Model building tutorial



Tutorial PDF: https://bit.ly/2XPsiox

Data: https://bit.ly/3ASQ411

AlphaFold add on: https://bit.ly/3KTo6qX

Model building and validation for cryoEM

Oliver Clarke

(@OliBClarke)





An atomic model is a compact interpretation of the density map in light of prior knowledge (both specific and general).

- Aim is to build a model that is consistent with **both** the density map and everything we independently know about the structure & composition of the macromolecule of interest, both specifically and in terms of our general knowledge of protein structure and chemistry.
- Tradeoff between available prior knowledge and required resolution for atomic modelling at the extremes, if a complete crystal structure is already available, 10Å data may be sufficient, while if no sequence/composition data is available even 3Å may not suffice.

Prior knowledge

- Protein sequence and derived info (secondary structure predictions, covariation/conservation, patterns of large/aromatic residues), disorder & contact prediction
- Crystal structures (+ homology & ML-derived models)
- Knowledge of protein structure, folding, chemistry, geometry.

Density map

- Resolution (+ local resolution, + map modification/sharpening)
- Patterns of large/small/absent sidechains
- Sharpening and density modification
- Conformational/compositional heterogeneity



- If possible, unique model that agrees with both density map and priors
- Otherwise (and per region), specify ambiguity (w/UNK residues and numbering or Ca only model)
- Validation not just (or even mostly) about overfitting.
- Identify, analyse, fix errors.
- Direction and register of sequence fit.
- Ligand identification/assignment.
- No model is or ever will be perfect. That's okay.

One extreme – at atomic resolution, the position of many atoms can be inferred without prior knowledge of the sequence



Yip, K.M., Fischer, N., Paknia, E. et al. Atomic-resolution protein structure determination by cryo-EM. Nature **587,** 157–161 (2020)

At 20 Å (here using cryoET), an informative model can be generated by taking advantage of external information – crystal structures, connectivity from crosslinking & MS, even when de novo building is not possible.



Kim, S., Fernandez-Martinez, J., Nudelman, I. et al. Integrative structure and functional anatomy of a nuclear pore complex. Nature **555,** 475–482 (2018)

Usually, we are somewhere in between the two – combining prior knowledge with inferences made from analyzing the density map.

To build a better/more reliable model, we can either get additional/better priors, or improve our density map (or part of it). Before you start – make sure your maps are appropriately sharpened and low pass filtered! (and consider whether building is justified or whether further improvement of the reconstruction is required first)

- Often it is helpful to build using multiple maps. Assuming 3-3.5Å global res, I would suggest using a map filtered to the global resolution, one filtered to the best local resolution, and one filtered to ~4-4.5 Å (to better visualize connectivity and mobile ligands/lipids).
- Try both simple B-factor sharpening and the approach used by *phenix.resolve_cryo_em*, which incorporates anisotropy removal and statistical density modification. In cases of **severe** anisotropy, deepEMhancer can be useful to assist map interpretation (**approach with caution**).
- Also, if your map doesn't "look like" 4 Å, trust your eyes! If it is nominally 4Å and there are no sidechains visible, or your helices look "stretched", assess orientation bias (3D-FSC server: <u>https://3dfsc.salk.edu</u>), local resolution variation, and double check sharpening and masking parameters (are you *sure* you're looking at the sharpened map? Is the mask used for FSC calculation sensible?)

Example of map anisotropy mitigated by masked refinement

- Map anisotropy hinders interpretation, even when resolution in "good" direction is high
- Can derive from either preferred orientation, or interdomain mobility (or combination).
- In latter case, masked refinement can improve local map quality to aid model building and map interpretation. Always better to improve the map than build in marginal density
- If anisotropy derives from preferred orientation, it is best to address this by improving the sample or data collection (tilt). If all else fails, ML-based map improvement using deepEMhancer can improve map interpretability.



Example of map anisotropy mitigated by masked refinement

- Map anisotropy hinders interpretation, even when resolution in "good" direction is high
- Can derive from either preferred orientation, or interdomain mobility (or combination).
- In latter case, masked refinement can improve local map quality to aid model building and map interpretation. Always better to improve the map than build in marginal density
- If anisotropy derives from preferred orientation, it is best to address this by improving the sample or data collection (tilt). If all else fails, ML-based map improvement using deepEMhancer can improve map interpretability.



DeepEMhancer

(Sanchez-Garcia R., 2021 Comm. Biol.)

Prep for model building - what can we learn from the sequence alone?

Your protein sequence contains a lot of useful information which you can use to aid model building:

- Start by identifying boundaries of conserved domains (NCBI CDD: <u>https://www.ncbi.nlm.nih.gov/Structure/cdd/</u>; DELTA-BLAST also performs CD-search by default)
- Then identify and/or generate suitable structural templates for building known domains: FUGUE, PHYRE2, MUSTER. trROSETTA useful in cases where sequence homology is limited (and Alphafold/ROSETTAfold!).
- Secondary structure, TM & disorder prediction (XtalPRED for overall summary; specific tools such as SPOT-DISORDER, SPIDER3 for best accuracy).
- Contact prediction from evolutionary couplings: EVFOLD & GREMLIN.
- Conservation analysis: Use favorite MSA algorithm (MUSCLE & CLUSTAL-OMEGA work well; TM-COFFEE, PRALINE-TM useful for membrane proteins) to create a sequence alignment of your protein with a few orthologs; gaps & insertions most commonly occur in loops/disordered regions. Useful as a guide during building.

Prep for model building - what can we learn from the sequence alone?

Your protein sequence contains a lot of useful information which you can use to aid model building:

- Start by identifying boundaries of conserved domains (NCBI CDD: <u>https://www.ncbi.nlm.nih.gov/Structure/cdd/</u>; DELTA-BLAST also performs CD-search by default)
- Then identify and/or generate suitable structural templates for building known domains: FUGUE, PHYRE2, MUSTER. trROSETTA useful in cases where sequence homology is limited (and Alphafold/ROSETTAfold!).
- Secondary structure, TM & disorder prediction (XtalPRED for overall summary; specific tools such as SPOT-DISORDER, SPIDER3 for best accuracy).
- Contact prediction from evolutionary couplings: EVFOLD & GREMLIN.
- Conservation analysis: Use favorite MSA algorithm (MUSCLE & CLUSTAL-OMEGA work well; TM-COFFEE, PRALINE-TM useful for membrane proteins) to create a sequence alignment of your protein with a few orthologs; gaps & insertions most commonly occur in loops/disordered regions. Useful as a guide during building.

CDD provides a guide to domain level architecture, including sequence alignments & representative structures.

3 NCBI			Conserved Domains			9955 779955	
ME SEARCH GUID	NewSe	earch Structure Hor	me 3D Macromol	lecular Structures	Conserved Domain	ns Pubchem B	ioSystems
Conserved (XYR1_HUMAN R) Protein Class SPRY1_RyR and F protein Constantion (domains yanodine re lification RR_TM4-6 dd	On [sp P21817] ceptor 1 OS=Homo sap	oiens OX=9606 GN=RYR	1 PE=1 SV=3 696388)		View Standard Resu	its 文 🛛
Graphical sur	nmary 📋	Zoom to residue level	show extra options »				2
	1	1000	2000	3000	40,00	5038	
Query seq.	1 I I						
Specific hits	HIR	SPRY SPR S SPRY SPR S SPR SPR SPR S	PRV RVDR_I SPRV			RR_TH Ion_t	
Non-specific hits	In						
Superfamilies	IIF	YDR_ SPRY Ry Ry SPR S	PRY RYDR_I	Ry Ry	RIH	RR_TH4-6 Ion_t	
	MIR s						
List of domai	n hits		Search for similar domain a		Refine search		
List of domai	n hits Accession	1	Search for similar domain a	rchitectures 2	Refine search	Interval	2 E-value
List of domai	n hits Accession pfam01365	RIH domain; The RIH (Ry	Search for similar domain a R and IP3R Homology) domai	In is an extracellular dor	Refine search	Interval 443-636	2.80e-83
List of domai	n hits Accession pfam01365	RIH domain; The RIH (Ry RIH domain; The RIH (Ry	Search for similar domain a R and IP3R Homology) domai R and IP3R Homology) domai	In is an extracellular dor in is an extracellular dor in is an extracellular dor	Refine search 7	Interval 443-636 2159-2369	E-value 2.80e-83 6.02e-82
List of domai	n hits Accession pfam01365 pfam01365 cd12878	RIH domain; The RIH (Ry RIH domain; The RIH (Ry SPRY domain 2 (SPRY2)	Search for similar domain a R and IP3R Homology domai of ryanodine receptor (RyR); Carbination	In is an extracellular dor in is an extracellular dor in is an extracellular dor This SPRY domain (SP	Refine search 2 main from two types main from two types RY2) is the second of	Interval 443-636 2159-2369 1072-1204	E-value 2.80e-83 6.02e-82 1.78e-81
List of domai Name H RYDR_ITPR H RYDR_ITPR SPRY2_RyR H RR_TM4-6	n hits Accession pfam01365 pfam01365 cd12878 pfam06459 pfam06459	RIH domain; The RIH (Ry RIH domain; The RIH (Ry SPRY domain 2 (SPRY2) Ryanodine Receptor TM 4 SPBY domain 2 (SPRY2)	Search for similar domain a R and IP3R Homology) domai R and IP3R Homology) domai R and IP3R Homology) domai of ryanodine receptor (R)R%) 46; This region covers TM reg dramandine corector (R)R%)	chitectures	Refine search 2 main from two types main from two types Reference of the second of he receptor 1	Interval 443-636 2159-2369 1072-1204 4383-4671 641-702	E-value 2.80e-83 6.02e-82 1.78e-81 3.36e-80
List of domai Name H RYDR_ITPR H RYDR_ITPR H SPRYZ_RYR RTM4-6 SPRY1_RYR SPRY1_RYR SPRY1_RYR	n hits Accession pfam01365 pfam01365 cd12878 pfam06459 cd12877 cd12870	RIH domain; The RIH (Ry RIH domain; The RIH (Ry SPRY domain 2 (SPRY2) Ryanodine Receptor TM SPRY domain 1 (SPRY1) SPRY domain 3 (SPRY3)	Search for similar domain a R and IP3R Homology) domai R and IP3R Homology) domai of ryanodine receptor (RyR); 145; This region covers TM reg of ryanodine receptor (RyR); of noncodine receptor (RyR);	rchitectures ^[2] I Description In is an extracellular dor This SPRY domain (SP ions 4-6 of the ryanodir This SPRY domain (SP This SPRY domain (SP This SPRY domain (SP	Refine search main from two types main from two types RY2) is the second of e frest of three PY3 is the entry of PY3 is the entry of	Interval 443-636 2159-2369 1072-1204 4383-4671 642-793 1418-1566	E-value 2.80e-83 6.02e-82 1.78e-81 3.36e-80 1.03e-79
List of domail Name RYDR_ITPR H RYDR_ITPR H SPRY2_R/R R TM4-6 SPRY1_R/R H SPRY3_R/R H ls182_R/R	n hits Accession pfam01365 cd12878 pfam06459 cd12877 cd12877 cd12879 pfam068709	RIH domain; The RIH (Ry RIH domain; The RIH (Ry SPRY domain 2 (SPRY2) Ryanodine Receptor TM 4 SPRY domain 3 (SPRY3) Institut 1 4 5-trisphoenbalt	Search for similar domain a R and IP3R Homology) domai R and IP3R Homology) domai dryanodine receptor (RyR): 46; This region covers TM reg of ryanodine receptor (RyR): for ryanodine receptor (RyR):	Pescription In is an extracellular dori This SPRY domain (SP ions 4-6 of the ryanodi This SPRY domain is th This SPRY domain is th This SPRY domain (SP	Refine search main from two types main from two types her receptor 1 to first of firee RY3) is the third of kinand hinding.	Interval 443-636 2159-2369 1072-1204 4383-4671 642-793 1418-1566 8-203	2 E-value 2.80e-83 6.02e-82 1.78e-81 3.36e-80 1.03e-79 4.00e-76 5.34e-76
List of domai Name H RYDR_ITPR H RYDR_ITPR H SPRY1_RVR H SPRY1_RVR H SPRY1_RVR H SPRY3_RVR H Ins145_P3_rec H MIR	n hits Accession pfam01365 pfam01365 cd12878 pfam06459 cd12877 cd12879 pfam08709 pfam08709	RIH domain; The RIH (Ry RIH domain; The RIH (Ry SPRY domain 2 (SPRY2) Ryanodine Receptor TM 4 SPRY domain 1 (SPRY1) SPRY domain 3 (SPRY3) Inositol 1,4,5-trisphosphat MIR domain: The MIR (br	Search for similar domain a R and IP3R Homology domain R and IP3R Homology domain Ar and IP3R Homology domain of ryandine receptor (RyR); dof, This regord covers (RyR); of ryandine receptor (RyR); ter/yandine receptor (RyR); ter/yandine receptor (RyR);	construction in is an extracellular dori in s an extracellular dori in s an extracellular dori This SPRY domain (SP ions 4-6 of the ryanodir This SPRY domain (SP main corresponds to th Sa and RvR) domain is	Refine search (2) main from two types main from two types RY(2) is the second of to first of three receptor 1 is domain that of is domain that may	Interval 443-636 2159-2369 1072-1204 4383-4671 642-793 1418-1566 8-203 211-389	2.80e-83 6.02e-82 1.78e-81 3.36e-80 1.03e-79 4.00e-76 5.34e-76 1.70e-73
List of domai Name RYDR_ITPR + RYDR_ITPR + SPRY2_RyR + SPRY1_RyR + SPRY3_RyR + Ins145_P3_rec + MIR + RR_MR	N hits Accession pfam01365 pfam01365 cd12878 pfam06459 cd12877 cd12877 pfam08709 pfam02815 ofam020206	RIH domain; The RIH (Ry RIH domain; The RIH (Ry SPRY domain 2 (SPRY2) SPRY domain 2 (SPRY2) SPRY domain 1 (SPRY1) SPRY domain; 1 (SPRY1) MIR domain; The MIR (pr RyR domain; The domain;	Search for similar domain a R and IP3R Homology) domai R and IP3R Homology) domai of ryanodine receptor (RyR); 145; This region covers TM reg of ryanodine receptor (RyR); 167; Anodine receptor (RyR); 167; Anodine receptor, This do tote mannosyltransferase, IP is called RyR for Ryanodine	chitectures Comparison	Refine search 2 main from two types main from two types Refront two types re receptor 1 re receptor 1 re receptor 1 receptor 1	Interval 443-636 2159-2369 1072-1204 4383-4671 642-793 1418-1566 8-203 211-389 850-940	E-value 2.80e-83 6.02e-82 1.78e-81 3.36e-80 1.03e-79 4.00e-76 5.34e-76 1.70e-73 1.97e-44
List of domai Name H RYDR_ITPR H SPRY2_R/R H RR_TM4-6 H SPRY1_R/R H SPRY3_R/R H Ins145_P3_rec H MIR H RyR H RyR	n hits Accession pfam01365 pfam01365 cd12878 pfam06459 cd12877 cd12877 pfam08709 pfam028709 pfam02026 pfam02026	RIH domain; The RIH (Ry SPRV domain ; The RIH (Ry SPRV domain 2 (SPRV2) SPRV domain 1 (SPRV1) Inositol 1 (4,5-trisphosphat RIR domain; The MIR (pr RyR domain; This domain RyR domain; This domain	Search for similar domain a R and IP3R Homology) domai R and IP3R Homology) domai of ryanodine receptor (RyR); 1-65; This region covers TM reg of ryanodine receptor (RyR); te/ryanodine receptor (RyR); te/ryanodine receptor (RyR); is called RyR for Ryanodine to is called RyR for Ryanodine to	construction c	Refine search 2 main from two types main from two types main from two types RY(2) is the second of e frist of three RY(3) is the third of ligand binding a domain that may found in four	Interval 443-638 2159-2389 1072-1204 4383-4671 642-793 1418-1566 8-203 211-389 850-940 2735-2825	E-value 2.80e-83 6.02e-82 1.78e-81 3.36e-80 1.03e-79 4.00e-76 5.34e-76 1.70e-73 1.97e-44 1.25e-41
List of domai	n hits Accession pfam01365 cd12878 pfam0459 cd12877 cd12877 cd12879 pfam02026 pfam02026 pfam02026 pfam02026	RIH domain; The RIH (Ry SPRY domain 2 (SPRY2) Ryanodine Receptor TM SPRY domain 1 (SPRY1) SPRY domain 3 (SPRY3) notatio 1,4,5-trisphesphat MR domain; The MR (pr RyR domain; This domain RyR domain; This domain RyR domain; This domain	Search for similar domain a R and IP3R Homology) domai R and IP3R Homology) domai of ryanodine receptor (RyR); 149; This region covers TM reg of ryanodine receptor (RyR); 164; Grandine receptor (RyR); 164; Grandine receptor, This do totein mannosyltransferase, IP is called RyR for Ryanodine r is called RyR for Ryanodine r is called RyR for Ryanodine r	contribution contreserve contreserve contreserve contreserve contr	Refine search Refine search Refine search Refine two types Refine two types Reform two types Reform two types Reform two types Reform the two types Reform	Interval 443-636 2159-2369 1072-1204 4383-4671 642-733 1418-1566 8-203 211-389 850-940 2735-7825 964-1054	2 E-value 2.80=83 6.02e-82 1.78e-81 3.36e-80 1.03e-79 4.00e-76 5.34e-76 1.70e-73 1.97e-44 1.25e-41 6.92e-38
List of domai Name H RVDR_ITPR H RYDR_ITPR H SPRY2_RyR H SPRY1_RyR H SPRY1_RyR H Ins145_P3_rec H Ins145_P3_rec H MIR H RyR H RyR H RyR	n hits Accession pfam01365 cd12877 cd12877 cd12877 cd12877 pfam02026 pfam02026 pfam02026 pfam02026	RIH domain; The RIH (Ry SPRV domain 2 (SPRV2) SPRV domain 2 (SPRV2) SPRV domain 3 (SPRV3) Inositol 1,4,5-trisphosphat SPRV domain 3 (SPRV3) Inositol 1,4,5-trisphosphat NR domain; This domain RyR domain; This domain RyR domain; This domain	Search for similar domain a R and IP3R Homology) domai R and IP3R Homology) domai of ryanodine receptor (RyR); 1-6; This region covers TM reg of ryanodine receptor (RyR); te/ryanodine receptor (RyR); is called RyR for Ryanodine r is called RyR for Ryanodine r	In is an extracellular dor in is an extracellular dor in is an extracellular dor This SPRY domain (SP This SPRY domain (SP main corresponds to th SR and RyR) domain (si receptor. The domain is receptor. The domain is receptor. The domain is receptor. The domain is	Refine search 2 main from two types main from two types RY(2) is the second of e frest of three RY(3) is the third of ligand binding a domain that may found in four found in four found in four found in four	Interval 443-636 2159-2369 1072-1204 4383-4671 642-793 11418-1566 8-203 211-389 850-940 2735-2825 964-1054 265-2839	E-value 2.80e-83 3.602e-82 1.78e-81 3.36e-80 1.03e-79 4.00e-76 5.34e-76 1.70e-73 1.97e-44 1.25e-41 6.92e-38 8.71e-36
List of domai Name + RYDR_ITPR + RYDR_ITPR + SPRY2_RyR + SPRY1_RyR + SPRY1_RyR + SPRY1_RyR + RYR + RYR + RyR + RyR + RyR + SPRY1 + SPRY + SPRY + RyR + RyR + SPRY + SPR	HIR # Accession pfam01365 pfam01365 cd12878 pfam06459 cd12877 cd12879 pfam08709 pfam02026 pfam02026 pfam02026 pfam02026 pfam02026	RIH domain: The RIH (Ry SPRY domain 2 (SPRY) Ryanodine Receptor TM SPRY domain 1 (SPRY1) SPRY domain 1 (SPRY1) SPRY domain 2 (SPRY1) Inostal 1 1,4 5,4 5(SPRY3) MIR domain; The MR (Ry RyR domain; The domain RyR domain; The domain	Search for similar domain a R and IP3R Homology) domai R and IP3R Homology) domai of ryanodine receptor (RyR); 49: This region covers TM reg of ryanodine receptor (RyR); 16: dynamic exceptor, This do otein mamosyltransferase, IP is called RyR for Ryanodine r is called RyR for Ryanodine r	Contraction Contracti	Refine search Refine search Refine search Refine two types hain from two types hain from two types hain from two types he receptor 1 hain that may found in four	Interval 443-835 2159-2369 1072-1204 4383-4671 642-733 1418-1566 8-203 211-389 850-940 2735-2825 964-1054 2855-2939 1084-1206	P E-value 2.80e-83 6.02e-82 1.78e-81 3.36e-80 1.03e-79 4.00e-76 5.34e-76 1.79e-44 1.25e-38 8.71e-36 6.98e-35
List of domai Name H RVDR ITPR H RVDR ITPR H SPRY2_RyR H RF_TM4-6 H SPRY1_RyR H SPRY1_RyR H SPRY3_RyR H RyR H RyR H RyR H RyR H SPRY H SPRY H SPRY	HIR 	RIH domain; The RIH (Ry SPRV domain 2 (SPRV2) SPRV domain 2 (SPRV2) SPRV domain 3 (SPRV3) Inositol 1,4.5-trisphosphal RR domain; The domain RyR domain; This domain RyR domain; This domain Domain in SPIa and the R SPRV domain; SPRV Dom	Search for similar domain a R and IP3R Homology) domain R and IP3R Homology) domain of ryanodine receptor (RyR); 145; This region covers TM reg of ryanodine receptor (RyR); te/ryanodine receptor (RyR); is called RyR for Ryanodine r is called RyR for Ryanodine r tyanobie Roceptor; Domain c main is named from SPla and I	chitectures	Refine search 2 main from two types main from two types RY2) is the second of e first of three RY3) is the third of bigand binding a domain that may found in four found in four	Interval 443-636 2159-2869 1072-1204 4383-4671 642-793 11418-1566 8-203 211-389 850-840 2735-2825 964-1054 265-2939 1084-1206 1084-1206	P E-value 2.80e-83 1.78e-81 3.36e-80 1.03e-79 4.00e-76 5.34e-76 1.70e-73 1.25e-41 6.92e-38 8.71e-36 6.63e-31
List of domai Name H RVDR_ITPR H SPRVTPR H SPRVT2_RyR H SPRVT2_RyR H Inst45_P3_rec H MRR H RyR H	HIR = Accession pfam01365 cd12878 cd12877 cd12877 cd12879 pfam02815 pfam02815 pfam02826 pfam02026 pfam02026 pfam020226 pfam00429 pfam00429 pfam00429	RiH domain: The RiH (Ry SPRV domain 2 (SPRV2) SPRV domain 2 (SPRV2) SPRV domain 3 (SPRV3) Incatol 1,4,5-trisphosphal MiR domain: The MIR (pr SPRV domain 3 (SPRV3) Incatol 1,4,5-trisphosphal MiR domain; The domain RyR domain; The domain RyR domain; The domain RyR domain; SPRV Dom SPRV domain; SPRV Dom	Search for similar domain a R and IP3R Homology domain and IP3R Homology domain of ryanodine receptor (RyR); 49; This region covers TM reg of ryanodine receptor (RyR); a ryanodine receptor (RyR); is called RyR for Ryanodine i is called RyR for Ryanodine i ryanodine receptor, This dow is called RyR for Ryanodine i is called RyR for Ryanodine i associated; This outAryotic do main is named from SPIa and 1 associated; This outAryotic do	rechitectures	Refine search Refine search Refine search RY2) is the second of RY2) is the second of RY2) is the second of is receptor 1 is domain that may found in four	Interval 443-835 2159-3389 1072-124041 642-793 1418-1566 8-203 211-388 850-840 2735-2825 964-1054 2255-2939 1084-1206 33773-9992	E-value 2.80e-83 6.02e-82 1.78e-81 3.36e-80 1.03e-79 4.00e-76 5.34e-76 1.70e-73 1.97e-44 1.25e-41 6.92e-38 8.71e-36 6.63e-35 6.63e-31 2.58e-30
List of domai RyDR_ITPR RYDR_ITPR SRYD2_FYR SRYD2_FYR SRYT1_FYR SRYT1_FYR SRYT1_FYR SRYT1_FYR SRYT1_FYR SRYT1_FYR SRYT1 SRYT1 SRYT1 SRYT1 SRYT1 SRYT1 SRYT1 SRYT1 SRYT1 SRYT1 SRYT1 SRYT1 SRYT1 SRYT1 SRYT1 SRYT1 SRYT1 SRYT1 SRYT1	HIR • Accession pfam01365 pfam01365 od12878 od12878 pfam08709 pfam02026 pfam02026 pfam02026 pfam02026 pfam02026 pfam02026 pfam02026 pfam02026 pfam02026 pfam00622 pfam08454 pfam08454	RIH domain; The RIH (Fty SPRV domain 2 (SPRV2) SPRV domain 2 (SPRV2) SPRV domain 3 (SPRV3) Inositol 1,4,5-trisphosphat SPRV domain; The domain RVR domain; The domain RVR domain; This domain Domain in SPIa and the 6 SPRV domain; SPRV Dor RVR and IP3R Homology SPRV domain; SPRV Dor RVR and IP3R Homology	Search for similar domain a R and IP3R Homology) domain R and IP3R Homology) domain of ryanodine receptor (RyR); 145; This region covers TM reg of ryanodine receptor (RyR); te/syanodine receptor (RyR); is called RyR for Ryanodine r is called RyR for Ryanodine r tyanoine Receptor, Domain c main is named from SPla and I associated; This eukaryotic d main is named from SPla mot	chitectures	Refine search 22 main from two types main from two types RY2) is the second of e receptor 1 e first of three RY3) is the third of bigand binding a domain that may found in four found in four for Comain of unknown for Domain of unknown	Interval 443-636 2159-2369 1072-1204 4383-4671 642-793 11418-1566 8-203 211-389 850-940 2735-2825 964-1054 265-2939 1084-1206 3873-3992 660-755	E-value 2.80e-83 6.02e-82 1.78e-81 3.36e-80 1.03e-79 4.00e-76 5.34e-76 1.70e-73 1.97e-44 1.25e-41 6.92e-38 8.71e-36 6.98e-35 6.63e-31 2.58e-30 7.17e-28
List of domai Name HYDR ITPR HYDR ITPR HYDR ITPR HSPRY2 RyR H SPRY1 RyR HSPRY1 RyR HSPRY HRYR HSPRY HSPRY HSPRY HSPRY HOLTANS	HIR = Accession pfam01365 pfam01365 pfam01365 qfam021365 qfam0249 pfam022815 pfam02206 pfam022026 pfam02202 pfam02222 pfam0494 pfam00429 pfam00622 pfam0622	RiH domain: The RiH (Ry SPRV domain 2 (SPRV2) SPRV domain 2 (SPRV2) SPRV domain 3 (SPRV3) Inositol 1 4,5 trisphosphat SPRV domain 3 (SPRV3) Inositol 1 4,5 trisphosphat RyR domain: This domain RyR domain: This domain RyR domain: This domain RyR domain: SPRV Dom SPRV domain; SPRV Horology SPRV domain; SPRV Don Ion transpot protein: This	Search for similar domain a R and IP3R Homology domain R and IP3R Homology domain Ar and IP3R Homology domain of ryanodine receptor (RyR); d ryanodine receptor (RyR); is called RyR for Ryanodine is called RyR for Ryanodine r vanodine receptor, This do sacolated; This eukaryotic main is named from SPIa and f associated; This eukaryotic main is named from SPIa and f and the solution of the solution of the solution of the solution of the solution of the solution of the solution and the solution of the solution of the solution of the solution of the solution of th	Trobitoctures 2 Description In is an octracellutar do This SPRY domain (SP Inis SPRY domain (SP) Inis SPRY domain (SP	Refine search Refine search RY2) is the second of Is a domain that may found in four	Interval 443-835 2159-3389 11072-1204 4383-4671 642-783 1148-1566 8-203 211-389 855-9409 2735-2825 964-1054 2255-2939 1004-1206 1008-1206 33773-5992 8660-759 4765-4499	2 E-value 2.80e-83 6.02e-82 1.78e-81 1.36e-80 1.03e-79 4.00e-76 5.34e-76 1.70e-73 1.97e-44 1.25e-41 6.92e-38 8.71e-36 6.63e-31 2.58e-30 7.17e-28 1.68e-25
List of domai Propagators RYDR ITPR RYDR ITPR RYDR ITPR RYDR ITPR RYDR ITPR SPRY2_RYR RR, TM4-5 SPRY1_RYR RYR RYR RYR SPRY1 SPRY	HIB • Accession pfam01365 pfam01365 of12878 pfam08499 of12879 pfam08709 pfam02266 pfam02026 pfam02020 pfam02020 pfam02020 pfam06222 pfam06202 pfam06202 pfam06202 pfam06202 pfam06202 pfam06202 pfam06202 pfam06202 pfam0622	RIH domain; The RIH (Fty SPRV domain 2 (SPRV2) SPRV domain 2 (SPRV2) SPRV domain 3 (SPRV3) Inositol 1,4,5-trisphosphat SPRV domain; The domain RyR domain; The domain RyR domain; The domain Domain in SPIa and the 6 SPRV domain; SPRV Dor RyR and IP3R Homology SPRV domain; SPRV Dor SPRV domain; SPRV Dor SPRV domain; SPRV Dor SPRV domain; SPRV Dor SPRV domain; SPRV Dor Jon transport protein; This Jona domain SPRV domain; SPRV Dor Jon transport protein; This Jona SPRV domain; SPRV Dor Jon transport protein; This Jona	Search for similar domain a R and IP3R Homology) domai R and IP3R Homology) domai of ryanodine receptor (RyR); 145; This region covers TM reg of ryanodine receptor (RyR); te/ryanodine receptor (RyR); is called RyR for Ryanodine r is called RyR for Ryanodine r tyanobian Receptor, Domain c main is named from SPla and family contains sodurim, potas main is named from SPla and	Productores 22 Control 10 Control	Refine search 22 main from two types main from two types RY2) is the second of e receptor 1 e first of three RY3) is the third of bigand binding a domain that may found in four found in four found four four four four four four four four	Interval 443-636 2159-2369 11072-1204 4383-4671 642-793 11418-1566 8-203 211-389 850-940 2735-2825 964-1054 2855-2939 1084-1206 3873-3992 660-795 4765-4499 1431-1565	E-value 2.80e-83 6.02e-82 1.76e-81 3.36e-80 1.03e-79 4.00e-76 5.34e-76 1.70e-73 1.97e-44 1.25e-41 6.98e-35 6.63e-31 2.58e-30 7.17e-28 1.68e-25 4.81e-24
List of domai Name H RYDR, ITPR H RYDR, ITPR H SPRY2, RyR H SRRY2, RyR H SRRY1, RyR H SPRY1, RyR H RyR H RyR H RyR H SPRY H SP	HIE = Accession pfam01365 pfam01365 pfam01365 qfam01365 qfam021365 qfam02026 pfam02026 pfam02026 pfam02026 pfam02022 pfam02022 pfam00222 pfam00429 pfam00620 pfam00620 pfam00622 pfam00620 pfam00622 pfam00620 pfam06620 pfam0660 p	RiH domain: The RiH (Ry RiH domain: The RiH (Ry SPRV domain 2 (SPRV2) SPRV domain 3 (SPRV3) Inositol 14,5-trisphosphat SPRV domain 3 (SPRV3) Inositol 14,5-trisphosphat RyR domain: This domain RyR domain: This domain Domain in SPR and the R SPRV domain; SPRV Dor SPRV domain; SPRV Dor SPRV domain; SPRV Dor Don transport protein: This SPRV domain; SPRV Dor Domain in SPR and the R	Search for similar domain a R and IP3R Homology) domain R and IP3R Homology) domain R and IP3R Homology domain of ryanodine receptor (RyR); 146; This region covers TM reg of ryanodine receptor (RyR); 161; This algoing the second receptor is called RyR for Ryanodine of is called RyR for Ryanodine of trandnice receptor, Domain of trandnice receptor, Domain of saccided RyR form Sylan and family contains sodium, polar and family contains sodium, polar and tranane from SPIa and t sanains in named from SPIa and t Yanodine Receptor, Domain of	Trachitectures 2 Construction In is an extractilutar don In is an extractilutar don In is an extractilutar don In is SPRY domain (sP main corresponds to the main co	Refine search Refine search Refine search Refine two types Refine two types Refine two types Refine the second of se receptor 1 found in four found in	Interval 443-836 2159-2389 1072-1204 4383-4671 442-733 1418-1586 8-203 211-389 255-8825 964-1024 964-1024 964-1024 1084-1206 1084-1206 1084-1206 1084-1205 4765-4949 1431-1568 1430-1563	2 E-value 2.80e-83 6.02e-82 1.78e-81 3.36e-80 1.03e-79 4.00e-76 5.34e-76 6.34e-76 6.32e-38 8.71e-36 6.92e-38 8.71e-36 6.92e-38 8.71e-36 6.92e-38 8.71e-36 6.98e-35 6.63e-31 2.58e-30 7.17e-28 1.88e-25 4.81e-24 4.81e-24 4.81e-24 3.22e-23 0.45e-45 1.48e-25 4.81e-24 1.48e-25 4.81e-24 1.48e-25 4.81e-26 1.48e-25 1.48
List of domai Name Propa TrpR RTDR, TrpR RTDR, TrpR RTDR, TrpR RTDR, TrpR RTTR, TRT	HIE Accession pfam01365 cd12877 pfam01485 cd12877 cd12877 cd12877 pfam02459 cd12877 pfam02459 pfam02459 pfam02459 pfam02206 pfam022026 pfam02020 pfam06222 pfam00622 pfam00622 pfam00622 pfam00522 pfam00622 pfam00522 pfam00622 pfam00522 pfam00522 pfam00522 pfam00522 pfam0524 pfam0622 pfam0525 pfam0524 pfam0525 pfam0525 pfam0525 pfam0525	RIH domain; The RIH (Fty SPRV domain 2 (SPRV2) SPRV domain 2 (SPRV2) SPRV domain 3 (SPRV3) Inositol 1,4,5-trisphosphat SPRV domain; The domain RyR domain; The domain RyR domain; The domain Domain in SPIa and the R SPRV domain; SPRV Dor RyR and IP3R Homology SPRV domain; SPRV Dor SPRV domain; SPRV Dor SPRV domain; SPRV Dor Domain in SPIa and the R Domain in SPIa and the R Domain in SPIa and the R	Search for similar domain a R and IP3R Homology) domain R and IP3R Homology) domain of ryanodine receptor (RyR); 145; This region covers TM reg of ryanodine receptor (RyR); te/syanodine receptor (RyR); is called RyR for Ryanodine r is called RyR for Ryanodine r tyanobine Receptor; Domain c family contains south from SPla and 1 family contains south from SPla and Vanodine Receptor; Domain c Yanodine Receptor; Domain c Yanodine Receptor; Domain c	The intervention of the second	Refine search 2 main from two types main from two types RY2) is the second of e receptor 1 e first of three RY3) is the third of bigand binding a domain that may found in four found in four f	Interval 443-636 2159-2369 11072-1204 4383-4671 642-793 11418-1566 8-203 211-389 850-940 2735-2825 964-1054 2855-2939 1084-1206 3873-3992 660-795 4765-4499 1431-1568 1430-1558 660-794	E-value 2.80e-83 6.02e-82 1.78e-81 3.36e-80 1.03e-79 4.00e-76 5.34e-76 1.70e-73 1.97e-44 1.25e-41 6.92e-38 8.71e-36 6.98e-35 6.63e-31 8.71e-36 6.98e-35 6.63e-31 1.68e-25 4.81e-24 3.22e-23 6.04e-19
List of domai Name H RYDR JTPR H RYDR JTPR H SPRY2 RyR H SRRY2 RyR H SRRY1 RyR H SPRY1 RyR H RyR H RyR H RyR H SPRY H	HIE Accession pfam01365 pfam01365 pfam01365 cd12878 pfam06479 pfam06479 pfam02026 pfam02026 pfam02226 pfam02026 pfam02226 pfam02026 pfam02226 pfam020226 pfam020226 pfam00422 pfam00422 pfam000520 pfam100439 smarti00448	RiH domain: The RiH (Ry RiH domain: The RiH (Ry SPRY domain 2 (SPRY2) SPRY domain 3 (SPRY3) Inositol 14,5-trisphosphat SPRY domain 3 (SPRY3) Inositol 14,5-trisphosphat RyR domain: This domain RyR domain: This domain RyR domain: This domain Domain in SPIs and the R SPRY domain: SPRY Do RyR and IP3R Homology SPRY domain: SPRY Do In transport protein: This SPRY domain: SPRY Do Tomain in SPIs and the R Domain in SPIs and the R	Search for similar domain a R and IP3R Homology) domain R and IP3R Homology) domain for anodine receptor (RyR); 145; This region covers TM reg of ryanodine receptor (RyR); 160; This region covers TM reg 16 called RyR for Ryanodine in 16 called RyR for Ryanodine in 18 called RyR for Ryanodine in 18 called RyR for Ryanodine in 18 called RyR for Ryanodine in 19 called RyR for Ryanodine in 20 called Ry	rechitectures	Refine search 2 main from two types main from two types RY(2) is the second of to first of three RY(3) is the third of bigand binding a domain that may found in four found in four	Interval 443-836 2159-2389 1072-1204 4383-4671 4383-4671 4383-4671 4842-793 211-389 250-3400 2785-8825 950-400 2785-8825 964-1026 1086-1206 1086-1206 1086-1206 1086-1206 1085-1205 4785-4949 1431-1558 6667-794 4083-4133	E-value 2.80e-83 1.76e-81 3.36e-80 1.03e-79 4.00e-76 5.34e-76 1.97e-44 1.25e-41 6.92e-38 8.71e-35 6.63e-31 2.56e-30 7.17e-28 1.86e-25 4.81e-24 4.81e-24 4.81e-24 3.22e-23 6.04e-19 5.82e-08

CDD provides a guide to domain level architecture, including sequence alignments & representative structures.

S NCBI		Conserved Store Lyde		
ME SEARCH GUID	NewSea	rch Structure Home 3D Macromolecular Structures Conserved Domains	Pubchem B	ioSystems
Conserved	domains	on [sp P21817] Vie	w Standard Resul	ts ᅌ 🛛
RYR1_HUMAN R	yanodine rec	eptor 1 OS=Homo sapiens OX=9606 GN=RYR1 PE=1 SV=3		
Protein Class	ification		Í.	()
SPRY1_RyR and I protein containing	RR_TM4-6 do domains RYDR_	main-containing protein (domain architecture ID 11696388) ITPR, SPRY1_RyR, SPRY2_RyR, and RR_TM4-6		
Graphical su	mmary 🗌	Zoom to residue level show extra options »		?
0	1	1000 2000 3000 4000	5038	
Non-specific hits	In HIR RY	S SPRY SPR SPRY SPRY SPRY SPRY SPRY SPRY	Ton_t	
Superfamilies	I I RY	RE_SPRY BY BY SPR SPRY RYDR_I BY RY RIH E BR_TH+	6 Ion_t	
		Search for similar domain architectures P Refine search 7		
List of domai	n hits			()
Name	Accession	Description	Interval	E-value
RYDR ITPR	pfam01365	RIH domain; The RIH (RyR and IP3R Homology) domain is an extracellular domain from two types	2159-2369	2.00e-03 6.02e-82
SPRY2 RyR	cd12878	SPRY domain 2 (SPRY2) of ryanodine receptor (RyR); This SPRY domain (SPRY2) is the second of	1072-1204	1.78e-81
RR_TM4-6	pfam06459	Ryanodine Receptor TM 4-6; This region covers TM regions 4-6 of the ryanodine receptor 1	4383-4671	3.36e-80
SPRY1_RyR	cd12877	SPRY domain 1 (SPRY1) of ryanodine receptor (RyR); This SPRY domain is the first of three	642-793	1.03e-79
SPRY3_RyR	cd12879	SPRY domain 3 (SPRY3) of ryanodine receptor (RyR); This SPRY domain (SPRY3) is the third of	1418-1566	4.00e-76
Ins145_P3_rec	pfam08709	Inositol 1,4,5-trisphosphate/ryanodine receptor; This domain corresponds to the ligand binding MIP domain: The MIP (protein manage/transforase, IP3P and PvP) domain is a domain that may	8-203	5.34e-76
RvR	pfam02026	RvR domain: This domain is called RvR for Rvanodine receptor. The domain is found in four	850-940	1.70e-73
RyR	pfam02026	RyR domain; This domain is called RyR for Ryanodine receptor. The domain is found in four	2735-2825	1.25e-41
RyR	pfam02026	RyR domain; This domain is called RyR for Ryanodine receptor. The domain is found in four	964-1054	6.92e-38
RyR	pfam02026	RyR domain; This domain is called RyR for Ryanodine receptor. The domain is found in four	2855-2939	8.71e-36
SPRT	nfam00622	SPRY domain: SPRY Domain is named from SPIa and the RYanodine Recentor. Distant nomologues are	1086-1206	6.630-31
RIH assoc	pfam08454	RyR and IP3R Homology associated: This eukaryotic domain is found in ryanodine receptors (RyR)	3879-3992	2.58e-30
SPRY	pfam00622	SPRY domain; SPRY Domain is named from SPIa and the RYanodine Receptor. Domain of unknown	660-795	7.17e-28
] lon_trans	pfam00520	Ion transport protein; This family contains sodium, potassium and calcium ion channels. This	4765-4949	1.68e-25
determines ion s membrane.	electivity. In sor	contains soduim, potassium and calcium ion channels. Inis tamily is transmembrane neices in which the last two n ne sub-families (e.g. Na channels) the domain is repeated four times, whereas in others (e.g. K channels) the protein f Pssm-ID: 334124 [Multi-domain] Cd Length: 237 Bit Score: 107.36 E-value: 1.68e-25	elices flank a loop w orms as a tetramer i	nich n the
sp P21 Cdd:pfa	817 4765 am00520 40	10 20 30 40 50 60 70 80 *		
	817 4830 am00520 120	90 100 110 120 130 140 150 160 SVENKGQLAVVYLLTVVAFNFFRFYNKSBEDERPENKCOMMTCYLFERKYVYSAGGGGDELEDAGDE 4909 SLIRSLKSLGNLLLLLLFLFIFAIIGYQLFGKFYTWENPDMGRTNFDNFPNAFLMLFQTWTTEGMCDILYDTIDGK 197		
sp P21 Cdd:pf				
sp P211 Cdd:pfa sp P211 Cdd:pfa	817 4910 am00520 198	170 180 190 200 *		
sp P21 Cdd:pfi Sp P21 Cdd:pfi	817 4910 am00520 198 pfam00622	170 180 190 200 * YELYRVVPDTPTPFVIULIATIOGLIDAPGERROOG 4949 GEWATIFVSTILGOFLILIARIOUTION/GUTERTE 237 SPRY domain; SPRY Domain is named from SPIa and the RYanodine Receptor. Domain of unknown	1431-1568	4.81e-24
sp P21 Cdd:pfi Sp P21 Cdd:pfi Cdd:pfi	817 4910 am00520 198 pfam00622 smart00449	170 180 190 200 YELYRWYDFITFFFFVUTLLALTIOGLTDAYGELDOOG 4949 GSFWATIFYFSFILGOPLLANEFGVTIDNYGELTBERTE 237 SPRY domain: SPRY Domain is named from SPIa and the RYanodine Receptor. Domain of unknown Domain in SPIa and the RYanodine Receptor. Domain of unknown function. Distant homologues are	1431-1568 1430-1568	4.81e-24 3.22e-23
sp P21; Cdd : pf; sp P21; Cdd : pf; d SPRY H SPRY H SPRY H SPRY H SPRY	817 4910 am00520 198 pfam00622 smart00449 smart00449	170 180 190 200 VELTAVEDITEFFUTUTILATION VELTAVEDITEFFUTUTILATION VELTAVEDITEFFUTUTILATION SPRY domain: SPRY Domain is named from SPIa and the RYanodine Receptor. Domain of unknown Domain in SPIa and the RYanodine Receptor; Domain of unknown function. Distant homologues are Domain in SPIa and the RYanodine Receptor; Domain of unknown function. Distant homologues are Domain in SPIa and the RYanodine Receptor; Domain of unknown function. Distant homologues are	1431-1568 1430-1568 660-794	4.81e-24 3.22e-23 6.04e-19
sp P211 Cdd : pfi Cdd : pfi Cdd : pfi H SPRY H SPRY H SPRY H SPRY H EF-hand_8 H MIR	817 4910 am00520 198 pfam00622 smart00449 smart00449 pfam13833 smart00472	170 180 190 200 YELKRWVPDITFFYVIVLIALIGGLITDARGELDOOG 4949 GSFWATIFFVSVILLAURIFUVTIONFORLPORTE 237 SPRV domain: SPRV Damain is mand from SPIs and the RYanodine Receptor. Domain of unknown Domain in SPIs and the RYanodine Receptor; Domain of unknown function. Distant homologues are EF-hand domain pair; Domain in contained and the RYanodine Receptor; Domain of unknown function. Distant homologues are EF-hand domain pair;	1431-1568 1430-1568 660-794 4083-4133 210-263	4.81e-24 3.22e-23 6.04e-19 5.82e-08 8.98e-08

CDD provides a guide to domain level architecture, including sequence alignments & representative structures.

	NCBI		k e T F 1 k d T Y ~ p Q L A ~ s Q L (~ s Q L (Conserved Domains	SH3	SH2	150 027207 1965 27375965		
ME SE/	ARCH GUIDE	NewSea	rch Structure Hom	ie 3D Macro	molecular Structures	Conserved	d Domains P	ubchem B	ioSystems
Cons	erved do	mains	on [sp P21817]				View	Standard Result	ts ᅌ 🛛
۲YR1_H	HUMAN Ryar	odine rec	eptor 1 OS=Homo sapi	iens OX=9606 GN=F	XYR1 PE=1 SV=3				
Prote	in Classifi	cation							?
orotein	_RyR and RR_ containing dom	_TM4-6 doi nains RYDR_	nain-containing protein ITPR, SPRY1_RyR, SPRY2_	(domain architecture ID _RyR, and RR_TM4-6	11696388)				
Grapi	hical sumn	nary 🔲	Zoom to residue level	show extra options »					()
	1		1000	2000	3000	400	ao	5038	
Juery Specif	seq. `ic hits	HIR P	R SPRY SPR SPR SPRY SPR SP SPR SPR SP	RY RYDR_I		•	R.TM	Ion_t	
lon-sp nits Superf	ecific I amilies	I I RM	IR_ SPRV RV RV SPR SPI	RY RYDR_I	Ry Ry	RIM	E RR_TH4-6	Ion_t	
				Search for similar doma	in architectures	Refine search			
List o	of domain	hits Accession			Description			Interval	() E-value
+] RYD	R_ITPR p	fam01365	RIH domain; The RIH (RyR	र and IP3R Homology) do	main is an extracellular	domain from two types		443-636	2.80e-83
+] RYDI	R_ITPR p V2 RvR c	lam01365	RIH domain; The RIH (RyR SPRY domain 2 (SPRY2) (₹ and IP3R Homology) do of rvanodine recentor (Rv	main is an extracellular	domain from two types . SPRY2) is the second o	ï	2159-2369	6.02e-82
RR 1	TM4-6 p	fam06459	Ryanodine Receptor TM 4-	-6; This region covers TM	I regions 4-6 of the ryance	dine receptor 1		4383-4671	3.36e-80
SPR	Y1_RyR o	d12877	SPRY domain 1 (SPRY1) o	of ryanodine receptor (Ry	R); This SPRY domain is	the first of three		642-798	1.030-79
SPR	Y3_RyR c	d12879	SPRY domain 3 (SPRY3) c	of ryanodine receptor (Ry	R); This SPRY domain (SPRY3) is the third of		1418-1566	4.000-76
	15_P3_rec p	/am08709 fam02815	MIR domain: The MIR (pro	/ryanodine receptor; This tein mannosyltransferase	domain corresponds to IP3R and RvR) domain	the ligand binding		8-203 211-389	5.34e-76
RvR	p	fam02026	RvR domain: This domain	is called RvR for Rvanodi	ine receptor. The domain	is found in four		850-940	1.97e-44
RyR	p	fam02026	RyR domain; This domain	is called RyR for Ryanodi	ine receptor. The domain	is found in four		2735-2825	1.25e-41
RyR	p	fam02026	RyR domain; This domain i	is called RyR for Ryanodi	ine receptor. The domain	n is found in four		964-1054	6.92e-38
RyR	p	fam02026	RyR domain; This domain i	is called RyR for Ryanod	ine receptor. The domain	is found in four		2855-2939	8.71e-36
+ SPR	Y si	mart00449	Domain in SPIa and the RY	ranodine Receptor; Doma	ain of unknown function.	Distant homologues are)	1084-1206	6.98e-35
I SPR	Y p	am00622	SPRY domain; SPRY Dom	ain is named from SPIa a	ind the RYanodine Rece	ptor. Domain of unknow	n	1086-1206	6.63e-31
H KIH_	assoc p	/am08454	SPRV domain: SPRV Dom	issociated; This eukaryoti	c domain is found in rya	nodine receptors (RyR)		3879-3992	2.588-30
TI CDD	p p	amooozz	or it'r domain, or it'r bom	ain is named nom of la a	ind the renarioune rece	DIOL DOILIGHT OF UTKIDW	71		· · · · · · · · · · · · · · · · · · ·
+] SPR	trans p	fam00520	Ion transport protein; This I	family contains sodium, p	otassium and calcium io	n channels. This	n	4765-4949	1.68e-25
+] SPR' -] lon_t lon tr deter mem	rans protein ransport protein rmines ion sele- brane.	fam00520 n; This family ctivity. In sor	Ion transport protein; This f contains sodium, potassiun ne sub-families (e.g. Na cha Pssm-ID: 2	family contains sodium, p n and calcium ion channe innels) the domain is repe 334124 [Multi-domain] C	otassium and calcium io els. This family is 6 transi eated four times, whereas d Length: 237 Bit Scor	n channels. This membrane helices in wh s in others (e.g. K chanr re: 107.36 E-value: 1.6	n hich the last two helice hels) the protein form 8e-25	4765-4949 es flank a loop w s as a tetramer in	1.68e-25 hich n the
+] SPR -] lon_t lon tr deter mem	rans port protein ransport protein rmines ion selec brane. sp P21817 Cdd:pfam0	fam00520 r; This family ctivity. In sor 4765 0520 40	Ion transport protein: This i contains sodium, potassium ne sub-families (e.g. Na cha pssm-ID: 10 20 *	family contains sodium, p and calcium ion channe nnels) the domain is repe 334124 [Multi-domain] C 0 30 * tFTDN- KrryLRSPwnildfvvvlp	All Stand And Calcium to botassum and calcium to botassum and calcium to bota standard standa	n channels. This membrane helices in wh s in others (e.g. K chanr re: 107.36 E-value: 1.6 60 70 	II ich the last two helici- hels) the protein form i8e-25 80 .* TLRTILS 4829 GLRTLVN 119	4765-4949 es flank a loop w s as a tetramer i.	1.68e-25 hich n the
+] SPR -] lon_t lon tr deter mem	rans protei ansport protei mines ion sele- brane. sp P21817 Cdd:pfam0 sp P21817 Cdd:pfam0	fam00520 ;; This family ctivity. In sor 	Ion transport protein: This contains sodium, potassium, es sub-families (e.g. Na cha Pssm-ID: 10 20 10 20 10 10 10 10 10 10 10 10 10 10 10 10 10	family contains sodium, p. and calcium ino channe mannels) the domain is repersively a solution of the domain is repersively a solution of the domain of the domain of the domain is repersively a solution of the domain of th	Adassum and calcum to bis. This family is 6 trans- ated four times, wherea d Length: 237 Bit Scor 40 50	channels. This mothrane helics in why s in others (e.g. K channels) fe: 60 70	ich the last two helic ich the jottein form i8e-25 80 TIRRIUS 4829 GIRTLIN EDEPAGDE 190 TUTKEX	4765-4949 es flank a loop w is as a tetramer i	1.68e-25 hich n the
+) SPR -) Ion_t Ion tr deter mem	rans proteir rmines ion sele- brane. sp P21817 Cdd:pfam0 sp P21817 Cdd:pfam0 sp P21817 Cdd:pfam0	fam00520 x; This family ctivity. In sor 4765 0520 40 4830 0520 120 4910 0520 198	Ion transport protein. This contains sodium, potassium, ne sub-families (e.g. Na cha Pesm-ID: 10 21 21 21 21 21 21 21 21 21 21 21 21 21	family contains sodium, p. family contains sodium, p. and calcium in channe mannels) the domain is reperiod. 334124 [Multi-domain] C 0 30 1	Addassum and calcum to blas. This family is 6 trans- stated four times, wherea the Length: 237 Bit Scot 40 50 10 5	n channels. This membrane helices in wh s in others (e.g. K chanr re: 107.36 E-value: 1.6 60 70 140 70 140 150 	H ich the last two helico- ich the protein form i8e-25 80 * TIRYILS 4829 GLRTLWN 119 160 * EDEPAGDE 4909 YDTIDGK 197	4765-4949 es flank a loop w s as a tetramer i	1.68e-25 hich n the
+) SPR -) Ion t Ion t deter mem +) SPR'	rans p ransort proteir mines ion sele- brane. sp P21817 Cdd:pfam0 sp P21817 Cdd:pfam0 y P21817	fam006220 1: This family 1:	Ion transport protein: This Ion transport protein: This contains sodium, potassium, potassi potassium, potassium, potassium, potassium, potassium,	family contains sodium. p and calcium in channe nnneis) the domain is repe 334124 [Multi-domain] C 0 30 1	Additional and calcium of a cal	n channels. This membrane helices in wh s in others (e.g. K channels (e.g. K channels) (e.g. K cha	n iich the last two helicit iich the last two helicit #86-25 80 * TXRTLS 4829 GLRTLVN 119 160 .* 160 .* 177 TVTIDGK 197 n	4765-4949 es flank a loop w s as a tetramer i 1431-1568	1.68e-25 hich n the 4.81e-24
+) SPR -) Ion tr lon tr deter mem +) SPR	rans p rans p ransport probin ransport probin sp P21817 Cdd:pfam0 sp P21817 Cdd:pfam0 sp P21817 Cdd:pfam0	fam00520 ;; This family ctivity. In sor 	Ion transport protein: This contains sodium, potassium, ne sub-families (e.g. Na cha Pssm-ID: 10 22 Pssm-ID: 10 22 10 22 10 22 10 20 20 20 20 20 20 20 20 20 20 20 20 20	family contains sodium; p and calcium in channe and calcium in channe and calcium in channe status in the second second associated with the second associated with the second second second second second the second second second second second second second second second the second se	Addagem and calcum to be This family is 6 trans- sated four times, wherea d Length: 237 Bit Scot 40 50 	n channels. This membrane helices in wh s in others (e.g. K chanr re: 107.36 E-value: 1.6 60 70 10	n ich the last two helico ich the rotain form i8e-25 80 *	4765-4949 es flank a loop w s as a tetramer l 1431-1568 1430-1568	1.68e-25 hich n the 4.81e-24 3.22e-23
+) SPR - lon t lon t deter mem + SPR + SPR + SPR + SPR	rans p ransort problem rmines ion sele- brane. sp P21817 Cdd:pfam0 sp P21817 Cdd:pfam0 Y Y Sort sele- sp P21817 Cdd:pfam0	fam00520 ; This family ctivity. In sor 4765 0520 40 520 120 4910 0520 138 fam00622 nart00449 nart00449 nart09449	Ion transport protein: This contains sodium, potassium, ne sub-families (e.g. Na cha Pesm-ID: 10 20 10 20 10 10 20 10 10 20 10 10 10 10 10 10 10 10 10 10 10 10 10	family contains sodium, p. family contains sodium, p. and calcium in channe and calcium in channe and calcium in channe 334124 [Multi-domain] C 0 30 1*	Addission and calcium of shorts and the standard standa	thannels. This mombrane helics in wh sin others (e.g. K chan re: 107.36 E-value: 1.6 60 70 70 70 70	n iich the last two helicio- lich the protein form iike-25 80 * TLRYILS 4829 GLRTLVN 119 * EEDPAGDE 4909 YDTIDGK 197 n	4765.4949 es flank a loop w a sa a teframer i 1431-1568 1432-1568 600-794	4.81e-24 4.81e-24 3.22e-25 6.04e-15
+ SPR' lon tr deter mem + SPR' + SPR' + SPR' + SPR' + SPR'	rans p ransort problem ransport problem sappl P21817 Cdd:pfam0 sp P21817 Cdd:pfam0 y p Y ss y ss and_8 p	fam00520 ; This family ctivity. In sor 47655 0520 40 0520 120 0520 120 0520 138 fam00622 mart00449 mart00449 mart00449 mart00449	Ion transport protein: This contains sodium, potassium, ne sub-families (e.g. Na cha Pssm-ID: 10 22 LLTRUKSIDVKYQINKFQCV VFTGIFTLEMLLKIITAGGP 90 100 	family contains sodium; p and calcium in channe nnnels) the domain is repe 334124 [Multi-domain] C 0 30 1** I*PDN= * I*PDN= * I*PDN= * I*.	A Second	n channels. This membrane helicos in wh s in others (e.g. K chanr re: 107.36 E-value: 1.6 60 70 INNEFFAARLEDIANGOV T.RELAVIRLEALTRELES 140 150 	n ich ihe last two helioo ich ihe protein form i8e-25 n 	4765-4949 es flank a loop was as a letramer i s as a tetramer i 1431-1568 1430-1568 660-704 4083-133 210-963	1.68e-22 hich n the 3.22e-23 6.04e-15 5.82e-02 8.98e-04

Once an initial trace is obtained for these regions, use DALI or PDBeFold to identify structural homologs that could not be identified by sequence alone.

Prep for model building - what can we learn from the sequence alone?

Your protein sequence contains a lot of useful information which you can use to aid model building:

- Start by identifying boundaries of conserved domains (NCBI CDD: <u>https://www.ncbi.nlm.nih.gov/Structure/cdd/</u>; DELTA-BLAST also performs CD-search by default)
- Then identify and/or generate suitable structural templates for building known domains: FUGUE, PHYRE2, MUSTER. trROSETTA useful in cases where sequence homology is limited (and Alphafold/ROSETTAfold!).
- Secondary structure, TM & disorder prediction (XtalPRED for overall summary; specific tools such as SPOT-DISORDER, SPIDER3 for best accuracy).
- Contact prediction from evolutionary couplings: EVFOLD & GREMLIN.
- Conservation analysis: Use favorite MSA algorithm (MUSCLE & CLUSTAL-OMEGA work well; TM-COFFEE, PRALINE-TM useful for membrane proteins) to create a sequence alignment of your protein with a few orthologs; gaps & insertions most commonly occur in loops/disordered regions. Useful as a guide during building.

XtalPRED is a great tool for summarizing predicted sequence properties.



Highlights predicted secondary structure, disorder, low complexity regions on sequence in an easily digestible format. Useful to print and consult while building. Also provides list of structural homologs. (http://ffas.burnham.org/XtalPred-cgi/xtal.pl)

(Also consider using some of the newer single purpose neural-network based classifiers; e.g. SPIDER-3 & SPOT-DISORDER-SINGLE from Yaoqi Zhou lab: <u>http://sparks-lab.org/index.php/Main/Services</u>)

Secondary structure prediction is a very useful guide when building.





Where is this motif in the sequence?

Secondary structure prediction is a very useful guide when building.



Secondary structure prediction is ~80% accurate. So if your model consistently disagrees with predicted secondary structure, look at it very closely!

What can we learn from the map alone?



What can we learn from the map alone?



Left handed! Obvious here – can be less clear at lower res, so be careful.

OK, that's better! What can we learn from the map alone?



Which direction does the helix point?



Which direction does the helix point?



Helices – alpha and 3₁₀





Alpha

- ~90%
- 3.6 residues per turn
- Fat

3₁₀

- ~10%. More common in TM? (e.g. S4 of VSD)
- 3 residues per turn. Triangular cross section.
- Skinny
- Can be tricky to identify at low resolution, can lead to register errors.













Notice that the **absence** of large sidechain densities at small residue positions is just as valuable in validating the fit as the fit of large sidechains to the density.



Also, note that the information content of local regions varies. Consider "VTVVAASSTVV" vs "FGAAYWVTRA" – which is more likely to be uniquely identifiable from the map?



CryoID can help when you don't even know the sequence!

- Similar approach codified and automated in the "cryoID" program – but in this case, starting from the density, with no sequence input!
- Split map into fragments
- Use reduced complexity pseudo-sequence to convert map fragments into motifs which can be used to search sequence database.
- Identify most likely candidate sequence, combine fragments and rebuild.
- Useful when purifying from endogenous sources, where composition may not be known.



(Ho et al., Nature Methods, 2020)
How to deal with uncertainty in sequence assignment and sidechain placement

- You will likely encounter situations where you cannot be certain of the local sequence register what to do?
- No clear consensus, but I suggest assigning residue code as "UNK" and numbering to "best guess" value. A more granular way to quantify/convey uncertainty would be helpful!
- Sidechain placement two main camps trim sidechains to density vs place them all (+/zero occ.). The former may sound more conservative, but it can hide errors during validation (during analysis of clashes). Either is acceptable, just be consistent, and preferably outline the approach taken when writing up the structure.



Prior knowledge can come in many forms – use any and all available info to guide model building.

Here, serendipitous identification of a conformational class of RyR1 lacking density for one subunit aided identification of protomer boundaries. In other cases, cross-linking data or NS data on subcomplexes or Fabcomplexes may be helpful. (Zalk et al, Nature 2015) In a similar manner, we can use locally aligned difference maps between holo and apo structures to locate ligands.



The three ligands are clustered around the C-terminal domain.



The three ligands are clustered around the C-terminal domain.



EM-specific considerations

- No unambiguous sequence markers at low resolution (no equivalent of SeMet).
- No feedback from phase improvement, but also no model bias WYSIWIG.
- Often substantial variation in local resolution different strategies and levels of detail required for different regions. Map sharpening essential.
- "Medium" resolution (4-6Å) much more common than for crystallography.
- Often have more than one map, with different composition or conformation (may be convenient to combine focused refinements in Chimera by taking max value at each voxel after alignment, e.g.: vop maximum #1,2 ongrid #1)



Building an initial model - where to start?

- If you have a crystal structure, of a fragment or a homology model of a domain, place it, and extend into density. (*Now, Alphafold & Rosettafold mean this is almost always the case*)
- Otherwise, identify structurally distinctive motifs in the sequence – for example, a strongly predicted helix with three aromatic residues near the N-term end – and identify candidate locations in the density map. Extend and see if hypothesis still holds.





Start with map and model.





Move model to approximate position (if known, to save computation)





Run fitmap with 'search' (here 100 orientations) and 'radius' (here 5 Å)



Using UCSF Chimera to fit solved domains

Chimera will return a list of candidate orientations, ranked by agreement with the map. Hopefully there will be a clear separation between the correct and incorrect solutions.

Using UCSF Chimera to fit solved domains



Chimera will return a list of candidate orientations, ranked by agreement with the map. Hopefully there will be a clear separation between the correct and incorrect solutions.

Using UCSF Chimera for voxel size calibration (of your map and others)

- Voxel size generally requires calibration against a crystal structure.
- Once calibrated, generally stable between samples/datasets at same magnification.
- Can calibrate by fitting in Chimera at range of nominal voxel sizes and measuring correlation.
- Incorrect voxel sizes are common in deposited maps - be aware of this when comparing structures. E.g. here there is a 3% difference – affects structural alignment, reported resolution (3.8 vs 3.9Å).





- Simple, intuitive interface for building and manipulating atomic models in density maps.
- Low computational requirements
- Extensive API easy to script or modify (using simple Python code)
- On-the-fly sharpening and low pass filtering (for MTZ).



(Try the latest nightly with new features for EM, improved RSR: http://www.ccpem.ac.uk/download.php) (*Emsley P. 2004, Acta Cryst. D; Casañal A. et al. 2020, Protein Science*)

- Simple, intuitive interface for ٠ building and manipulating atomic models in density maps.
- Low computational ٠ requirements
- Extensive API easy to script or ٠ modify (using simple Python code)
- On-the-fly sharpening and low pass filtering (for MTZ).

••• 0	olibclarke/coot-trimmings: Py × +					
🗲 🛈 🔒 GitHut	, Inc. (US) https://github.com/olibclarke/coot-trimmin	ngs C	얒 Search	☆ 自 💟 🦊 🎓 🦂 🚇	* `ਰ' ≡	
Most Visited -	🛞 Read Later 🛛 🧕 Getting Started 🛛 🕵 NEBcutter	- ORF S 🛞 Lab Book 🍀 cry	oSPARC 🛛 🏶 cryoSPARC (home)			
	📮 olibclarke / coot-trimmings		⊙ Unwatch ◄	1 🛨 Star 0 😵 Fork	0	
	<> Code ① Issues 0 ① Pull req	uests 0 III Projects 0 III	Wiki 🧄 Pulse 🔝 Graphs	🌣 Settings		
	Python customizations for the macro Add topics	molecular model building softw	are Coot.	Ed	lit	
	Image: Beach and Second Sec		\heartsuit O releases	L 1 contributor		
	Branch: master - New pull request		Create new file Upload fi	les Find file Clone or download	•	
	olibclarke Fixed a bug in find_sequence.			Latest commit dc0e68b 12 days ago		
	README.md	Update README.md	2 months ago equence. 12 days ago			
	oli_custom.py	Fixed a bug in find_sequence.				
	I README.md					
	COOT-trimmings Python customizations for the F Copy to ~/.coot-preferences (h You should see a new menu ("C (e.g. "sequence context").	nacromolecular model building idden dir, copy on command lii Custom") and a bunch of new k	software Coot. ne) and restart coot. ey bindings, as well as a couple	e of new toolbar buttons		

• •

- Simple, intuitive interface for building and manipulating atomic models in density maps.
- Low computational requirements
- Extensive API easy to script or modify (using simple Python code)
- On-the-fly sharpening and low pass filtering (for MTZ).

O olibclarke/coot-trimmings: Py × +					
GitHub, Inc. (US) https://github.com/olibclarke/	coot-trimmings	C Search	☆ 自 ♥ 🖡 ⋒ 🦂 🐵 → 🔶	ੱਡ =	
it Visited 👻 🛞 Read Later 🛛 😻 Getting Started 🏾 🌔	🕅 NEBcutter - ORF S 🛞 Lab Book 🛞 🗱 ci	ryoSPARC 🛛 🎇 cryoSPARC (home)			
📮 olibclarke / coot-trimm	ings	O Unwate	ch ▼ 1 ★ Star 0 % Fork 0		
<> Code () Issues ()	🕅 Pull requests 0 🔳 Projects 0 🛙	🗉 Wiki 🔸 Pulse 🔟 Graph	is 🔅 Settings		
Python customizations for the Add topics	ne macromolecular model building soft	ware Coot.	Edit		
🕞 8 commits	₽ 1 branch	\bigcirc O releases	a contributor		
Branch: master - New pull requ	est	Create new file Uploa	ad files Find file Clone or download -		
olibclarke Fixed a bug in find	sequence	Latest commit dc0e68b 12 d			
README.md	Update README.md		2 months ago		
icustom.py	Fixed a bug in find_sequence.		12 days ago		
I README.md					
coot-trimmi	ngs				
Python customization:	Python customizations for the macromolecular model building software Coot.				
Copy to ~/.coot-prefe	rences (hidden dir, copy on command	line) and restart coot.			
You should see a new (e.g. "sequence conte	You should see a new menu ("Custom") and a bunch of new key bindings, as well as a couple of new toolbar buttons (e.g. "sequence context").				

Any Python (or Scheme) file you put in ~/.coot-preferences will be executed when starting Coot. Can use this for extra key bindings, scripts, custom functions.

- Simple, intuitive interface for building and manipulating atomic models in density maps.
- Low computational requirements
- Extensive API easy to script or modify (using simple Python code)
- On-the-fly sharpening and low pass filtering (for MTZ).

def mutate_by_entered_code(): def mutate_single_letter(X): entry=str(X).upper() mol_id=active_residue()[0] ch_id=active_residue()[1] resno=active_residue()[2] ins_code=active_residue()[3] resname=residue_name(mol_id,ch_id,resno,ins_code) map_id=imol_refinement_map() aa_dic={'A':'ALA','R':'ARG','N':'ASN','D':'ASP','C':'CYS','E':'GLU','Q':'GLN','G':'GLY','H':'HIS','I':'ILE','L':'LEU','K':'LY nt_list=['A','C','T','G','U'] if (resname in aa_dic.values()) and (aa_dic.get(entry,0)!=0): mutate(mol_id,ch_id,resno,ins_code,aa_dic.get(entry,0)) elif (resname in nt_list) and (entry in nt_list): mutate_base(mol_id, ch_id, resno, ins_code, entry) else: info_dialog("Invalid target residue! Must be protein or nucleic acid, and entered code must be single letter.") generic_single_entry("New residue? (single letter code)","A","Mutate by single-letter code",mutate_single_letter)

#mutate active residue to entered residue code (upper or lower case single-letter)
add_key_binding("Mutate by single letter code","M",
lambda: mutate_by_entered_code())

- Simple, intuitive interface for building and manipulating atomic models in density maps.
- Low computational requirements
- Extensive API easy to script or modify (using simple Python code)
- On-the-fly sharpening and low pass filtering (for MTZ).

ef mutate_by_entered_code(): def mutate_single_letter(X): entry=str(X).upper() mol_id=active_residue()[0] ch_id=active_residue()[1] resno=active_residue()[2] ins_code=active_residue()[3] resname=residue_name(mol_id,ch_id,resno,ins_code) map_id=imol_refinement_map() aa_dic={'A':'ALA','R':'ARG','N':'ASN','D':'ASP','C':'CYS','E':'GLU','Q':'GLN','G':'GLY','H':'HIS','I':'ILE','L':'LEU','K':'LY nt_list=['A','C','T','G','U'] if (resname in aa_dic.values()) and (aa_dic.get(entry,0)!=0): mutate(mol_id,ch_id,resno,ins_code,aa_dic.get(entry,0)) elif (resname in nt_list) and (entry in nt_list): mutate_base(mol_id,ch_id,resno,ins_code,entry) else: info_dialog("Invalid target residue! Must be protein or nucleic acid, and entered code must be single letter.") generic_single_entry("New residue? (single letter code)","A","Mutate by single_letter code",mutate_single_letter)

#mutate active residue to entered residue code (upper or lower case single-letter)
add_key_binding("Mutate by single letter code","M",
lambda: mutate by entered code())

Many pre-packaged functions available in COOT API. Mostly documented in online manual. Very easy to write your own! Useful e.g. for scripting domain-wise rigid body refinement.

- Simple, intuitive interface for building and manipulating atomic models in density maps.
- Low computational requirements
- Extensive API easy to script or modify (using simple Python code)
- On-the-fly sharpening and low pass filtering (for MTZ).

Coot User Manual × +							
🔄 🛈 🔒 https://www2.mrc-lmb.cam.ac.uk/Personal/pemsley/coot/web/docs/coot.html 🖾 C 🔍 Search 🟠 🖨 💟 🦊 🏠 🥥 🔶 🦉 🚍							
🙍 Most Visited = 🕐 Read Later 👋 Getting Started 🎼 NEBcutter - ORF S 🕐 Lab Book 🛞 🔆 cryoSPARC 🎄 cryoSPARC (home)							
Next: ncs, Previous: filter, Up: Scheme Scripting Functions							
13.11 coot-gui							
— procedure: run-gtk-pending-events — procedure: coot-gui							
Fire up the coot scripting gui. This function is called from the main C++ code of coot. Not much use if you don't have a gui to type functions in to start with.							
- procedure: handle-smiles-go tlc-entry smiles-entry							
The callback from pressing the Go button in the smiles widget, an interface to run libcheck.							
— procedure: smiles-gui							
smiles GUI							
- procedure: generic-single-entry function-label entry-1-default-text go-button-label handle-go-function							
Generic single entry widget							
Pass the hint labels of the entries and a function that gets called when user hits "Go". The handle-go-function accepts one argument that is the entry text when the go button is pressed.							
- procedure: generic-double-entry label-1 label-2 entry-1-default-text entry-2-default-text check-button-label handle-check-button-function go-button-label handle-go-function							
handle-go-function takes 3 arguments, the third of which is the state of the check button.							
if check-button-label not a string, then we don't display (or create, even) the check-button. If it *is* a string, create a check button and add the callback handle-check-button- function which takes as an argument the active-state of the the checkbutton.							
- procedure: generic-multiple-entries-with-check-button entry-info-list check-button-info go-button-label handle-go-function							
generic double entry widget, now with a check button							
OLD:							

Many pre-packaged functions available in COOT API. Mostly documented in online manual.

- Simple, intuitive interface for building and manipulating atomic models in density maps.
- Low computational requirements
- Extensive API easy to script or modify (using simple Python code)
- On-the-fly sharpening and low pass filtering (for MTZ).



Successfully read coordinates file /Users/olibclarke/Dropbox/ryr_models_paper2/cam/reprocess/best_tet_resampled_apr12_rs2_real_space_refined.pdb...

Lots of key bindings, and easy to define custom keys. Learn them. They make everything much faster.

- Simple, intuitive interface for building and manipulating atomic models in density maps.
- Low computational requirements
- Extensive API easy to script or modify (using simple Python code)
- On-the-fly sharpening and low pass filtering (for MTZ).



Successfully read coordinates file /Users/olibclarke/Dropbox/ryr_models_paper2/cam/reprocess/best_tet_resampled_apr12_rs2_real_space_refined.pdb... 🖉

Lots of key bindings, and easy to define custom keys. Learn them. They make everything much faster.

- Semi-automated helix placement
- Place cursor at the center of the helix and trigger "Place helix here" (I suggest via a key binding - "h" with coottrimmings)
- Coot will attempt to automatically determine the length and direction of the helix.
- Trim/extend, adjust weights, then refine using real-space refine zone. Drag into density to adjust fit.



- Semi-automated helix placement
- Place cursor at the center of the helix and trigger "Place helix here" (I suggest via a key binding - "h" with coottrimmings)
- Coot will attempt to automatically determine the length and direction of the helix.
- Trim/extend, adjust weights, then refine using real-space refine zone. Drag into density to adjust fit.





- Semi-automated helix placement
- Place cursor at the center of the helix and trigger "Place helix here" (I suggest via a key binding - "h" with coottrimmings)
- Coot will attempt to automatically determine the length and direction of the helix.
- Trim/extend, adjust weights, then refine using real-space refine zone. Drag into density to adjust fit.





- Semi-automated helix placement
- Place cursor at the center of the helix and trigger "Place helix here" (I suggest via a key binding - "h" with coottrimmings)
- Coot will attempt to automatically determine the length and direction of the helix.
- Trim/extend, adjust weights, then refine using real-space refine zone. Drag into density to adjust fit.



(mol. no: 6) CA /1//-1 ALA occ: 1.00 bf: 50.00 ele: C pos: (251.69,255.83,193.56)

- Semi-automated helix placement
- Place cursor at the center of the helix and trigger "Place helix here" (I suggest via a key binding - "h" with coottrimmings)
- Coot will attempt to automatically determine the length and direction of the helix.
- Trim/extend, adjust weights, then refine using real-space refine zone. Drag into density to adjust fit.



(mol. no: 6) CA /1//-1 ALA occ: 1.00 bf: 50.00 ele: C pos: (251.69,255.83,193.56)



- Semi-automated helix placement
- Place cursor at the center of the helix and trigger "Place helix here" (I suggest via a key binding - "h" with coottrimmings)
- Coot will attempt to automatically determine the length and direction of the helix.
- Trim/extend, adjust weights, then refine using real-space refine zone. Drag into density to adjust fit.





- Semi-automated helix placement
- Place cursor at the center of the helix and trigger "Place helix here" (I suggest via a key binding - "h" with coottrimmings)
- Coot will attempt to automatically determine the length and direction of the helix.
- Trim/extend, adjust weights, then refine using real-space refine zone. Drag into density to adjust fit.





- Sequence assignment.
- Adjust numbering to match expected position in sequence.
- Mutate to match sequence
- Fill sidechains manually.
- Adjust sequence register to optimize local fit to sidechain densities.





- Sequence assignment.
- Adjust numbering to match expected position in sequence.
- Mutate to match sequence
- Fill sidechains manually.
- Adjust sequence register to optimize local fit to sidechain densities.



- Sequence assignment.
- Adjust numbering to match expected position in sequence.
- Mutate to match sequence
- Fill sidechains manually.
- Adjust sequence register to optimize local fit to sidechain densities.





- Sequence assignment.
- Adjust numbering to match expected position in sequence.
- Mutate to match sequence
- Fill sidechains manually.
- Adjust sequence register to optimize local fit to sidechain densities.





- Sequence assignment.
- Adjust numbering to match expected position in sequence.
- Mutate to match sequence
- Fill sidechains manually.
- Adjust sequence register to optimize local fit to sidechain densities.



(mol. no: 6) CA /1//12 ALA occ: 1.00 bf: 50.00 ele: C pos: (241.31,258.23,210.37)

Use 'Add Terminal residue' to extend chain.

ISOLDE

- Interactive molecular dynamics flexible fitting, implemented as plugin for ChimeraX
- Useful during "polishing" stage of generating a final model, identifying and fixing otherwise difficult to correct errors in geometry, non-bonded contacts. Physically realistic simulation guided by map, user input.
- Complementary to COOT COOT better for de novo building and assembly, ligand placement, ISOLDE very useful for final round of real space fitting.

(Croll, 2018, Acta. Cryst. D)

Types of errors in macromolecular models

- Identity (e.g. wrong domain)
- Directionality
- Topology/connectivity
- Register
- Rotamer
- Backbone torsion
- Ligand identification and placement

Types of errors in macromolecular models


Types of errors in macromolecular models

- Identity (e.g. wrong domain)
- Directionality
- Topology/connectivity
- Register
- Rotamer
- Backbone torsion
- Ligand identification and placement



Types of errors in macromolecular models

- Identity (e.g. wrong domain)
- Directionality
- Topology/connectivity
- Register
- Rotamer
- Backbone torsion
- Ligand identification and placement



Strategy for identifying and correcting errors.

- Analyse as you go "sanity checks" on chemistry, nonbonded interactions, surface composition. Use Molprobity for clashes, Chimera or pymol to check e.g. for buried polars, exposed hydrophobics. Monitor agreement with secondary structure, disorder predictions.
- Use EM-ringer (or Q-scores) to identify errors in backbone and rotamer geometry.
- Look at everything! Manually check and recheck the fit of every residue. Tedious but necessary.
- Sometimes, you just can't tell the right answer. Don't be afraid to specify sequence ambiguity (use UNKs).
- Half-map FSCs are only really useful to analyse overfitting they tell you little about the local quality or correctness of the model.

Finally...

"ALL MODELS ARE WRONG, BUT SOME ARE USEFUL" - George P. Box

*

It should be remembered that just as the Declaration of Independence promises the <u>pursuit</u> of happiness rather than happiness itself, so the iterative scientific model building process offers only the pursuit of the perfect model. For even when we feel we have carried the model building process to a conclusion some new initiative may make further improvement possible. Fortunately to be useful a model does not have to be perfect.

George P. Box, "Robustness in the Strategy of Scientific Model Building", 1979

Thank you for listening!



Model building tutorial



Tutorial PDF: https://bit.ly/2XPsiox

Data: https://bit.ly/3ASQ411

AlphaFold add on: https://bit.ly/3KTo6qX