

= Rosetta @ home =

Rosetta @ home is a distributed computing project for protein structure prediction on the Berkeley Open Infrastructure for Network Computing (BOINC) platform , run by the Baker laboratory at the University of Washington . Rosetta @ home aims to predict protein ? protein docking and design new proteins with the help of about sixty thousand active volunteered computers processing at over 110 teraFLOPS on average as of November 20 , 2015 . Foldit , a Rosetta @ Home videogame , aims to reach these goals with a crowdsourcing approach . Though much of the project is oriented towards basic research on improving the accuracy and robustness of the proteomics methods , Rosetta @ home also does applied research on malaria , Alzheimer 's disease and other pathologies .

Like all BOINC projects , Rosetta @ home uses idle computer processing resources from volunteers ' computers to perform calculations on individual workunits . Completed results are sent to a central project server where they are validated and assimilated into project databases . The project is cross @-@ platform , and runs on a wide variety of hardware configurations . Users can view the progress of their individual protein structure prediction on the Rosetta @ home screensaver .

In addition to disease @-@ related research , the Rosetta @ home network serves as a testing framework for new methods in structural bioinformatics . These new methods are then used in other Rosetta @-@ based applications , like RosettaDock and the Human Proteome Folding Project , after being sufficiently developed and proven stable on Rosetta @ home 's large and diverse collection of volunteer computers . Two particularly important tests for the new methods developed in Rosetta @ home are the Critical Assessment of Techniques for Protein Structure Prediction (CASP) and Critical Assessment of Prediction of Interactions (CAPRI) experiments , biannual experiments which evaluate the state of the art in protein structure prediction and protein ? protein docking prediction , respectively . Rosetta @ home consistently ranks among the foremost docking predictors , and is one of the best tertiary structure predictors available .

= = Computing platform = =

Both the Rosetta @ home application and the BOINC distributed computing platform are available for the Microsoft Windows , Linux and Macintosh platforms (BOINC also runs on several other platforms , e.g. FreeBSD) . Participation in Rosetta @ home requires a central processing unit (CPU) with a clock speed of at least 500 MHz , 200 megabytes of free disk space , 512 megabytes of physical memory , and Internet connectivity . As of July 20 , 2016 , the current version of the Rosetta Mini application is 3 @.@ 73 . The current recommended BOINC program version is 7 @.@ 6 @.@ 22 . Standard HTTP (port 80) is used for communication between the user 's BOINC client and the Rosetta @ home servers at the University of Washington ; HTTPS (port 443) is used during password exchange . Remote and local control of the BOINC client use port 31416 and port 1043 , which might need to be specifically unblocked if they are behind a firewall . Workunits containing data on individual proteins are distributed from servers located in the Baker lab at the University of Washington to volunteers ' computers , which then calculate a structure prediction for the assigned protein . To avoid duplicate structure predictions on a given protein , each workunit is initialized with a random number seed . This gives each prediction a unique trajectory of descent along the protein 's energy landscape . Protein structure predictions from Rosetta @ home are approximations of a global minimum in a given protein 's energy landscape . That global minimum represents the most energetically favorable conformation of the protein , i.e. its native state .

A primary feature of the Rosetta @ home graphical user interface (GUI) is a screensaver which shows a current workunit 's progress during the simulated protein folding process . In the upper @-@ left of the current screensaver , the target protein is shown adopting different shapes (conformations) in its search for the lowest energy structure . Depicted immediately to the right is the structure of the most recently accepted . On the upper right the lowest energy conformation of the current decoy is shown ; below that is the true , or native , structure of the protein if it has already

been determined . Three graphs are included in the screensaver . Near the middle , a graph for the accepted model 's free energy is displayed , which fluctuates as the accepted model changes . A graph of the accepted model 's root mean square deviation (RMSD) , which measures how structurally similar the accepted model is to the native model , is shown far right . On the right of the accepted energy graph and below the RMSD graph , the results from these two functions are used to produce an energy vs. RMSD plot as the model is progressively refined .

Like all BOINC projects , Rosetta @ home runs in the background of the user 's computer using idle computer power , either at or before logging into an account on the host operating system . Rosetta @ home frees resources from the CPU as they are required by other applications so that normal computer usage is unaffected . To minimize power consumption or heat production from a computer running at sustained capacity , the maximum percentage of CPU resources that Rosetta @ home is allowed to use can be specified through a user 's account preferences . The times of day during which Rosetta @ home is allowed to do work can also be adjusted , along with many other preferences , through a user 's account settings .

Rosetta , the software that runs on the Rosetta @ home network , was rewritten in C + + to allow easier development than that offered by its original version , which was written in Fortran . This new version is object @-@ oriented , and was released on February 8 , 2008 . Development of the Rosetta code is done by Rosetta Commons . The software is freely licensed to the academic community and available to pharmaceutical companies for a fee .

= = Project significance = =

With the proliferation of genome sequencing projects , scientists can infer the amino acid sequence , or primary structure , of many proteins that carry out functions within the cell . To better understand a protein 's function and aid in rational drug design , scientists need to know the protein 's three @-@ dimensional tertiary structure .

Protein 3D structures are currently determined experimentally through X @-@ ray crystallography or nuclear magnetic resonance (NMR) spectroscopy . The process is slow (it can take weeks or even months to figure out how to crystallize a protein for the first time) and comes at high cost (around \$ 100 @,@ 000 USD per protein) . Unfortunately , the rate at which new sequences are discovered far exceeds the rate of structure determination ? out of more than 7 @,@ 400 @,@ 000 protein sequences available in the NCBI non @-@ redundant (nr) protein database , fewer than 52 @,@ 000 proteins ' 3D structures have been solved and deposited in the Protein Data Bank , the main repository for structural information on proteins . One of the main goals of Rosetta @ home is to predict protein structures with the same accuracy as existing methods , but in a way that requires significantly less time and money . Rosetta @ home also develops methods to determine the structure and docking of membrane proteins (e.g. , GPCRs) , which are exceptionally difficult to analyze with traditional techniques like X @-@ ray crystallography and NMR spectroscopy , yet represent the majority of targets for modern drugs .

Progress in protein structure prediction is evaluated in the biannual Critical Assessment of Techniques for Protein Structure Prediction (CASP) experiment , in which researchers from around the world attempt to derive a protein 's structure from the protein 's amino acid sequence . High scoring groups in this sometimes competitive experiment are considered the de facto standard @-@ bearers for what is the state of the art in protein structure prediction . Rosetta , the program on which Rosetta @ home is based , has been used since CASP5 in 2002 . In the 2004 CASP6 experiment , Rosetta made history by being the first to produce a close to atomic @-@ level resolution , ab initio protein structure prediction in its submitted model for CASP target T0281 . Ab initio modeling is considered an especially difficult category of protein structure prediction , as it does not use information from structural homology and must rely on information from sequence homology and modeling physical interactions within the protein . Rosetta @ home has been used in CASP since 2006 , where it was among the top predictors in every category of structure prediction in CASP7 . These high quality predictions were enabled by the computing power made available by Rosetta @ home volunteers . Increasing computational power allows Rosetta @ home to sample

more regions of conformation space (the possible shapes a protein can assume) , which , according to Levinthal 's paradox , is predicted to increase exponentially with protein length .

Rosetta @ home is also used in protein docking prediction , which determines the structure of multiple complexed proteins , or quaternary structure . This type of protein interaction affects many cellular functions , including antigen ? antibody and enzyme ? inhibitor binding and cellular import and export . Determining these interactions is critical for drug design . Rosetta is used in the Critical Assessment of Prediction of Interactions (CAPRI) experiment , which evaluates the state of the protein docking field similar to how CASP gauges progress in protein structure prediction . The computing power made available by Rosetta @ home 's project volunteers has been cited as a major factor in Rosetta 's performance in CAPRI , where its docking predictions have been among the most accurate and complete .

In early 2008 , Rosetta was used to computationally design a protein with a function never before observed in nature . This was inspired in part by the retraction of a high @-@ profile paper from 2004 which originally described the computational design of a protein with improved enzymatic activity compared to its natural form . The 2008 research paper from David Baker 's group describing how the protein was made , which cited Rosetta @ home for the computational resources it made available , represented an important proof of concept for this protein design method . This type of protein design could have future applications in drug discovery , green chemistry , and bioremediation .

= = Disease @-@ related research = =

In addition to basic research in predicting protein structure , docking and design , Rosetta @ home is also used in immediate disease @-@ related research . Numerous minor research projects are described in David Baker 's Rosetta @ home journal . As of February 2014 , information on recent publications and a short description of the work are being updated on the forum .

= = = Alzheimer 's disease = = =

A component of the Rosetta software suite , RosettaDesign , was used to accurately predict which regions of amyloidogenic proteins were most likely to make amyloid @-@ like fibrils . By taking hexapeptides (six amino acid @-@ long fragments) of a protein of interest and selecting the lowest energy match to a structure similar to that of a known fibril forming hexapeptide , RosettaDesign was able to identify peptides twice as likely to form fibrils as are random proteins . Rosetta @ home was used in the same study to predict structures for amyloid beta , a fibril @-@ forming protein that has been postulated to cause Alzheimer 's disease . Preliminary but as yet unpublished results have been produced on Rosetta @-@ designed proteins that may prevent fibrils from forming , although it is unknown whether it can prevent the disease .

= = = Anthrax = = =

Another component of Rosetta , RosettaDock , was used in conjunction with experimental methods to model interactions between three proteins ? lethal factor (LF) , edema factor (EF) and protective antigen (PA) ? that make up anthrax toxin . The computational model accurately predicted docking between LF and PA , helping to establish which domains of the respective proteins are involved in the LF ? PA complex . This insight was eventually used in research resulting in improved anthrax vaccines .

= = = Herpes simplex virus 1 = = =

RosettaDock was used to model docking between an antibody (immunoglobulin G) and a surface protein expressed by the cold sore virus , herpes simplex virus 1 (HSV @-@ 1) which serves to degrade the antiviral antibody . The protein complex predicted by RosettaDock closely agreed with

the particularly difficult @-@ to @-@ obtain experimental models , leading researchers to conclude that the docking method has potential in addressing some of the problems that X @-@ ray crystallography has with modeling protein ? protein interfaces .

== = HIV == =

As part of research funded by a \$ 19 @.@ 4 million grant by the Bill & Melinda Gates Foundation , Rosetta @ home has been used in designing multiple possible vaccines for human immunodeficiency virus (HIV) .

== = Malaria == =

In research involved with the Grand Challenges in Global Health initiative , Rosetta has been used to computationally design novel homing endonuclease proteins , which could eradicate *Anopheles gambiae* or otherwise render the mosquito unable to transmit malaria . Being able to model and alter protein ? DNA interactions specifically , like those of homing endonucleases , gives computational protein design methods like Rosetta an important role in gene therapy (which includes possible cancer treatments) .

== Development history and branches ==

Originally introduced by the Baker laboratory in 1998 as an ab initio approach to structure prediction , Rosetta has since branched into several development streams and distinct services . The Rosetta platform derives its name from the Rosetta Stone , as it attempts to decipher the structural " meaning " of proteins ' amino acid sequences . More than seven years after Rosetta 's first appearance , the Rosetta @ home project was released (i.e. announced as no longer beta) on October 6 , 2005 . Many of the graduate students and other researchers involved in Rosetta 's initial development have since moved to other universities and research institutions , and subsequently enhanced different parts of the Rosetta project .

== = RosettaDesign == =

RosettaDesign , a computational approach to protein design based on Rosetta , began in 2000 with a study in redesigning the folding pathway of Protein G. In 2002 RosettaDesign was used to design Top7 , a 93 @-@ amino acid long ? / ? protein that had an overall fold never before recorded in nature . This new conformation was predicted by Rosetta to within 1 @.@ 2 Å RMSD of the structure determined by X @-@ ray crystallography , representing an unusually accurate structure prediction . Rosetta and RosettaDesign earned widespread recognition by being the first to design and accurately predict the structure of a novel protein of such length , as reflected by the 2002 paper describing the dual approach prompting two positive letters in the journal Science , and being cited by more than 240 other scientific articles . The visible product of that research , Top7 , was featured as the RCSB PDB 's ' Molecule of the Month ' in October 2006 ; a superposition of the respective cores (residues 60 ? 79) of its predicted and X @-@ ray crystal structures are featured in the Rosetta @ home logo .

Brian Kuhlman , a former postdoctoral associate in David Baker 's lab and now an associate professor at the University of North Carolina , Chapel Hill , offers RosettaDesign as an online service .

== = RosettaDock == =

RosettaDock was added to the Rosetta software suite during the first CAPRI experiment in 2002 as the Baker laboratory 's algorithm for protein ? protein docking prediction . In that experiment , RosettaDock made a high @-@ accuracy prediction for the docking between streptococcal pyogenic

exotoxin A and a T cell α - β receptor ? α - β chain , and a medium accuracy prediction for a complex between porcine ? α - β amylase and a camelid antibody . While the RosettaDock method only made two acceptably accurate predictions out of seven possible , this was enough to rank it seventh out of nineteen prediction methods in the first CAPRI assessment .

Development of RosettaDock diverged into two branches for subsequent CAPRI rounds as Jeffrey Gray , who laid the groundwork for RosettaDock while at the University of Washington , continued working on the method in his new position at Johns Hopkins University . Members of the Baker laboratory further developed RosettaDock in Gray 's absence . The two versions differed slightly in side α - β chain modeling , decoy selection and other areas . Despite these differences , both the Baker and Gray methods performed well in the second CAPRI assessment , placing fifth and seventh respectively out of 30 predictor groups . Jeffrey Gray 's RosettaDock server is available as a free docking prediction service for non α - β commercial use .

In October 2006 , RosettaDock was integrated into Rosetta @ home . The method used a fast , crude docking model phase using only the protein backbone . This was followed by a slow full α - β atom refinement phase in which the orientation of the two interacting proteins relative to each other , and side α - β chain interactions at the protein ? protein interface , were simultaneously optimized to find the lowest energy conformation . The vastly increased computational power afforded by the Rosetta @ home network , in combination with revised " fold α - β tree " representations for backbone flexibility and loop modeling , made RosettaDock sixth out of 63 prediction groups in the third CAPRI assessment .

=== Robetta ===

The Robetta server is an automated protein structure prediction service offered by the Baker laboratory for non α - β commercial ab initio and comparative modeling . It has participated as an automated prediction server in the biannual CASP experiments since CASP5 in 2002 , performing among the best in the automated server prediction category . Robetta has since competed in CASP6 and 7 , where it did better than average among both automated server and human predictor groups .

In modeling protein structure as of CASP6 , Robetta first searches for structural homologs using BLAST , PSI α - β BLAST , and 3D α - β Jury , then parses the target sequence into its individual domains , or independently folding units of proteins , by matching the sequence to structural families in the Pfam database . Domains with structural homologs then follow a " template α - β based model " (i.e. , homology modeling) protocol . Here , the Baker laboratory 's in α - β house alignment program , K * sync , produces a group of sequence homologs , and each of these is modeled by the Rosetta de novo method to produce a decoy (possible structure) . The final structure prediction is selected by taking the lowest energy model as determined by a low α - β resolution Rosetta energy function . For domains that have no detected structural homologs , a de novo protocol is followed in which the lowest energy model from a set of generated decoys is selected as the final prediction . These domain predictions are then connected together to investigate inter α - β domain , tertiary α - β level interactions within the protein . Finally , side α - β chain contributions are modeled using a protocol for Monte Carlo conformational search .

In CASP8 , Robetta was augmented to use Rosetta 's high resolution all α - β atom refinement method , the absence of which was cited as the main cause for Robetta being less accurate than the Rosetta @ home network in CASP7 .

=== Foldit ===

On May 9 , 2008 , after Rosetta @ home users suggested an interactive version of the distributed computing program , the Baker lab publicly released Foldit , an online protein structure prediction game based on the Rosetta platform . As of September 25 , 2008 , Foldit has over 59 @, @ 000 registered users . The game gives users a set of controls (e.g. " shake " , " wiggle " , " rebuild ") to manipulate the backbone and amino acid side chains of the target protein into more energetically

favorable conformations . Users can work on solutions individually as " soloists " or collectively as " evolvers " , accruing points under either category as they improve their structure predictions . Users can also individually compete with other users through a " duel " feature , in which the player with the lowest energy structure after 20 moves wins .

= = Comparison to similar distributed computing projects = =

There are several distributed computed projects which have study areas similar to those of Rosetta @ home , but differ in their research approach :

= = = Folding @ home = = =

Of all the major distributed computing projects involved in protein research , Folding @ home is the only one not to use the BOINC platform . Both Rosetta @ home and Folding @ home study protein misfolding diseases such as Alzheimer 's disease , but Folding @ home does so much more exclusively . Folding @ home almost exclusively uses all @-@ atom molecular dynamics models to understand how and why proteins fold (or potentially misfold , and subsequently aggregate to cause diseases) . In other words , Folding @ home 's strength is modeling the process of protein folding , while Rosetta @ home 's strength is computational protein design and prediction of protein structure and docking .

Some of Rosetta @ home 's results are used as the basis for some Folding @ home projects . Rosetta provides the most likely structure , but it is not definite if that is the form the molecule takes or whether or not it is viable . Folding @ home can then be used to verify Rosetta @ home 's results , but can also provide additional atomic @-@ level information , as well as details into how the molecule changes shape .

The two projects also differ significantly in their computing power and host diversity . Averaging about 6 @, @ 650 teraFLOPS from a host base of CPUs , GPUs and PS3s , Folding @ home has nearly 108 times more computing power than Rosetta @ home .

= = = World Community Grid = = =

Both Phase I and Phase II of the Human Proteome Folding Project (HPF) , a subproject of World Community Grid , have used the Rosetta program to make structural and functional annotations of various genomes . Although he now uses it to create databases for biologists , Richard Bonneau , head scientist of the Human Proteome Folding Project , was active in the original development of Rosetta at David Baker 's laboratory while obtaining his PhD . More information on the relationship between the HPF1 , HPF2 and Rosetta @ home can be found on Richard Bonneau 's website .

= = = Predictor @ home = = =

Like Rosetta @ home , Predictor @ home specialized in protein structure prediction . While Rosetta @ home uses the Rosetta program for its structure prediction , Predictor @ home used the dTASSER methodology . In 2009 , Predictor @ home shut down .

Other protein related distributed computing projects on BOINC include QMC @ home , Docking @ home , POEM @ home , SIMAP , and TANPAKU . RALPH @ home , the Rosetta @ home alpha project which tests new application versions , work units , and updates before they move on to Rosetta @ home , runs on BOINC as well .

= = Volunteer contributions = =

Rosetta @ home depends on computing power donated by individual project members for its research . As of April 18 , 2014 , about 26 @, @ 900 users from 150 countries were active members of Rosetta @ home , together contributing idle processor time from about 66 @, @ 000 computers

for a combined average performance of over 83 teraFLOPS .

Users are granted BOINC credits as a measure of their contribution . The credit granted for each workunit is the number of decoys produced for that workunit multiplied by the average claimed credit for the decoys submitted by all computer hosts for that workunit . This custom system was designed to address significant differences between credit granted to users with the standard BOINC client and an optimized BOINC client , and credit differences between users running Rosetta @ home on Windows and Linux operating systems . The amount of credit granted per second of CPU work is lower for Rosetta @ home than most other BOINC projects . Rosetta @ home is thirteenth out of over 40 BOINC projects in terms of total credit .

Rosetta @ home users who predict protein structures submitted for the CASP experiment are acknowledged in scientific publications regarding their results . Users who predict the lowest energy structure for a given workunit are featured on the Rosetta @ home homepage as ' Predictor of the Day ' , along with any team of which they are a member . A ' User of the Day ' is chosen at random each day to be on the homepage as well from users who have made a Rosetta @ home profile .