

[Artificia](https://doi.org/10.1016/j.ailsci.2023.100072)l[IntelligenceintheLifeSciences3(2023)100072](https://doi.org/10.1016/j.ailsci.2023.100072)

|  |  |  |
| --- | --- | --- |
|  | Contentslistsavailableat[ScienceDirect](http://www.ScienceDirect.com) |  |
| ArtificialIntelligenceintheLifeSciences |
| journalhomepage:[www.elsevier.com/locate/ailsci](http://www.elsevier.com/locate/ailsci) |

Bayesianoptimizationforternarycomplexprediction(BOTCP)

ArjunRao1,∗,TinM.Tunjic1,MichaelBrunsteiner1,MichaelMüller,HoseinFooladi, ChiaraGasbarri,NoahWeber2

*CelerisTherapeuticsGmbH,Schmiedlstraße3,Graz8042,Austria*

|  |  |  |
| --- | --- | --- |
| article | info | abstract |
| *Keywords:*  Bayesianoptimization  Activelearning  Machinelearning  Ternarycomplexprediction Targetedproteindegradation | | Proximity-inducingcompounds(PICs)areanemergentdrugtechnologythroughwhichaproteinofinterest(POI), oftenadrugtarget,isbroughtintothevicinityofasecondproteinwhichmodifiesthePOI’sfunction,abundance orlocalisation,givingrisetoatherapeuticeffect.Oneofthebest-knownexamplesforsuchcompoundsarehet-erobifunctionalmoleculesknownasproteolysistargetingchimeras(PROTACs).PROTACsreducetheabundance ofthetargetproteinbyestablishingproximitytoanE3ligasewhichlabelstheproteinfordegradationviathe ubiquitin-proteasomalpathway.DesignofPROTACsinsilicorequiresthecomputationalpredictionoftheternary complexconsistingofPOI,PROTACmolecule,andtheE3ligase. |

Wepresentanovelmachinelearning-basedmethodforpredictingPROTAC-mediatedternarycomplexstructures usingBayesianoptimization.Weshowhowafitnessscorecombininganestimationofprotein-proteininter-actionswithPROTACconformationenergycalculationsenablesthesample-efficientexplorationofcandidate structures.Furthermore,ourmethodpresentstwonovelscoresforfilteringandrerankingwhichtakePROTAC stability(Autodock-VinabasedPROTACstabilityscore)andproteininteractionrestraints(theTCP-AIRscore) intoaccount.WeevaluateourmethodusingDockQscoresonanumberofavailableternarycomplexstructures (includingpreviouslyunevaluatedcases)anddemonstratethatevenwithaclusteringthatrequiresmembersto haveahighsimilarity,i.e.,withsmallerclusters,wecanassignhighrankstothoseclustersthatcontainposesclose totheexperimentallydeterminednativestructureoftheternarycomplexes.Wealsodemonstratetheresultant improvedyieldofnear-nativeposes3intheseclusters.

**1.Introduction**

Targetedproteindegradationisanemergingtherapeuticmodality which,insteadofinhibitingtheactivityofadrugtarget,actsbyinduc-ingdegradationofthetargetproteinitself[1,2].Itemploysso-called monofunctionaldegraderssuchasmoleculargluesorheterobifunctional degraderssuchasProteolysisTargetingChimeras(PROTACs)whichact asproximity-inducingcompounds.PICsbringtogethertheproteinof interest(POI)andanE3ubiquitinligasetoinducesubsequentubiquiti-nationanddegradationofthetargetprotein(SeeFig.1)[3–6].

TargetedproteindegradationusingPROTACsoffersseveraladvan-tagesovertraditionaloccupancy-basedinhibitorssuchasCiulliand Trainor[7]:

•Expandingthedruggablespace:thereisnoneedfortightbindingto thefunctionalsiteofthePOI,andasaresult,PROTACscantarget

|  |
| --- |
| ∗Corresponding[author.](mailto:a.rao@celeristx.com)  *E-mailaddress:*[a.rao@celeristx.com](mailto:a.rao@celeristx.com)(A.Rao).  1Equalcontribu[tion.](mailto:a.rao@celeristx.com)  2ProjectSupervisor. 3Thenear-nativeposeisdefinedasaposethathasDockQscore≥0*.*23. |

proteinsmoreshallowbindingpockets(e.g.,transcriptionfactors andscaffoldingproteins).

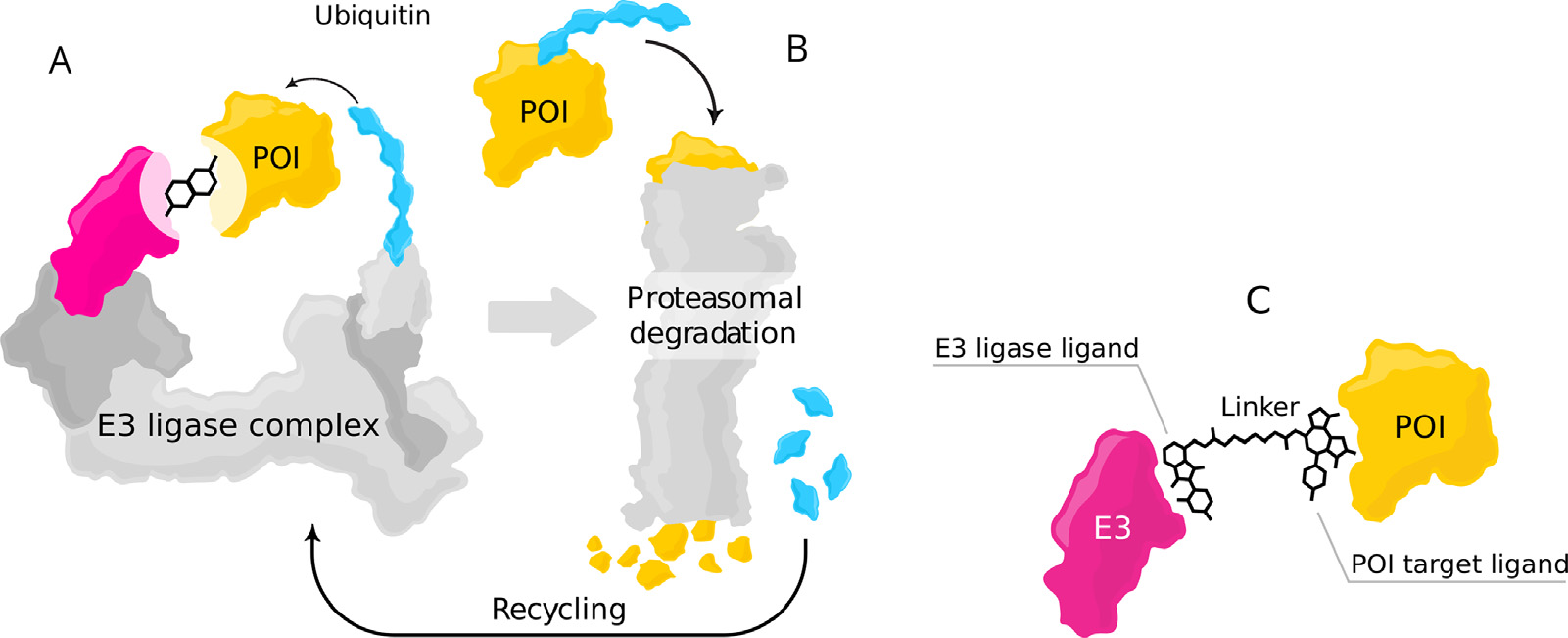
•Workingatasubstoichiometricconcentration:AsinglePROTAC moleculecaninducedegradationofmultiplePOImoleculesover time.

•Improvedefficacy:Recoveryoftargetproteinabundance(andthere-foreoverallactivity)islinkedtoproteinsynthesisincellsandnot justtotheabsenceofthePROTAC.

•Increasednumberofbindingsites:PROTACsaretypicallylargerthan traditionalinhibitors.Thismeansthat,albeitwithaslightriskof side-effectsduetolessselectivebinding,bindingsitesthataretra-ditionallyinaccessabletoothersmallmoleculescanbeutilized.The promiscousbindingofPROTACfragmentsistypicallycompensated forbythefactthatPROTACefficiencyisveryspecifictothelinker designed.

<https://doi.org/10.1016/j.ailsci.2023.100072>  
[Received21November2022;Receivedinrevi](https://doi.org/10.1016/j.ailsci.2023.100072)sedform8March2023;Accepted5April2023   
Availableonline19April2023   
2667-3185/© 2023PublishedbyElsevierB.V.ThisisanopenaccessarticleundertheCCBY-NC-NDlicense(<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

*A.Rao,T.M.Tunjic,M.Brunsteineretal.*  *ArtificialIntelligenceintheLifeSciences3(2023)100072*



**Fig.1.**Modeofactionfortargetedproteindegradation:(A)APOIisshowninaproximitywithanE3ligasecomplex– thereceptorpartofE3ligaseishighlighted inmagenta– aheterobifunctionaldegradersuchasPROTACbringstogetherthePOIandE3ligase,whichinitiatethedegradationprocess.(B)Afterformation ofaternarycomplex,thePOIismarkedwithapolyubiquitinchain(blue),whichsubsequentlyleadstoproteindegradationthroughUPS.Afterdegradation,the ubiquitinmoleculesaredetachedandrecycledbytheUPS.Moreover,thePROTACmoleculeisdetachedandcaninducedegradationofmultiplePOIs(C)Ternary complexformedfromtheinteractionofthePOI,receptorpartoftheE3ligaseandheterobifunctionaldegrader.(Forinterpretationofthereferencestocolorinthis figurelegend,thereaderisreferredtothewebversionofthisarticle.)

Structurally,aPROTACconsistsofawarheadthatbindstothePOI, anE3ligandthatbindsintotheE3ligase,andalinkerthatconnects thesetwoparts.CurrentPROTACsutilizealimitednumberofE3ligases. ThemostprominentE3ligasesarecereblon(CRBN)andVon-Hippel-LindauTumorSuppressor(VHL)duetotheavailabilityofhigh-affinity bindersandfavourablepropertiesforbothproteins.ForCRBNbinders, immunomodulatorydrugssuchasthalidomide[8],lenalidomide[9], andpomalidomide[10]arewidelyused.ForVHL,binderswereini-tiallyderivedfromanaturaldegronpeptideandevolvedtohighlypo-tentsmallmolecules[11,12].Inaddition,someotherE3ligasessuchas cellinhibitorsofapoptosisprotein(cIAP)[13],murinedoubleminute 2(MDM2)[14],andKEAP1[15]havebeenharnessedfortargetedpro-teindegradation.Thereisanongoingeffortonexpandingthelibraryof E3ligasesandfindingsuitablepotentligandsagainstotherE3ligases. WhilethesetofE3ligasesusedfortargetedproteindegradationwith PROTACshassofarbeenlimited,theyhavebeenusedtotargetadiverse rangeofPOIsfromdifferentproteinfamilies.Thereareprominentex-ampleswithgoodevidenceofsafetyandefficacyininitialclinicaltrials, includingdegraderstargetingtheandrogenreceptor(AR)[16,17],es-trogenreceptor(ER)[18],IRAK4,BCL-XL[19],Helios(IKZF2)[20],and GSPT1.Furthermore,invivostudiesforPROTACtargetingBRD4[21], BTK[22,23],RIPK2[24],andSMARCA2[25]haveshownpromising results.ThesefindingshighlightthepotentialofPROTACforthedegra-dationofdiverseproteinfamilies.

JudgingtheefficacyofacandidatePROTACmoleculein-silicoisa dauntingtaskowingtoaplethoraofinteractionsthataredifficultto evaluate.ForaPROTACtobeeffictive,itmustat-leastsatisfythefol-lowingrequirements.Firstly,itmustbeabletopenetratethecellmem-braneandstaystableinsidethecell.Secondly,itsinteractionswiththe E3ligaseandthePOImustresultintheformationofastableternary complexwherethePOIisheldneartheE3ligase.Finally,thiscomplex mustbeonethatlendsitselftobeingtaggedbyubiquitinsothatthePOI maybedegraded.

Inlightoftheabovesteps,itisclearthataccurateandefficientpre-dictionofthestructureoftheternarycomplexisimperativeforimprove-mentsatmanystagesofthePROTACdevelopmentprocess.Forinstance, computationalmodelingofternarycomplexesallowsthescreeningof multipledifferentE3ligasesforacandidatetarget,maximizingthesta-bilityoftheternarycomplex[26].Knowingthestructureoftheternary complexisthefirststeptocalculatingitsstability,andexperimentalev-idenceshowsthattheformationofastableternarycomplexcanleadto moreefficientdegradation[26].

2

*A.Rao,T.M.Tunjic,M.Brunsteineretal.*

complexesforafewavailablePDBs.Infollow-upwork[31],Drummond etal.addedtwoadditionalmethodswhichimprovedthepreviousper-formance.Thisworkintroducesclusteringofthegeneratedstructures, anessentialstepinobtaininggoodresults.Zaidmanetal.[32]pro-posedanothermethod,PRosettaC,whichaddsconsecutivestepssuch asglobalprotein-proteindockingandlocaldockingrefinement,gener-atingconformationofPROTACsthatcanfitinthedockingsolutions andfinalclustering.Althoughtheyhavebeenabletoreproduceagood amountofcrystalstructuresofternarycomplexes,theyareinitiating theirpredictionfromboundstructures,whichdoesnotreflectreal-world conditions.Boundstructurescomefromthedecompositionofalready availableternarycomplexes,butaprioriwedonothaveaccesstothis information.Inoneoftherecentworks,Wengetal.[33]proposeda methodagainbasedonglobalprotein-proteindocking,localrefinement withRosettaDock[34],generatingPROTACsusingRDkit,andapplying filteringandre-rankingcriteriaforpredictingnear-nativeternarycom-plexes.Theystartedfromunboundstructuresandshowedthatgoodre-sultscouldbeachievedwiththesemethods.Furthermore,Lietal.[35], andLiaoetal.[36]offerdifferentapproachestorankternarycomplexes wheretheyuselongerMDsimulations.Theirapproachesresultingood outcomesbutareunsuitableforareal-worldscenario,wherewemight tryalargenumberofPROTACs,sincetheyrequirealotofcomputing resourcesandextendedsimulationtime.

Thefirstcontributionofourcurrentworkisthedemonstrationofthe successfuluseofaBayesianoptimization(BO)[37]frameworktoper-formasample-efficientexplorationofthespaceofternarycomplexcan-didates.BOisanestablishedmethodthatallowsoptimizingfunctions thatarecostlytoevaluatebyproposingpromisingevaluationcandidates (here:candidateternarycomplexstructures).Thisleadstosampleeffi-ciency,i.e.,fewercandidatesneedtobeevaluated(e.g.,byadocking software)comparedto(e.g.,MonteCarlo-)searchmethods,andallows theuseofpotentiallycomplexfitnessfunctionswhichareneededto characterizethemultipleinteractionswithinaternarycomplex.Using thisapproach,wecanefficientlysampleposesthatareviablecandidates forthestructureoftheternarycomplex,withtheviabilityscoredbya weightedsumofprotein-proteininteraction(PPI)scoreandaPROTAC score.ThePPIscorequantifiestheinteractionsbetweentheproteinsin aparticularcandidate.Incontrast,thePROTACscoreconsidersthefea-sibilityofembeddingaconformationofthePROTACmolecule,given thepositionofthebindingpocketsinthisconfigurationofproteins. Thesecondcontributionofourworkliesindesigningoftwonovel scoresspecializedforfiltering,andranking.Thefirstscoreisusedfor filteringfavourablecomplexesbyimposingconstraintsonthedistances oftheproteinsurfacesfromoneanother– aversionoftheambigu-ousinteractionrestraintsenergy(hereabbreviatedas“TCP-AIREn-ergy”)[38]thathasbeenmodifiedspecificallyforternarycomplexes (seeFig.5).ThesecondscoreisthePROTACstabilityscore,whichis calculatedusingamodifiedversionofAutodock-Vina.Givenaparticu-larconfigurationofproteins,itpacksthePROTACbetweentheproteins andindicatesitsstabilityandenergeticfavourabilityinthisposition. Thisscoreisusefulforranking.

Weshowinourcurrentwork(SeeTable2)that,withefficientsam-plingwiththeBOloop,combinedwithfiltering,clustering,andranking usingtheabovescores,weareabletoeffectivelyfindhighlyranked clusterswithnear-nativeposes.Specifically,ourclusteringismorecon-finedcomparedtopreviousworks,andeachclustercontainsonlyasmall numberofconformers.Ourabilitytoranktheseeffectivelygivesusclus-terswithahighyieldofnear-nativescores.

**2.Methods**

Inthispaper,weproposetheBayesianOptimizationforTernary ComplexPrediction(BOTCP)method.AtitscoreliesaBayesianopti-mization(BO)loopcapableoffindingappropriateensemblesofternary complexesbyoptimizingacombinedfitnessfunctionwhichincludes afitnessdescribingthestrengthoftheinteractionofthetwoproteins

3

*A.Rao,T.M.Tunjic,M.Brunsteineretal.*  *ArtificialIntelligenceintheLifeSciences3(2023)100072*

**Table1**   
Detailsoftheproteinstructuresusedinthiswork.ColumnsshowthePDBIDforeachternarycomplexaswellasthechainidentifiers(complex),thegene name,thePDBIDofthestructureusedasinputforprediction(template),thechainIDsofE3ligaseandPOI,thefirstandlastresiduenumberincludedinto theusedmodels(N-term,C-term).ForthePROTAC,thecolumsshowtheresidueID,thenatureofanyionizablegroup,andthenetchargeusedinmolecular models.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Complex | | E3ligase | | | | | | | POI | | | | | PROTAC | | | | | | | | | | |
| PDBID | Name | Template | Chain | N-term | | C-term | | Name | | | Template | | Chain | N-term | | C-term | | | | Residue | | Ionizable | | Netcharge |
| 5T35\_1  5T35\_2  6BN7  6BOY  6HAX\_1 6HAX\_2 6HAY\_1 6HAY\_2 6HR2\_1 6HR2\_2 6SIS\_1  6SIS\_2  6W7O\_1 6W7O\_2 6W8I\_1  6W8I\_2  6W8I\_3  6ZHC  7JTO  7JTP  7KHH  7Q2J | VHL  VHL  CRBN  CRBN  VHL  VHL  VHL  VHL  VHL  VHL  VHL  VHL  BIRC2  BIRC2  BIRC2  BIRC2  BIRC2  VHL  VHL  VHL  VHL  VHL | 4W9H-I 4W9H-I 4TZ4-C  4TZ4-C  4W9H-I 4W9H-I 4W9H-I 4W9H-I 4W9H-I 4W9H-I 4W9H-I 4W9H-I 6W74-A 6W74-A 6W74-A 6W74-A 6W74-A 4W9H-I 4W9H-I 4W9H-I 4W9H-I 4W9H-I | D  H  B  B  B  F  B  F  B  F  D  H  C  D  D  E  F  A  L  L  C  C | 62  62  48  48  62  62  62  62  62  62  62  62  266  266  266  266  266  62  62  62  62  62 | 204  204  426  426  204  204  204  204  204  204  204  204  349  349  349  349  349  204  204  204  204  204 | | BRD4-2  BRD4-2  BRD4-1  BRD4-1  SMARCA2 SMARCA2 SMARCA2 SMARCA2 SMARCA4 SMARCA4 BRD4-2  BRD4-2  BTK  BTK  BTK  BTK  BTK  BCL2L1  WDR5  WDR5  BRD4-1  WDR5 | | | | | 5UEU-A 5UEU-A 3MXF-A 3MXF-A 6HAZ-A 6HAZ-A 6HAZ-A 6HAZ-A 6ZS2-A  6ZS2-A  5UEU-A 5UEU-A 5P9J-A  5P9J-A  5P9J-A  5P9J-A  5P9J-A  4QVX-A 4QL1-A  4QL1-A  3MXF-A 4QL1-A | A  E  C  C  A  E  A  E  A  E  A  E  A  B  A  B  C  D  B  A  D  D | 349  349  44  44  1378  1378  1378  1378  1449  1449  349  349  396  396  396  396  396  2  32  32  44  32 | 457  457  168  168  1490  1490  1490  1490  1569  1569  457  457  656  656  656  656  656  197  333  333  168  333 | | | | 759  759  RN3  RN6  FWZ  FWZ  FX8  FX8  FWZ  FWZ  LFE  LFE  TL7  TL7  TKY  TKY  TKY  QL8  VKA  X6M  WEP  8KH | | no  no  imine  imine  piperazine  piperazine  piperazine  piperazine  piperazine  piperazine  secamine  secamine  secamine  secamine  secamine  secamine  secamine  carboxylicacid 2piperazines  piperazine  no  piperazine | | | 0  0  0  0  1  1  1  1  1  1  1  1  1  1  1  1  1 -1  2  1  0  1 |
|  |
| **Table2**  ResultsoftheBOTCPmethodforternarycomplexpredictiononunboundstructuresbeforethestructuralrefinement(usinggreedyclustering2.4).1Thebestrank containingatleastonemodelwithDockQ≥0.23.2Thetotalnumberofclusters.3Thepercentageofthenear-nativeposesinthatspecificrankedcluster.Foreach ternarycomplex,weselectedaclusterwhichincludesanRRTwithahighDockQscore.ColumnsalsoshowthePDBIDandcharacterizationofthebestRRTin termsof*𝑓*nat,*𝐼*RMSD,*𝐿*RMSDandDockQscore(seetext). | | | | | | | | | | | | | | | | | | | | | | | | |
| PDBID | Rankofcluster1 | | | Numberofclusters2 | | | | | | %Near-native3 | | | | *𝑓*nat | | | *𝐼*RMSD | | | | | | *𝐿*RMSD | DockQ |
| 6HAY\_1 6HAY\_2 7JTO  6BN7  7Q2J  6W7O\_1 6W7O\_2 5T35\_1  5T35\_2  6W8I\_1  6W8I\_2  6W8I\_3  6HAX\_1 6HAX\_2 7KHH  6BOY  6SIS\_1  6SIS\_2  6ZHC  7JTP  6HR2\_1 6HR2\_2 | 2  5  15  15  7  11  4  76  51  13  79  6  8  14  None  19  1  8  142  8  26  54 | | | 914  946  403  917  322  269  271  765  763  362  364  365  816  795  None  720  339  365  670  170  727  679 | | | | 100  100  50  6.7  79.1  9.4  29.2  94.1  88.9  96.6  25  84.6  100  100  None  80  100  100  75  88.1  41.7  25 | | | | | | 0.61  0.61  0.5  0.14  1.0  0.17  0.35  0.73  0.67  1.0  0.52  0.81  0.57  0.83  None  0.61  0.53  0.65  0.82  0.68  0.44  0.14 | | | | 1.07  1.77  3.05  3.33  2.32  3.77  2.92  2.09  1.81  2.248  2.61  1.85  2.61  1.39  None  2.27  1.72  2.14  3  2.52  2.27  3.51 | | | | | 3.19  3.03  14.82  10.28  5.86  7.31  7.99  4.92  8.36  4.41  11.35  3.7  5.34  8.95  None  10.04  5.62  4.56  19.15  5.54  12.95  8.93 | **0.72**  **0.64**  **0.31**  **0.24**  **0.66**  **0.29**  **0.36**  **0.61**  **0.53**  **0.7**  **0.38**  **0.68**  **0.51**  **0.62**  **None**  **0.44**  **0.55**  **0.59**  **0.4**  **0.55**  **0.35**  **0.26** |

3.WecroppedflexibleN-andC-terminaltails(usuallywithpoorcon- fidencelevelsinAlphaFoldpredictions)asspecifiedin1.

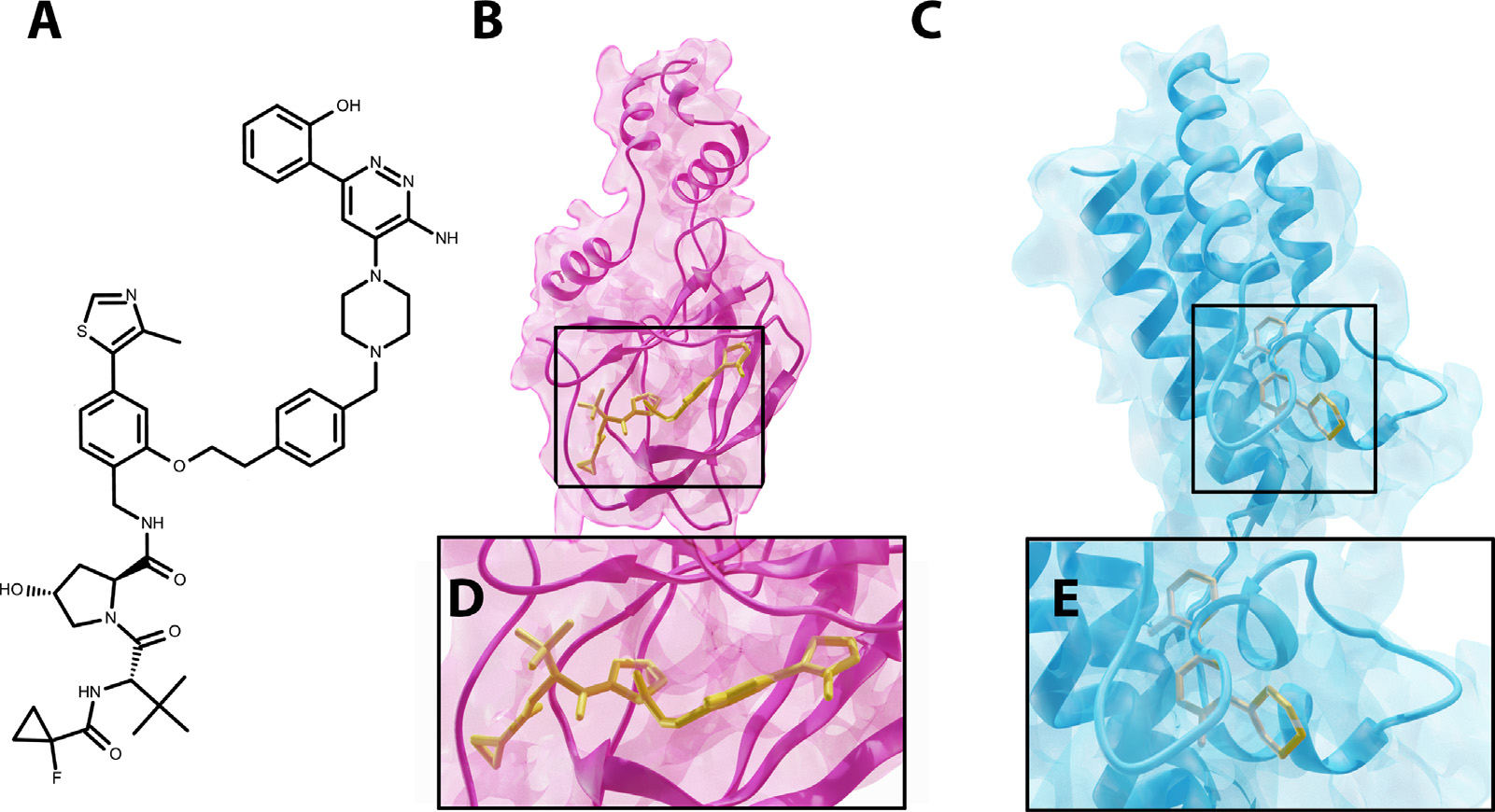
4.MissingatomswerereplacedusingthePDBfixertool[44].

5.WeusedReduce[45]toprotonate/addhydrogens,and,whereap- propriate,flipasparagine,glutamine,orhistidinesidechains.

ThePROTACmoleculeswereextractedfromtheoriginalPDBfile andconvertedtoSDFformatusingOpenBabel[46].Ifrequired,bond ordersandprotonationwerecorrectedmanually.Hereweassumedthat allweakbases(secondaryandtertiaryamines)andacids(carboxylic acid)are(de)protonatedtoformanionsorcations.Sinceallmodels andforcefieldsweuseworkwithfixedprotonationstatesonly,we

4

*A.Rao,T.M.Tunjic,M.Brunsteineretal.*  *ArtificialIntelligenceintheLifeSciences3(2023)100072*



**Fig.2.**Exampleoftherequiredinputsfortheternarycomplex6HAX(PDBID).(A)PROTACmolecule.(B)and(D)E3ligasedockedtowarhead.(C)and(E)POI

dockedtothewarhead.

Ourmethodusesthreeinputs:thePOIwiththedockedwarhead,the E3ligasewiththedockedE3binder,andthePROTACSMILES(seeFig.2 foranexample),whichareprovidedtotheBayesianoptimizationloop. Itismoresuitabletoaccesstheco-crystalstructureofPOI+warhead andE3ligase+binder,butifthisinformationisnotavailable,itis requiredfirsttodockthecorrespondingligandsintotheirproteinsand usethesestructuresaftermoleculardockingforthenextstepinout BOTCPpipeline.

Weapplyourmethodtothestructuresofproteinsanddockedligands extractedfromtheco-crystalstructureofknownPROTACternarycom-plexes(so-calledboundproteindata)andthemuchmorechallenging caseofunbounddata,wheretheco-crystalstructureofknownPROTAC ternarycomplexesisnotavailable.

*2.2.RigidposesamplingviaBayesianoptimization*

Inordertodeterminethestructureoftheternarycomplex,wefirst recognizethatthetwoproteinscaninteractinavarietyofposes.The firststepinourpipelineistosampleasetofsuchposesthatareviable candidatesfortheternarycomplex.TodosoweconsidertheinputPOI andE3ligaseasrigidstructures.Inthiscase,eachposeischaracterised bytherelativepositionandorientationofthePOIwithrespecttothe E3Ligase.Wethensearchthroughthespaceofposes,andsampleposes thatareviablestructuresfortheternarycomplex.

*2.2.1.Representationofposes*   
 Inordertoencodeapose,weconsiderabaselineposewhereboth thePOIandE3Ligaseareshiftedsothattheircenterofmasseslineup attheorigin.KeepingthepositionoftheE3-ligasefixed,eachposethen correspondstoatranslationandrotationofthePOI.Thisisdenotedby a7-Dvector**𝐱**∈ ℝ7thatrepresentstherelativerotationandtranslation (RRT)ofthePOIwithrespecttoitspositioninthebaselinepose,where **𝐱**1…3istherelativetranslation,and**𝐱**4…7isaquaternionrepresentinga relativerotation.Duetothefactthataprotein-proteinconfigurationcan berepresentedassuch,fortheremainderofthispaper,wewillrefertoa specificprotein-proteinconfigurationusingthetermRRT.Ourobjective abovethenbecomes,tosampleasetofRRT’sthatareviablecandidates forthestructureoftheternarycomplex.

*2.2.2.Fitnessofapose(RRT)*   
 HerewemakeclearwhatwemeanbyaparticularRRTbeinga‘vi-ablecandidate’forthestructureoftheternarycomplex.Foraparticular

5

*A.Rao,T.M.Tunjic,M.Brunsteineretal.*

simpleUFFforcefieldisbecauseitallowsustoconstrainatomstoatoms tospecificpositionsbyusingaquadraticpotentialthatpenalizesdevia-tionfromthesetargetpositions.Thisquadraticpotentialprovesuseful whenapproximatingthePROTACscore(seefollowingsection)   
 Aftergenerating4candidateconformers,wechoosetheoutputtobe theconformerwiththelowestUFFenergy.Thechoiceof4conformers isbecauseitisthesmallestnumberofconformerswefindisneededto reliablyreporttheminimumenergyconstrainedconformer.ThePRO-TACfitnessisdefinedasthenegativesquare-rootoftheminimizedUFF energyofthisconformer.

NotethatwhencalculatingthePROTACscoreabove,wedonotcon-sidertheinteractionsofthePROTACwiththeproteins,thus,itcannot beusedtofilteroutcaseswithstericclashesbetweenthelinkerandthe proteins.ThesecasesaretakenintoaccountbythePROTACstability score(seeSection2.3).

*ApproximatingPROTACfitnesswithaneural-network*   
 ThePROTACfitnessforaparticularRRTisafunctionoftheenergy obtainedafterminimizationoftheUFFforce-fieldenergy.Thismini-mizationiscomputationallyexpensive,andtheresultoftheminimiza-tionisnotananalyticfunctionoftheRRT,whichmeansthatitisex-pensivetooptimizetheRRTwithrespecttothisscore.Asasolution,we havefoundthatwecantrainaneuralnetworkmodelthatcaneffectively approximatethePROTACfitnessforaspecificRRT.

ForaparticularPOI,E3ligase,andPROTAC,wetrainaneuralnet-workmodelwhichmapsthe7DRRTvector**𝐱** tothePROTACfitness valuesgeneratedusingRDKit(asdescribedabove).Themodelconsists oftwocomponents:asimplemodelwhichlearnstopredictasymptotic componentoftheenergy,whichisasymptoticallylinearw.r.tthedis-tancebetweenthebindingpocketsofthePROTACfragmentsforthe particularRRT,anda3-hiddenlayerneuralnetworkwhichlearnsto predicttheresidualcomponent,capturingthemorecomplexaspectsof thescoresuchasself-collisionsthatmayariseduetotheRRTsbringing thebindingpocketstooclose.

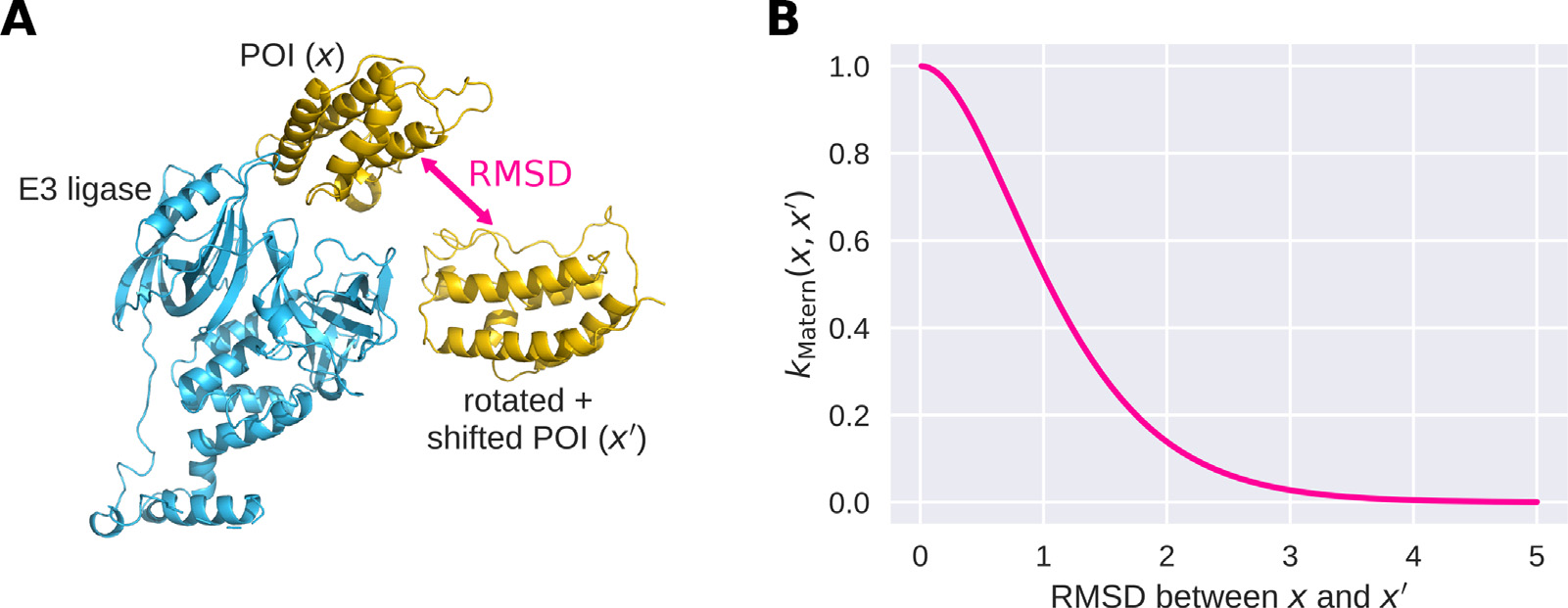
PriortotheBOloop,wepre-trainthismodelusingadatasetofRDKit-generatedconformersfordifferentRRT,andusethismodeltocalculate anapproximationtothePROTACfitnessforthe**𝐱** valuesproposedby theBayesianoptimizationloop.Theadvantageisthat,afterbeingpre-trained,computingthefitnessforanewRRTusingthemodelrequires justoneforwardpass,makingthefitnesscomputationveryfast.Wefind thateventhoughwerequireadditionalcomputationtotrainthemodel, theresultantspeedgainmakesupforitwhenoptimizingthePROTAC scoreintheBOloop.

*Localfitnessascent*   
 Asisoftenseeninbiologicalsystems,smallchangesinpositioncan drasticallyaffectfitnessvaluesforexample,asaresultofstericclashes. Thismakesthefitnesslandscapehighlyirregularwithregionsoflow fitnessdirectlyadjacenttoregionsofthehighestfitnesses.Inorderto effectivelylearnthisfitnesslandscapeviatheBayesianoptimization,we modifythefitnessfunctionasfollows.ForaparticularRRT**𝐱**,weper-formalocalascentinthespaceofRRTsusingablackboxoptimization algorithm(Nelder-meadsimplexascent[50]for15steps),andtakethe finaloptimizedfitness*𝑦*opt(**𝐱**)tobethefitnesscorrespondingto**𝐱**.This optimizationmeansthatevenRRTswithstericclasheswillbemappedto nearbyRRTsthathavehighfitnesses,makingthelandscapealotmore regular.Additionally,themaximaof*𝑦*opt(**𝐱**)correspondtothemaxima of*𝑦*(**𝐱**)meaningthatwecanmaximizeoneinsteadoftheother.This approachisdescribedinAlgorithm1.

*2.2.3.Bayesianoptimization*   
 WhensamplingviableRRTs,WewishtoselecteachRRT**𝐱** byop-timizingtheblackboxfunction*𝑦*(**𝐱**)describedabove.However,while wecanevaluate*𝑦*(**𝐱**)atanyRRT**𝐱**,theevaluationsarecostly,andthe functionhasnoanalyticalform,i.e.wedonothaveanyinformation aboutthegradient,thusbecomingaproblemofgradient-freeorblack-boxoptimization.Inthiswork,wechoosetoapproachthisproblemvia thetechniqueofBayesianOptimization[51,52]asitisknowntobe

6

*A.Rao,T.M.Tunjic,M.Brunsteineretal.*  *ArtificialIntelligenceintheLifeSciences3(2023)100072*



**Fig.4.**IllustrationoftheMaternkernel,usedascovariancekernelfortheGaussianprocess.(A)RMSDbetweendifferentPOIpositionsrelativetotheE3ligase(i.e.,

differentRRTs**𝐱**).(B)Maternkernel(seeEq.(4))asafunctionoftheRMSD.

|  |
| --- |
| **Algorithm2**TheBOloop.  *𝑋*train←*𝑁*initrandomlysampledRRTs  **𝐲**train←correspondingfitnessvalues  **repeat**   Fitsurrogatemodel(GP)using*𝑋*train*,𝑦*train *𝑋*new←AcquirePoints()   **𝐲**new←EvaluateFitness(*𝑋*new)   *𝑋*train←Concatenate(*𝑋*train,*𝑋*new)   **𝐲**train←Concatenate(**𝐲**train,**𝐲**new)  **until**stoppingconditionismet |
| and/orhighuncertainty(exploration).Wethenevaluatethefitnesses forthesepointsandretrainthesurrogatemodelandrepeattheproce-dure.Thedetailsofthesurrogatemodelandacquisitionstrategyare describedbelow.  *Surrogatemodel*   WhileinprincipleBayesianoptimizationcanutilizedifferentsur-rogatemodels,weemployedthewidelyusedGaussianprocess(GP) [53,54].AGPisamodelthatplacesaprioronthespaceoffunctions thatcanbeusedtodescribethedata,byspecifyingaprioronthemean ofthefunctionvalue*𝑚*(**𝐱**)andviaacovariancekernel*𝑘*(**𝐱***,***𝐱**′)whichde-scribesthecovariancebetweenthefunctionvaluesatanytwopoints**𝐱** and**𝐱**′,i.e.,  *𝑓*(**𝐱**)∼ *𝑃* Weuseaconstantlearnedvalue*𝑚*(**𝐱**)≡ *𝑚*0aspriormean.Ascovariance ( *𝑚*(**𝐱**)*,𝑘*(**𝐱***,***𝐱**′) ) *.*  (3)  kernel,weusetheawell-knownMaternkernel[54](seeFig.4),i.e.,  *𝑘*(**𝐱***,***𝐱**′)=21−*𝜈* Γ(*𝜈*) (√2*𝜈𝑑* )*𝜈 𝐾𝜈* (√2*𝜈𝑑* ) (4)  where*𝑑* isthedistancebetween**𝐱** and**𝐱**′,*𝜈* isasmoothnessparameter (takingonvalues1 2,3 2,or5 2,withlargervaluesincreasingsmoothness), and*𝐾𝜈* isamodifiedBesselfunction.  Inouroptimization,Weuse*𝜈* =5 2,andasdistance*𝑑*,weusethe RMSDbetweentheatomiccoordinatesoftheatomsofthePOIscaledby alearnedparameterΘ.WemakeuseofthefactthattheRMSDbetween tworigidtransformationsofthePOIcanbeefficientlycalculatedbypre-calculatingthemoment-of-inertiatensorforthePOI,thusnotrequiring ustotransformallthecoordinatesforeachRRT.Thismakesthiskernel verycomputationallyefficienttocompute.  *Acquisitionstrategy*   Bayesianoptimizationrequiresastrategybywhichweselectnew candidateRRTsforwhichweevaluatethefitness.Intuitively,wewantto performexplorationbysamplingpointsforwhichthesurrogatemodel hasnoinformationabouttheirpotentialvalue,whilealsoperforming exploitationbysamplingpointsforwhichthesurrogatemodelesti- |

7

*A.Rao,T.M.Tunjic,M.Brunsteineretal.*

performedusinggradientascent(usingADAM[56]).Weusetheopen-sourceBoTorchpackage[57]tosolvethisoptimizationproblemand implementtheBOloop.

*2.3.LocaloptimizationwithsimulatedannealingofthePROTACstability score*

Fromtheabovepoints,wefilteroutRRTswithinfeasiblePROTAC scoresorPPIscores,i.e.weselectonlycomplexesthathavePPIscores thatareat-least6.0(theclippedvalue),andPROTACscoresthatcorre-spondtoaUFFenergylesserthan1.5timestheminimumUFFenergy. WenextlocallyoptimizedtheremainingRRTsusingacombinationof thePPIscoreandaPROTACstabilityscore.Weperformedtheoptimiza-tionwiththesimulatedannealing.

Thesethresholdsabovewerechosenheuristicallytooptimizethe numberofnear-nativestructuresthatpassthefilterforthecompounds 6HAX\_1,6HAX\_2,6HAY\_1,6HAY\_2.Thesewerenotoptimizedsubse-quentlybasedonthefinalrankinginordertoavoidoverfitting.Anypa-rameterthatismentionedinthefollowingsectionsissetbasedonthe datageneratedwhenrunningtheBOTCPprotocolfortheabovefour compounds.

*Simulatedannealing*WeusedasumofthePROTACstabilityscore, describedunderneath,andthePPIscoretoperformlocaloptimization oftheremainingcandidatesusingsimulatedannealing[58].Forany RRT,simulatedannealingproceedsviaasequenceof10steps,where ineachstepwerandomlysampleanRRTnearthecurrentRRT.We evaluatetheweightedsumofthePROTACstabilityandPPIscoresand stochasticallyacceptandrejectitbasedonthat.Note,thatinordertonot wastecomputationoninfeasibleRRTs,therandomlygeneratedsamples abovearepre-optimizedtohaveareasonablePPIandPROTACscore. AsthisprocessallowsthePROTACconformationtochange(both explicitlythroughsimulatedannealingandimplicitlywithinAutodock-Vina),itrepresentsastructuralrefinement/packingofthePROTAC conformer.

*PROTACstabilityscore*   
 WecalculatedaPROTACstabilityscoreusinganextensionof AutoDock-Vina[59].Thisscoretakesintoaccountthestabilityofthe PROTACwithinthecontextoftheproteinconfigurationcorresponding toaparticularRRT(whichisignoredinthecalculationofthePROTAC fitness,seeabove).

Inordertodothis,weextendedAutodock-Vinabyimplementingan additionalrestoringenergy,thatallowsonetoconstraintheposition ofasubsetofatomsoftheligand(whichisbeingdocked)tospecific coordinates.Considerforexample,aparticularligandwhereasubset ofatoms*𝑆𝑐*areconstrainedtospecificcoordinates.Nowlet’sdenote foraparticularconformationoftheligand,thecoordinatesofatom*𝑗* by*𝑟𝑗*∈ ℝ3,anditsconstrainedpositionby*𝑟*0*,𝑗*∈ ℝ3,thentherestoring energy*𝐸𝑅*isgivenby

*𝐸𝑅*=*𝑘𝑅*∑ Thisenergygrowsquadratically,andimposesaquadraticpenaltyon‖‖‖*𝑟𝑗*−*𝑟*0*,𝑗*‖‖‖ (6) 2

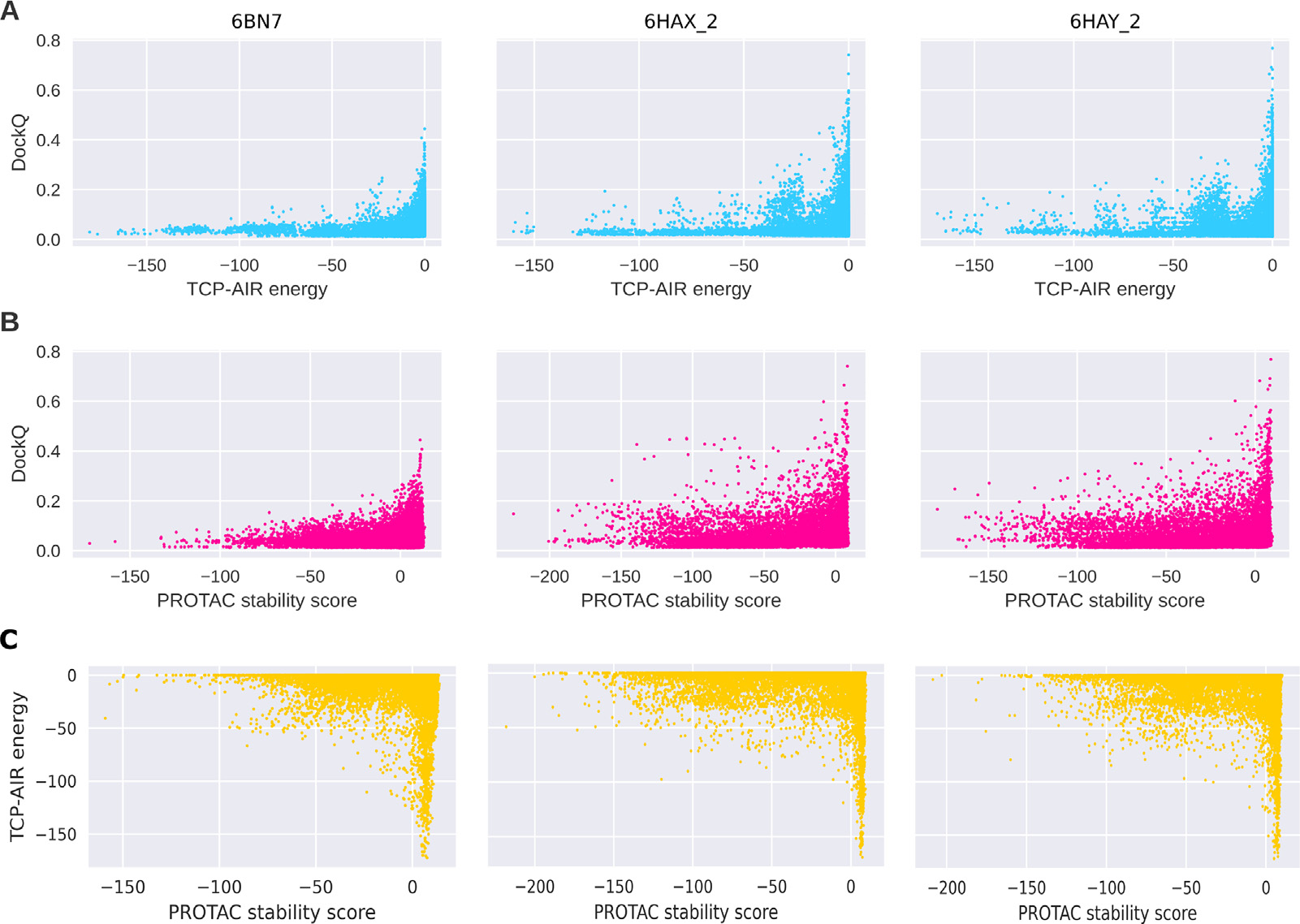
thedeviationofanatomspositionfromit’sconstrainedposition.The gradientofthisrestoringenergyistermedtherestoringforce.Weused thisforcetoconstraintheE3binderandthewarheadtohavethecoordi-natesoftherespectivebindingpocketsgiventhecurrentprotein-protein RRT.TocalculatethePROTACstabilityscoreforaparticularRRT,we startfrom10randomlygeneratedinitialPROTACconformations,and foreachofthemperformthefollowing:

1.WeoptimizethePROTACmolecule’sconformationunderonlythe restoringforceandtheintra-molecularpartoftheVinaforcefield, creatingaviableconformerwiththebindingfragmentsheldattheir positions.

2.Wethenturnedoff therestoringforce(i.e.set*𝑘𝑅*=0)andperform anoptimizationusingthefull(intra-andinter-molecular)Vinaforce field.

8

*A.Rao,T.M.Tunjic,M.Brunsteineretal.*  *ArtificialIntelligenceintheLifeSciences3(2023)100072*



**Fig.5.**UsefulnessofTCP-AIRenergyandthePROTACstabilityscoreforfiltering.A)CorrelationbetweentheTCP-AIRenergyandtheDockQscorefordifferent exampleternarycomplexes.AshighDockQscoresareonlyachievedwhentheTCP-AIRenergyis0orcloseto0,thelatercanbeusedasafilteringcriterion.B) SameasA,butforthePROTACstabilityscore.C)CorrelationbetweenthePROTACstabilityscoreandtheTCP-AIRenergyfor6BN7(*𝑟*=−0*.*01,*𝜌* =0*.*09),6HAX\_2 (*𝑟*=−0*.*07,*𝜌* =−0*.*08)and6HAY\_2(*𝑟*=−0*.*05,*𝜌* =−0*.*07),with*𝑟*beingthepearsoncorrelationcoefficientand*𝜌* thespearman’srankcorrelationcoefficient.Thelow correlationindicatesthatthetwoscorescaptureindependentaspectsofthenear-nativepose,thusjustifyingtheuseofbothscores.

Fortheseresidues,wecalculatetheeffectivedistances(asdonein previouswork[38]).TheseareusedtocalculatetheTCP-AIRenergy (usingtheNOEenergyformula,[60])usingthreethresholdvalues*𝐿*, *𝑈*,and*𝑆*.*𝐿*and*𝑈* arethelowerandupperboundsontheeffectivedis-tance,while*𝑆* isthevaluefromwherethecomputationoftheenergy changesfromquadratictolinear.Theseparametersneedtobespecified foreachternarycomplexseparatelysinceeachstructurehasdifferent distancesbetweentwoactivesitesoftheproteins(duetothespaceoc-cupiedbythePROTACmolecule).Tosetthevalues,wecalculatedthe averagedistancesofeachactiveresidueofoneproteinwiththeactive andpassiveresiduesoftheotherprotein,resultingintwodistancesfor eachcandidate.WeconsideredallofoursampleswithPROTACstabil-ityscore*>*5.Foreachofthem,wetookanaverageofthetwoproteins’distancesanddefined*𝐿*asthe25thquantile.Wethendefined*𝑈* =*𝐿*+ 3Å and*𝑆* =*𝑈* +2Å.Whenfiltering,wepreservedallclusterswithat leastonecandidatewithTCP-AIRenergylargerthanthe90thquantile ofallAIRenergies(seeFig.5).

Alloftheparametersdefinedaboveweredefinedandoptimizedus-ingfourternarystructures6HAX\_1,6HAX\_2,6HAY\_1,and6HAY\_2.The activeandpassivedistancesweredecidedbyinspectingthestructuresof theseternarycomplex,whichshowedittobeareasonableassumption thatthenativecomplexwillcontainasizeablefractionofinteracting residueswithin5Angstromsfromthebindingsite,andalmostcertainly 7.5Afromit.TheparametersL,U,S,weremanuallytunedtomaximize thenumberofnear-nativecomplexesfilteredforthese4cases.

*Rankingofclusters*   
 Afterfilteringtheclusters,weranktheremainingclustersusingthe PROTACstabilityscore(seeabove,Section2.3).Clustersareranked accordingtothelargestPROTACstabilityscoreoftheirmembers.We thencalculatetherankofthenear-nativeclusters(Table2).Forthe structuralrefinement,wegenerateadiversesetofcomplexesbyselect-ingthestructurewiththehighestPROTACstabilityscoreoutofeach ofthe100top-rankedclusters.Foreachofthese,weusethe10PRO-

9

*A.Rao,T.M.Tunjic,M.Brunsteineretal.*

conformationalspacetoidentifyanear-localminimumandnotforex-haustivesampling.Weassumethattheprevioussteps,namelyBayesian optimizationsamplingandlocaloptimizationofthePROTACstability score,enableustoattainlarger-scalesampling.

*2.6.Re-clusteringandre-ranking*

Therefinedstructures(1000ofthem)wereclusteredbythefraction ofcommoncontact(FCC)clusteringmethod[68].Afterwards,theclus-terswererankedbasedonsize(theclusterwiththehighestnumberof elementsbeingthefirst-ranked).Theclusteringthresholdwassetto0.5 andminimumnumberofelementsinaclusterwassetto2.Therestof non-clusteredstructureswerenottakenintoconsideration.

*2.7.Evaluationmetric*

ForevaluatingthequalityofthecandidatesgeneratedbytheBOloop (Table2),weusedtheDockQscore[69],whichisbasedonthemetrics underlyingthewell-knowncriticalassessmentofpredictedinteractions (CAPRI)score[70]:*𝑓*nat,*𝐼*RMSD,and*𝐿*RMSD.TheDockQscoreforapar-ticularcandidateprovidesaquantitativemeasureofhowcloselythe interactionofproteinsinthiscandidatestructureresemblesthatofpro-teinsinareferencenativestructure.Itisafunctionofthefollowing metrics   
 *𝑓*natisthefractionofprotein’snativecontactsthatacandidatere-covers.*𝐼*RMSDiscalculatedoverallbackboneatomsforthoseresidues foundintheinterfaceofthereferencestructure(residueswithinaradius of5Å).*𝐿*RMSDiscalculatedbasedonthebackboneatomsofthetarget protein.BothRMSDvaluesarecalculatedbetweenthecandidateand nativemodel(afteraligningthereceptorstructures,i.e.,theE3ligase). TheDockQscorecombinesthesemeasuresintoasinglescalarvalue between0and1(whichisthescoreofthereferencestructure),with highervaluescorrespondingtoternarycomplexeswhichresemblethe originalstructuremoreclosely.ThedistributionoftheDockQscoreshas beendefinedasfollows[69]:

•0*<*DockQ*<*0.23-Incorrectcomplex•0.23≤DockQ*<*0.49-Acceptablequalitycomplex•0.49≤DockQ*<*0.80-Mediumqualitycomplex•0.80≤DockQ*<*=1.0-Highqualitycomplex

AllcandidateposesthatachieveaDockQ*>*0*.*23comparedtothe nativecrystalstructurearetermedasnear-nativeposes

*2.8.Fitnesslandscapeanalysis*

Using5exemplaryternarycomplexesfromPDB(boundstructures), weextractedtheRRTcorrespondingtothenativeposeandmodified themtoinvestigatetheresultingchangeofthefitness*𝑦*(**𝐱**).Foreach ternarycomplex,wegenerated4000randomnear-nativesamplesby firstrandomlyshiftingthePOIbeforerotatingthem.Theshiftdirections weredrawnfromavonMises–Fisherdistributionwiththemeansetto theoffsetofthePOI(i.e.,pointingawayfromtheE3ligase)andthe directionalparameter*𝜅* =1(i.e.,shiftsmoderatelyconcentratedinthis direction).Theshiftdistancesweredrawnfromanexponentialdistribu-tionwith*𝜆* =1Å.Therotationsweregeneratedbydrawingarandom rotationdirectionfromauniformdistribution(nodirectionsfavored) andarotationangledrawnfromavonMisesdistributionwithmean0 and*𝜅* =10(i.e.,smallrotationsfavored).

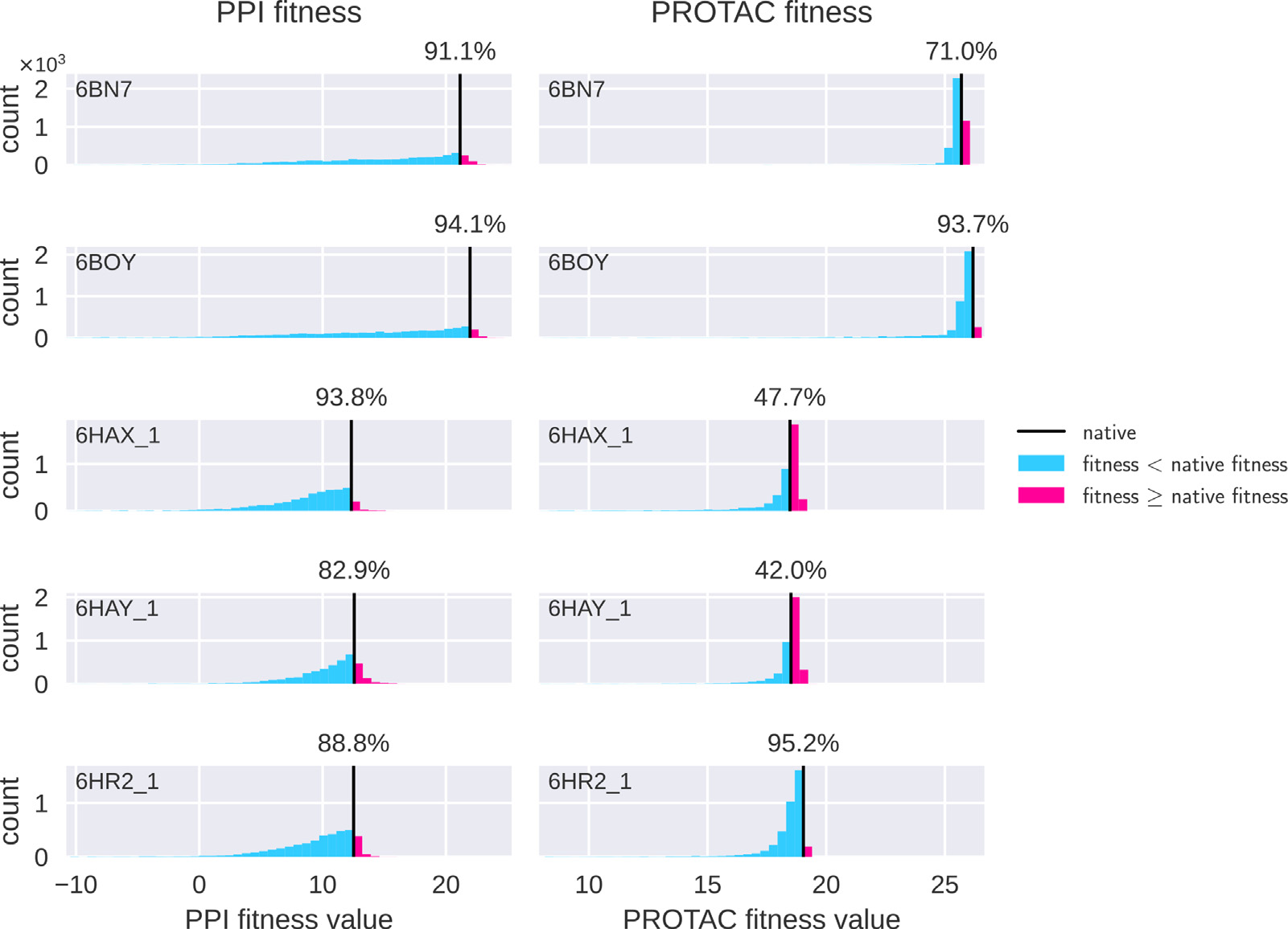
**3.Results**

*3.1.Neural-networkbasedapproximationofPROTACscore*

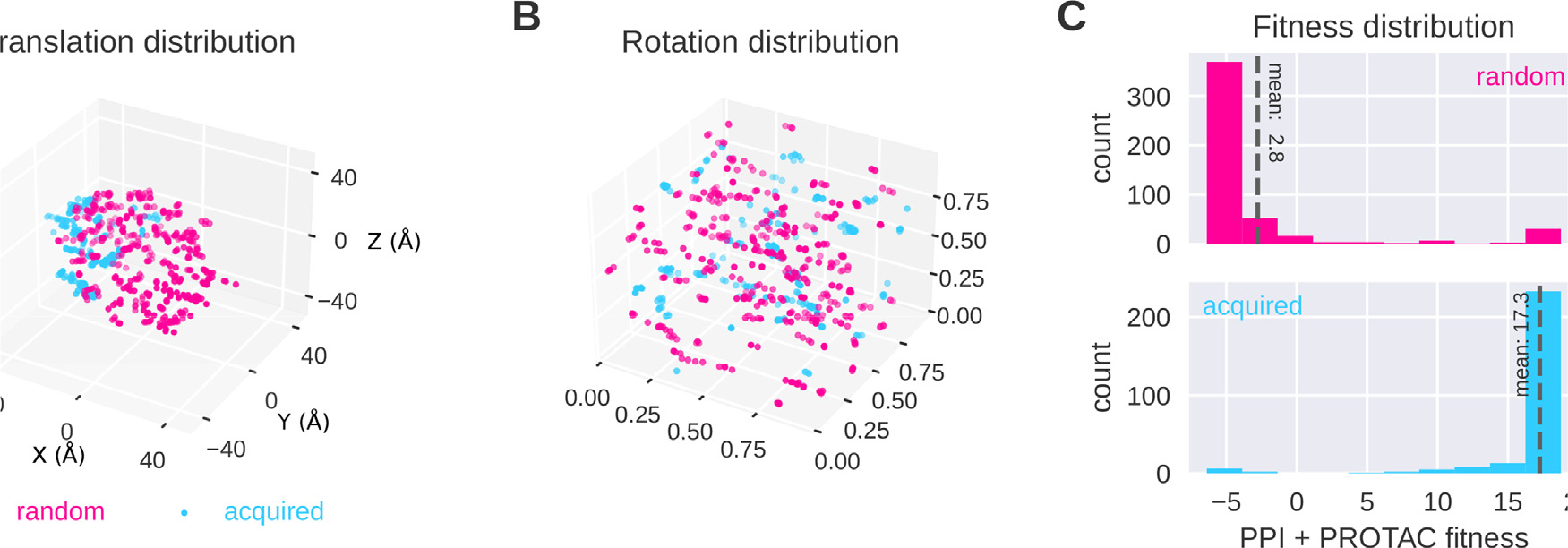
ThePROTACscoreisthenegativesquarerootoftheenergyachieved afterminimizingtheUFFforce-field.ThevalueofthisPROTACscore achievesforanun-strainedconformationofthePROTACrangesfrom

10

*A.Rao,T.M.Tunjic,M.Brunsteineretal.*  *ArtificialIntelligenceintheLifeSciences3(2023)100072*



**Fig.6.**Nativecomplexeshavehighfitnessvalues.Thecolumnsshow(forfivesampleternarycomplexes)histogramsofPPIfitness(left)andPROTACfitness(right) forrandomnear-nativeRRTs,createdbyrandomlyshiftingandrotatingthePOI(seetextfordetails).Eachplotcontainsthehistogramof4096suchRRTs,withthe percentileofthefitnessofthenativeRRTdisplayedabovetheblackline.Theactualnativecomplex(blackline)achievesacomparitivelyhighfitnessinmostcases. Inafewcaseswheretheconstraint-fitnesshasalowerpercentile,wefindthatthisiscompensatedforbythehigherpercentileofthePPIscore.



**Fig.7.**Bayesianoptimizationresultsinhigherfitnessvaluesthanrandomsearch.Shownherearesamplesfromtheacquisitionfunctionatiteration15oftheBO loopforthe6HAXcomplex,comparedtorandomsamplespickedwithproteinsclosetoeachother.(A)TheTranslationcoordinatesoftheRRTs(inÅ)fromrandom sampling(pink)andproposedbyBayesianoptimization(blue).Thelatterleadstoacandidateconcentrationinapromisingregion.Thespreadofpointsalsosuggests thattheBOloopsufficientlyexploresthespacewithoutconvergingprematurelytoanyspecificminima.PointssampledviabothrandomsamplingandBOaresubject toalocaloptimization(B)TheRotationquaternionsfromtheRRTs,mappeduniformlyontoa3-DunitcubefromrandomsamplingandBayesianoptimization.This mappingisdescribedintheSupplementSectionS1.Thediversityofrotationquaternionsshowstheextentofexploration.(C)Resultingfitnesshistograms.While randomsamplesfindsomeconformationswithhighfitness,thebestfitnessacquiredbytheBayesianoptimizationishigher,andthemeanfitnessissubstantially better,confirmingthatoursamplesefficientlysearchthroughthespaceofhigherfitnessvalues.(Forinterpretationofthereferencestocolorinthisfigurelegend, thereaderisreferredtothewebversionofthisarticle.)

near-nativeRRTs.ThusinTable2,weseenonear-nativeposeswere sampledfor7KHH.Thisissuehoweverispartiallyaddressedbystruc-turalrefinementwhichisfreetorefinethePROTACposition(Table4). Forthecaseof6W8I,thevarianceinperformancecanbeattributed tothefactthatitconsistsoftwosetsofverydifferentposes.6W8I\_1and 6W8I\_2arequalitativelysimilar(RMSDof7Å)comparedto6W8I\_3 (RMSDof33Å w.r.t6W8I\_1).Thereasonforthelowerperformance in6W8I\_2isduetothefactthattheorientationoftheLYSresiduein

11

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *A.Rao,T.M.Tunjic,M.Brunsteineretal.* | | | *ArtificialIntelligenceintheLifeSciences3(2023)100072* | | | | | | | | |
| **Table3**  ResultsandthecomparisonbetweenBOTCPandWengetal.(FRODOCK-BasedProtocol)methodsforternarycomplexpredictiononunboundstructuresbefore thestructuralrefinement.Themodelsqualityismarkedwith∗(∗∗:high,∗:mediumandno∗:acceptable).1Thebestrankcontainingatleastonemodelwith DockQ≥0.23.2Thetotalnumberofclusters.3Thepercentageofthenear-nativeposesinthatspecificrankedcluster(availableonlyforBOTCP).4Thebestpossible dockqscorewithrigiddocking(wealignanunboundtargettoanativecomplexandcalculatetheDockQscore).Foreachternarycomplex,weselectedacluster whichincludesanRRTwithahighDockQscore.Noneindicatestherewerenonear-nativecomplexesfoundandanemptyfieldthatamethodwasnotrunforthat complex.ThecolumnsalsoshowthePDBIDandcharacterizationofthebestRRTintermsofDockQscore(seetext). | | | | | | | | | | | |
| Wengetal. | | | BOTCP | | | | | | | | |
| PDBID | Clusterrank1 | #Clusters2 | DockQ | | Clusterrank1 | | #Clusters2 | | DockQ | %Near-native3 | BestDockQ4 |
| 6HAY\_1 6HAY\_2 6HAX\_1 6HAX\_2 6W7O\_1 6W7O\_2 5T35\_1  5T35\_2  6BN7  6BN8  6BN9  6BNB  6BOY  6SIS\_1  6SIS\_2  7KHH  7KHH\_2 6HR2\_1 6HR2\_2 6ZHC  6W8I\_1  6W8I\_2  6W8I\_3  7JTO  7Q2J  7JTP | 15 | 24 | 0.46 | **3∗** **1∗** **7∗** **6**/21∗ **1**/9∗ **1**/4∗ 5/15∗ **2**/27∗ 30 | | 374  377  300  283  102  106  370  315  606 | | **0.72**  **0.55**  **0.63**  **0.39**  0.48  **0.32**  0.39  **0.39**  0.31 | | 98.94  30.47  1.04  0.14  35.79  17.24  0.84  31.45  3.33 | 0.89  0.87  0.85  0.85  0.84  0.83  0.85  0.88  0.8 |
| 8 | 16 | 0.44 |
| 2∗ | 15 | **0.68** |
| **4∗∗** | 35 | **0.83** |
| **3**  2  3  1  **3**  2∗ | 25  31  17  1  30  14 | 0.26  **0.26**  **0.27**  **0.47**  0.39  **0.75** |
| 12∗ **1∗** **1**  **3** | | 423  161  189  162 | | **0.69**  0.59  **0.36**  0.23 | | 77.78  17.76  2.35  0.44 | 0.81  0.86  0.83  0.92 |
| 5  3∗ 21 | 29  15  26 | **0.43**  **0.65**  **0.48** |
| **5**/18∗ **4**/21∗ 8  **5∗** **93**  **1**/16∗ **4**  **8∗∗** **2**/5∗ | | 302  327  194  191  196  188  199  69  49 | | 0.28  **0.27**  0.39  **0.57**  **0.38**  **0.27**  **0.31**  **0.8**  **0.39** | | 1.11  0.60  5.9  16.67  4  8.7  4.49  1.1  0.43 | 0.81  0.84  0.91  0.84  0.79  0.86  0.76  0.88  0.84 |
| **2∗** None  None  None | 22  None  None  None | **0.58**  None  None  None |

near-nativecluster,atleast50%oftheposesinitarenear-nativeposes (seeTable4).

*3.5.Computationalperformance*

OurresultsdonotrequiretheuseofGPUstocomputethescoresand, using128standardx64CPUcores(16AWSEC2c5.4xlargeinstances), foreachcomplex,ourresultsincludingfilteringandre-rankingexclud-ingstructuralrefinement,typicallytakelessthan2h.Thisrepresentsa significantimprovementoverconventionaltechniques,thatattemptto handlemanyinteractionsinaternarycomplexviaGPUintensivemolec-ulardynamicssimulationswhichtypicallytakedaysorweekstoreport anear-nativepose.Notethatthiscodeisaprototypewithsignificant roomtooptimizethevariouscomponentsfurtherstill.

*3.6.Comparisontopreviouswork*

Ternarycomplexpredictionmodelshavefrequentlybeenevaluated onboundstructures(e.g.,PRosettaC[32]).Wehavedemonstratedin thisworkthateffectiverankingofpredictedstructuresinthemorechal-lengingscenarioofunboundstructuresispossiblewithourmethod.We areawareofoneotherapproachbasedonunboundstructureshasbeen presentedbyWengetal.[33],whohaveshownsuccessfulrankingofthe near-nativeclusterswithinthetop-15clusters(Tables3and4).How-ever,Wengetal.usedclusteringwithFCCusingathresholdof0.5, whichisrelativelylow,leadingtoalargespreadofposeswithinthe resultingclusters.Thesamethresholdwasusedinourcase,butonly afterrefinement,resultinginmuchbiggernear-nativepercentageof therankedclusters.Whenweusedfccwiththreshold0.5beforere-finement,wegotmuchworseresultswithsomenear-nativepercent-agesbeinglowerthan1%(Table3).Eventhoughtheremaybeanear-nativeposeinsuchacluster,athresholdof0.5impliesthatthefrac-

12

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *A.Rao,T.M.Tunjic,M.Brunsteineretal.* | | | *ArtificialIntelligenceintheLifeSciences3(2023)100072* | | | | | | |
| **Table4**  ResultsandthecomparisonbetweenBOTCPandWengetal.(RosettaDock-BasedRefinement)methodsforternarycomplexpredictiononunboundstructures afterthestructuralrefinement.Themodelsqualityismarkedwith∗(∗∗:high,∗:mediumandno∗:acceptable).1Thebestrankcontainingatleastonemodel withDockQ≥0.23.2Thetotalnumberofclusters.3Thepercentageofthenear-nativeposesinthatspecificrankedcluster(BOTCP).Foreachternarycomplex,we selectedaclusterwhichincludesanRRTwithahighDockQscore.Noneindicatestherewerenonear-nativecomplexesfoundandanemptyfieldthatamethod wasnotrunforthatcomplex.ThecolumnsalsoshowthePDBIDandcharacterizationofthebestRRTintermsofDockQscore(seetext). | | | | | | | | | |
| Wengetal. | | | BOTCP | | | | | | |
| PDBID | Clusterrank1 | #Clusters2 | DockQ | | Clusterrank1 | | #Clusters2 | DockQ | %Near-native3 |
| 6HAY\_1  6HAY\_2  6HAX\_1  6HAX\_2  6W7O\_1  6W7O\_2  5T35\_1  5T35\_2  6BN7  6BN8  6BN9  6BNB  6BOY  6SIS\_1  6SIS\_2  7KHH  7KHH\_2  6HR2\_1  6HR2\_2  6ZHC  6W8I\_1  6W8I\_2  6W8I\_3  7JTO  7Q2J  7JTP | 7 | 88 | 0.48 | **1∗** **1∗** **1**/10∗ **1**/10∗ 42  **42**  **1∗** **1∗** 8 | | 117  117  129  129  99  99  89  89  130 | | **0.6**  **0.59**  0.35  **0.36**  0.44  **0.45**  0.54  **0.51**  0.27 | 100  100  100  100  40  40  100  100  30 |
| 2 | 51 | **0.45** |
| **9∗** | 37 | **0.56** |
| 3∗ | 102 | **0.72** |
| **1∗** **1∗** **1∗** **1∗** **4∗** **1∗** | 69  80  48  1  78  15 | **0.66**  **0.7**  **0.67**  **0.5**  **0.68**  **0.72** |
| 11  2∗ **2∗** 28 | | 133  70  70  121 | | 0.40  0.56  **0.56**  0.34 | 83  94  96.5  66.7 |
| **24∗** **8∗** **6** | 99  52  75 | **0.5**  **0.65**  0.43 |
| 8∗ **8**  None  **10∗** **86**  21∗ **13**  **6∗**/34∗∗ **7**/21∗ | | 92  92  None  113  128  113  98  82  46 | | **0.56**  **0.48**  None  **0.51**  **0.35**  **0.51**  **0.48**  **0.51**  **0.40** | 100  100  None  100  100  100  82  100  15.8 |
| **15∗** 13  None  **13** | 75  35  None  35 | **0.58**  0.31  None  0.28 |

turalrefinementofthestructuresprovidedfromtheBOTCPmodulecur-rentlytakesabout25hbutcouldbefurtheroptimized.Forinstanceby runningthestructuralrefinementforlessthan1000posesorbyusing asmallernumberofPROTACconformationsperstructure.Whilethe structuralrefinementisstillcomputationallyintensive,weseethatthe resultswithoutstructuralrefinementsuggestthatthecurrentscoresand samplingtechniquecanbebeneficialinassessingtheinteractionsfora giventernarycomplex.

Asmentionedintheintroduction,thecurrentworkpredictsthe ternarycomplexstructuremosteffectivelyunderasetofspecificas-sumptions,i.e.thattheproteinsdonotundergosignificantbackbone conformationalchanges,andthatwehaveinformationregardingthe E3-ligasebinder,thewarheadandtherespectivebindingpockets.Cur-rently,thePROTACstabilityscoreandthePPIscoreareaffectedbecause theside-chainsaremisalignedwiththesidechainsasintheoriginal crystalstructure.Theselimitationsaffectusforinstanceinthecaseof 7KHHand6W8I\_2,wheremisalignedbindersandside-chainsinthein-putimpactperformance.

FutureworkcanincorporatemoreinformativePPIandconstraint scoresintheBOloopwhichwillimprovethefilterationofnon-native poses.Oursimulationsshowthatthenativestructureconsistentlyhas aPROTACstabilityfitnessthatisamongthehighestobservedforthat complex,whichpromisessignificantimprovementsifthisscoreincludes side-chainflexibility(e.g.,usingAutodock-Vina).

Thecurrentapproachtovalidateternarycomplexpredictionishin-deredbytheseverelackofavailabilityofstructuresofternarycom-plexes.Apromisingdirectionofresearchisthustodevelopascorethat isderivedfromtheposessampledbyBOTCP,whichcanbevalidated againstmoreabundantlyavailabledataonthedissociationconstantof aternarycomplex.

Tosummarize,wedemonstrateinthisworkthesuccessfulapplica-tionofBayesianoptimizationaswellasthedesignoftwospecialized scoresthatcaneffectivelysampleandrankhighlyclustersofnear-native posesforTernaryComplexes.

13

*A.Rao,T.M.Tunjic,M.Brunsteineretal.*

**References**

[[1]SchapiraM,CalabreseMF,BullockAN,CrewsCM.Targetedproteindegradation:](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0001)  [expandingthetoolbox.NatRevDrugDiscov2019;18(12):949–63.](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0001)

[[2]BaiN,RichingKM,MakajuA,WuH,AckerTM,OuS-C,etal.ModelingtheCRL4A](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0001) [ligasecomplextopredicttargetproteinubiquitinationinducedbycereblon-recruit-](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0002)[ingPROTACs.JBiolChem2022;298:101653.](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0003)

[[3]](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0004)[SchreiberSL.Theriseofmolecularglues.Cell2021;184(1):3–9.](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0003)

[[4]](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0004)[DongG,DingY,HeS,ShengC.Moleculargluesfortargetedp](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0003)[roteindegradation:](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0004)  [fromserendipitytorationaldiscovery.JMedChem2021;64(15):10606–20.](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0004)

[[5]NeklesaTK,WinklerJD,CrewsCM.TargetedproteindegradationbyPROTACs.](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0004)  [PharmacolTher2017;174:138–44.](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0005)

[[6]PetterssonM,CrewsCM.Proteolysistargetingchimeras(PROTACs)—Past,present](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0005)  [andfuture.DrugDiscovToday2019;31:15–27.](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0006)

[[7]CiulliA,TrainorN.ABeginner’sguidetoPROTACsandtargetedproteindegrada-](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0006) [tion.Biochemist2021;43(5):74–9.](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0007)

[[8]FischerES,BöhmK,LydeardJR,YangH,StadlerMB,CavadiniS,NagelJ,SerlucaF,](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0007) [AckerV,LingarajuGM,etal.StructureoftheDDB1–CRBNE3ubiquitinligasein complexwiththalidomide.Nature2014;512(7512):49–53.](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0008)

[[9]FinkEC,EbertBL.Thenovelmechanismoflenalidomideactivity.Blood,JAmSoc](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0008)  [Hematol2015;126(21):2366–9.](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0010)

[[10]GaoS,WangS,SongY.Novelimmunomodulatorydrugsandneo-substrates.](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0010)  [BiomarkRes2020;8(1):1–8.](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0010)

[[11]BuckleyDL,VanMolleI,GareissPC,TaeHS,MichelJ,NoblinDJ,JorgensenWL,](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0010) [CiulliA,CrewsCM.TargetingthevonHippel–LindauE3ubiquitinligaseus-ingsmallmoleculestodisrupttheVHL/HIF-1](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0011)*[𝛼](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0011)* [interaction.JAmChemSoc](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0011) [2012;134(10):4465–8.](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0012)

[[12]GaldeanoC,GaddMS,SoaresP,Scaffi](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0012)di[S,VanMolleI,BircedI,etal.Struc-ture-guideddesignandoptimizationofsmallmoleculestargetingtheprotein–pro-teininteractionbetweenthevonHippel–Lindau(VHL)E3ubiquitinligaseandthe hypoxiainduciblefactor(HIF)alphasubunitwithinvitronanomolaraffinities.J MedChem2014;57(20):8657–63.](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0012)

[[13]SchiemerJ,HorstR,MengY,MontgomeryJI,XuY,FengX,etal.Snapshotsand](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0012) [ensemblesofBTKandcIAP1proteindegraderternarycomplexes.NatChemBiol 2021;17(2):152–60.](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0013)

[[14]HinesJ,LartigueS,DongH,QianY,CrewsCM.MDM2-recruitingPROTACoffer](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0013)s [superior,synergisticantiproliferativeactivityviasimultaneousdegradationofBRD4 andstabilizationofP53.CancerRes2019;79(1):251–62.](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0014)

[[15]WeiJ,MengF,ParkK-S,YimH,VelezJ,KumarP,etal.HarnessingtheE3ligase](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0014)  [KEAP1fortargetedproteindegradation.JAmChemSoc2021;143(37):15073–83. [16]PetrylakD.P.,GaoX.,VogelzangN.J.,Garfiel](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0015)d[M.H.,TaylorI.,DouganMooreM.,](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0015)  PeckR.A.,BurrisH.A.III.First-in-humanphaseIstudyofARV-110,anandro- genreceptor(AR)PROTACdegraderinpatients(PTS)withmetastaticcastrate- resistantprostatecancer(MCRPC)followingenzalutamide(ENZ)and/orabiraterone [(ABI)2020.](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0016)

[[17]NeklesaT,SnyderLB,WillardRR,VitaleN,PizzanoJ,GordonDA,etal.ARV-110: anoralandrogenreceptorPROTACdegraderforprostatecancer.JClinOncol 2019;37(7):14–16.](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0016)

[[18]FlanaganJ.,QianY.,GoughS.,AndreoliM.,BookbinderM.,CadelinaG.,Bradley](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0016) J.,RousseauE.,WillardR.,PizzanoJ.,etal.AbstractP5-04-18:ARV-471,anoral [estrogenreceptorPROTACdegraderforbreastcancer.2019.](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0017)

[[19]HeY,KochR,BudamaguntaV,ZhangP,ZhangX,KhanS,ThummuriD,OrtizYT, ZhangX,LvD,etal.DT2216–aBcl-xL-specificdegraderishighlyactiveagainst Bcl-xL-dependentTcelllymphomas.JHematolOncol2020;13(1):1–13.](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0017)

[[20]HansenJD,CorreaM,NagyMA,AlexanderM,PlantevinV,GrantV,etal.Discovery](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0017) [ofCRBNE3ligasemodulatorCC-92480forthetreatmentofrelapsedandrefractory multiplemyeloma.JMedChem2020;63(13):6648–76.](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0018)

[[21]ZhouB,HuJ,XuF,ChenZ,BaiL,Fernandez-SalasE,etal.Discoveryofasmal-](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0018)[l-moleculedegraderofbromodomainandextra-terminal(BET)proteinswithpico-molarcellularpotenciesandcapableofachievingtumorregression.JMedChem 2018;61(2):462–81.](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0019)

[[22]ZorbaA,NguyenC,XuY,StarrJ,BorzilleriK,SmithJ,ZhuH,FarleyKA,DingW,](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0019) [SchiemerJ,etal.DelineatingtheroleofcooperativityinthedesignofpotentPRO-TACsforBTK.ProcNatlAcadSci2018;115(31):E7285–92.](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0020)

[[23]DobrovolskyD,WangES,MorrowS,LeahyC,FaustT,NowakRP,DonovanKA,](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0020) [YangG,LiZ,FischerES,etal.Brutontyrosinekinasedegradationasatherapeutic](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0021) [strategyforcancer.Blood,JAmSocHematol2019;133(9):952–61.](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0022)

[[24]MaresA,MiahAH,SmithIE,RackhamM,ThawaniAR,CryanJ,HailePA,VottaBJ, BealAM,CapriottiC,etal.Extendedpharmacodynamicresponsesobservedupon PROTAC-mediateddegradationofRIPK2.CommunBiol2020;3(1):1–13.](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0022)

[[25]Kofin](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0022)k[C,TrainorN,MairB,WöhrleS,WurmM,MischerikowN,BaderG,RumpelK,](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0022) [GerstbergerT,CuiY,etal.AselectiveandorallybioavailableVHL-recruitingPRO-TACachievesSMARCA2degradationinvivo.ChemRxiv2022.doi:10.26434/chem-rxiv-2022-q63s3.](https://doi.org/10.26434/chemrxiv-2022-q63s3)

[[26]HughesSJ,CiulliA.Molecularrecognitionofternarycomplexes:anewdi-](https://doi.org/10.26434/chemrxiv-2022-q63s3)[mensioninthestructure-guideddesignofchemicaldegraders.EssaysBiochem 2017;61(5):505–16.](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0024)

[[27]BemisTA,LaClairJJ,BurkartMD.Unravelingtheroleoflinkerdesigninproteolysis](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0024)  [targetingchimeras:miniperspective.JMedChem2021;64(12):8042–52.](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0025)

[[28]GaddMS,TestaA,LucasX,ChanK-H,ChenW,LamontDJ,etal.Structuralbasisof](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0025) [PROTACcooperativerecognitionforselectiveproteindegradation.NatChemBiol 2017;13(5):514–21.](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0026)

[[29]TroupRI,FallanC,BaudMG.CurrentstrategiesforthedesignofPROTAClinkers:](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0026)  [acriticalreview.ExplorTargetAnti-TumorTher2020;1:273–312.](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0027)

[[30]DrummondML,WilliamsCI.InsilicomodelingofPROTAC-mediatedternarycom-](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0027) [plexes:validationandapplication.JChemInfModel2019;59(4):1634–44.](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0028)

*ArtificialIntelligenceintheLifeSciences3(2023)100072*

[31][DrummondML,HenryA,LiH,WilliamsCI.ImprovedaccuracyformodelingPRO-TAC-mediatedternarycomplexformationandtargetedproteindegradationvianew insilicomethodologies.JChemInfModel2020;60(10):5234–54.](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0029)

[32][ZaidmanD,PriluskyJ,LondonN.Prosettac:rosettabasedmodelingofPROTAC](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0029)  [mediatedternarycomplexes.JChemInfModel2020;60(10):4894–903.](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0030)

[33][WengG,LiD,KangY,HouT.IntegrativemodelingofPROTAC-mediatedternary](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0030)  [complexes.JMedChem2021;64(21):16271–81.](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0032)

[34][MarzeNA,RoyBurmanSS,Sheffl](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0032)er[W,GrayJJ.Efficie](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0032)nt[flexibl](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0032)e[backbonepro-](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0032) [tein–proteindockingforchallengingtargets.Bioinformatics2018;34(20):3461–9.](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0032) [35][LiW,ZhangJ,GuoL,WangQ.Importanceofthree–bodyproblemsandprotein–pro-](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0032) [teininteractionsinproteolysis-targetingchimeramodeling:insightsfrommolecular](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0033)  [dynamicssimulations.JChemInfModel2022;25(3):523–32.](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0033)

[36][LiaoJ,NieX,UnartaIC,EricksenSS,TangW.Insilicomodelingandscoringof](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0033) [PROTAC-mediatedternarycomplexposes.JMedChem2022;65(8):6116–32.](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0034)

[37][BrochuE.,CoraV.M.,DeFreitasN..AtutorialonBayesianoptimizationofexpensive](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0034) costfunctions,[withapplicationtoactiveuser](http://arxiv.org/abs/10122599)modelingandhierarchicalreinforce-[mentlearning.arXivpreprintarXiv:101225992010.](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0035)

[38][DominguezC,BoelensR,BonvinAM.Haddock:aprotein-proteindocking approachbasedonbiochemicalorbiophysicalinformation.JAmChemSoc 2003;125(7):1731–7.](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0035)

[39][BermanHM.Theproteindatabank.NucleicAcidsRes2000;28(1):235–42.](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0035)  doi[:10.1093/nar/28.1.235.](https://doi.org/10.1093/nar/28.1.235)

[40]Mc[NuttAT,FrancoeurP,](https://doi.org/10.1093/nar/28.1.235)AggarwalR,MasudaT,MeliR,RagozaM,etal.GN-INA[1.0:moleculardockingwith](https://doi.org/10.1186/s13321-021-00522-2)deeplearning.JCheminform2021;13(1):1–20. doi[:10.1186/s13321-021-00522-2.](https://doi.org/10.1186/s13321-021-00522-2)

[41]Jum[perJ,EvansR,PritzelA,Gr](https://doi.org/10.1186/s13321-021-00522-2)eenT,FigurnovM,RonnebergerO,Tunyasuvu-nakoolK,BatesR,ŽídekA,PotapenkoA,BridglandA,MeyerC,KohlSAA,Bal-lardAJ,CowieA,Romera-ParedesB,NikolovS,JainR,AdlerJ,BackT,PetersenS, ReimanD,ClancyE,ZielinskiM,SteineggerM,PacholskaM,BerghammerT,Boden-steinS,SilverD,VinyalsO,SeniorAW,KavukcuogluK,KohliP,HassabisD.Highly accura[teproteinstructureprediction](https://doi.org/10.1038/s41586-021-03819-2)withAlphaFold.Nature2021;596(7873):583–9.doi[:10.1038/s41586-021-03819-2.](https://doi.org/10.1038/s41586-021-03819-2)

[42]Varad[iM,AnyangoS,Deshpande](https://doi.org/10.1038/s41586-021-03819-2)M,NairS,NatassiaC,YordanovaG,etal.

AlphaFoldproteinstructuredatabase:massivelyexpandingthestructuralcov-erageofprotein-sequence[spacewithhigh-accura](https://doi.org/10.1093/nar/gkab1061)cymodels.NucleicAcidsRes 2022;50(D1):D439–44.doi:[10.1093/nar/gkab1061.](https://doi.org/10.1093/nar/gkab1061)

[43]PettersenEF,GoddardTD,[HuangCC,CouchGS,](https://doi.org/10.1093/nar/gkab1061)GreenblattDM,MengEC,etal.

UCSFChimera—Avisualizationsyste[mforexploratoryres](https://doi.org/10.1002/jcc.20084)earchandanalysis.JCom-putChem2004;25(13):1605[–12.doi](https://github.com/pandegroup/pdbfixer)[:10.1002/jcc.20084.](https://doi.org/10.1002/jcc.20084)

[44]EastmanP..PDBFixer.2019.[https://](https://github.com/pandegroup/pdbfixer)[github.com/pandeg](https://doi.org/10.1002/jcc.20084)[roup/pdbfixe](https://github.com/pandegroup/pdbfixer)r[.](https://github.com/pandegroup/pdbfixer)

[45]WilliamsCJ,HeaddJJ,Mori[artyNW,PrisantMG,VideauLL,DeisLN](https://github.com/pandegroup/pdbfixer),etal.Mol-Probity:moreandbetterreferenced[ataforimprovedal](https://doi.org/10.1002/pro.3330)l-atomstructurevalidation. ProteinSci2018;27(1):293–315.doi[:10.1002/pro.3330.](https://doi.org/10.1002/pro.3330)

[46]O’BoyleNM,BanckM,James[CA,MorleyC,V](https://doi.org/10.1002/pro.3330)andermeerschT,Hutchi-son[GR.Openbabel:anop](https://doi.org/10.1186/1758-2946-3-33)enchemicaltoolbox.JCheminform2011;3(1):33. [doi](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0043)[:10.1186/1758-2946-3-33.](https://doi.org/10.1186/1758-2946-3-33)

[47][Zha](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0043)[ngC,LiuS,ZhuQ,](https://doi.org/10.1186/1758-2946-3-33)[ZhouY.Aknowledge-basedenergyfunctionfor protein-ligand,protein-protein,andprotein-dnacomplexes.JMedChem 2005;48(7):2325–35.](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0043)

[48][Rappé AK,CasewitCJ,ColwellK,GoddardWAIII,Skiff WM](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0043).[Uff](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0043),[afullperiodic](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0043) [tableforcefieldformolecularmechanicsandmoleculardynamicssimulations.J AmChemSoc1992;114(25):10024–35.](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0044)

[49][LandrumG..Rdkit:open-sourcecheminformaticssoftware2016;https://github.](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0044)  [com/rdkit/rdkit/releases/tag/Release\_2016\_09\_4.](https://github.com/rdkit/rdkit/releases/tag/Release_2016_09_4)

[50][NelderJA,MeadR.Asimplexmethodforfunctionminimization.ComputJ](https://github.com/rdkit/rdkit/releases/tag/Release_2016_09_4)  [1965;7(4):308–13.doi:10.1093/comjnl/7.4.308.](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0046)

[51][MockusJ,TiesisV,ZilinskasA.TheapplicationofBayesianmethodsforseekingthe](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0046)  [extremum.TowardsGlobOptim1978;2:117–29.](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0047)

[52][AgnihotriA,BatraN.ExploringBayesianoptimization.Distill2020;5(5):e26.](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0047) [53][RasmussenCE.Gaussianprocessesinmachinelearning.In:Summerschoolo](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0047)[nma-](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0048) [chinelearning.Springer;2003.p.63–71.](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0048)

[54][KanagawaM.,HennigP.,SejdinovicD.,SriperumbudurB.K..Gaussianprocesses](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0048) [andkernelmethods:areviewonconnectionsandequivalences.arXivpreprint](http://arxiv.org/abs/180702582) [arXiv:1807025822018.](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0049)

[55][AuerP.Usingconfidenc](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0049)e[boundsforexploitation-explorationtrade-offs](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0049).[JMach](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0049)  [LearnRes2002;3(Nov):397–422.](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0049)

[56][KingmaD.P.,BaJ..Adam:amethodforstochasticoptimization.arXivpreprint](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0049)  [arXiv:141269802014.](http://arxiv.org/abs/14126980)

[57][BalandatM,KarrerB,JiangDR,DaultonS,LethamB,WilsonAG,BakshyE.BoTorch:](http://arxiv.org/abs/14126980)  aframeworkforefficientMonte–CarloBayesi[anoptimization.Advancesinneu](http://arxiv.org/abs/1910.06403)ral [informationprocessingsystems,vol33;2020.](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0051)<http://arxiv.org/abs/1910.06403>[58][KirkpatrickS,GelattJrCD,VecchiMP.Optim](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0051)[izationbysimulatedannealing.S](http://arxiv.org/abs/1910.06403)[ci-](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0051) [ence1983;220(4598):671–80.](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0051)

[59][EberhardtJ,Santos-MartinsD,TillackAF,ForliS.Autodockvina1.2.0:new](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0051) [dockingmethods,expandedforcefield,andpythonbindings.JChemInfModel 2021;61(8):3891–8.](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0052)

[60][NilgesM.Acalculationstrategyforthestructuredeterminationofsymmetricdemers](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0052)  [by1HNMR.Proteins1993;17(3):297–309.](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0053)

[61][MaierJA,MartinezC,KasavajhalaK,WickstromL,HauserKE,SimmerlingC.ff14SB](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0053): improvingtheaccuracyofproteinsidechainandba[ckboneparametersfromf](https://doi.org/10.1021/acs.jctc.5b00255)f99SB. JChemTheoryComput2015;11(8):3696–713.doi[:10.1021/acs.jctc.5b00255.](https://doi.org/10.1021/acs.jctc.5b00255)

[62]WangJ,WolfRM,CaldwellJW,KollmanPA[,CaseDA.Development](https://doi.org/10.1021/acs.jctc.5b00255)and test[ingofagenerala](https://doi.org/10.1002/jcc.20035)mberforcefield.JComputChem2004;25(9):1157–74. doi[:10.1002/jcc.20035.](https://doi.org/10.1002/jcc.20035)

[63]Jak[alianA,JackDB,](https://doi.org/10.1002/jcc.20035)BaylyCI.Fast,efficientgenerationofhigh-qualityatomic charges.AM1-BCCmodel[:II.Parameterizati](https://doi.org/10.1002/jcc.10128)onandvalidation.JComputChem 2002;23(16):1623–41.doi[:10.1002/jcc.10128.](https://doi.org/10.1002/jcc.10128)

14

*A.Rao,T.M.Tunjic,M.Brunsteineretal.*

[64]CaseD.,AktulgaH.,BelfonK.,Ben-ShalomI.,BrozellS.,CeruttiD.S.,CheathamT.E.

III,CisnerosG.,CruzeiroV.,DardenT.,DukeR.,GiambasuG.,GilsonM.,Gohlke H.,GoetzA.,HarrisR.,IzadiS.,IzmailovS.,JinC.,KaK..Amber.2021.

[65]EastmanP,SwailsJ,ChoderaJD,McGibbonRT,ZhaoY,BeauchampKA, [etal.OpenMM7:rapiddevelopmentofhighperformancealgorithmsfor moleculardynamics.PLoSComputBiol2017;13(7):e1005659.doi:10.1371/jour-nal.pcbi.1005659.](https://doi.org/10.1371/journal.pcbi.1005659)

[[66]NguyenH,RoeDR,SimmerlingC.Improvedgeneralizedbornsolventmodel](https://doi.org/10.1371/journal.pcbi.1005659) par[ametersforprotein](https://doi.org/10.1021/ct3010485)simulations.JChemTheoryComput2013;9(4):2020–34. doi[:10.1021/ct3010485.](https://doi.org/10.1021/ct3010485)

[67]Ekb[ergV,RydeU.O](https://doi.org/10.1021/ct3010485)ntheuseofinteractionentropyandrelatedmethods to[estimatebindingentropi](https://doi.org/10.1021/acs.jctc.1c00374)es.JChemTheoryComput2021;17(8):5379–91. doi[:10.1021/acs.jctc.1c00374.](https://doi.org/10.1021/acs.jctc.1c00374)

*ArtificialIntelligenceintheLifeSciences3(2023)100072*

[68][RodriguesJPGLM,TrelletM,SchmitzC,KastritisP,KaracaE,MelquiondASJ, etal.Clusteringbiomolecularcomplexesbyresiduecontactssimilarity.Proteins 2012;80(7):1810–17.](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0060)

[69][BasuS,WallnerB.Dockq:aqualitymeasureforprotein-proteindockingmodels.](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0060)  [PLoSOne2016;11(8):e0161879.](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0061)

[70][LensinkMF,NadzirinN,VelankarS,WodakSJ.Modelingprotein-protein,protein-](http://refhub.elsevier.com/S2667-3185(23)00016-8/sbref0061)peptide,andprotein-o[ligosaccharidecomp](https://doi.org/10.1002/prot.25870)lexes:CAPRI7thedition.Proteins 2020;88(8):916–38.doi[:10.1002/prot.25870.](https://doi.org/10.1002/prot.25870)

15