

# 1

## Supplementary materials

### 1.1 Amino acid sequence alignments

---

For the four amino acid sequence alignments below, **5GJV\_A\_Fasta** represents the amino acid sequence (in fasta format) for the Chain A in the protein structure (PDB ID: 5GJV), while **5GJW\_A\_Fasta** represents the amino acid sequence (in fasta format) for the Chain A in the protein structure (PDB ID: 5GJW).

**5gjb.pdb\_1276** represents the amino acid sequence (in fasta format) for the Chain A in the protein structure (PDB ID: 5GJV) extracted from the PDB file 5gjb.pdb, and 1276 represents the total number of amino acid residues whose atomic coordinates were experimentally determined.

**5gjw.pdb\_1291** represents the amino acid sequence (in fasta format) for the Chain A in the protein structure (PDB ID: 5GJW) extracted from the PDB file 5gjw.pdb, and 1291 represents the total number of amino acid residues whose atomic coordinates were experimentally determined.

5GJV_A_Fasta	MEPSSPQDEGLRKKQPKKPLPEVLPRPPRALFCLT	35
5GJW_A_Fasta	MEPSSPQDEGLRKKQPKKPLPEVLPRPPRALFCLT	35
5GJV_A_Fasta	LQNPLRKACISIVEWKPFETIILLTIFANCVALAV	70
5GJW_A_Fasta	LQNPLRKACISIVEWKPFETIILLTIFANCVALAV	70
5GJV_A_Fasta	YLPMPEDDNNNSLNLGLEKLEYFFLTVFSIEAAMKI	105
5GJW_A_Fasta	YLPMPEDDNNNSLNLGLEKLEYFFLTVFSIEAAMKI	105
5GJV_A_Fasta	IAYGFLFHQDAYLRSGWNVLDFFIIVFLGVFTAILE	140
5GJW_A_Fasta	IAYGFLFHQDAYLRSGWNVLDFFIIVFLGVFTAILE	140
5GJV_A_Fasta	QVNVIQSNTAPMSSKGAGLDVKALRAFRVLRPLRL	175
5GJW_A_Fasta	QVNVIQSNTAPMSSKGAGLDVKALRAFRVLRPLRL	175
5GJV_A_Fasta	VSGVPSLQVVLNSIFKAMLPLFHIALLVLFMVIIY	210
5GJW_A_Fasta	VSGVPSLQVVLNSIFKAMLPLFHIALLVLFMVIIY	210
5GJV_A_Fasta	AIIGLELFGKGMHKTCYYIGTDIVATVENEKPSPC	245
5GJW_A_Fasta	AIIGLELFGKGMHKTCYYIGTDIVATVENEKPSPC	245
5GJV_A_Fasta	ARTGSGRPCTINGSECRGGWPGPNHGITHFDNFGF	280
5GJW_A_Fasta	ARTGSGRPCTINGSECRGGWPGPNHGITHFDNFGF	280
5GJV_A_Fasta	SMLTVYQCITMEGWTDVLYWVND AIGNEW PWIYFV	315
5GJW_A_Fasta	SMLTVYQCITMEGWTDVLYWVND AIGNEW PWIYFV	315
5GJV_A_Fasta	TLILLGSFFILNLVLGVLSGEFTKEREKAKSRGTF	350
5GJW_A_Fasta	TLILLGSFFILNLVLGVLSGEFTKEREKAKSRGTF	350
5GJV_A_Fasta	QKLREKQOLEEDLRGYMSWITQGEVMDVEDLREGK	385
5GJW_A_Fasta	QKLREKQOLEEDLRGYMSWITQGEVMDVEDLREGK	385
5GJV_A_Fasta	LSLEEGGSDTESLYEIEGLNKIIQFIRHWRQWNRV	420
5GJW_A_Fasta	LSLEEGGSDTESLYEIEGLNKIIQFIRHWRQWNRV	420
5GJV_A_Fasta	FRWKCHDLVKSRVFYWLVLILIVALNTLSIASEHHN	455
5GJW_A_Fasta	FRWKCHDLVKSRVFYWLVLILIVALNTLSIASEHHN	455
5GJV_A_Fasta	QPLWLTHLQDIANRVLLSLFTIEMLLKMYGLGLRQ	490
5GJW_A_Fasta	QPLWLTHLQDIANRVLLSLFTIEMLLKMYGLGLRQ	490
5GJV_A_Fasta	YFMSIFNRFDCFVVCSGILELLLVESGAMTPLGIS	525
5GJW_A_Fasta	YFMSIFNRFDCFVVCSGILELLLVESGAMTPLGIS	525
5GJV_A_Fasta	VLRCIRLLRLFKITKYWTSLSNLVASLLNSIRSIA	560
5GJW_A_Fasta	VLRCIRLLRLFKITKYWTSLSNLVASLLNSIRSIA	560

5GJV_A_Fasta	SLLLLLFLFIIIFALLGMQLFGGRYDFEDTEVRRS	595
5GJW_A_Fasta	SLLLLLFLFIIIFALLGMQLFGGRYDFEDTEVRRS	595
5GJV_A_Fasta	NFDNFPQALISVFQVLTGEDWNSVMYNGIMAYGGP	630
5GJW_A_Fasta	NFDNFPQALISVFQVLTGEDWNSVMYNGIMAYGGP	630
5GJV_A_Fasta	SYPGVLVCIYFIIILFVCGNYILLNVFLAIAVDNLA	665
5GJW_A_Fasta	SYPGVLVCIYFIIILFVCGNYILLNVFLAIAVDNLA	665
5GJV_A_Fasta	EAESLTSAQKAKAEERKRRKMSRGLPDKTEEEKSV	700
5GJW_A_Fasta	EAESLTSAQKAKAEERKRRKMSRGLPDKTEEEKSV	700
5GJV_A_Fasta	MAKKLEQKPKGEGIPTTAKLKVDEFESNVNEVKDP	735
5GJW_A_Fasta	MAKKLEQKPKGEGIPTTAKLKVDEFESNVNEVKDP	735
5GJV_A_Fasta	YPSADFPGDDEEDEPEIPVSPRPRPLAELQLKEKA	770
5GJW_A_Fasta	YPSADFPGDDEEDEPEIPVSPRPRPLAELQLKEKA	770
5GJV_A_Fasta	VPIPEASSFFIFSPTNKVRVLCHRIVNATWFTNFI	805
5GJW_A_Fasta	VPIPEASSFFIFSPTNKVRVLCHRIVNATWFTNFI	805
5GJV_A_Fasta	LLFILLSSAALAAEDPIRAESVRNQILGYFDIAFT	840
5GJW_A_Fasta	LLFILLSSAALAAEDPIRAESVRNQILGYFDIAFT	840
5GJV_A_Fasta	SVFTVEIVLKMTTYGAFLHKGSGFCRNYFNILDLLV	875
5GJW_A_Fasta	SVFTVEIVLKMTTYGAFLHKGSGFCRNYFNILDLLV	875
5GJV_A_Fasta	VAVSLISMGLSSTISVVKILRVLRVLRPLRAINR	910
5GJW_A_Fasta	VAVSLISMGLSSTISVVKILRVLRVLRPLRAINR	910
5GJV_A_Fasta	AKGLKHVVQCVFVAIRTIIGNIVLVTTLLQFMFACI	945
5GJW_A_Fasta	AKGLKHVVQCVFVAIRTIIGNIVLVTTLLQFMFACI	945
5GJV_A_Fasta	GVQLFKGKFFSCNDLSKMTEEECRGYYYVYKDGD	980
5GJW_A_Fasta	GVQLFKGKFFSCNDLSKMTEEECRGYYYVYKDGD	980
5GJV_A_Fasta	TQELRPRQWIHNDHFHDNVLSAMMSLFTVSTFEG	1015
5GJW_A_Fasta	TQELRPRQWIHNDHFHDNVLSAMMSLFTVSTFEG	1015
5GJV_A_Fasta	WPQLLYRAIDSNEEDMGPVYNNRVEMAIFFFIIYII	1050
5GJW_A_Fasta	WPQLLYRAIDSNEEDMGPVYNNRVEMAIFFFIIYII	1050
5GJV_A_Fasta	LIAFFMMNIFVGFVIVTFQEQGETEYKNCELDKNQ	1085
5GJW_A_Fasta	LIAFFMMNIFVGFVIVTFQEQGETEYKNCELDKNQ	1085
5GJV_A_Fasta	RQCVQYALKARPLRCYIPKNPYQYQVWYVVTSSYF	1120
5GJW_A_Fasta	RQCVQYALKARPLRCYIPKNPYQYQVWYVVTSSYF	1120

5GJV_A_Fasta	EYLMFALIMLNTICLGMQHYHQSEEMNHISDILNV	1155
5GJW_A_Fasta	EYLMFALIMLNTICLGMQHYHQSEEMNHISDILNV	1155
5GJV_A_Fasta	AFTIIFTLEMILKLLAFKARGYFGDPWNVDFDLIV	1190
5GJW_A_Fasta	AFTIIFTLEMILKLLAFKARGYFGDPWNVDFDLIV	1190
5GJV_A_Fasta	IGSIIDVILSEIDTFLASSGGLYCLGGGCGNVDPD	1225
5GJW_A_Fasta	IGSIIDVILSEIDTFLASSGGLYCLGGGCGNVDPD	1225
5GJV_A_Fasta	ESARISSAFFRLFRVMRLIKLLSRAEGVRTLLWTF	1260
5GJW_A_Fasta	ESARISSAFFRLFRVMRLIKLLSRAEGVRTLLWTF	1260
5GJV_A_Fasta	IKSFQALPYVALLIVMLFFIYAVIGMQMFGKIALV	1295
5GJW_A_Fasta	IKSFQALPYVALLIVMLFFIYAVIGMQMFGKIALV	1295
5GJV_A_Fasta	DGTQINRNNNFQTFPQAVLLLFRCATGEAWQEILL	1330
5GJW_A_Fasta	DGTQINRNNNFQTFPQAVLLLFRCATGEAWQEILL	1330
5GJV_A_Fasta	ACSYGKLCDPESDYAPGEEYTCGTNFAYYYFISFY	1365
5GJW_A_Fasta	ACSYGKLCDPESDYAPGEEYTCGTNFAYYYFISFY	1365
5GJV_A_Fasta	MLCAFLIINLFFVAVIMDNFDYLTRDWSILGPHHLD	1400
5GJW_A_Fasta	MLCAFLIINLFFVAVIMDNFDYLTRDWSILGPHHLD	1400
5GJV_A_Fasta	EFKAIWAEYDPEAKGRIKHLDVVTLLRRIQPPLGF	1435
5GJW_A_Fasta	EFKAIWAEYDPEAKGRIKHLDVVTLLRRIQPPLGF	1435
5GJV_A_Fasta	GKFCPHRVACKRLVGMMNPLNSDGTVTFNATLFAL	1470
5GJW_A_Fasta	GKFCPHRVACKRLVGMMNPLNSDGTVTFNATLFAL	1470
5GJV_A_Fasta	VRTALKIKTEGNFEQANEELRAIIKKIWKRTSMKL	1505
5GJW_A_Fasta	VRTALKIKTEGNFEQANEELRAIIKKIWKRTSMKL	1505
5GJV_A_Fasta	LDQVIPPIGDDEVTVGKFYATFLIQEHFRKFMKRO	1540
5GJW_A_Fasta	LDQVIPPIGDDEVTVGKFYATFLIQEHFRKFMKRO	1540
5GJV_A_Fasta	EEYYGYRPKKDTVQIQAGLRTIEEEAAPEIRRTIS	1575
5GJW_A_Fasta	EEYYGYRPKKDTVQIQAGLRTIEEEAAPEIRRTIS	1575
5GJV_A_Fasta	GDLTAEELERAMVEAAMEERIFRRTGGLFGQVDT	1610
5GJW_A_Fasta	GDLTAEELERAMVEAAMEERIFRRTGGLFGQVDT	1610
5GJV_A_Fasta	FLERTNSLPPVMANQRPLQFAEIEEMEELESPVFLE	1645
5GJW_A_Fasta	FLERTNSLPPVMANQRPLQFAEIEEMEELESPVFLE	1645
5GJV_A_Fasta	DFPQDARTNPLARANTNNANANVAYGNSNHSNNQM	1680
5GJW_A_Fasta	DFPQDARTNPLARANTNNANANVAYGNSNHSNNQM	1680

5GJV_A_Fasta	FSSVHCEREFPGEAETPAAGRGALSHSHRALGPHS	1715
5GJW_A_Fasta	FSSVHCEREFPGEAETPAAGRGALSHSHRALGPHS	1715
5GJV_A_Fasta	KPCAGKLNQGLVQPGMPINQAPPAPCQQPSTDPPE	1750
5GJW_A_Fasta	KPCAGKLNQGLVQPGMPINQAPPAPCQQPSTDPPE	1750
5GJV_A_Fasta	RGQRRTSLTGSLQDEAPQRRSSEGSTPRRPAPATA	1785
5GJW_A_Fasta	RGQRRTSLTGSLQDEAPQRRSSEGSTPRRPAPATA	1785
5GJV_A_Fasta	LLIQEALVRGGLDTLAADAGFVTATSQALADACQM	1820
5GJW_A_Fasta	LLIQEALVRGGLDTLAADAGFVTATSQALADACQM	1820
5GJV_A_Fasta	EPEEVEVAATELLKARESVQGMASVPGSLSRSSL	1855
5GJW_A_Fasta	EPEEVEVAATELLKARESVQGMASVPGSLSRSSL	1855
5GJV_A_Fasta	GSLDQVQGSQETLIPRP	1873
5GJW_A_Fasta	GSLDQVQGSQETLIPRP	1873

☐ non conserved  
☒  $\geq 50\%$  conserved

**Figure 1.1** Amino acid sequence alignment of PDB ID 5GJV (fasta format) and PDB ID 5GJW (fasta format).

5g jv.pdb_1276	.....FCLT	4
5GJV_A_Fasta	MEPSSPQDEGLRKKQPKKPLPEVLPRPPRALFCLT	35
5g jv.pdb_1276	LQNPLRKACISIVEWKPFETIILLTIFANCVALAV	39
5GJV_A_Fasta	LQNPLRKACISIVEWKPFETIILLTIFANCVALAV	70
5g jv.pdb_1276	YLPMPEDDNNSLNLGLEKLEYFFLTVFSIEAAMKI	74
5GJV_A_Fasta	YLPMPEDDNNSLNLGLEKLEYFFLTVFSIEAAMKI	105
5g jv.pdb_1276	IAYGFLFHQDAYLRSGWNVLDFFIIVFLGVFTAIL.	108
5GJV_A_Fasta	IAYGFLFHQDAYLRSGWNVLDFFIIVFLGVFTAIL	140
5g jv.pdb_1276	.....VKALRAFRVLRPLRL	123
5GJV_A_Fasta	QVNVIQSNTPMSSKGAGLDVKALRAFRVLRPLRL	175
5g jv.pdb_1276	VSGVPSLQVVLNSIFKAMLPLFHIALLVLFMVIIY	158
5GJV_A_Fasta	VSGVPSLQVVLNSIFKAMLPLFHIALLVLFMVIIY	210
5g jv.pdb_1276	AIIGLELFKKGKMHKTCYYIGTDIVATVENEKPSPC	193
5GJV_A_Fasta	AIIGLELFKKGKMHKTCYYIGTDIVATVENEKPSPC	245
5g jv.pdb_1276	ARTGSGRPCTINGSECRGGWPGPNHGITHFDNFGF	228
5GJV_A_Fasta	ARTGSGRPCTINGSECRGGWPGPNHGITHFDNFGF	280
5g jv.pdb_1276	SMLTVYQCITMEGWTDVLYWVNDAIGNEWPIYFV	263
5GJV_A_Fasta	SMLTVYQCITMEGWTDVLYWVNDAIGNEWPIYFV	315
5g jv.pdb_1276	TLILLGSFFILNLVLGVLSGEFTKEREKAKSRGTF	298
5GJV_A_Fasta	TLILLGSFFILNLVLGVLSGEFTKEREKAKSRGTF	350
5g jv.pdb_1276	QKLREKQQLEEDLRGYMSWITQGEVM.....	324
5GJV_A_Fasta	QKLREKQQLEEDLRGYMSWITQGEVMDVEDLREGK	385
5g jv.pdb_1276	.....WNRV	328
5GJV_A_Fasta	LSLEEGGSDTESLYEIEGLNKIIQFIRHWRQWNRV	420
5g jv.pdb_1276	FRWKCHDLVKSRVFYWLVLILIVALNTLSIASEHNN	363
5GJV_A_Fasta	FRWKCHDLVKSRVFYWLVLILIVALNTLSIASEHNN	455
5g jv.pdb_1276	QPLWLTHLQDIANRVLLSLFTIEMLLKMYGLGLRQ	398
5GJV_A_Fasta	QPLWLTHLQDIANRVLLSLFTIEMLLKMYGLGLRQ	490
5g jv.pdb_1276	YFMSIFNRFDCFVVCSGILELLLVESGAMTPLGIS	433
5GJV_A_Fasta	YFMSIFNRFDCFVVCSGILELLLVESGAMTPLGIS	525
5g jv.pdb_1276	VLRCIRLLRLFKITKYWTSLSNLVASLLNSIRSIA	468
5GJV_A_Fasta	VLRCIRLLRLFKITKYWTSLSNLVASLLNSIRSIA	560

5g jv.pdb_1276	SLLLLLFLFIIIFALLGMQLFGGRYDFEDTEVRRS	503
5GJV_A_Fasta	SLLLLLFLFIIIFALLGMQLFGGRYDFEDTEVRRS	595
5g jv.pdb_1276	NFDNFPQALISVFQVLTGEDWNSVMYNGIMAYGGP	538
5GJV_A_Fasta	NFDNFPQALISVFQVLTGEDWNSVMYNGIMAYGGP	630
5g jv.pdb_1276	SYPGVLVCIYFIILFVCGNYILLNVFLAIAVDNLA	573
5GJV_A_Fasta	SYPGVLVCIYFIILFVCGNYILLNVFLAIAVDNLA	665
5g jv.pdb_1276	EAES.....	577
5GJV_A_Fasta	EAESLTSAQKAKAEERKRRKMSRGLPDKTEEEKSV	700
5g jv.pdb_1276	.....	577
5GJV_A_Fasta	MAKKLEQKPKGEGIPTTAKLKVDEFESNVNEVKDP	735
5g jv.pdb_1276	.....	577
5GJV_A_Fasta	YPSADFPGDDEEDEPEIPVSPRPRPLAELQLKEKA	770
5g jv.pdb_1276	.....VRVLCHRIVNATWFTNFI	595
5GJV_A_Fasta	VPIPEASSFFIFSPTNKVRVLCHRIVNATWFTNFI	805
5g jv.pdb_1276	LLFILLSSAALAAEDPIRAESVRNQILGYFDIAFT	630
5GJV_A_Fasta	LLFILLSSAALAAEDPIRAESVRNQILGYFDIAFT	840
5g jv.pdb_1276	SVFTVEIVLKMTTYGAFLHKGSFCRNYFNILDLLV	665
5GJV_A_Fasta	SVFTVEIVLKMTTYGAFLHKGSFCRNYFNILDLLV	875
5g jv.pdb_1276	VAVSLISMGL.....VVKILRVLRLRPLRAINR	694
5GJV_A_Fasta	VAVSLISMGLESSTISVVKILRVLRLRPLRAINR	910
5g jv.pdb_1276	AKGLKHVVQCVFVAIRTIGNIVLVTTLLQFMFACI	729
5GJV_A_Fasta	AKGLKHVVQCVFVAIRTIGNIVLVTTLLQFMFACI	945
5g jv.pdb_1276	GVQLFKGKFFSCNDLSKMTEEECRGYYYVYKDGP	764
5GJV_A_Fasta	GVQLFKGKFFSCNDLSKMTEEECRGYYYVYKDGP	980
5g jv.pdb_1276	TQMELRPRQWIHNDHFHDNVLSAMMSLFTVSTFEG	799
5GJV_A_Fasta	TQMELRPRQWIHNDHFHDNVLSAMMSLFTVSTFEG	1015
5g jv.pdb_1276	WPQLLYRAIDSNEEDMGPVYNNRVEMAIFFIIYII	834
5GJV_A_Fasta	WPQLLYRAIDSNEEDMGPVYNNRVEMAIFFIIYII	1050
5g jv.pdb_1276	LIAFFMMNIFVGFVIVTFQEQGETEYKNCELDKNQ	869
5GJV_A_Fasta	LIAFFMMNIFVGFVIVTFQEQGETEYKNCELDKNQ	1085
5g jv.pdb_1276	RQCVQYALKARPLRCYIPKNPYQYQVWYVVTSSYF	904
5GJV_A_Fasta	RQCVQYALKARPLRCYIPKNPYQYQVWYVVTSSYF	1120



5g jv.pdb_1276	EYLMFALIMLNTICLGMQHYHQSEEMNHISDILNV	939
5GJV_A_Fasta	EYLMFALIMLNTICLGMQHYHQSEEMNHISDILNV	1155
5g jv.pdb_1276	AFTIIIFTLEMILKLLAFKARGYFGDPWNVDFDLIV	974
5GJV_A_Fasta	AFTIIIFTLEMILKLLAFKARGYFGDPWNVDFDLIV	1190
5g jv.pdb_1276	IGSIIDVILSEIDTF.....	989
5GJV_A_Fasta	IGSIIDVILSEIDTF LASSGGLYCLGGGCGNVDPD	1225
5g jv.pdb_1276	...RISSAFFRLFRVMRLIKLLSRAEGVRTLLWTF	1021
5GJV_A_Fasta	ESARISSAFFRLFRVMRLIKLLSRAEGVRTLLWTF	1260
5g jv.pdb_1276	IKSFQALPYVALLIVMLFFIYAVIGMQMFGKIALV	1056
5GJV_A_Fasta	IKSFQALPYVALLIVMLFFIYAVIGMQMFGKIALV	1295
5g jv.pdb_1276	DGTQINRNNNFQTFPQAVLLLFRCATGEAWQEILL	1091
5GJV_A_Fasta	DGTQINRNNNFQTFPQAVLLLFRCATGEAWQEILL	1330
5g jv.pdb_1276	ACSYGKLCDPESDYAPGEEYTCGTNFAYYYFISFY	1126
5GJV_A_Fasta	ACSYGKLCDPESDYAPGEEYTCGTNFAYYYFISFY	1365
5g jv.pdb_1276	MLCAFLIINLFFVAVIMDNFDYLTRDWSILGPHHLD	1161
5GJV_A_Fasta	MLCAFLIINLFFVAVIMDNFDYLTRDWSILGPHHLD	1400
5g jv.pdb_1276	EFKAIWAEYDPEAKGRIKHLDVVTLLRRIQPPLGF	1196
5GJV_A_Fasta	EFKAIWAEYDPEAKGRIKHLDVVTLLRRIQPPLGF	1435
5g jv.pdb_1276	GKFCPHRVACKRLVGMNMPLNSDGTVTFNATLFAL	1231
5GJV_A_Fasta	GKFCPHRVACKRLVGMNMPLNSDGTVTFNATLFAL	1470
5g jv.pdb_1276	VRTALKIKTEGNFEQANEELRAIIKKIWKRTSMKL	1266
5GJV_A_Fasta	VRTALKIKTEGNFEQANEELRAIIKKIWKRTSMKL	1505
5g jv.pdb_1276	LDQVIPPIGD.....	1276
5GJV_A_Fasta	LDQVIPPIGDDEVTVGKFYATFLIQEHFRKFMKRQ	1540
5g jv.pdb_1276	.....	1276
5GJV_A_Fasta	EEYYGYRPPKDTVQIQAGLRTIEEEAAPEIRRTIS	1575
5g jv.pdb_1276	.....	1276
5GJV_A_Fasta	GDLTAEELERAMVEAAMEERIFRRTGGLFGQVDT	1610
5g jv.pdb_1276	.....	1276
5GJV_A_Fasta	FLERTNSLPPVMANQRPLQFAEIEEMEELESPVFLE	1645
5g jv.pdb_1276	.....	1276
5GJV_A_Fasta	DFPQDARTNPLARANTNNANANVAYGNSNHSNNQM	1680

```

5g jv.pdb_1276      ..... 1276
5GJV_A_Fasta      FSSVHCEREFPGEAETPAAGRGALSHSHRALGPHS 1715

5g jv.pdb_1276      ..... 1276
5GJV_A_Fasta      KPCAGKLNQQLVQPGMPINQAPPAPCQQPSTDPPE 1750

5g jv.pdb_1276      ..... 1276
5GJV_A_Fasta      RGQRRTSLTGSLQDEAPQRRSSEGSTPRRPAPATA 1785

5g jv.pdb_1276      ..... 1276
5GJV_A_Fasta      LLIQEALVRGGLDTLAADAGFVTATSQALADACQM 1820

5g jv.pdb_1276      ..... 1276
5GJV_A_Fasta      EPEEVEVAATELLKARESVQGMASVPGSLSRSSL 1855

5g jv.pdb_1276      ..... 1276
5GJV_A_Fasta      GSLDQVQGSQETLIPRP 1873

```

☐ non conserved

☒ ≥ 50% conserved

**Figure 1.2** Amino acid sequence alignment of PDB ID **5GJV** (fasta format) and its sequence extracted from the PDB file with an in-house python script, supplementary file **extract-aa.py**.

5g jw.pdb_1291	.....FCLT	4
5GJW_A_Fasta	MEPSSPQDEGLRKKQPKKPLPEVLPRPPRALFCLT	35
5g jw.pdb_1291	LQNPLRKACISIVEWKPFETIILLTIFANCVALAV	39
5GJW_A_Fasta	LQNPLRKACISIVEWKPFETIILLTIFANCVALAV	70
5g jw.pdb_1291	YLPMPEDDNNSLNLGLEKLEYFFLTVFSIEAAMKI	74
5GJW_A_Fasta	YLPMPEDDNNSLNLGLEKLEYFFLTVFSIEAAMKI	105
5g jw.pdb_1291	IAYGFLFHQDAYLRSGWNVLDFFIIVFLGVFTAIL.	108
5GJW_A_Fasta	IAYGFLFHQDAYLRSGWNVLDFFIIVFLGVFTAIL	140
5g jw.pdb_1291	.....VKALRAFRVLRPLRL	123
5GJW_A_Fasta	QVNVIQSNTPMSSKGAGLDVKALRAFRVLRPLRL	175
5g jw.pdb_1291	VSGVPSLQVVLNSIFKAMLPLFHIALLVLFMVIIY	158
5GJW_A_Fasta	VSGVPSLQVVLNSIFKAMLPLFHIALLVLFMVIIY	210
5g jw.pdb_1291	AIIGLELFKKGKMHKTCYYIGTDIVATVENEKPSPC	193
5GJW_A_Fasta	AIIGLELFKKGKMHKTCYYIGTDIVATVENEKPSPC	245
5g jw.pdb_1291	ARTGSGRPCTINGSECRGGWPGPNHGITHFDNFGF	228
5GJW_A_Fasta	ARTGSGRPCTINGSECRGGWPGPNHGITHFDNFGF	280
5g jw.pdb_1291	SMLTVYQCITMEGWTDVLYWVNDAIGNEWPWIIYFV	263
5GJW_A_Fasta	SMLTVYQCITMEGWTDVLYWVNDAIGNEWPWIIYFV	315
5g jw.pdb_1291	TLILLGSFFILNLVLGVLSGEFTKEREKAKSRGTF	298
5GJW_A_Fasta	TLILLGSFFILNLVLGVLSGEFTKEREKAKSRGTF	350
5g jw.pdb_1291	QKLREKQQLEEDLRGYMSWITQGEVM.....	324
5GJW_A_Fasta	QKLREKQQLEEDLRGYMSWITQGEVMDVEDLREGK	385
5g jw.pdb_1291	.....NRV	327
5GJW_A_Fasta	LSLEEGGSDTESLYEIEGLNKIIQFIRHWRQWNRV	420
5g jw.pdb_1291	FRWKCHDLVKSRVFYWLVLILIVALNTLSIASEHNN	362
5GJW_A_Fasta	FRWKCHDLVKSRVFYWLVLILIVALNTLSIASEHNN	455
5g jw.pdb_1291	QPLWLTHLQDIANRVLLSLFTIEMLLKMYGLGLRQ	397
5GJW_A_Fasta	QPLWLTHLQDIANRVLLSLFTIEMLLKMYGLGLRQ	490
5g jw.pdb_1291	YFMSIFNRFDCFVVCSGILELLLVESEGAMTPLGIS	432
5GJW_A_Fasta	YFMSIFNRFDCFVVCSGILELLLVESEGAMTPLGIS	525
5g jw.pdb_1291	VLRCIRLLRLFKITKYWTSLSNLVASLLNSIRSIA	467
5GJW_A_Fasta	VLRCIRLLRLFKITKYWTSLSNLVASLLNSIRSIA	560

5g jw.pdb_1291	SLLLLLFLFIIIFALLGMQLFGGRYDFEDTEVRRS	502
5GJW_A_Fasta	SLLLLLFLFIIIFALLGMQLFGGRYDFEDTEVRRS	595
5g jw.pdb_1291	NFDNFPQALISVFQVLTGEDWNSVMYNGIMAYGGP	537
5GJW_A_Fasta	NFDNFPQALISVFQVLTGEDWNSVMYNGIMAYGGP	630
5g jw.pdb_1291	SYPGVLVCIYFIILFVCGNYILLNVFLAIAVDNLA	572
5GJW_A_Fasta	SYPGVLVCIYFIILFVCGNYILLNVFLAIAVDNLA	665
5g jw.pdb_1291	EAESLTSAQKAKAEERKRRKM.....	593
5GJW_A_Fasta	EAESLTSAQKAKAEERKRRKMSRGLPDKTEEEKSV	700
5g jw.pdb_1291	.....	593
5GJW_A_Fasta	MAKKLEQKPKGEGIPTTAKLKVDEFESNVNEVKDP	735
5g jw.pdb_1291	.....	593
5GJW_A_Fasta	YPSADFPGDDEEDEPEIPVSPRPRPLAELQLKEKA	770
5g jw.pdb_1291	.....VRVLCHRIVNATWFTNFI	611
5GJW_A_Fasta	VPIPEASSFFIFSPTNKVRVLCHRIVNATWFTNFI	805
5g jw.pdb_1291	LLFILLSSAALAAEDPIRAESVRNQILGYFDIAFT	646
5GJW_A_Fasta	LLFILLSSAALAAEDPIRAESVRNQILGYFDIAFT	840
5g jw.pdb_1291	SVFTVEIVLKMTTYGAFLHKGSFCRNYFNILDLLV	681
5GJW_A_Fasta	SVFTVEIVLKMTTYGAFLHKGSFCRNYFNILDLLV	875
5g jw.pdb_1291	VAVSLISMG.....VVKILRVLRLRPLRAINR	709
5GJW_A_Fasta	VAVSLISMGLESSTISVVKILRVLRLRPLRAINR	910
5g jw.pdb_1291	AKGLKHVVQCVFVAIRTIGNIVLVTTLLQFMFACI	744
5GJW_A_Fasta	AKGLKHVVQCVFVAIRTIGNIVLVTTLLQFMFACI	945
5g jw.pdb_1291	GVQLFKGKFFSCNDLSKMTEEECRGYYYVYKDGP	779
5GJW_A_Fasta	GVQLFKGKFFSCNDLSKMTEEECRGYYYVYKDGP	980
5g jw.pdb_1291	TQMELRPRQWIHNDHFHDNVLSAMMSLFTVSTFEG	814
5GJW_A_Fasta	TQMELRPRQWIHNDHFHDNVLSAMMSLFTVSTFEG	1015
5g jw.pdb_1291	WPQLLYRAIDSNEEDMGPVYNNRVEMAIFFIIYII	849
5GJW_A_Fasta	WPQLLYRAIDSNEEDMGPVYNNRVEMAIFFIIYII	1050
5g jw.pdb_1291	LIAFFMMNIFVGFVIVTFQEQGETEYKNCELDKNQ	884
5GJW_A_Fasta	LIAFFMMNIFVGFVIVTFQEQGETEYKNCELDKNQ	1085
5g jw.pdb_1291	RQCVQYALKARPLRCYIPKNPYQYQVWYVVTSSYF	919
5GJW_A_Fasta	RQCVQYALKARPLRCYIPKNPYQYQVWYVVTSSYF	1120

5g jw.pdb_1291	EYLMFALIMLNTICLGMQHYHQSEEMNHISDILNV	954
5GJW_A_Fasta	EYLMFALIMLNTICLGMQHYHQSEEMNHISDILNV	1155
5g jw.pdb_1291	AFTIIIFTLEMILKLLAFKARGYFGDPWNVDFDLIV	989
5GJW_A_Fasta	AFTIIIFTLEMILKLLAFKARGYFGDPWNVDFDLIV	1190
5g jw.pdb_1291	IGSIIDVILSEIDTF.....	1004
5GJW_A_Fasta	IGSIIDVILSEIDTF LASSGGLYCLGGGCGNVDPD	1225
5g jw.pdb_1291	...RISSAFFRLFRVMRLIKLLSRAEGVRTLLWTF	1036
5GJW_A_Fasta	ESARISSAFFRLFRVMRLIKLLSRAEGVRTLLWTF	1260
5g jw.pdb_1291	IKSFQALPYVALLIVMLFFIYAVIGMQMFGKIALV	1071
5GJW_A_Fasta	IKSFQALPYVALLIVMLFFIYAVIGMQMFGKIALV	1295
5g jw.pdb_1291	DGTQINRNNNFQTFPQAVLLLFRCATGEAWQEILL	1106
5GJW_A_Fasta	DGTQINRNNNFQTFPQAVLLLFRCATGEAWQEILL	1330
5g jw.pdb_1291	ACSYGKLCDPESDYAPGEEYTCGTNFAYYYFISFY	1141
5GJW_A_Fasta	ACSYGKLCDPESDYAPGEEYTCGTNFAYYYFISFY	1365
5g jw.pdb_1291	MLCAFLIINLFFVAVIMDNFDYLTRDWSILGPHHLD	1176
5GJW_A_Fasta	MLCAFLIINLFFVAVIMDNFDYLTRDWSILGPHHLD	1400
5g jw.pdb_1291	EFKAIWAEYDPEAKGRIKHLDVVTLLRRIQPPLGF	1211
5GJW_A_Fasta	EFKAIWAEYDPEAKGRIKHLDVVTLLRRIQPPLGF	1435
5g jw.pdb_1291	GKFCPHRVACKRLVGMNMPLNSDGTVTFNATLFAL	1246
5GJW_A_Fasta	GKFCPHRVACKRLVGMNMPLNSDGTVTFNATLFAL	1470
5g jw.pdb_1291	VRTALKIKTEGNFEQANEELRAIIKKIWKRTSMKL	1281
5GJW_A_Fasta	VRTALKIKTEGNFEQANEELRAIIKKIWKRTSMKL	1505
5g jw.pdb_1291	LDQVIPPIGD.....	1291
5GJW_A_Fasta	LDQVIPPIGDDEVTVGKFYATFLIQEHFRKFMKRQ	1540
5g jw.pdb_1291	.....	1291
5GJW_A_Fasta	EEYYGYRPPKDTVQIQAGLRTIEEEAAPEIRRTIS	1575
5g jw.pdb_1291	.....	1291
5GJW_A_Fasta	GDLTAEELERAMVEAAMEERIFRRTGGLFGQVDT	1610
5g jw.pdb_1291	.....	1291
5GJW_A_Fasta	FLERTNSLPPVMANQRPLQFAEIEEMEELESPVFLE	1645
5g jw.pdb_1291	.....	1291
5GJW_A_Fasta	DFPQDARTNPLARANTNNANANVAYGNSNHSNNQM	1680

```

5g jw.pdb_1291      . . . . . 1291
5GJW_A_Fasta      FSSVHCEREFPGEAETPAAGRGALSHSHRALGPHS 1715

5g jw.pdb_1291      . . . . . 1291
5GJW_A_Fasta      KPCAGKLNGLVQPGMPINQAPPAPCQQPSTDPPE 1750

5g jw.pdb_1291      . . . . . 1291
5GJW_A_Fasta      RGQRRTSLTGSLQDEAPQRRSSEGSTPRRPAPATA 1785

5g jw.pdb_1291      . . . . . 1291
5GJW_A_Fasta      LLIQEALVRGGLDTLAADAGFVTATSQALADACQM 1820

5g jw.pdb_1291      . . . . . 1291
5GJW_A_Fasta      EPEEVEVAATELLKARESVQGMASVPGSLSRSSL 1855

5g jw.pdb_1291      . . . . . 1291
5GJW_A_Fasta      GSLDQVQGSQETLIPRP 1873

```

☐ non conserved  
☒ ≥ 50% conserved

**Figure 1.3** Amino acid sequence alignment of PDB ID **5GJW** (fasta format) and its sequence extracted from the PDB file with an in-house python script, supplementary file **extract-aa.py**.

5g jv.pdb_1276	FCLTLQNPLRKACISIVEWKPFETIILLTIFANCV	35
5g jw.pdb_1291	FCLTLQNPLRKACISIVEWKPFETIILLTIFANCV	35
5g jv.pdb_1276	ALAVYLPMPEDDNNSLNLGLEKLEYFFLTVFSIEA	70
5g jw.pdb_1291	ALAVYLPMPEDDNNSLNLGLEKLEYFFLTVFSIEA	70
5g jv.pdb_1276	AMKIIAYGFLFHQDAYLRSGWNVLDfIIIVFLGVFT	105
5g jw.pdb_1291	AMKIIAYGFLFHQDAYLRSGWNVLDfIIIVFLGVFT	105
5g jv.pdb_1276	AILVKALRAFRVLRPLRLVSGVPSLQVVLNSIFKA	140
5g jw.pdb_1291	AILVKALRAFRVLRPLRLVSGVPSLQVVLNSIFKA	140
5g jv.pdb_1276	MLPLFHIALLVLFMVIIYAIIGLELFFKGMHKTCY	175
5g jw.pdb_1291	MLPLFHIALLVLFMVIIYAIIGLELFFKGMHKTCY	175
5g jv.pdb_1276	YIGTDIVATVENEKPSPCARTGSGRPCTINGSECR	210
5g jw.pdb_1291	YIGTDIVATVENEKPSPCARTGSGRPCTINGSECR	210
5g jv.pdb_1276	GGWPGPNHGITHFDNFGFSMLTVYQCITMEGWTDV	245
5g jw.pdb_1291	GGWPGPNHGITHFDNFGFSMLTVYQCITMEGWTDV	245
5g jv.pdb_1276	LYWVNDAIGNEWPWIYFVTLLLLGSFFILNLVLGV	280
5g jw.pdb_1291	LYWVNDAIGNEWPWIYFVTLLLLGSFFILNLVLGV	280
5g jv.pdb_1276	LSGEFTKEREKAKSRGTFQKLREKQOLEEDLRGYM	315
5g jw.pdb_1291	LSGEFTKEREKAKSRGTFQKLREKQOLEEDLRGYM	315
5g jv.pdb_1276	SWITQGEVMWNRVFRWKCHDLVKSRVFYWLVLIV	350
5g jw.pdb_1291	SWITQGEVMNRVFRWKCHDLVKSRVFYWLVLIV	349
5g jv.pdb_1276	ALNTLSIASSEHHNQPLWLTHLQDIANRVLLSLFTI	385
5g jw.pdb_1291	ALNTLSIASSEHHNQPLWLTHLQDIANRVLLSLFTI	384
5g jv.pdb_1276	EMLLKMYGLGLRQYFMSIFNRFDCFVVCSGILELL	420
5g jw.pdb_1291	EMLLKMYGLGLRQYFMSIFNRFDCFVVCSGILELL	419
5g jv.pdb_1276	LVESGAMTPLGISVLR CIRLLRLFKITKYWTSLSN	455
5g jw.pdb_1291	LVESGAMTPLGISVLR CIRLLRLFKITKYWTSLSN	454
5g jv.pdb_1276	LVASLLNSIRS IASLLLLLFLFIIIFALLGMQLFG	490
5g jw.pdb_1291	LVASLLNSIRS IASLLLLLFLFIIIFALLGMQLFG	489
5g jv.pdb_1276	GRYDFEDTEVRRSNFDNFPQALISVFQVLTGEDWN	525
5g jw.pdb_1291	GRYDFEDTEVRRSNFDNFPQALISVFQVLTGEDWN	524
5g jv.pdb_1276	SVMYNGIMAYGGPSYPGVLVCIYFIILFVCGNYIL	560
5g jw.pdb_1291	SVMYNGIMAYGGPSYPGVLVCIYFIILFVCGNYIL	559

5g jv.pdb_1276	LNVFLAIAVDNLAEAES.....V	578
5g jw.pdb_1291	LNVFLAIAVDNLAEESLTSAQKAKAEERKRRKMV	594
5g jv.pdb_1276	RVLCHRIVNATWFTNFILLFILLSSAALAAEDPIR	613
5g jw.pdb_1291	RVLCHRIVNATWFTNFILLFILLSSAALAAEDPIR	629
5g jv.pdb_1276	AESVRNQILGYFDIAFTSVFTVEIVLKMTTYGAFL	648
5g jw.pdb_1291	AESVRNQILGYFDIAFTSVFTVEIVLKMTTYGAFL	664
5g jv.pdb_1276	HKGSFCRNYFNILDLLVAVSLISMGLVVKILRVL	683
5g jw.pdb_1291	HKGSFCRNYFNILDLLVAVSLISMGLVVKILRVL	698
5g jv.pdb_1276	RVLRLRAINRAKGLKHVVQCVFVAIRTIGNIVLV	718
5g jw.pdb_1291	RVLRLRAINRAKGLKHVVQCVFVAIRTIGNIVLV	733
5g jv.pdb_1276	TTLLOFMFACIGVQLFKGKFFSCNDLSKMTEEECR	753
5g jw.pdb_1291	TTLLOFMFACIGVQLFKGKFFSCNDLSKMTEEECR	768
5g jv.pdb_1276	GYYYVYKDGDPQTQELRPRQWIHNDHFHFNVL SAM	788
5g jw.pdb_1291	GYYYVYKDGDPQTQELRPRQWIHNDHFHFNVL SAM	803
5g jv.pdb_1276	MSLFTVSTFEGWPQLLYRAIDSNEEDMGPVYNNRV	823
5g jw.pdb_1291	MSLFTVSTFEGWPQLLYRAIDSNEEDMGPVYNNRV	838
5g jv.pdb_1276	EMAIFFIIYIILIAFFMMNIFVGFVIVTFQEQGET	858
5g jw.pdb_1291	EMAIFFIIYIILIAFFMMNIFVGFVIVTFQEQGET	873
5g jv.pdb_1276	EYKNCELDKNQRCVQYALKARPLRCYIPKNPYQY	893
5g jw.pdb_1291	EYKNCELDKNQRCVQYALKARPLRCYIPKNPYQY	908
5g jv.pdb_1276	QVWYVVTSSYFEYLMFALIMLNTICLGMQHYHQSE	928
5g jw.pdb_1291	QVWYVVTSSYFEYLMFALIMLNTICLGMQHYHQSE	943
5g jv.pdb_1276	EMNHISDILNVAFTIIIFTLEMILKLLAFKARGYFG	963
5g jw.pdb_1291	EMNHISDILNVAFTIIIFTLEMILKLLAFKARGYFG	978
5g jv.pdb_1276	DPWNVDFDLIVIGSIIDVILSEIDTFRISSAFFRL	998
5g jw.pdb_1291	DPWNVDFDLIVIGSIIDVILSEIDTFRISSAFFRL	1013
5g jv.pdb_1276	FRVMRLIKLLSRAEGVRTLLWTFIKSFQALPYVAL	1033
5g jw.pdb_1291	FRVMRLIKLLSRAEGVRTLLWTFIKSFQALPYVAL	1048
5g jv.pdb_1276	LIVMLFFIYAVIGMQMFGKIALVDGTQINRNNNFQ	1068
5g jw.pdb_1291	LIVMLFFIYAVIGMQMFGKIALVDGTQINRNNNFQ	1083
5g jv.pdb_1276	TFPQAVLLLFRCATGEAWQEILLACSYGKLCDPES	1103
5g jw.pdb_1291	TFPQAVLLLFRCATGEAWQEILLACSYGKLCDPES	1118



5g jv.pdb_1276	DYAPGEEYTCGTNFAYYYFISFYMLCAFLIINLFV	1138
5g jw.pdb_1291	DYAPGEEYTCGTNFAYYYFISFYMLCAFLIINLFV	1153
5g jv.pdb_1276	AVIMDNFDYLTRDWSILGPHHLDEFKAIWAEYDPE	1173
5g jw.pdb_1291	AVIMDNFDYLTRDWSILGPHHLDEFKAIWAEYDPE	1188
5g jv.pdb_1276	AKGRIKHLDVVTLLRRIQPPLGFGKFCPHRVACKR	1208
5g jw.pdb_1291	AKGRIKHLDVVTLLRRIQPPLGFGKFCPHRVACKR	1223
5g jv.pdb_1276	LVGMNMPLNSDGTVTFNATLFALVRTALKIKTEGN	1243
5g jw.pdb_1291	LVGMNMPLNSDGTVTFNATLFALVRTALKIKTEGN	1258
5g jv.pdb_1276	FEQANEELRAIIKKIWKRTSMKLLDQVIPPIGD	1276
5g jw.pdb_1291	FEQANEELRAIIKKIWKRTSMKLLDQVIPPIGD	1291

☐ non conserved

☒ ≥ 50% conserved

**Figure 1.4** Amino acid sequence alignment of PDB IDs **5GJV** and **5GJV** (both extracted from their respective PDB files with an in-house python script, supplementary file **extract-aa.py**).

The protein data bank (PDB) currently holds over 150,000 (As of March 20, 2019) biomolecular structures and continues to release new structures on a weekly basis. The PDB is an essential resource to the structural bioinformatics community to develop software that mine, use, categorize, and analyze such data. Here, with L-type voltage-dependent calcium channel  $\text{Ca}_v1.1$  as an example, the four amino acid sequence alignments above raised a few points below,

1. Overall, the  $\text{Ca}_v1.1$  structure (in PDB) is incomplete (as of March 20, 2019), neither for 5GJV (1276 versus 1873), nor for 5GJW (1291 versus 1873).
2. From Figure 1.1, it can be seen that the template sequences (in fasta format) are identical for 5GJV and 5GJW.
3. From Figure 1.2, with **5GJV\_A\_Fasta** as the template sequence, there are a few gaps in the experimentally determined structure of  $\text{Ca}_v1.1$ . For example, in PDB file 5gju.pdb, the atomic coordinates are missing for  $\text{Ca}_v1.1$  residues from Met376 to Trp417 ( $\Delta = 41$ ), and from Ser669 to Val788 ( $\Delta = 119$ ), and from Asp1515 to the end of  $\text{Ca}_v1.1$  ( $\Delta = 358$ ), where  $\Delta$  represents the size of the hole of specific residues, whose atomic coordinates are missing in the PDB files.
4. From Figure 1.3, with **5GJW\_A\_Fasta** as the template sequence, the situation is by and large similar, where there are structurally undetermined regions of  $\text{Ca}_v1.1$ , too.
5. From Figure 1.4, it can be seen that the experimentally determined fragments of  $\text{Ca}_v1.1$  are mostly identical, except for a gap (lines 578 and 594, Figure 1.4) and a few minor genetic variations, too.