

# Welcome to iProtein – the proteome in the sky!

iProtein provides access to Eidogen-Sertanty's Target Informatics Platform ([TIP](#)) - the world's first protein structure informatics system and knowledgebase enabling structural [interrogation](#) of the druggable genome.

TIP amplifies the rapidly expanding body of experimental protein structural information found in the Protein Data Bank ([PDB](#)) by generating high accuracy protein structural models across sequences found in Swiss-Prot, RefSeq, Ensembl, IPI, etc. TIP leverages protein structural information across family members and corresponding crystal structure templates using an [algorithm](#) that has been validated in several top pharmaceutical companies and academic research sites. As such, TIP contains the world's largest repository of protein structures and models.

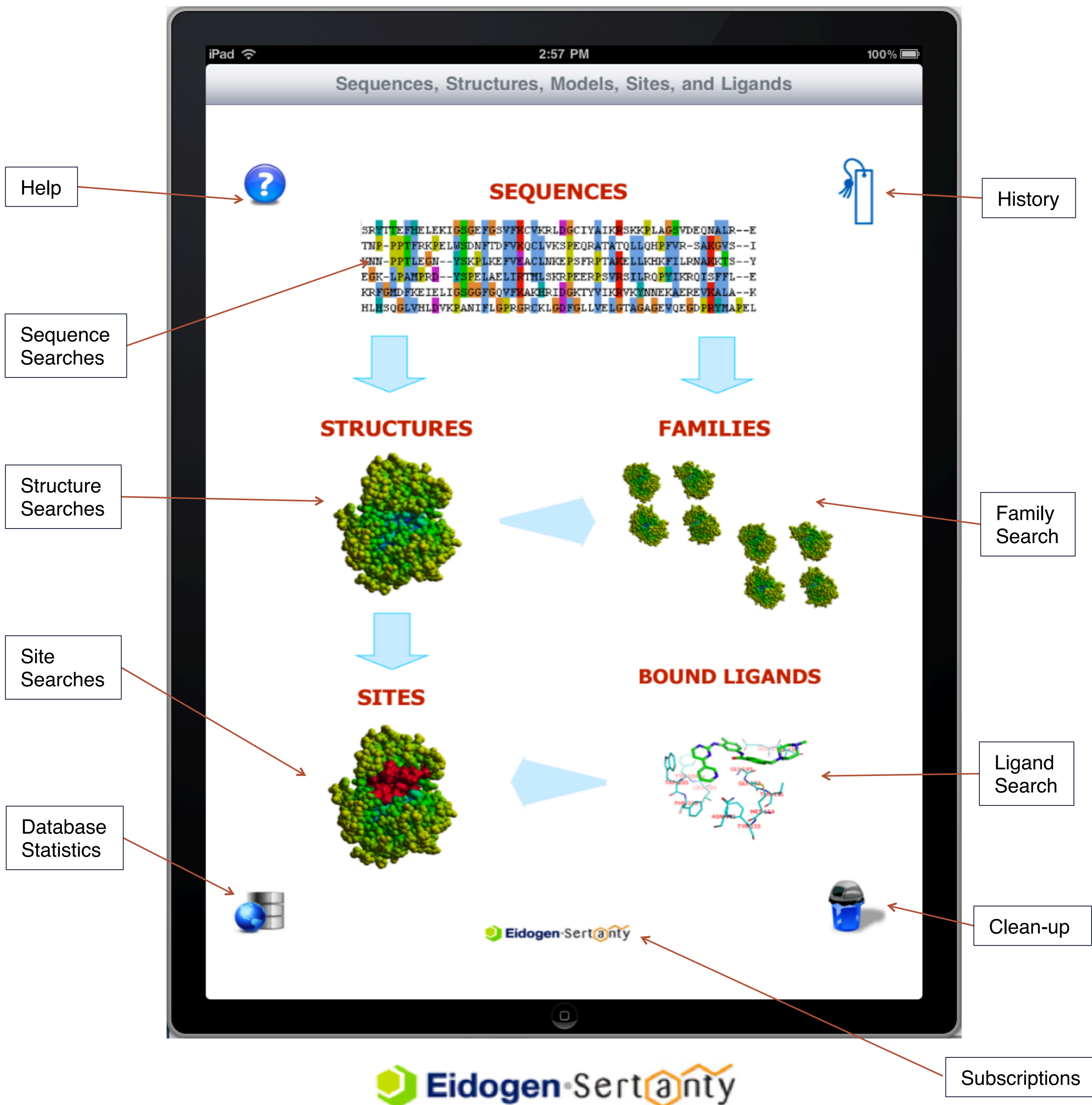
## [The TIP Database](#)

The Target Informatics Platform's content includes high resolution protein structures with reliably annotated small molecule binding sites, including >200,000 chains and comparative models of human proteins, covering every major drug target family including proteases, kinases, phosphatases, phosphodiesterases, nuclear receptors, and GPCRs.

In addition to TIP's high quality structural annotation, it is the only database of its kind that stores all similarity relationships between every sequence, structure, and binding site, making it an incredibly powerful system for structural and comparative proteomics.

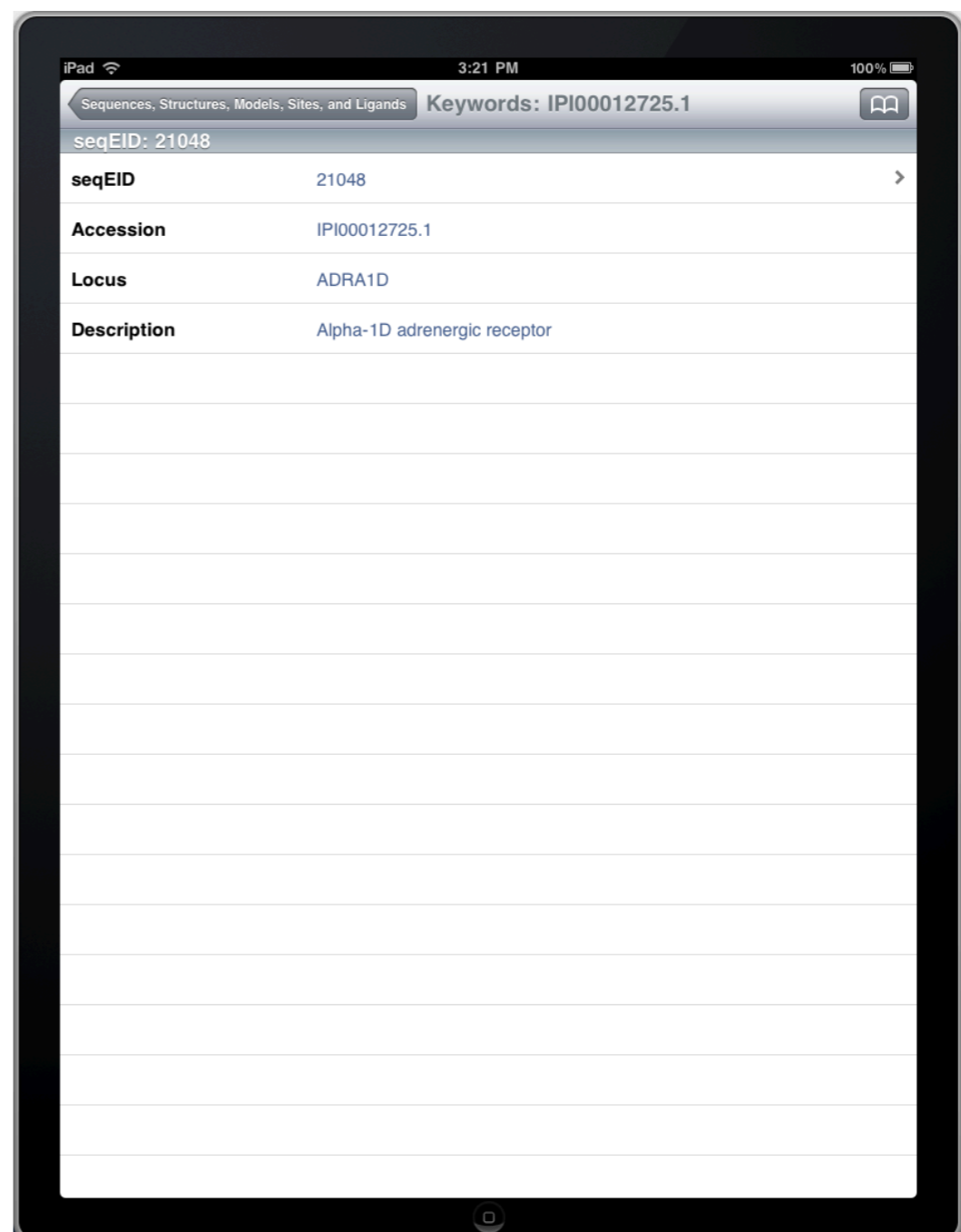
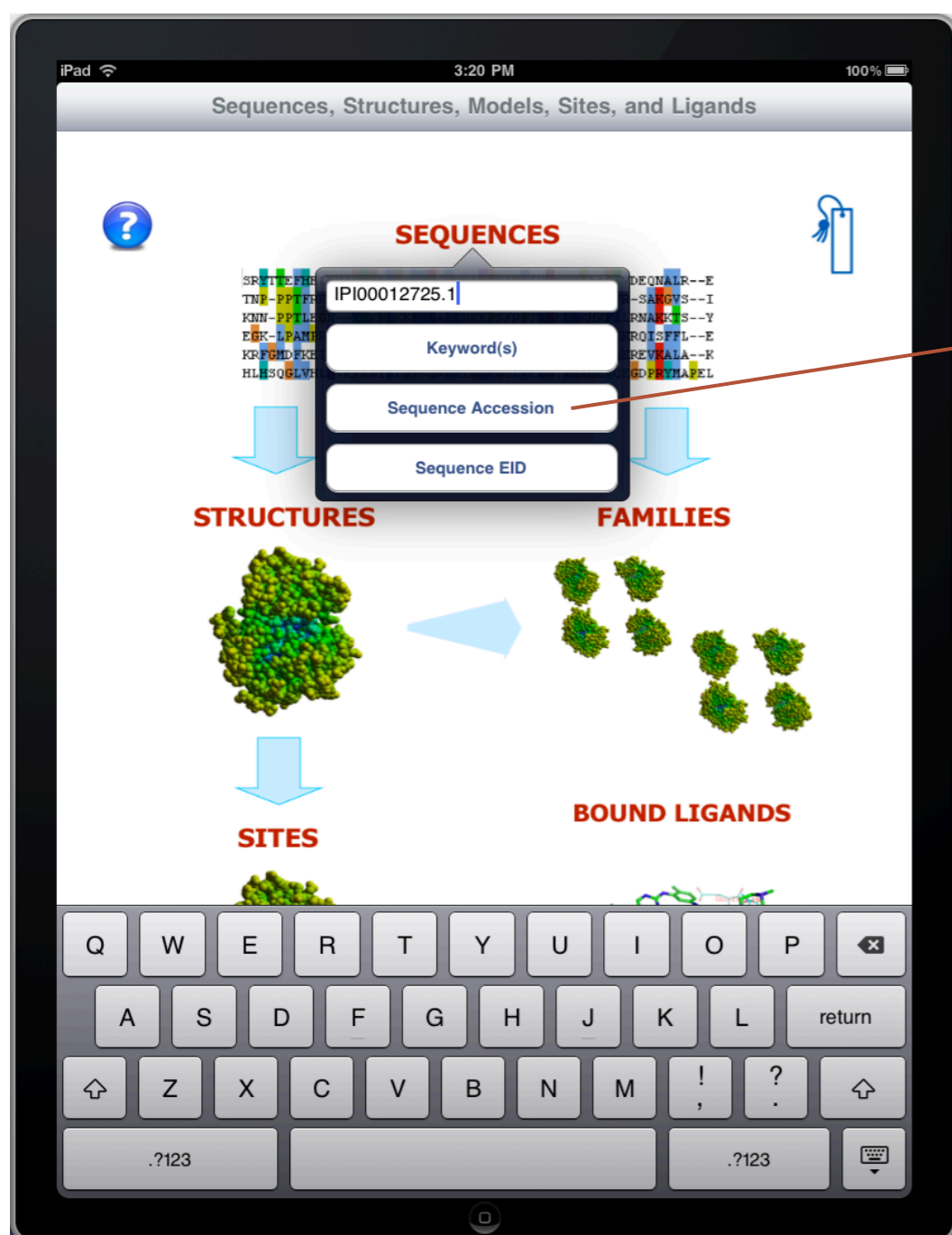
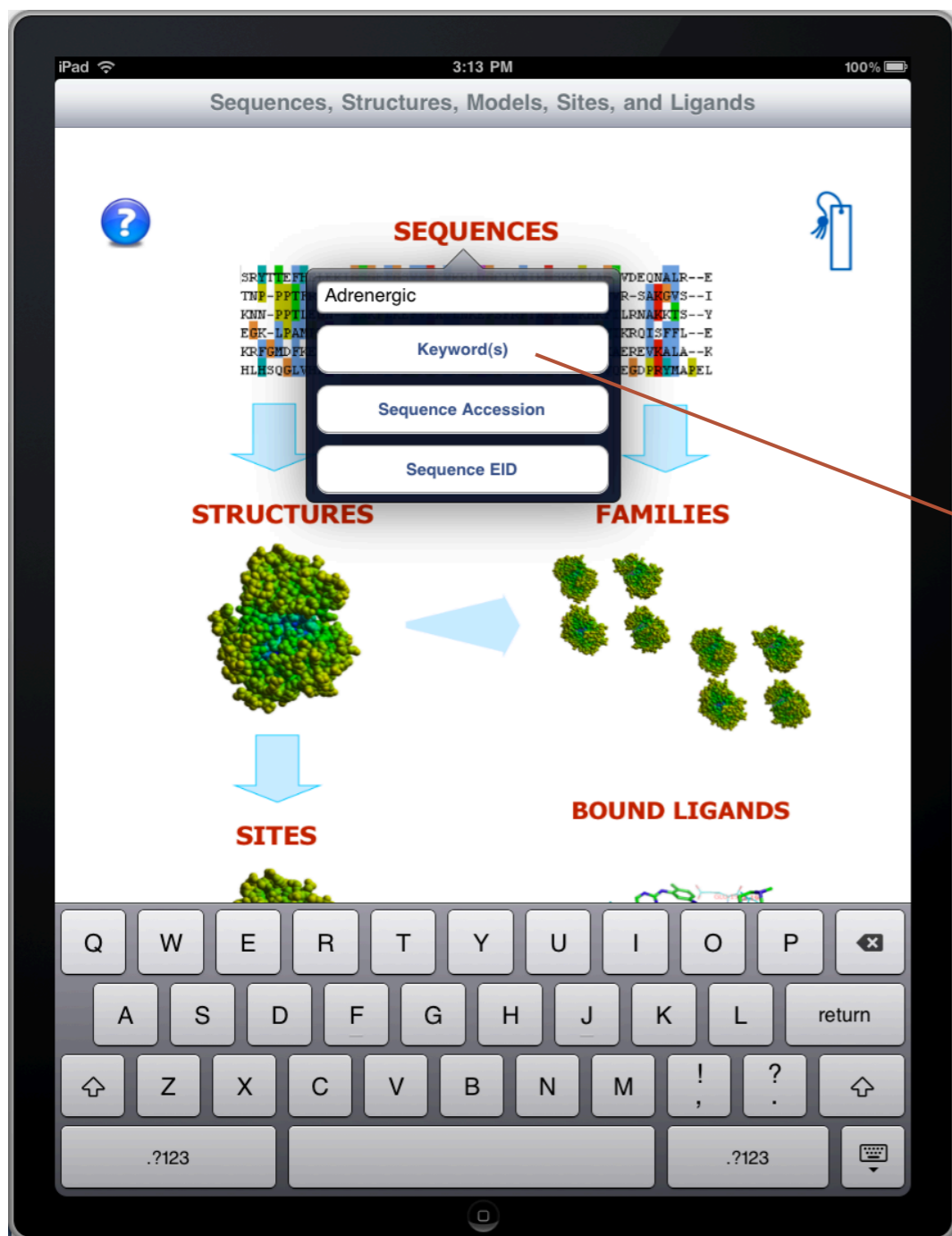
# Accessing TIP content through iProtein

Through iProtein, the TIP database can be surveyed by Sequence, Structure/Model, Site, head-nodes (i.e. protein family), and by bound ligand structure searches. Simply click on any image in the main page to initiate a search. Future versions of iProtein may enable more complex searching – e.g. protein structure search, site-search, etc. So stay tuned....



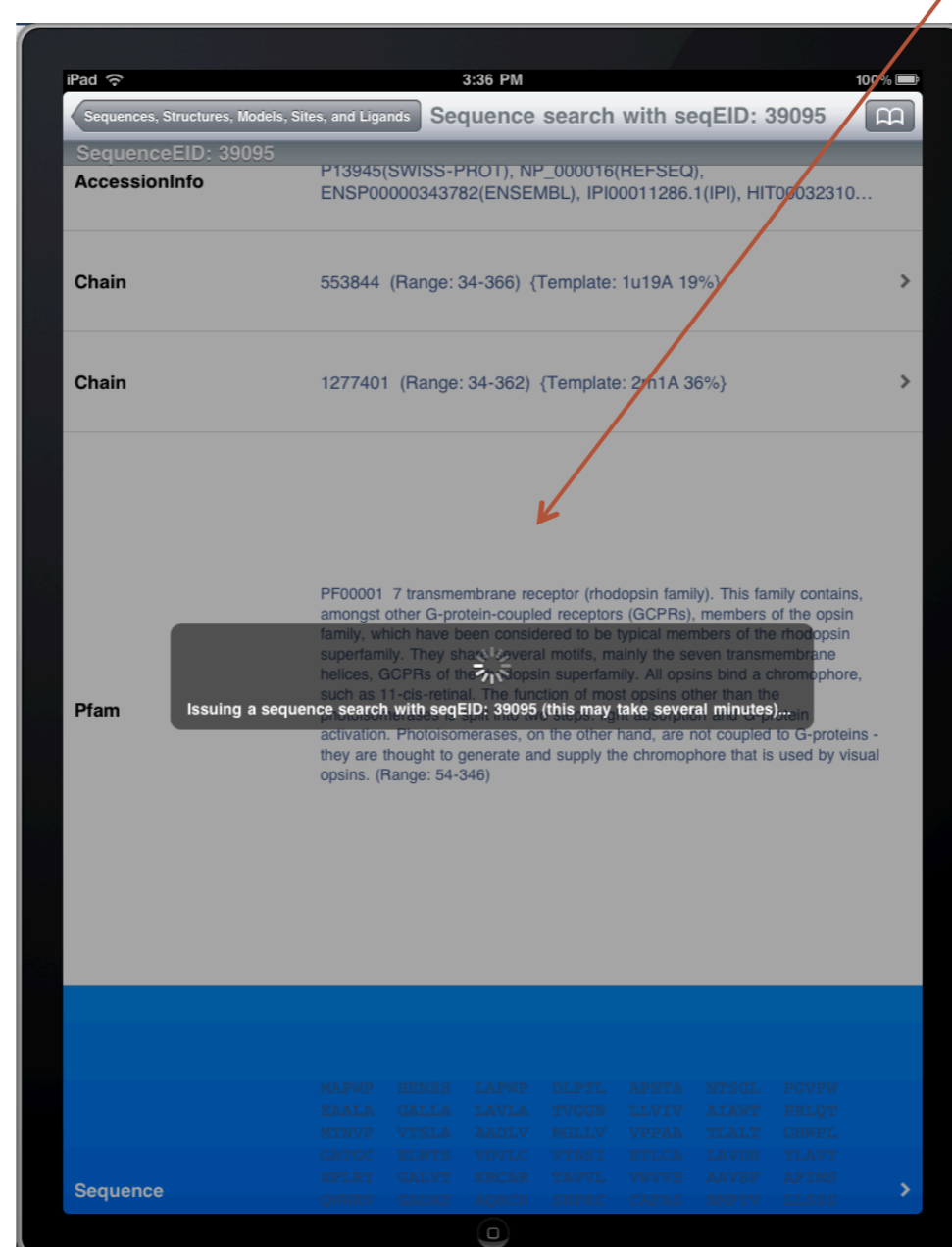
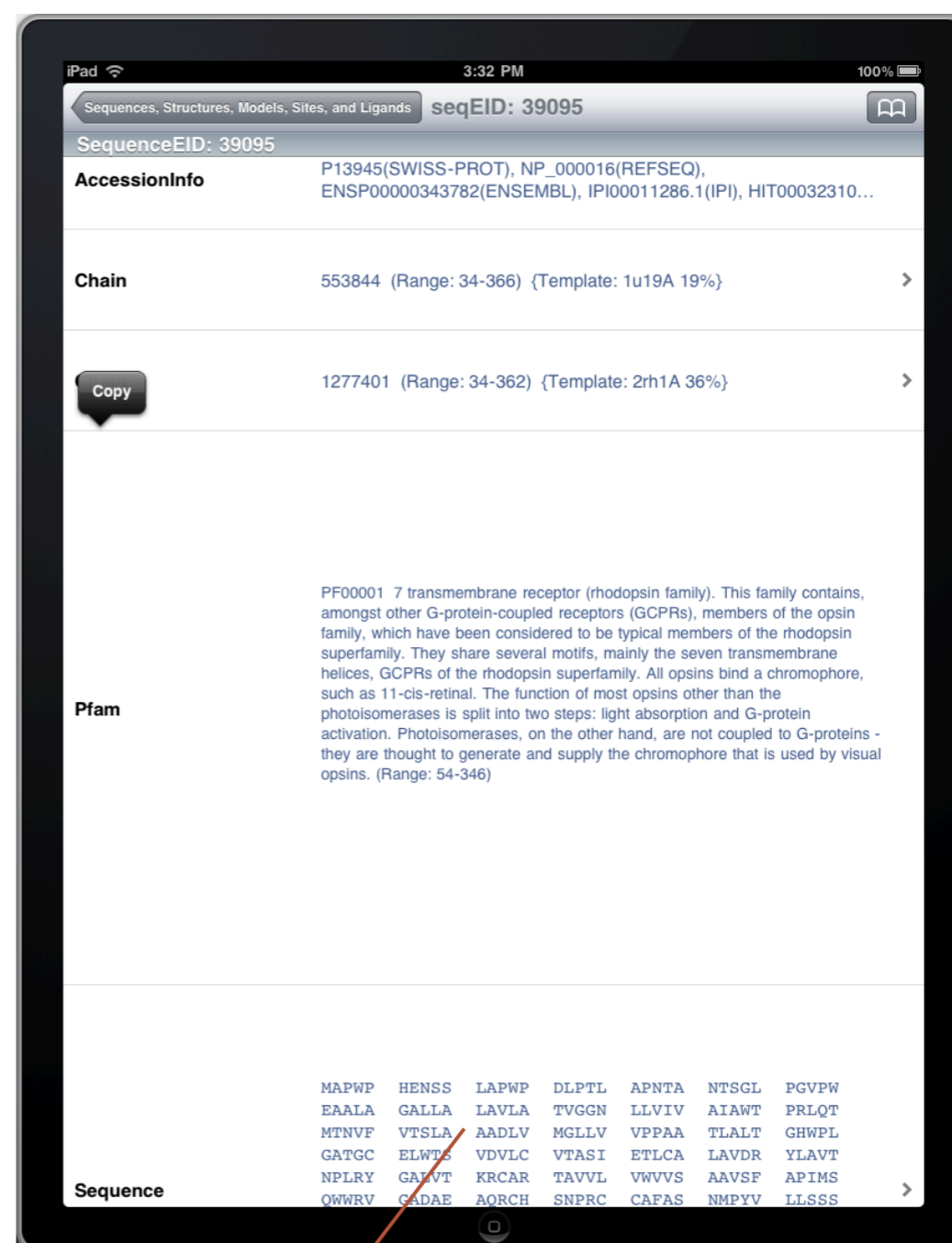
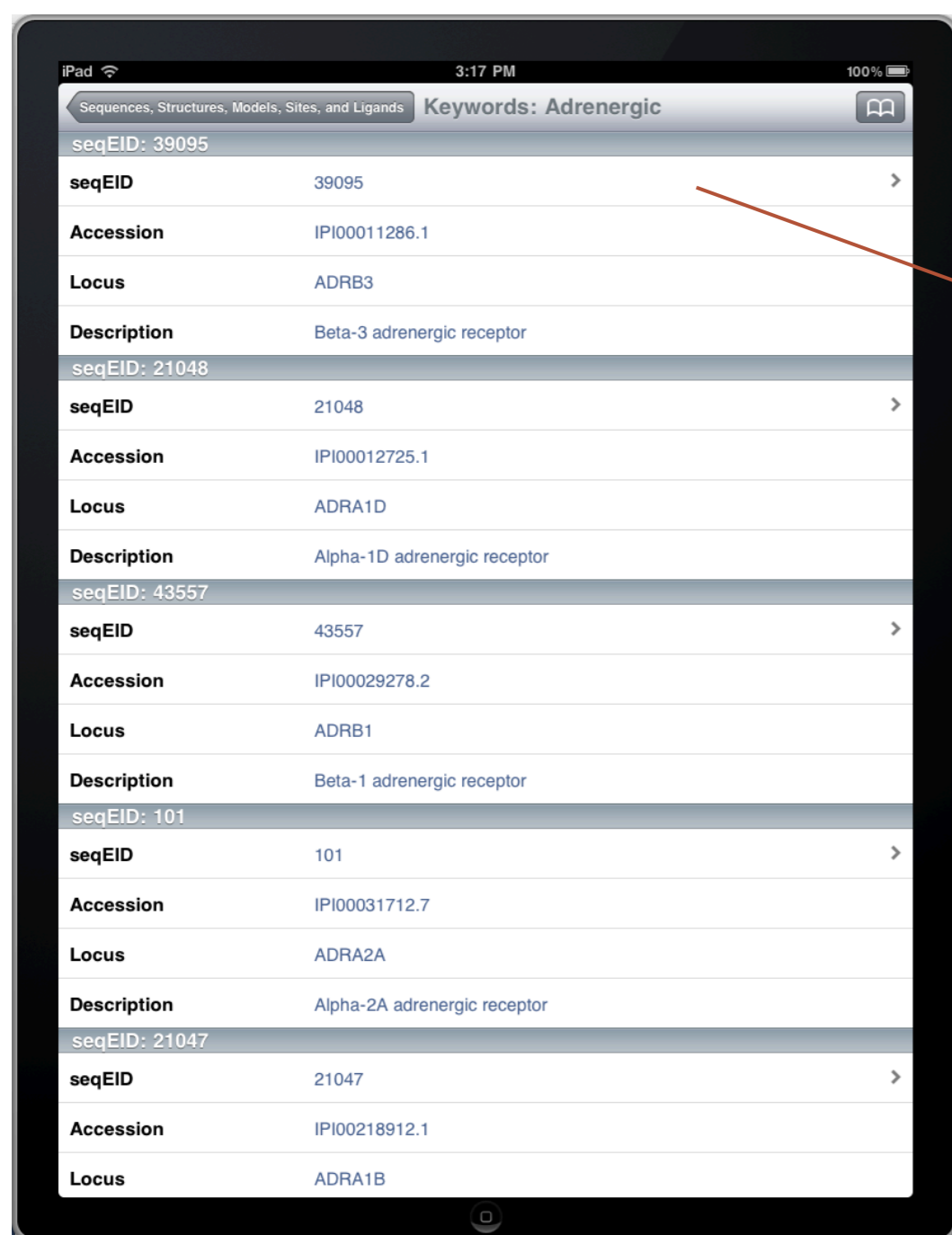
# Sequence Searches

Within iProtein you can locate sequences by Keywords, Sequence Accessions (e.g. Swiss-Prot, RefSeq, Ensembl, IPI, etc.), and by Eidogen Identifiers (EID). Below are example queries:



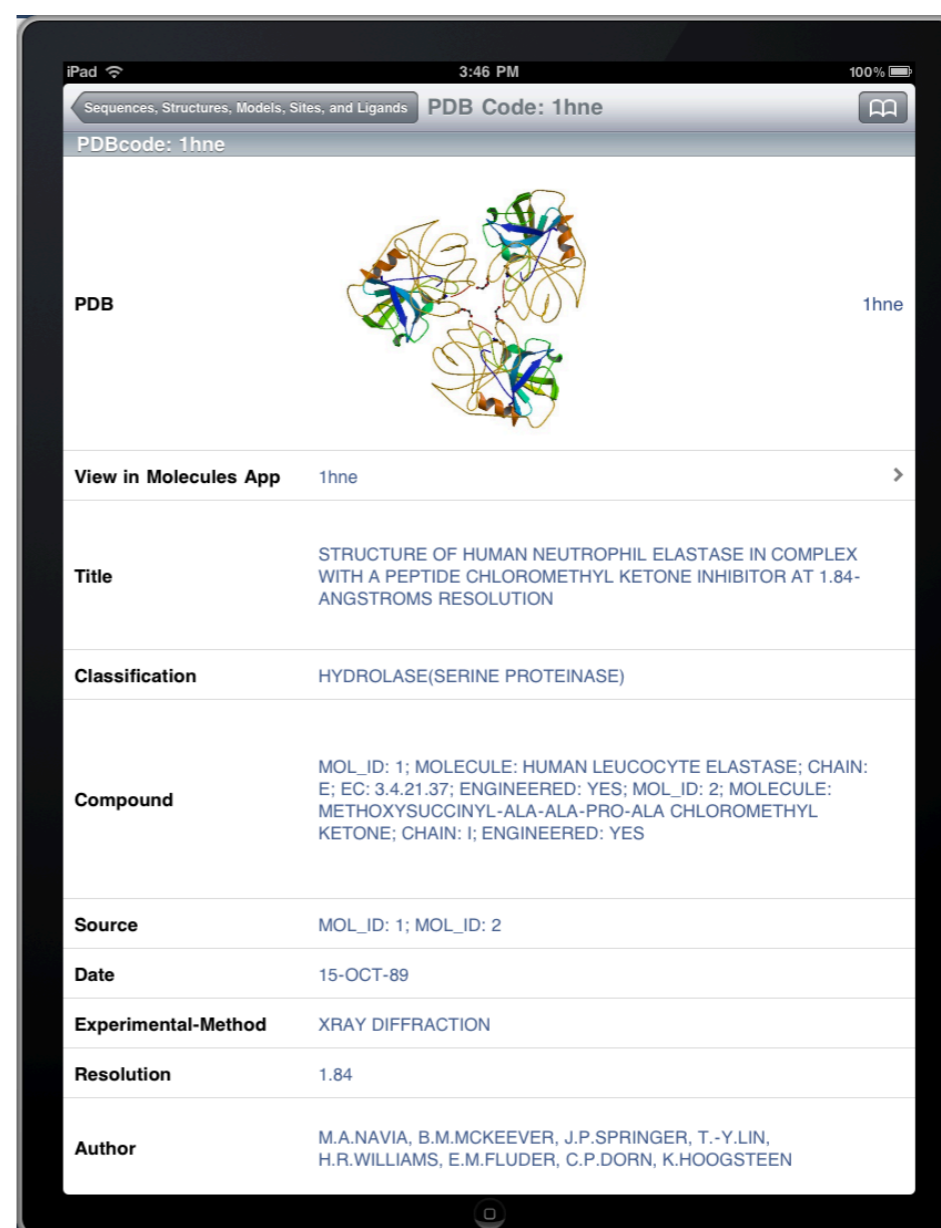
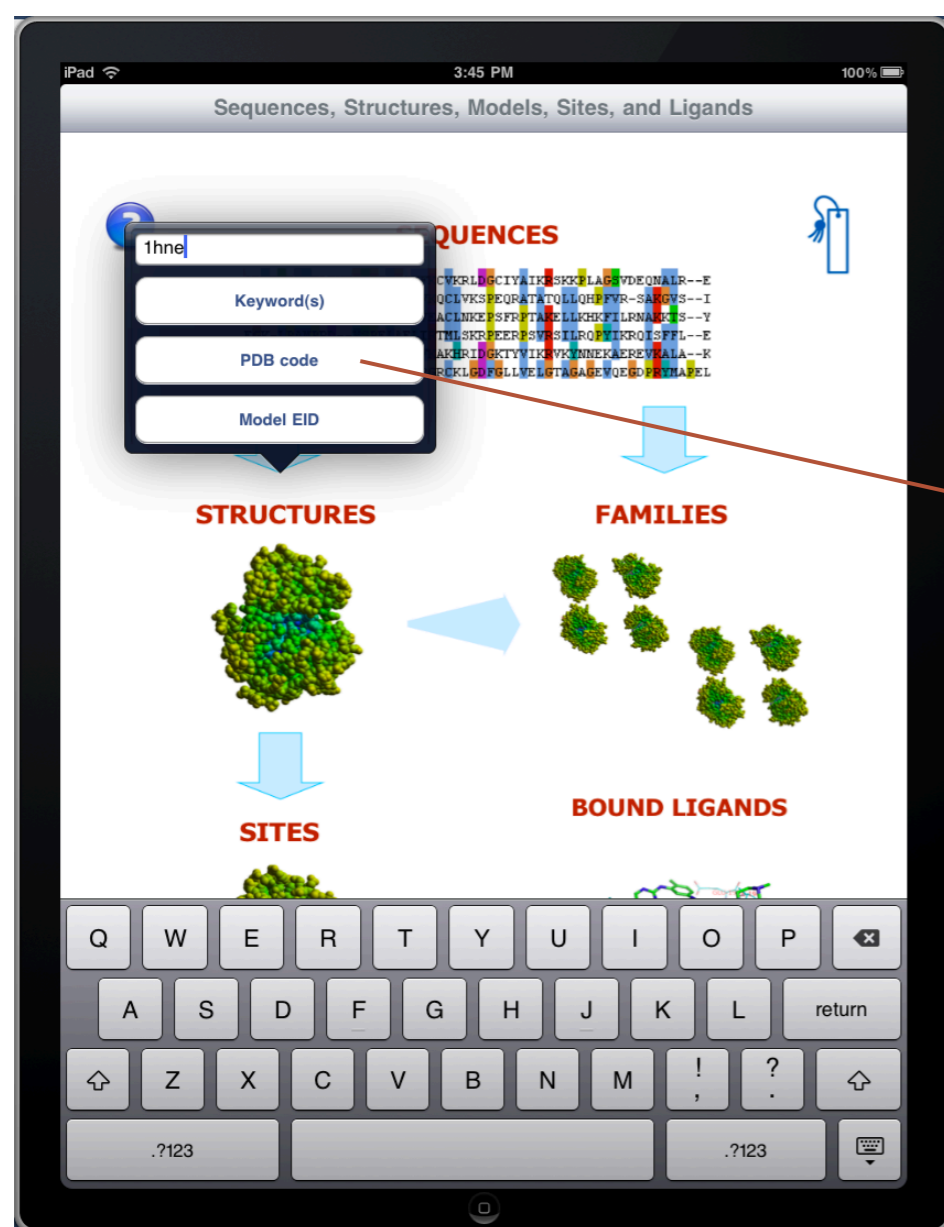
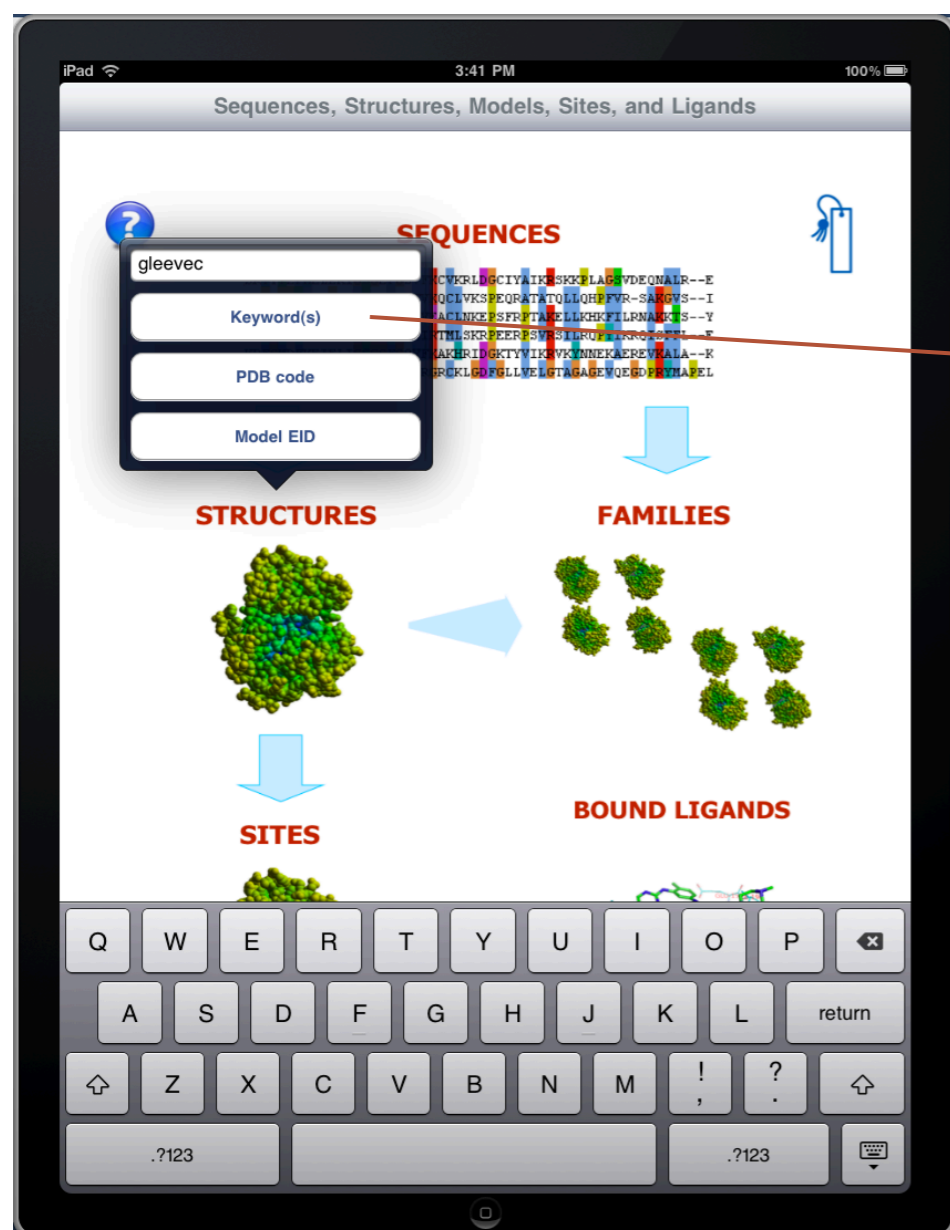
# Sequence Searches (cont.)

Click on the arrow (when present) to see more data. You can double-tap on any row without an arrow (“>”) to initiate a copy-to-clipboard action or follow the arrow to see more detail. By clicking on the full sequence, a sequence search will be issued (homology  $\geq 80\%$ ). Please note, sequence searches may take a several minutes, so please be patient...



# Structure/Model Searches

Within iProtein you can locate structures by Keywords, PDB codes, and by Eidogen Identifiers (EIDs). Below are a couple example queries:



# Structure/Model Searches (cont)

Click on the arrow (when present) to see more data. You can click the arrow (“>”) to see more detail. If there is a bound ligand, you can initiate ligand-based searches.

Sequences, Structures, Models, Sites, and Ligands Keywords: gleevec  
PDB: 1xbb

**View More Detail** 1xbb >

**Title** A NOVEL MODE OF GLEEVEC BINDING IS REVEALED BY THE STRUCTURE OF SPLEEN TYROSINE KINASE

**Classification** TRANSFERASE

**Compound** TYROSINE-PROTEIN KINASE SYK; CHAIN: A; SYNONYM: SPLEEN TYROSINE KINASE; EC: 2.7.1.112; ENGINEERED: YES

**Source** ORGANISM\_SCIENTIFIC: HOMO SAPIENS; ORGANISM\_COMMON: HUMAN; GENE: SYK; EXPRESSION\_SYSTEM: SPODOPTERA FRUGIPERDA; EXPRESSION\_SYSTEM\_COMMON: FALL ARMYWORM; EXPRE...

**Date** 02-NOV-04

**Experimental-Method** XRAY DIFFRACTION

**Resolution** 1.57

**Reference** J.BIOL.CHEM. V. 279 55827 2004

**score** 9.89

Sequences, Structures, Models, Sites, and Ligands PDB: 1xbb

**PDBcode:** 1xbb

**Source** MOL\_ID: 1; ORGANISM\_SCIENTIFIC: HOMO SAPIENS; ORGANISM\_COMMON: HUMAN; GENE: SYK; EXPRESSION\_SYSTEM: SPODOPTERA FRUGIPERDA; EXPRESSION\_SYSTEM\_COMMON: FALL ARMYWORM; EXPRE...

**Date** 02-NOV-04

**Experimental-Method** XRAY DIFFRACTION

**Resolution** 1.57

**Author** V.L.NIENABER, S.ATWELL, J.M.ADAMS, J.BADGER, M.D.BUCHANAN, I.K.FEIL, K.J.FRÖNING, X.GAO, J.HENDLE, K.KEEGAN, B.C.LEON, H.J.MÜLLER-DEICKMANN, B.W.NOLAND, K.POST, K.R.RAJASHANKAR, A.RAMOS, M.RUSSELL, S.K.BURL...

**Journal-Title** A NOVEL MODE OF GLEEVEC BINDING IS REVEALED BY THE STRUCTURE OF SPLEEN TYROSINE KINASE

**Journal-Reference** J.BIOL.CHEM. V. 279 55827 2004

**chains** 1xbbA(2)

**Site-1** 521165 NumResidues: 22 (PDB Co-crystal) STI: 4-(4-METHYL-PIPERAZIN-1-YLMETHYL)-N-[4-METHYL-3-(4-PYRIDIN-3-YL-PYRIMIDIN-2-YLAMINO)-PHENYL]-BENZAMIDE >

**Site-2** 521166 NumResidues: 40 (SiteSeeker) Predicted Site >

Sequences, Structures, Models, Sites, and Ligands Site: 521165

siteID: 521165

**View Animated Site** 521165 >

**View-Site-In-Molecules** 521165 >

**Source** PDB Co-crystal

**Confidence** 100%

**Resolution** XRAY

**Description** STI: 4-(4-METHYL-PIPERAZIN-1-YLMETHYL)-N-[4-METHYL-3-(4-PYRIDIN-3-YL-PYRIMIDIN-2-YLAMINO)-PHENYL]-BENZAMIDE

**View Parent** 1xbbA >

**Ligand-Structure-Search** >

**Site Residue** LEU Chain: A Index: 15 Pos: 377

**Site Residue** GLY Chain: A Index: 16 Pos: 378

**Site Residue** SER Chain: A Index: 17 Pos: 379

**Site Residue** GLY Chain: A Index: 18 Pos: 380

**Site Residue** VAL Chain: A Index: 23 Pos: 385

**Site Residue** ALA Chain: A Index: 38 Pos: 400

**Site Residue** LYS Chain: A Index: 40 Pos: 402

**Site Residue** GLU Chain: A Index: 58 Pos: 420

Ligand Structure Search: STI

Draw a structure and touch the button to issue a SSSearch

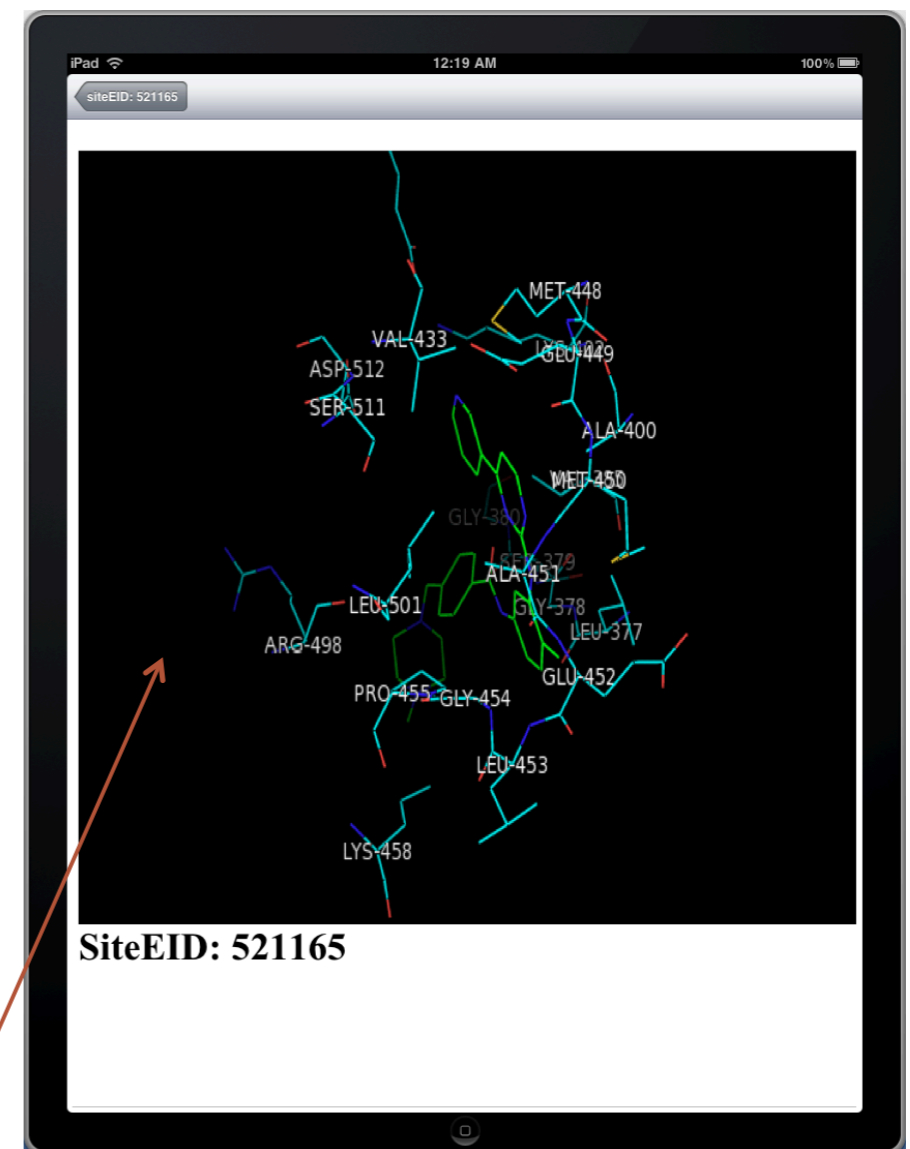
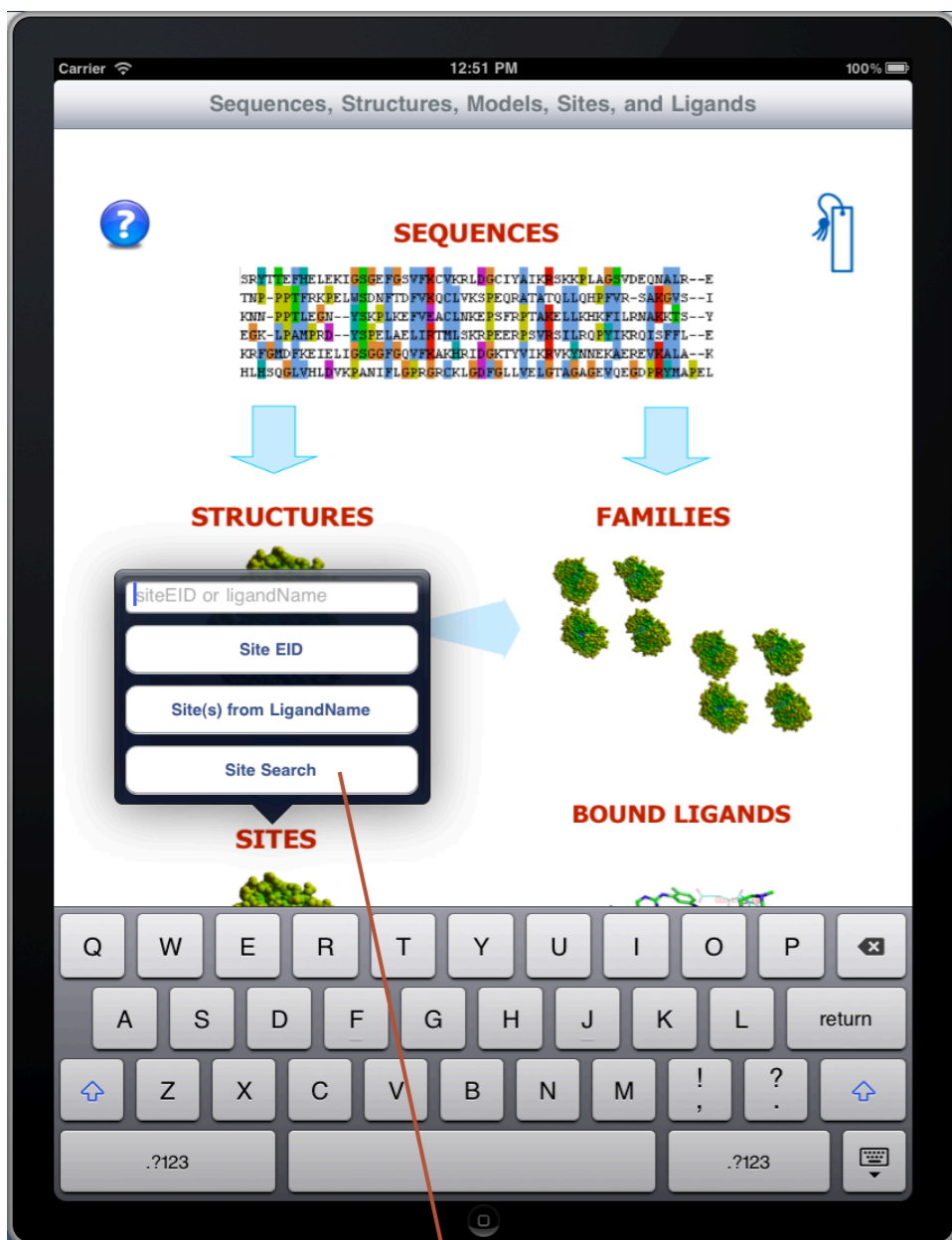
H  
C  
N  
O  
S  
P  
F  
Cl  
Br  
...

Chemical structure of the ligand is shown in the center.

Site	PDB	Site Hit
114219	2825	HUM
101401	1241	29M
118201	3125	IT2

# Site Searches

Within iProtein you can locate sites by Eidogen Identifiers (EIDs), PDB assigned ligand name and/or site ligand common name. Sites are annotated in both PDB structures and models. Ligand structures are shown if present in PDB co-complexes or inferred in models, given the template used in the modeling. Predicted sites within PDBs and models are also annotated. With version 2 of iProtein, Eidogen's powerful Site-Similarity searching capability is now accessible:



View Animated Site

The screenshot shows the details page for Site EID: 521165. The page includes the following information:

View Animated Site	521165
SiteSimilarity Search	521165
Source	PDB Co-crystal
Confidence	100%
Resolution	XRAY
Description	STI: 4-(4-METHYL-PIPERAZIN-1-YLMETHYL)-N-(4-METHYL-3-(4-PYRIDIN-3-YL-PYRIMIDIN-2-YLAMINO)-PHENYL)-BENZAMIDE
View Parent	1xbbA
Ligand-Structure-Search	[Ligand Structure]
Site Residue	LEU Chain: A Index: 15 Pos: 377
Site Residue	GLY Chain: A Index: 16 Pos: 378
Site Residue	SER Chain: A Index: 17 Pos: 379
Site Residue	GLY Chain: A Index: 18 Pos: 380
Site Residue	VAL Chain: A Index: 23 Pos: 385
Site Residue	ALA Chain: A Index: 38 Pos: 400
Site Residue	LYS Chain: A Index: 40 Pos: 402
Site Residue	GLU Chain: A Index: 58 Pos: 420

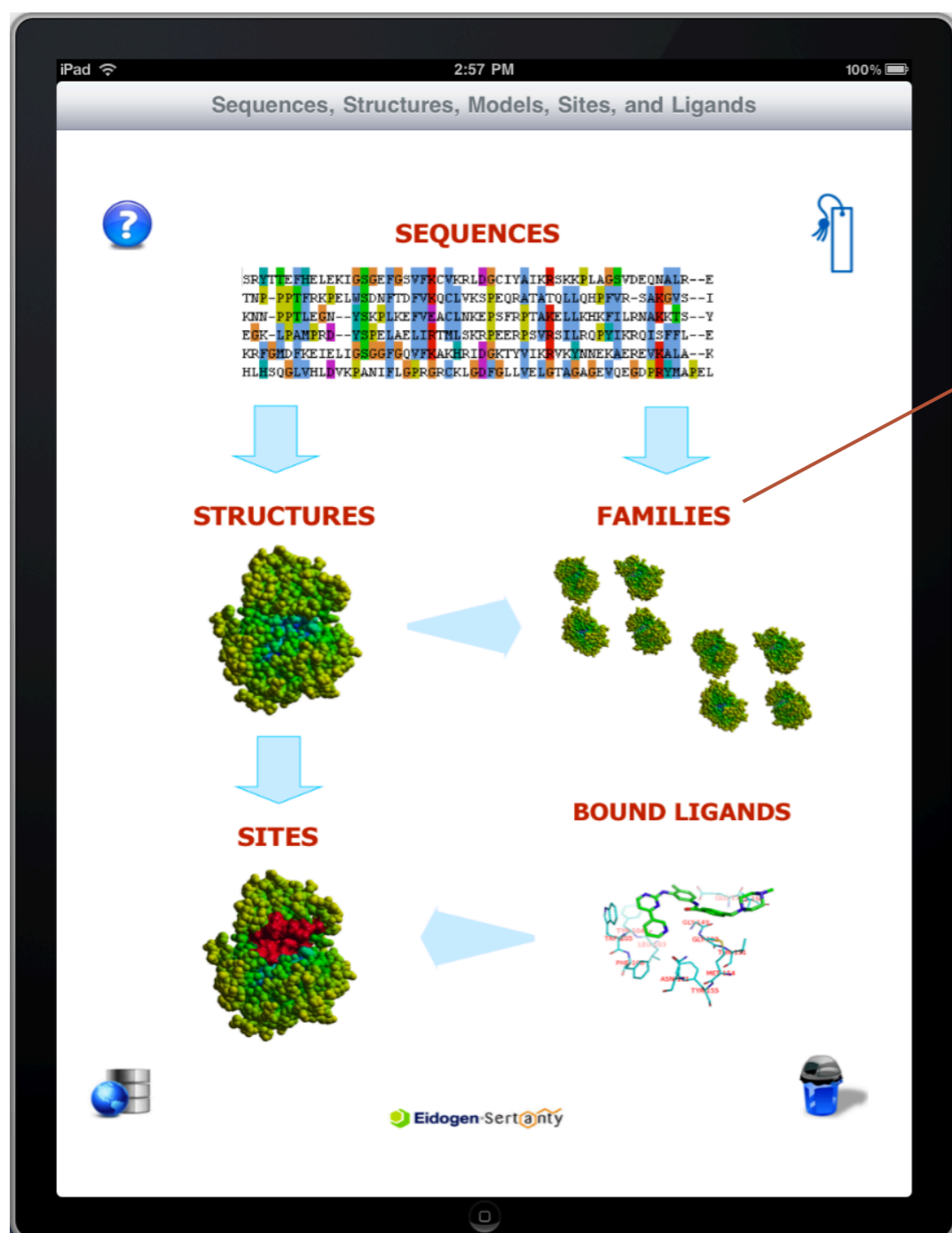
The screenshot shows the SiteSimilarity Search results page for Site EID: 521165. The page includes the following information:

View Animated Site	521165
Source	PDB Co-crystal
Description	TYROSINE-PROTEIN KINASE SYK
View Parent	1xbbA
Ligand-Structure-Search	[Ligand Structure]
AA-Similarity	100
SiteSorter-Score	100
Contact-Similarity	100
siteEID: 1154340	
View Animated Site	1154340
Source	PDB Co-crystal
Description	TYROSINE-PROTEIN KINASE SYK
View Parent	1xbbA
Ligand-Structure-Search	[Ligand Structure]
AA-Similarity	100

SiteSimilarity Search  
(Note: These can take several minutes)

# Family Searches

The TIP system clusters structurally similar protein structures. Representative members of each cluster are termed “headnodes.” Headnodes and corresponding family members are used in the sequence-to-structure modeling process. By touching “Families,” the current list of headnodes is downloaded. You can show representative family members by touching respective headnodes:



The screenshot shows the "Headnodes" list. At the top, it says "Sequences, Structures, Models, Sites, and Ligands" and "Headnodes". Below this is a search bar. The list contains the following entries:

Headnode ID	Description
1A6M	1A6M OXY-MYOGLOBIN, ATOMIC RESOLUTION; FamilyId = 2136; PDBs in Family = 222
1A8R	1A8R GTP CYCLOHYDROLASE I (H112S MUTANT) IN COMPLEX WITH GTP; FamilyId = 2279; PDBs in Famili...
1AD2	1AD2 RIBOSOMAL PROTEIN L1 MUTANT WITH SERINE 179 REPLACED BY CYSTEINE; FamilyId = 2152;...
1BEB	1BEB BOVINE BETA-LACTOGLOBULIN, LATTICE X; FamilyId = 2426; PDBs in Family = 41
1CEX	1CEX STRUCTURE OF CUTINASE; FamilyId = 3173; PDBs in Family = 45
1CQM	1CQM PROTEIN AGGREGATION AND ALZHEIMER'S DISEASE: CRYSTALLOGRAPHIC ANALYSIS OF THE...
1D0C	1D0C BOVINE ENDOTHELIAL NITRIC OXIDE SYNTHASE HEME DOMAIN COMPLEXED WITH 3-BROMO-7-...
1D4A	1D4A CRYSTAL STRUCTURE OF HUMAN NAD[P]H-QUINONE OXIDOREDUCTASE AT 1.7 A RESOLUTION;...
1DBF	1DBF CHORISMATE MUTASE FROM BACILLUS SUBTILIS AT 1.30 ANGSTROM; FamilyId = 3264; PDBs in...
1DGF	1DGF HUMAN ERYTHROCYTE CATALASE; FamilyId = 2831; PDBs in Family = 47
1DPS	1DPS THE CRYSTAL STRUCTURE OF DPS, A FERRITIN HOMOLOG THAT BINDS AND PROTECTS DNA;...
1DW9	1DW9 STRUCTURE OF CYANASE REVEALS THAT A NOVEL DIMERIC AND DECAMERIC ARRANGEMENT...
1DY5	1DY5 DEAMIDATED DERIVATIVE OF BOVINE PANCREATIC RIBONUCLEASE; FamilyId = 2199; PDBs in Fa...
1E7W	1E7W ONE ACTIVE SITE, TWO MODES OF REDUCTION CORRELATE THE MECHANISM OF LEISHMANIA...
1EJD	1EJD CRYSTAL STRUCTURE OF UNLIGANDED MURA (TYPE1); FamilyId = 2194; PDBs in Family = 60
1ETO	1ETO THE CRYSTAL STRUCTURE OF E. COLI FIS MUTANT R71L; FamilyId = 5456; PDBs in Family = 47
1EXR	1EXR THE 1.0 ANGSTROM CRYSTAL STRUCTURE OF CA+2 BOUND CALMODULIN; FamilyId = 3369; PDB...
1F41	1F41 CRYSTAL STRUCTURE OF HUMAN TRANSTHYRETIN AT 1.5A RESOLUTION; FamilyId = 2613; PDBs...
1F52	1F52 CRYSTAL STRUCTURE OF GLUTAMINE SYNTHETASE FROM SALMONELLA TYPHIMURIUM CO-CR...
1F58	1F58 IGG1 FAB FRAGMENT (58.2) COMPLEX WITH 24-RESIDUE PEPTIDE (RESIDUES 308-333 OF HIV-1...
1F8D	

The screenshot shows the "1CEX/Family Members" page. At the top, it says "Headnodes" and "1CEX/Family Members". Below this is a list of family members:

seqEID	Title	Classification
263484	STRUCTURE OF CUTINASE	SERINE ESTERASE
265751	CUTINASE, N172K MUTANT	HYDROLASE (SERINE ESTERASE)
265755		



# Ligand Searches

iProtein provides a chemical drawing interface that enables chemical structure drawing with your finger(s). After structures are drawn (or edited), you can issue substructure searches by tapping the iProtein icon.

The screenshot shows the iProtein application interface on an iPad. At the top, the status bar displays 'iPad', signal strength, '3:18 PM', and '100%' battery. Below the status bar is a navigation bar with the text 'Ligand-Structure-Search: STI'. The main drawing area contains a toolbar with various icons for drawing and editing. Below the toolbar is a drawing area containing two chemical structures: a benzene ring with substituents R and A, and a fused bicyclic system with two nitrogen atoms. A callout box points to a green cube icon in the toolbar with the text 'Tap to issue a substructure-search'.

Tap to issue a substructure-search

# Ligand Searches (cont)

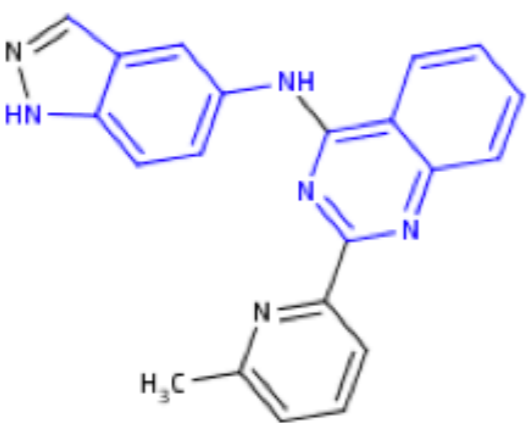
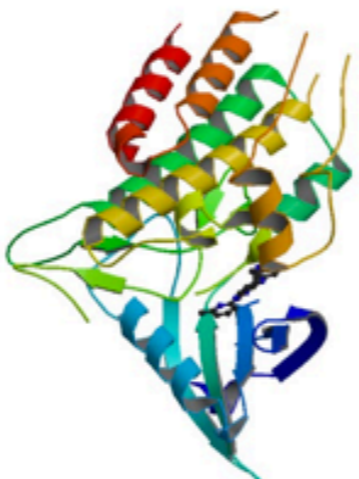
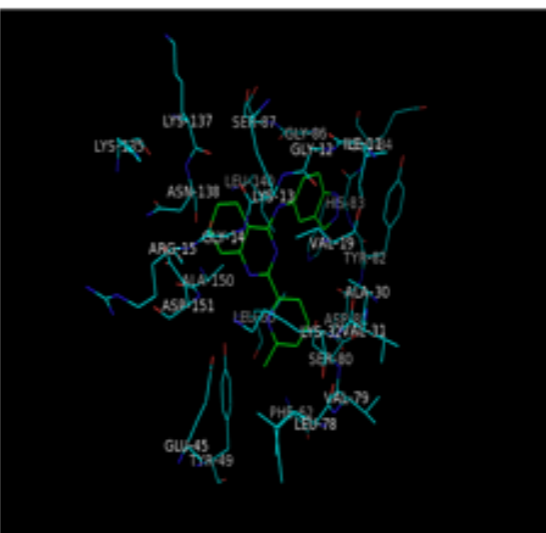
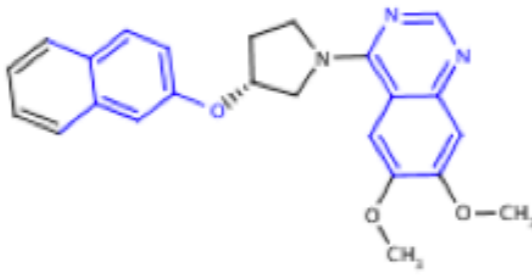
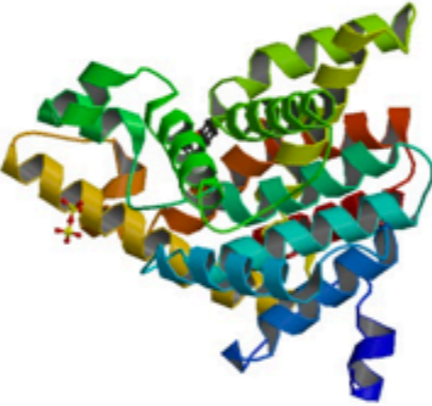
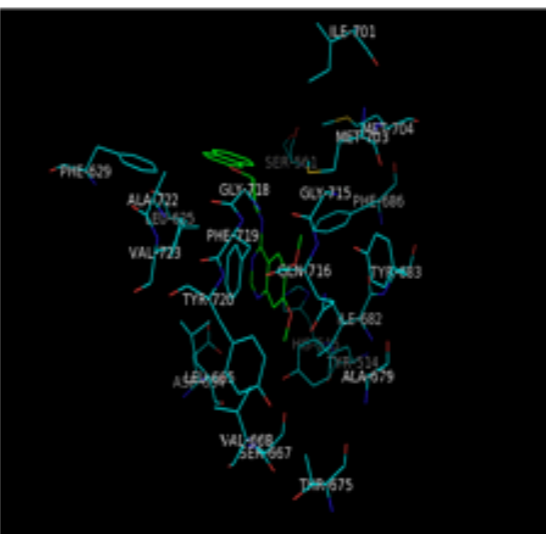
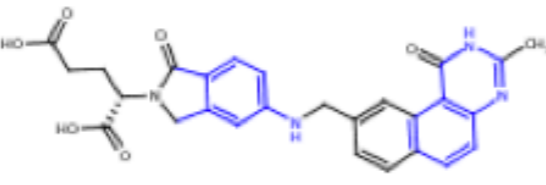
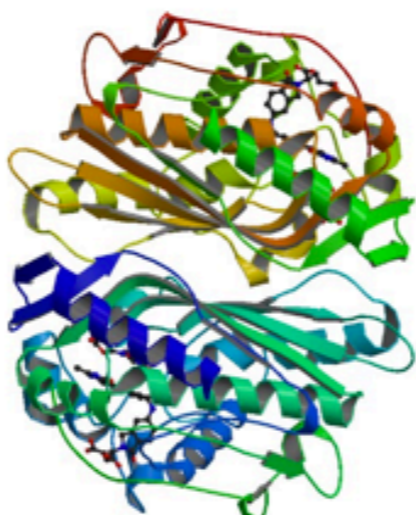
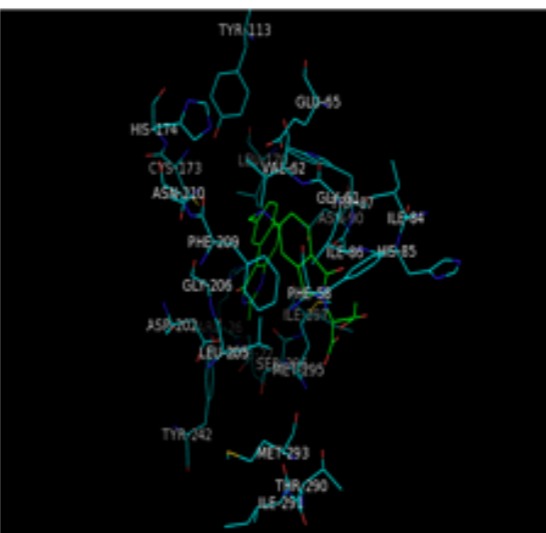
Ligand search results will tabulate the substructure hit(s), a ribbon diagram of the parent protein complex, and present an animated rendering of the co-complexed ligand within the receptor site. Site residues are annotated. You can click on the PDB code or the site EID to drill into more detail:

iPad 3:19 PM 100%

Ligand-Structure-Search: STI

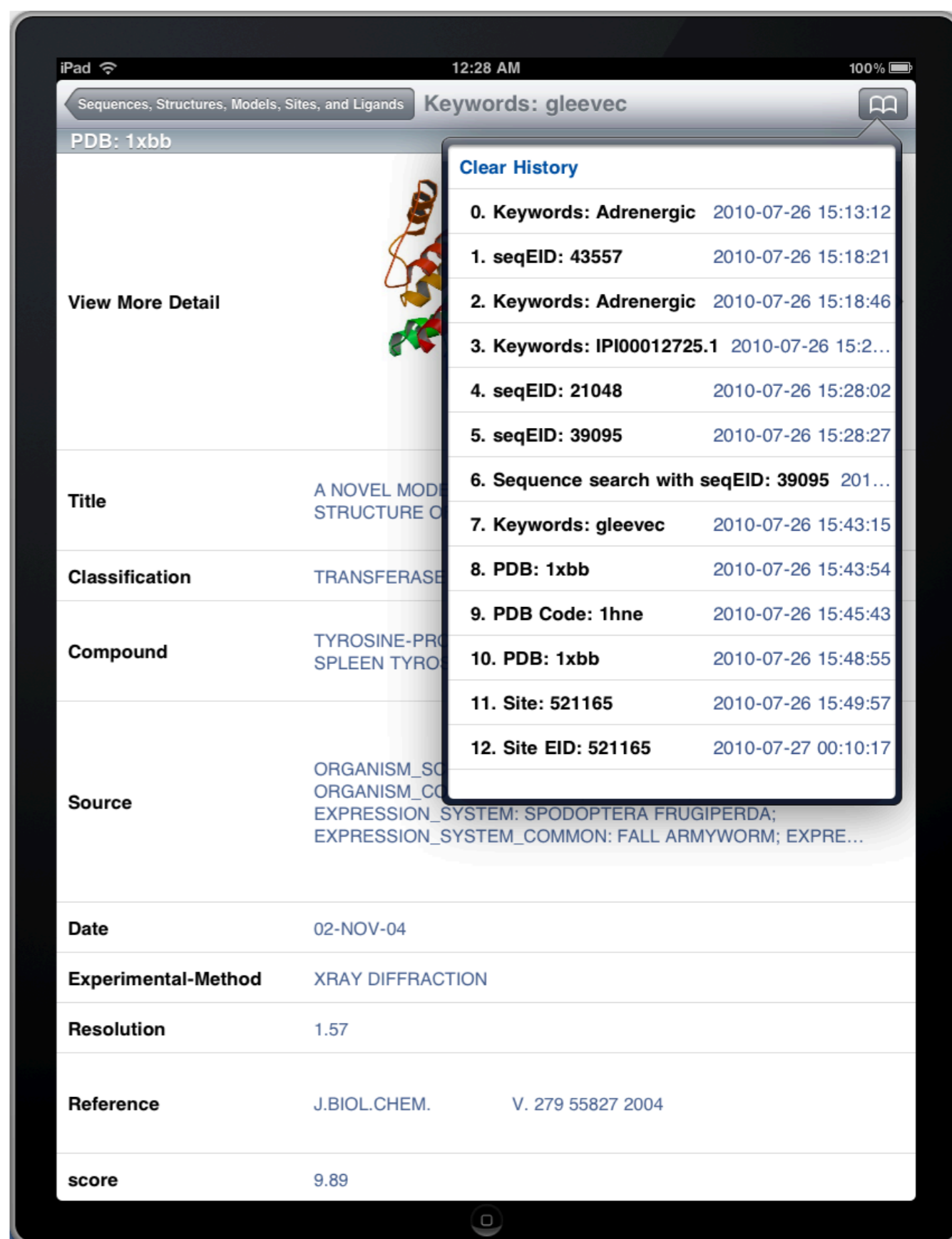
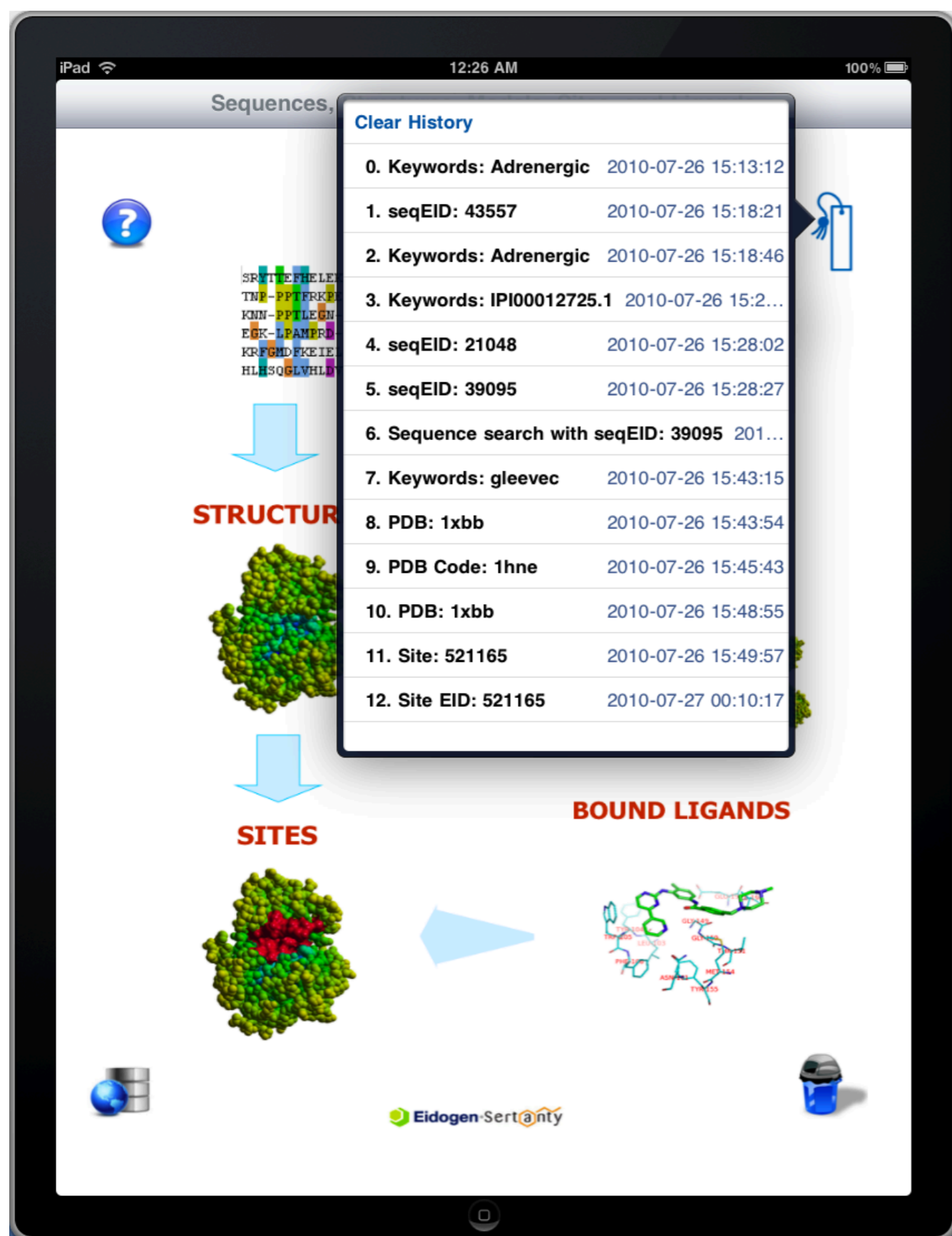
Go Again

Searching...

Ligand Hit	PDB	Site
 QIG	 <a href="#">3gxl</a>	 <a href="#">1554672</a>
 PFH	 <a href="#">2ovv</a>	 <a href="#">917479</a>
 F89	 <a href="#">1f28</a>	 <a href="#">370095</a>

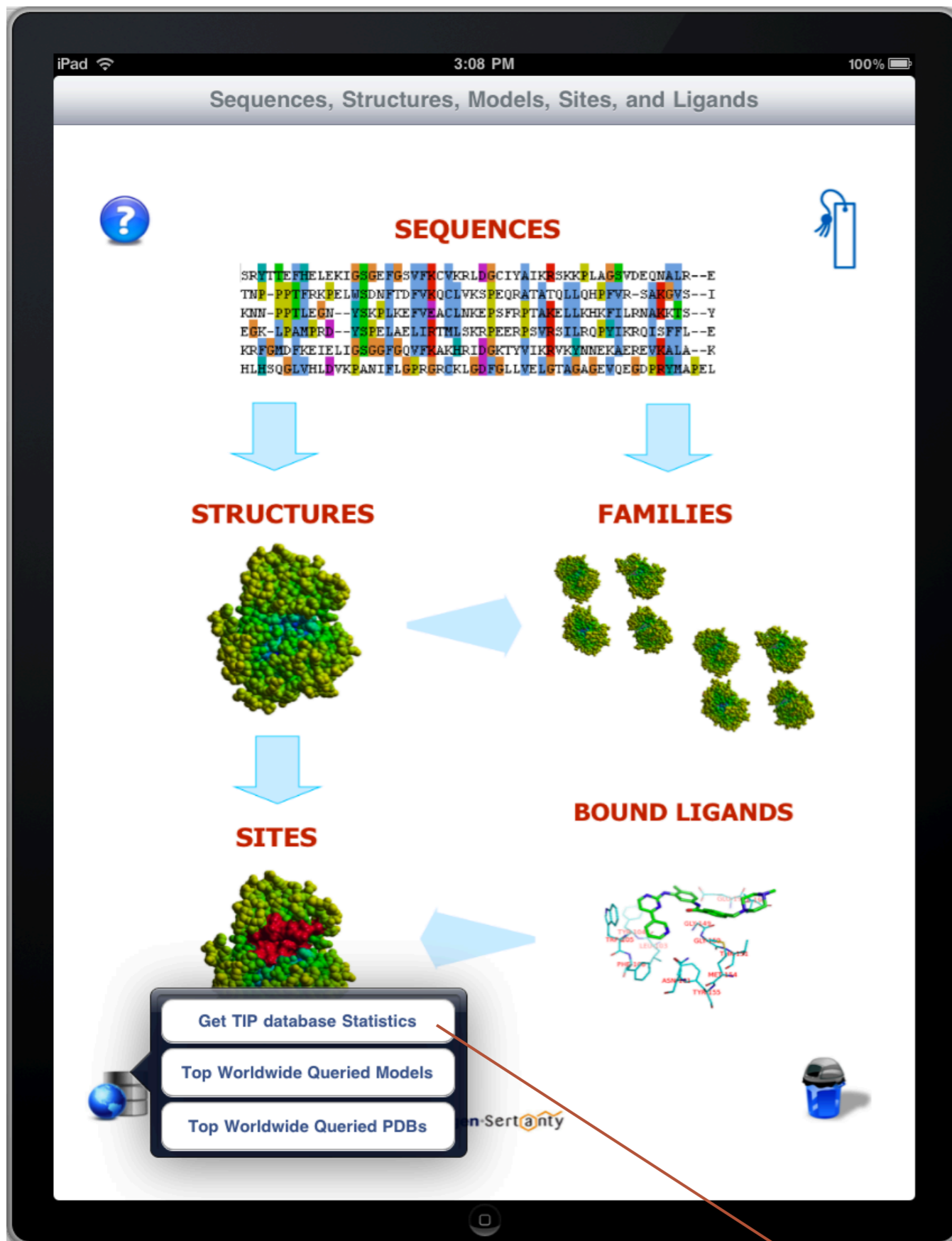
# History

As you survey the proteome with iProtein, the results of your searches are stored locally on your device and accessible through the bookmark icons. You can use this history to quickly step backwards through past searches as you continue to survey the proteome:



# Database Statistics

The [current TIP statistics](#) can be retrieved from the mainpage, along with the top worldwide queried models and PDBs. These data will change weekly!



TIP™ Database Statistics as of 07/27/2010

Number of Sequences	417511
<b>Human Sequences per Knowledge Level</b>	
Co-crystallized Ligand	4732
Experimental Structure	5048
Predicted Site	16862
Predicted Structure	4771
No Structure	46641
Number of PDB Proteins	63189
Number of PDB Chains	155597
Number of Models	47721
<b>Models per %ID to Template</b>	
71%-100%	13309
41%-70%	10369
31%-40%	7462
21%-30%	10381
0%-20%	6200
Number of Co-crystal Sites	266386
Number of Predicted Sites on Models	77663
Number of Predicted Sites on PDBs	171703
<b>Predicted Sites per %Confidence</b>	
100%	25254
90%-99%	51375
80%-89%	32765

# Clean-Up

All images, PDB structures, Models, Sites, and historical results (i.e. bookmarks) can be cleared from your device by touching the garbage can in the lower right:

