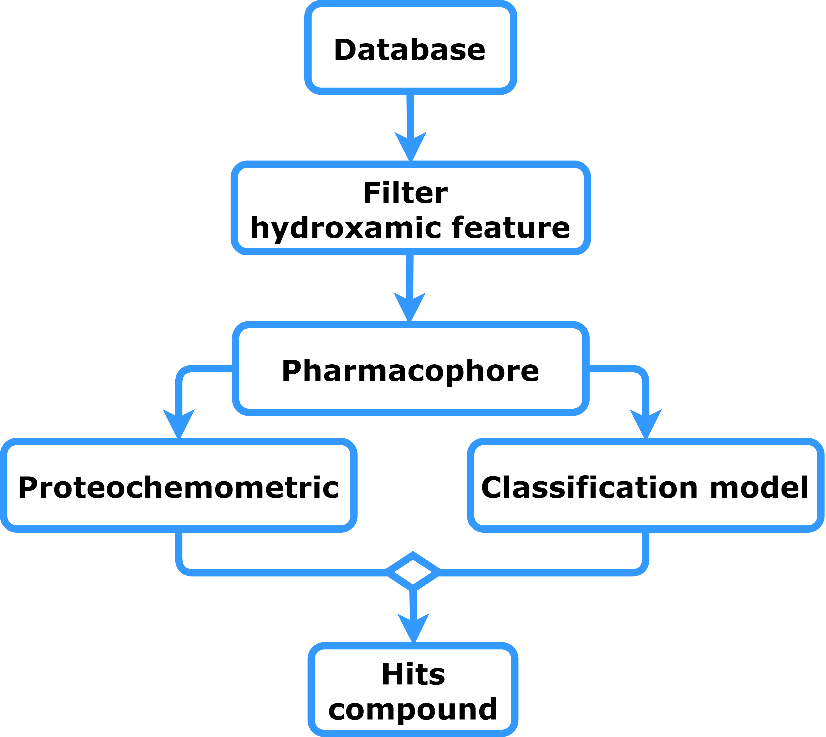
Mô tả theo Z-scales: physiochemical

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | z-1 | z-2 | z-3 | z-4 | z-5 |
| HDAC1 | 0.4374066 | -0.2691701 | -0.1757676 | -0.4429046 | 0.1501660 |
| HDAC2 | 0.4805738 | -0.3130943 | -0.1608607 | -0.4022541 | 0.1475205 |
| HDAC3 | 0.08691589 | -0.28817757 | -0.02742991 | -0.42373832 | 0.15936916 |
| HDAC4 | 0.3126337 | -0.5322222 | -0.2510391 | -0.3972016 | 0.3802263 |
| HDAC5 | 0.3270053 | -0.6286720 | -0.2883244 | -0.4949554 | 0.3635740 |
| HDAC6 | 0.1872675 | -0.7015720 | -0.1390206 | -0.4406584 | 0.2620988 |
| HDAC7 | 0.2401468 | -0.5711740 | -0.1348637 | -0.3090776 | 0.4185430 |
| HDAC8 | -0.01098143 | -0.55771883 | -0.10777188 | -0.39628647 | 0.17408488 |
| HDAC9 | 0.3299604 | -0.5120772 | -0.3010781 | -0.4433531 | 0.3878042 |
| HDAC10 | -0.08893871 | -0.75215247 | -0.14388640 | -0.35532138 | 0.38403587 |
| HDAC11 | 0.01195965 | -0.45625360 | -0.41435159 | -0.34256484 | 0.23492795 |

2.3.3. Machine learning algorithms

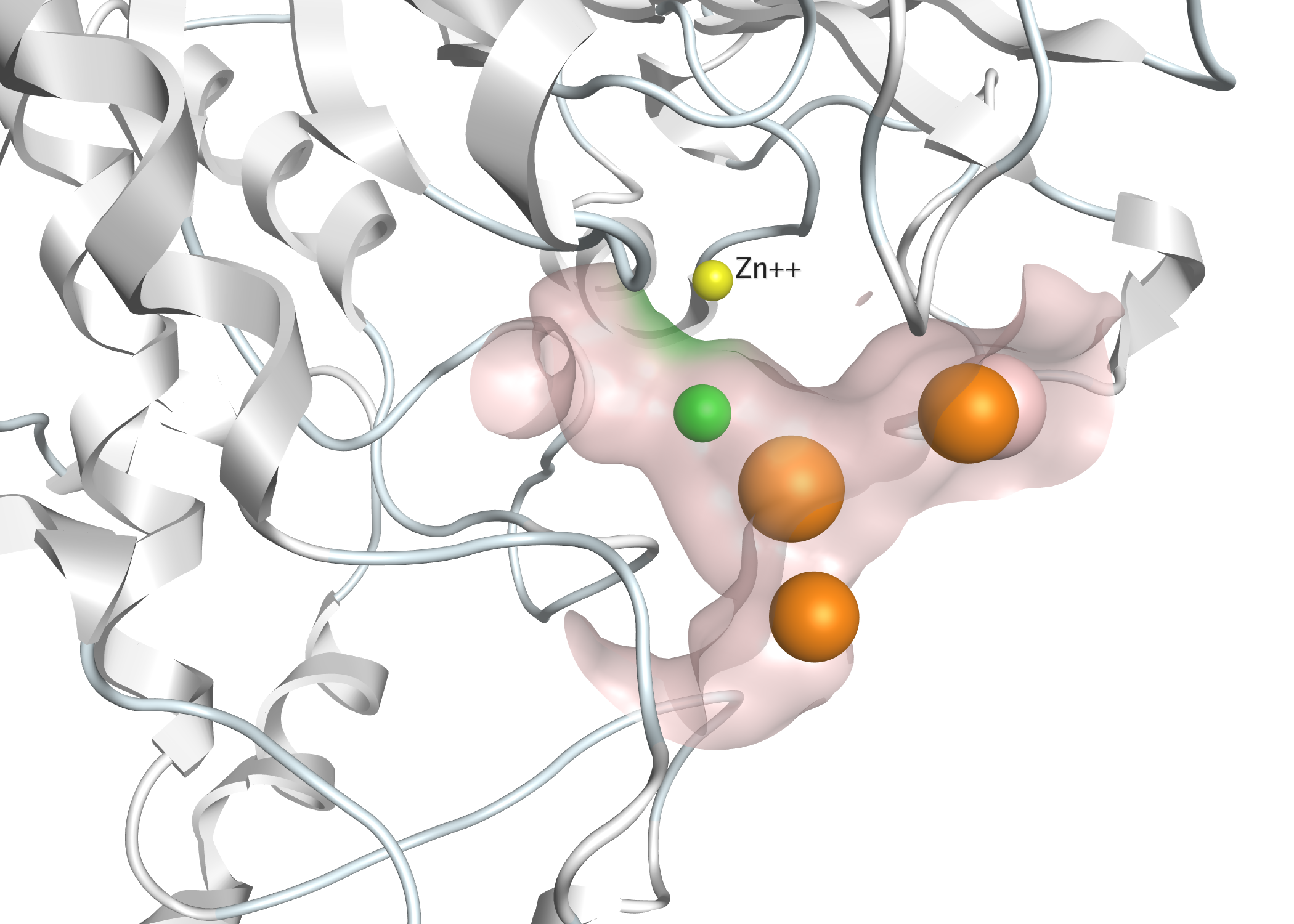
So sánh giữa thuật toán hồi quy tuyến tính và phi tuyến tính: Linear Regression và RandomForest

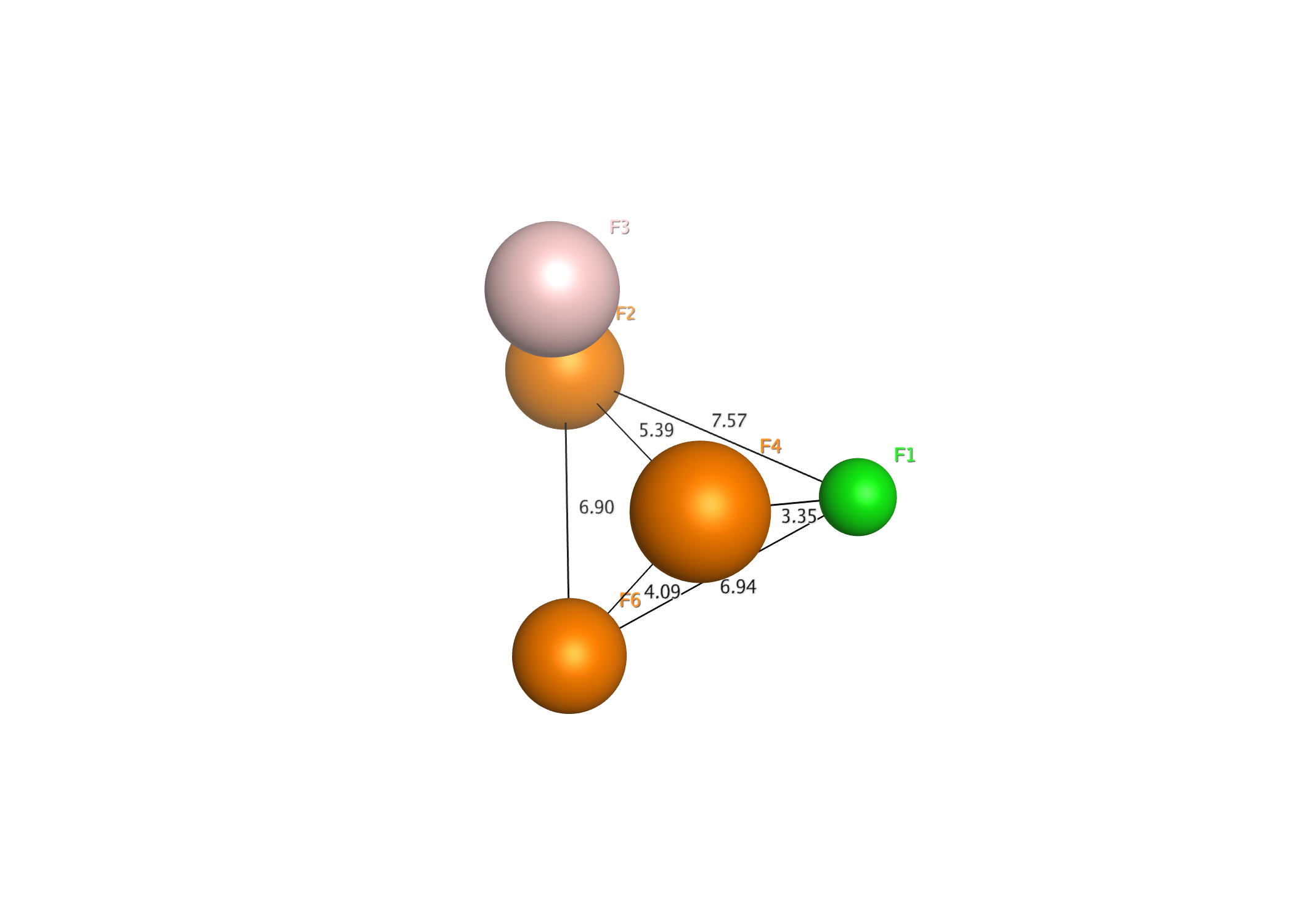
2.5. Virtual screening



3. Result

3.1. Pharmacophore features





Kết quả của mô hình ph4 được thể hiện qua các thành phần F1…

Validate pharmacophore model: with 371 active HDAC6’s inhibitors (IC50 < 100 nM)

Hits list: 368 (99.2%) molecules => chứng tỏ khả năng nhận diện được active của mô hình là tốt => có thể sử dụng để thực hiện sàng lọc ảo bước đầu.

3.2. Proteochemometric

|  |  |  |  |
| --- | --- | --- | --- |
| Model | Ligand descriptors | Protein descriptors | Cross-term descriptors |
| M1 | MOE\_2D | z-scales |  |
| M2 | MOE\_2D | Sequence similarity |  |
| M3 | MACCS key fingerprints | Z-scales |  |
| M4 | MACCS key fingerprints | Sequence similarity |  |
| M5 | MOE\_2D | z-scales + Sequence similarity |  |
| M6 | MOE\_2D + MACCS key fingerprints | z-scales + Sequence similarity |  |
| M1\_CP | MOE\_2D | z-scales | Protein-protein |
| M2\_CP | MOE\_2D | Sequence similarity | Protein-protein |
| M3\_CP | MACCS key fingerprints | Z-scales | Protein-protein |
| M4\_CP | MACCS key fingerprints | Sequence similarity | Protein-protein |
| M5\_CP | MOE\_2D | z-scales + Sequence similarity | Protein(z)-Protein(s)  Protein(z)-Protein(z)  Protein(s)-Protein(s) |
| M6\_CP | MOE\_2D + MACCS key fingerprints | z-scales + Sequence similarity | Protein(z)-Protein(s)  Protein(z)-Protein(z)  Protein(s)-Protein(s) |
| M1\_CPL | MOE\_2D | z-scales | Protein-Ligand |
| M2\_CPL | MOE\_2D | Sequence similarity | Protein-Ligand |
| M3\_CPL | MACCS key fingerprints | Z-scales | Protein-Ligand |
| M4\_CPL | MACCS key fingerprints | Sequence similarity | Protein-Ligand |
| M5\_CPL | MOE\_2D | z-scales + Sequence similarity | Protein(z)-Ligand  Protein(s)-Ligand |
| M6\_CPL | MOE\_2D + MACCS key fingerprints | z-scales + Sequence similarity | Protein(z)-Ligand(moe)  Protein(s)-Ligand(moe)  Protein(z)-Ligand(maccs)  Protein(s)-Ligand(maccs) |
| M1\_CPLonly |  |  | Protein(z)-Ligand(moe) |
| M2\_CPLonly |  |  | Protein(s)-Ligand(moe) |
| M3\_CPLonly |  |  | Protein(z)-Ligand(maccs) |
| M4\_CPLonly |  |  | Protein(s)-Ligand(maccs) |
| M5\_CPLonly |  |  | Protein(z)-Ligand(moe)  Protein(s)-Ligand(moe) |
| M6\_CPLonly |  |  | Protein(z)-Ligand(moe)  Protein(s)-Ligand(moe)  Protein(z)-Ligand(maccs)  Protein(s)-Ligand(maccs) |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

RandomForest:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | M1 | M2 | M3 | M4 | M5 | M6 |  |
| R2 train | 0.9777 | 0.9821 | 0.9651 | 0.9653 | 0.982 | 0.9821 |  |
| Q2 test | 0.7753 | 0.8287 | 0.751 | 0.7567 | 0.8293 | 0.8219 |  |
| MAE test | 0.5117 | 0.4618 | 0.5447 | 0.541 | 0.4632 | 0.4715 |  |
| RMSE test | 0.7021 | 0.6334 | 0.7325 | 0.7269 | 0.6325 | 0.6436 |  |
| Q2 CV | 0.7648 | 0.8058 | 0.7215 | 0.7275 | 0.8111 | 0.8051 |  |
| MAE cv | 0.5252 | 0.4881 | 0.5687 | 0.5647 | 0.4845 | 0.489 |  |
| RMSE cv | 0.7074 | 0.6565 | 0.7591 | 0.7531 | 0.6515 | 0.6578 |  |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | M1\_CP | M2\_CP | M3\_CP | M4\_CP | M5\_CP | M6\_CP |  |
| R2 train | 0.9812 | 0.9834 | 0.9645 | 0.9637 | 0.984 | 0.9837 |  |
| Q2 test | 0.8163 | 0.8411 | 0.7425 | 0.7305 | 0.8394 | 0.8424 |  |
| MAE test | 0.4768 | 0.4497 | 0.5541 | 0.5633 | 0.4531 | 0.4498 |  |
| RMSE test | 0.6498 | 0.6123 | 0.7437 | 0.7572 | 0.6178 | 0.6142 |  |
| Q2 CV | 0.8003 | 0.8276 | 0.7183 | 0.7087 | 0.8247 | 0.822 |  |
| MAE cv | 0.4917 | 0.4664 | 0.5736 | 0.5788 | 0.4726 | 0.4739 |  |
| RMSE cv | 0.6656 | 0.627 | 0.7633 | 0.7734 | 0.6356 | 0.6386 |  |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | M1\_CPL | M2\_CPL | M3\_CPL | M4\_CPL | M5\_CPL | M6\_CPL |  |
| R2 train | 0.9825 | 0.9825 | 0.9601 | 0.96 | 0.9824 | 0.9817 |  |
| Q2 test | 0.7964 | 0.7954 | 0.6742 | 0.6641 | 0.7955 | 0.7944 |  |
| MAE test | 0.5065 | 0.5044 | 0.6043 | 0.6098 | 0.5065 | 0.5071 |  |
| RMSE test | 0.688 | 0.6863 | 0.8173 | 0.8275 | 0.6874 | 0.6901 |  |
| Q2 CV | 0.7879 | 0.7933 | 0.6374 | 0.6399 | 0.7898 | 0.7819 |  |
| MAE cv | 0.5142 | 0.5106 | 0.6265 | 0.6263 | 0.5121 | 0.5181 |  |
| RMSE cv | 0.6924 | 0.6859 | 0.8427 | 0.8405 | 0.6918 | 0.7007 |  |

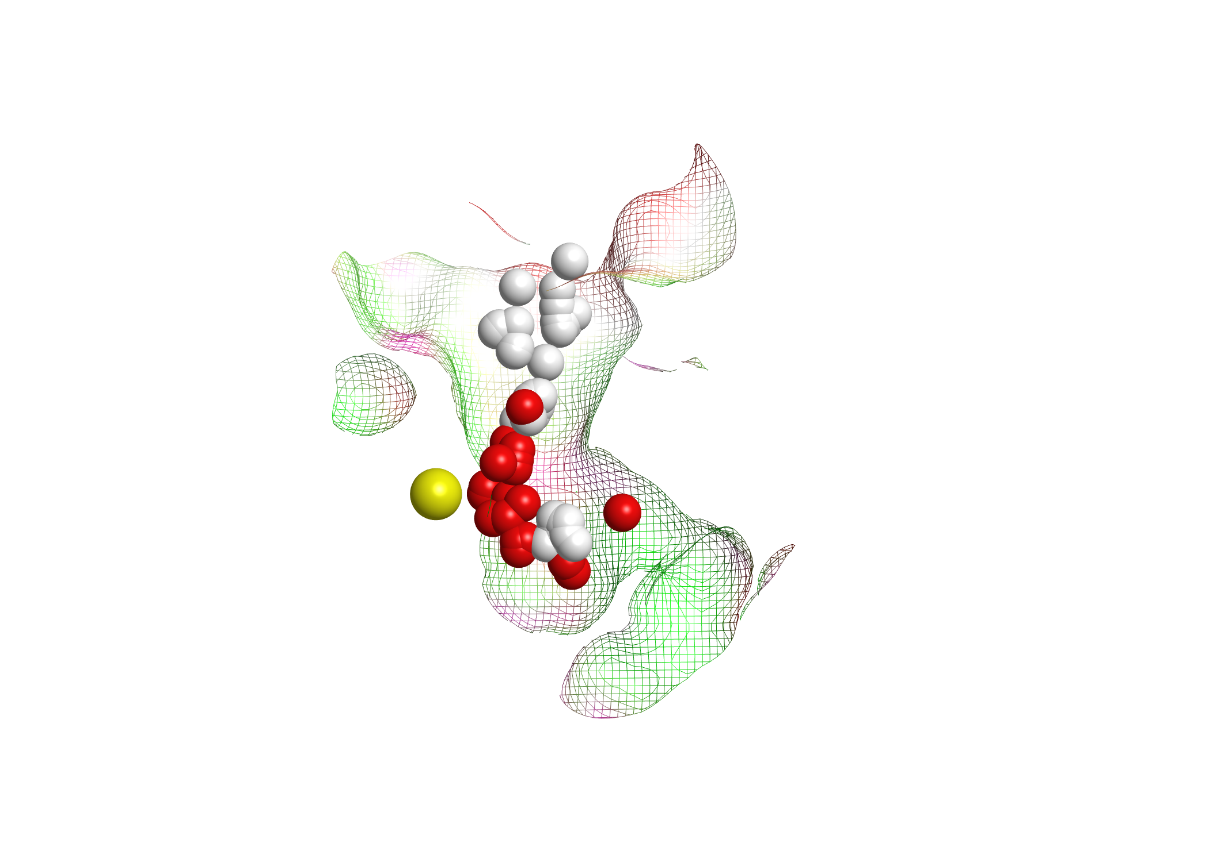
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | M1\_CPLonly | M2\_CPLonly | M3\_CPLonly | M4\_CPLonly | M5\_CPLonly | M6\_CPLonly |
| R2 train | 0.982 |  |  |  |  |  |
| Q2 test | 0.7874 |  |  |  |  |  |
| MAE test | 0.517 |  |  |  |  |  |
| RMSE test | 0.6998 |  |  |  |  |  |
| Q2 CV | 0.7773 |  |  |  |  |  |
| MAE cv | 0.5247 |  |  |  |  |  |
| RMSE cv | 0.7073 |  |  |  |  |  |

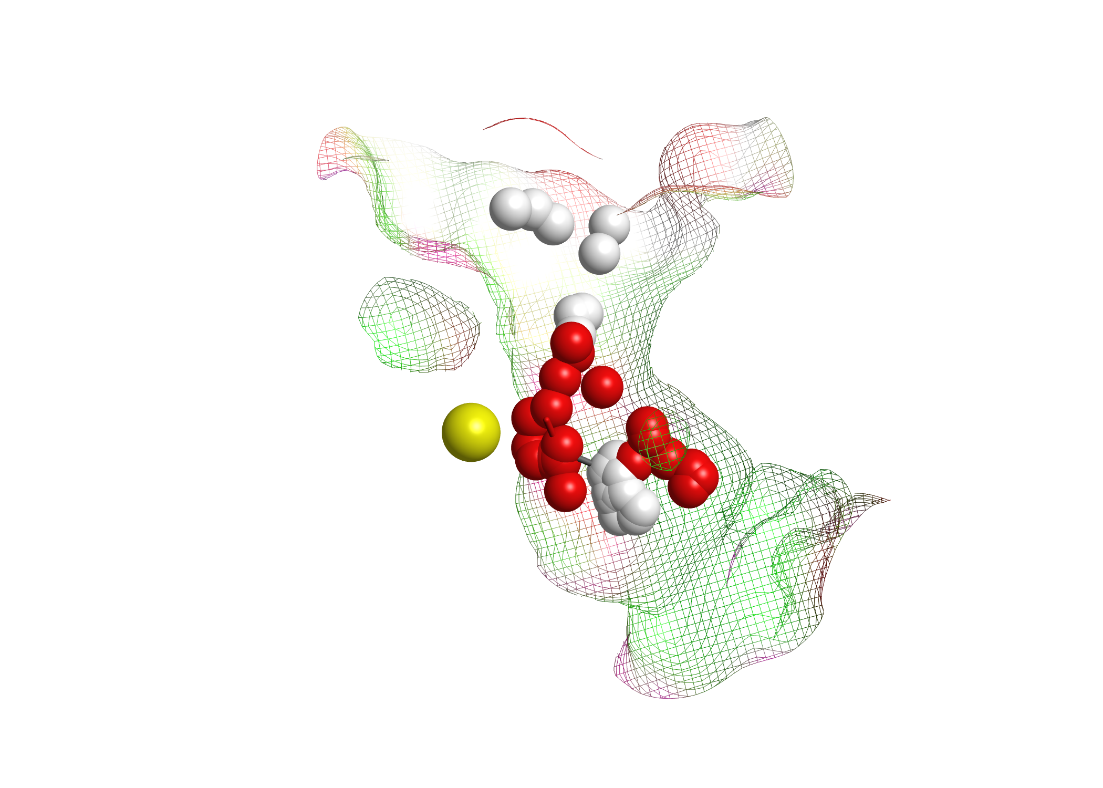
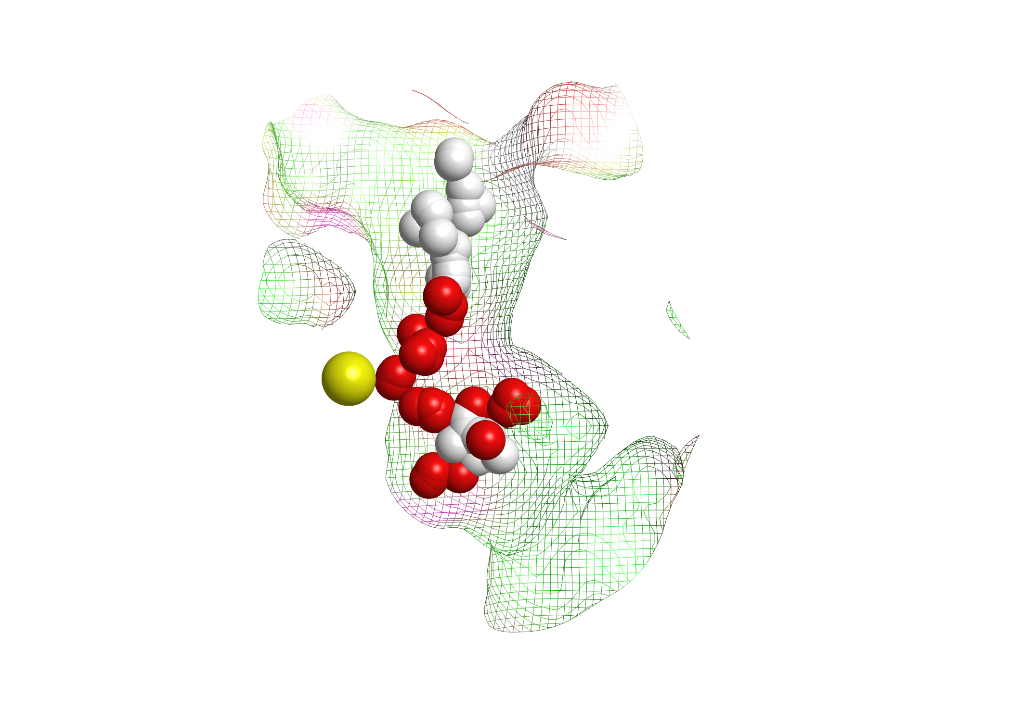
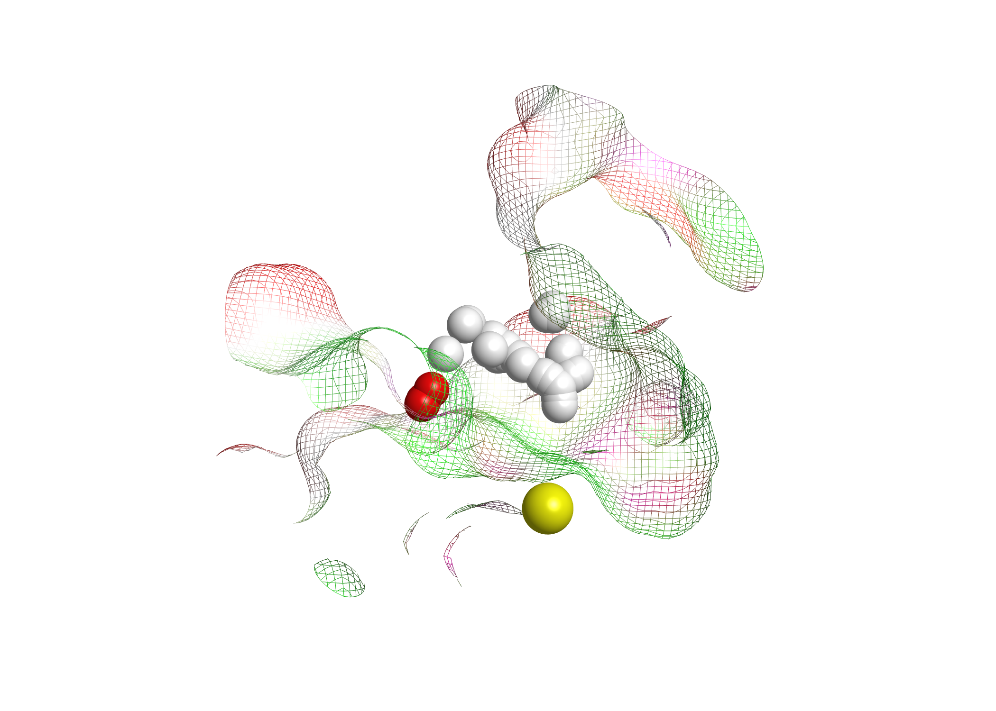
3.3. Classification model

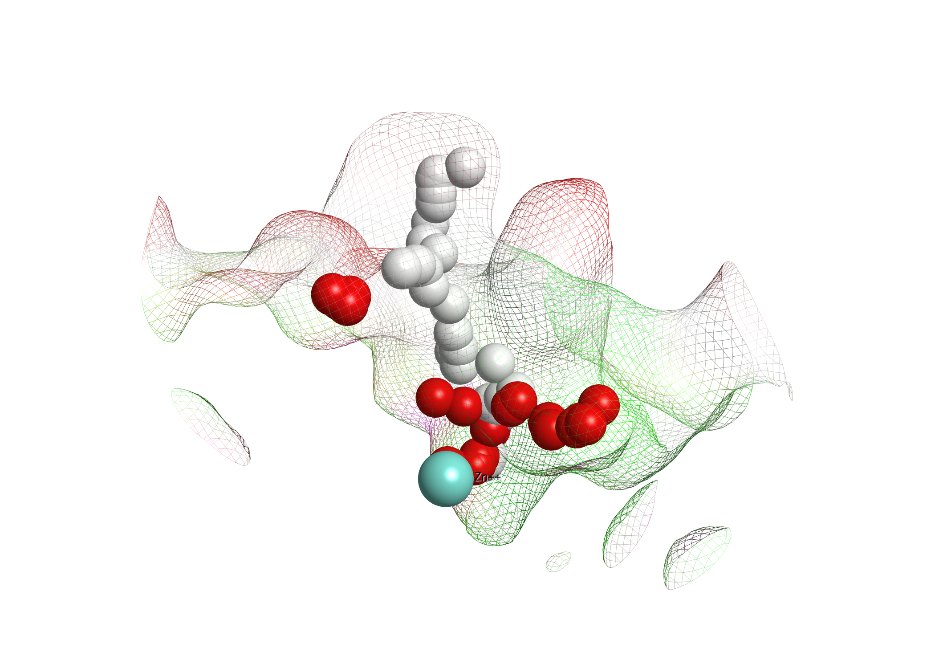
|  |  |
| --- | --- |
| C1 | imbalanced classification |
| C2 | balance duplicate Active |
| C3 | balance SMOTE |
|  |  |

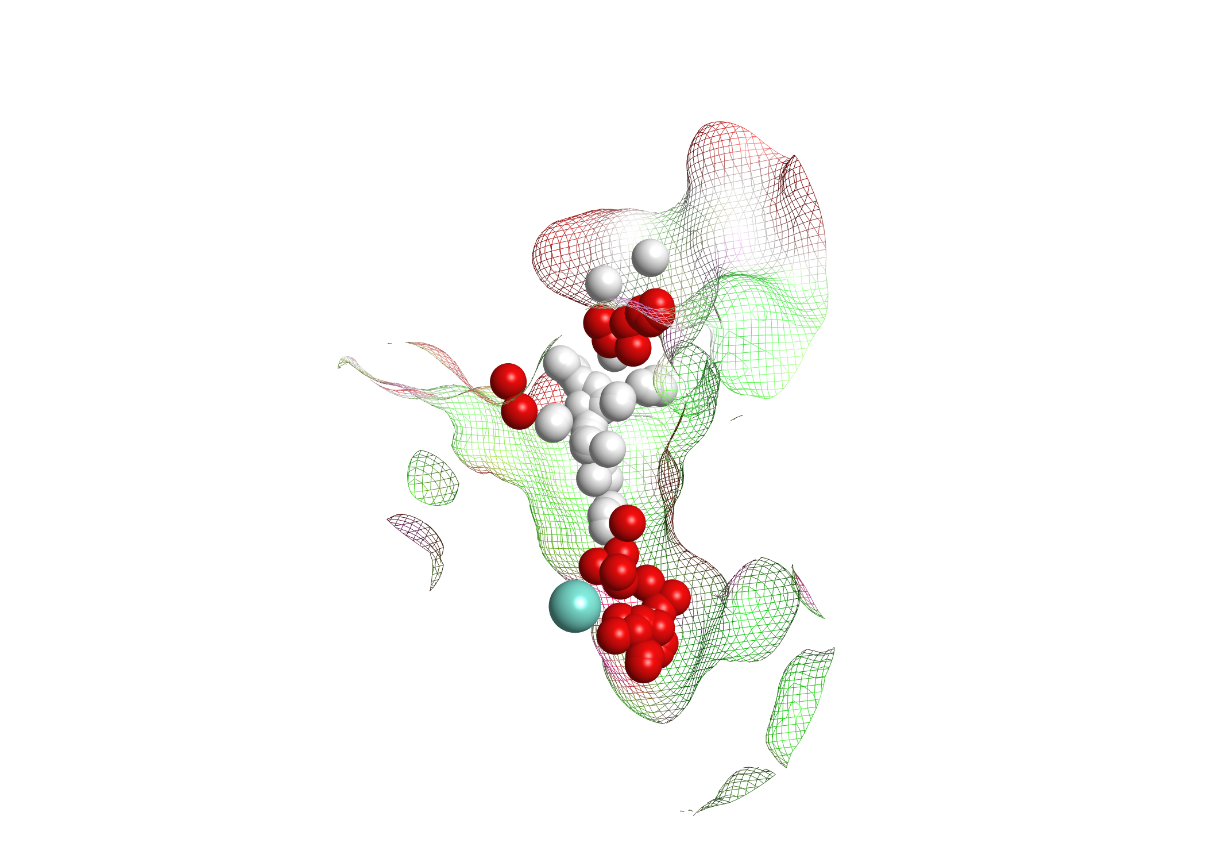
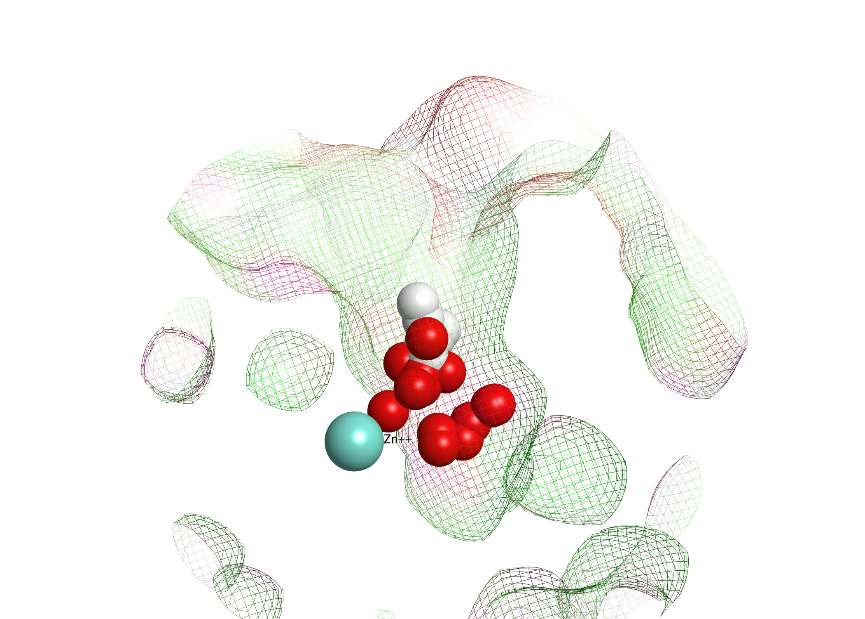
3.4. Virtual screening

3.5. Phylogenetic Trees

3.6. Binding site analysis







|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Volume | Area | Hydrophobicity | Buriedness | Aromatic | DLID | Radius | Nonsphericity |
| 1 | 252.1 | 244.9 | 0.6678 | 0.8558 | 0.2813 | 0.2706 | 3.919 | 1.269 |
| 2 | 301.3 | 270 | 0.7079 | 0.8923 | 0.3369 | 0.6369 | 4.159 | 1.242 |
| 3 | 297 | 267.1 | 0.6603 | 0.8713 | 0.3175 | 0.4357 | 4.139 | 1.241 |
| 4 | 476.4 | 478.7 | 0.5488 | 0.7175 | 0.1259 | -0.07488 | 4.845 | 1.623 |
| 6 | 270.1 | 256.2 | 0.6663 | 0.8304 | 0.3438 | 0.2177 | 4.01 | 1.268 |
| 7 | 362.9 | 288.3 | 0.7659 | 0.8849 | 0.247 | 0.8768 | 4.425 | 1.172 |
| 8 | 229 | 210.8 | 0.7437 | 0.9239 | 0.3761 | 0.6403 | 3.795 | 1.164 |
| 10 | 290.2 | 250.2 | 0.7125 | 0.8718 | 0.321 | 0.5388 | 4.107 | 1.18 |

4. Conclusion

1. Allfrey VG, Faulkner R, Mirsky AE. Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. Proc. Natl Acad.Sci. USA 51, 786–794 (1964). [↑](#endnote-ref-1)
2. Biel M, Wascholowski V, Giannis A. Epigenetics an epicenter of gene regulation: Histones and histone-modifying enzymes. Angew Chem Int Ed Engl. 2005; 44:3186-3216. [↑](#endnote-ref-2)
3. Schafer S, Jung M. Chromatin modifications as targets for new anticancer drugs. Arch Pharm (Weinheim). 2005; 338:347-357. [↑](#endnote-ref-3)
4. Yang, X.J. & Seto, E. The Rpd3/Hda1 family of lysine deacetylases: from bacteria and yeast to mice and men. Nat. Rev. Mol. Cell Biol.9, 206–218 [↑](#endnote-ref-4)
5. Grunstein M. Histone acetylation in chromatin structure and transcription. Nature. 1997; 389:349-352. [↑](#endnote-ref-5)
6. Berger, S. L., Kouzarides, T., Shiekhattar, R., and Shilatifard, A.(2009) An operational definition of epigenetics.Genes Dev. 23, 781−783. [↑](#endnote-ref-6)
7. Delcuve, G. P., Khan, D. H., and Davie, J. R. (2012) Roles ofhistone deacetylases in epigenetic regulation: emerging paradigmsfrom studies with inhibitors.Clin. Epigenet. 4,5. [↑](#endnote-ref-7)
8. Choudhary, C., Weinert, B. T., Nishida, Y., Verdin, E., andMann, M. (2014) The growing landscape of lysine acetylation linksmetabolism and cell signalling.Nat. Rev. Mol. Cell Biol. 15, 536−550. [↑](#endnote-ref-8)
9. Wang, Q., Zhang, Y., Yang, C., Xiong, H., Lin, Y., Yao, J., Li, H.,Xie, L., Zhao, W., Yao, Y., Ning, Z.-B., Zeng, R., Xiong, Y., Guan, K.-L., Zhao, S., and Zhao, G.-P. (2010) Acetylation of metabolicenzymes coordinated carbon source utilization and metabolic flux.Science 327, 1004−1007. [↑](#endnote-ref-9)
10. Zhao, S., Xu, W., Jiang, W., Yu, W., Lin, Y., Zhang, T., Yao, J.,Zhou, L., Zeng, Y., Li, H., Li, Y., Shi, J., An, W., Hancock, S. M., He,F., Qin, L., Chin, J., Yang, P., Chen, X., Lei, Q., Xiong, Y., and Guan,K. L. (2010) regulation of cellular metabolism by protein lysineacetylation.Science 327, 1000−1004. [↑](#endnote-ref-10)
11. Marcus AI, Zhou J, O’Brate A, Hamel E, Wong J, Nivens M, et al. The synergistic combination of the farnesyl transferase inhibitor lonafarnib and paclitaxel enhances tubulin acetylation and requires a functional tubulin deacetylase. Cancer Res. 2005;65(9):3883–3893. doi: 10.1158/0008-5472.CAN-04-3757. [↑](#endnote-ref-11)
12. Saji S, Kawakami M, Hayashi S-i, Yoshida N, Hirose M, Horiguchi S-i, et al. Significance of HDAC6 regulation via estrogen signaling for cell motility and prognosis in estrogen receptor-positive breast cancer. Oncogene. 2005;24(28):4531–4539. doi: 10.1038/sj.onc.1208646. [↑](#endnote-ref-12)
13. Carew JS, Giles FJ, Nawrocki ST. Histone deacetylase inhibitors: mechanisms of cell death and promise in combination cancer therapy. Cancer Lett. 2008;269(1):7–17. doi: 10.1016/j.canlet.2008.03.037.  [↑](#endnote-ref-13)
14. Michael S, Cristina F, Mario D, Marc D. Epigenetic alterations as a universal feature of cancer hallmarks and a promising target for personalized treatments. Curr Top Med Chem. 2016;16(7):745–776. [↑](#endnote-ref-14)
15. Weidle UH, Grossmann A. Inhibition of histone deacetylases: A new strategy to target epigenetic modifications for anticancer treatment. Anticancer Res. 2000; 20:1471-1485. [↑](#endnote-ref-15)
16. Witt O, Deubzer HE, Milde T, Oehme I. HDAC family: what are the cancer relevant targets? Cancer Lett. 2009;277(1):8–21. doi: 10.1016/j.canlet.2008.08.016. [↑](#endnote-ref-16)