HHS Public Access

Author manuscript

Trends Cell Mol Biol. Author manuscript; available in PMC 2016 March 21.

Published in final edited form as:

Trends Cell Mol Biol. 2015 ; 10: 49–75.

Comparative analysis of the human and zebrafish kinomes: focus on the development of kinase inhibitors

Nathan Wlodarchak[1], **Rehan Tariq**[1], and **Rob Striker**[1,2,*]

[1]Department of Medicine, School of Medicine and Public Health, University of Wisconsin-Madison, Microbial Sciences Building, 1550 Linden Drive, Madison, WI 53706, USA

[2]W. S. Middleton Memorial Veteran's Administration Hospital, 2500 Overlook Terrace, Madison, WI 53705, USA

Abstract

Targeting kinases with semi-selective kinase inhibitors is one of the most successful drug development strategies of the 21st century. Zebrafish have become an increasingly useful model for pharmaceutical development. Water-soluble compounds can be screened for zebrafish phenotypes in a high throughput format against a living vertebrate, and cell-signaling events can be imaged in transparent living fish. Despite zebrafish being a more relevant model than more distantly related systems such as the well-annotated kinome of yeast and drosophila, there is no comparative analysis of the human and zebrafish kinome. Furthermore most approved kinase inhibitors, often called 'DFG in' ATP competitive inhibitors, act on conserved active site residues in the kinase. Since the active site residues can be identified by examining the primary sequence, primary sequence identity can be a rough guide as to whether a particular inhibitor will have activity against another kinase. There is a need to evaluate the utility of zebrafish as a drug development model for active site inhibitors of kinases. Here we offer a systematic comparison of the catalytic domains of classical human kinases with the catalytic domains of all annotated zebrafish kinases. We found a high degree of identity between the catalytic domains of most human kinases and their zebrafish homologs, and we ranked 504 human kinase catalytic domains by order of similarity. We found only 23 human kinases with no easily recognizable homologous zebrafish catalytic domain. On the other hand we found 78 zebrafish kinase catalytic domains with no close human counterpart. These 'additional kinase active sites' could represent potential mediators of zebrafish toxicity that may not be relevant to human kinase inhibitors. We used two clinically approved human kinase inhibitors, one targeting a highly homologous target and one targeting a lesser homologous target, and we compared the known human kinase target structures with modeled zebrafish target structures. As expected, the homologous target had high structural identity, but even the less homologous target had high structural identity in residues contacted by the inhibitor. Overall this analysis should help guide researchers interested in studying human kinases and their inhibitors in more tractable systems.

*Corresponding author: rtstriker@wisc.edu.

CONFLICT OF INTEREST STATEMENT

The authors have no conflicts of interest to declare.

Keywords

kinome; drug development; kinase inhibitors; human kinases; zebrafish

INTRODUCTION

Protein phosphorylation is the primary signal transduction mechanism in eukaryotic cells. This is accomplished through the opposing action of kinases and phosphatases, and the proper regulation of these enzymes is critical for cell homeostasis [1–3]. Inhibiting kinases has become one of the most reliable pharmacologic strategies in the last decade with several new kinase targeting drugs currently in use [4]. Initially kinase inhibitors were intensively studied and approved as cancer drugs, but now the scope of their potential use has widened to immune regulation, anti-infectives, and cardiovascular diseases [4, 5]. Tofacitinib is currently approved for rheumatoid arthritis, and other Jak family inhibitors are entering clinical trials for rheumatoid arthritis and psoriasis [6]. Inhibitors for p38 MAP kinases are under investigation for treatment of atherosclerosis and chronic obstructive pulmonary disease (COPD) [7]. These and other successes have reinforced the search for small molecule kinase inhibitors which could selectively inhibit kinases with minimal side effects. There are 518 human kinases and inhibition of any of these kinases can potentially elicit a potent therapeutic effect because of the intrinsic cascading mechanism each kinase possesses [5]. Initially there was concern and skepticism that small molecules could confer selectivity for a single kinase. It is now clear that sufficient selectivity is achievable [5, 8]. Therapeutics inhibiting more than one kinase are sometimes desirable, and potentially more than 518 different effects could be achieved by inhibiting different combinations of kinases. Currently 25 kinase inhibitors are FDA approved, most of which are oral [4].

While the so-called ‘atypical’ kinases lack canonical primary sequence motifs, there is a common evolutionary and structural basis for protein kinases [9]. Protein kinases have a bilobed structure with catalysis occurring in the cleft between lobes [10]. The cleft contains key conserved features critical to phosphorylation: an ATP binding pocket, substrate binding residues, and a divalent cation—typically Mg^{2+} [10, 11]. Throughout the kinome there are well-recognized conserved motifs that collectively line the cleft, coordinate the metal cofactor and transfer the phosphate from ATP to the substrate [10, 11]. Steven Hanks and others identified several domains which were conserved among almost all kinases and were essential for their kinase activity [10, 12]. These domains became subsequently known as Hanks domains (Figure 1). As the ancestral kinases diversified, many different subfamilies evolved specializing in, for example, tyrosine phosphorylation or serine/threonine phosphorylation [13]. The overall human kinome has been divided into subfamilies based on structural and functional differences, and this classifying information is annotated in online databases [1, 14]. Prominent families include serine/threonine kinases such as AGC and CAMK as well as exclusive tyrosine kinases (TKs) and tyrosine kinase like (TKL) kinases which phosphorylate tyrosine as well as serine and threonine [15]. How these kinase families evolved has been examined in model organisms across a vast swath of divergent organisms from Saccharomyces to mice to humans, but more recent model organisms such as zebrafish have not been analyzed in detail [16].

Despite extensive diversification, the Hanks domains have remained largely unchanged, particularly inside the catalytic domain of the N-lobe between the glycine rich 'P-loop' and the DFG motif [10]. In fact, even small changes in these motifs lead to the categorization of kinases as likely pseudokinases [17]. Some classified pseudokinases retain catalytic activity, but non-canonical motifs at least suggest functions other than phospho-transfer [17, 18]. Although the Hanks domains in the cleft are highly conserved, some variation near the motifs may be present among species, particularly if their substrates diverged greatly. To better utilize zebrafish as a high throughput model for kinase inhibitor drug development, it is necessary to evaluate how similar the clefts are for both the human and zebrafish kinome.

Zebrafish (*Danio rerio*) is now a well-established model organism for developmental biologists [19]. It has many advantages including low cost, tissue transparency, fecundity, and short generation time [20]. These advantages have also made zebrafish an attractive model for drug development [21]. The zebrafish genome is sequenced and it has 71% homology to the human genome [22]. Genome duplication events in the teleost lineage complicate genetic comparisons between zebrafish and humans, since some genes were duplicated several times and others were not at all [23]. A common tool to probe the specific genetics of zebrafish has been the use of antisense morpholino oligomers to knock down specific genes; however, the specificity of this has been challenged recently [24]. The use of small molecule inhibitors may be a complementary approach if key inhibitor binding sites are well conserved between humans and zebrafish. The zebrafish kinome has yet to be extensively described, but if the kinome parallels the genome in homology, then the zebrafish may potentially serve as a functional model organism for kinase inhibitor development. While zebrafish are already a model for testing drug toxicity [21], expanding zebrafish utility as a kinase-specific model for toxicity and efficacy will require more detailed knowledge about the zebrafish kinome.

Understanding the relatedness of human and zebrafish kinases will improve the reliability of zebrafish as a model organism for studying kinase inhibitors. Testing drugs for human kinases with no clear zebrafish homolog might yield false negatives, and conversely, zebrafish kinases which have clear homologs to human kinases but with several specialized isoforms not found in humans may also provide data that is ultimately not helpful in the drug development process. Knowledge of these relationships would streamline drug development by identifying candidate kinase targets which are viable models in zebrafish and allowing more high-throughput testing to be done earlier in the drug development process. Here we did not examine the overall kinase homology *per se* but specifically analyzed how well the active sites match since most ATP competitive kinase inhibitors rely on critical residues in the Hank's domains.

METHODS

The human kinase domain sequences were retrieved from the KinBase database [1]. There are 531 non-redundant human kinases annotated in Kinbase, of which 25 are classified as atypical. Manning and colleagues have identified 15 additional atypical kinases and 13 pseudokinase domains, of which the former were excluded, and the latter were included in our analysis [1]. Approximately 350 zebrafish kinases were obtained from ensemble [25]

and approximately 550 were obtained from zfin [26] using pfam 00069 for a total of ~900 zebrafish sequences from both sources. These kinases were visually examined, and redundant sequences were removed to arrive at a final list of 692. The shortest canonical human kinases from each family were used as a model to identify zebrafish kinase catalytic domains by performing sequence alignment and trimming zebrafish sequences where they align to the human catalytic domains. MultAlign [27] was used to align the kinase domains of the organisms up within their respective subfamilies in order to identify the homologous kinase domains in zebrafish kinases, and analysis was restricted to the kinase domain. Residues prior to the start of the P-loop, usually indicated by the presence of a nearby glycine rich region, and residues subsequent to the DFG motif were truncated, based on the alignment to the canonical kinase domain. This region will be subsequently referred to as the 'catalytic domain' (CD) (Figure 1).

Clustal Omega [28] was used to perform a protein BLAST analysis and generate a phylogenetic tree that depicted the relationship between the human and zebrafish kinase CDs. The data were organized by pairing kinases grouped between species and unique CDs that have no orthologs between humans and zebrafish. The maximum matched percent identity of each zebrafish CD and of each human CD was extracted using Excel (Microsoft) and condensed into a master table (Table S1). A frequency distribution of the maximum percent identities was calculated using Prism (GraphPad software). A cutoff of 40% identity was chosen based on the frequency distribution of all maximum CD percent identities (Figure S1) as well as based on visual examination of the dendogram. Below 40% identity there was not clear pairing of human and zebrafish kinase active sites.

Crystal structures of FDA approved kinase inhibitors were downloaded from the PDB (3OG7 and 3LXK) [29, 30] and visualized in Pymol (Schrodinger). Predicted zebrafish kinase models were generated using I-Tasser [31]. Structural figures were also rendered in Pymol.

RESULTS AND DISCUSSION

The zebrafish and human kinase catalytic domains share high identity

The protein BLAST analysis revealed generally high percent identities between the zebrafish and human kinase catalytic domains (CD). Four hundred and eighty one human CDs matched to 614 zebrafish CDs with a percent identity greater than 40 (Figure 2). The extra zebrafish kinases are primarily due to additional isoforms for highly homologous forms of certain kinases—chiefly the PIM kinases and Aurora kinases (Table S1). Twenty four unique or divergent human CDs were identified (Figure 2, Table 1) and 78 unique zebrafish CDs were identified (Figure 2, Table 2). The average percent identity of CDs between both kinomes including unique kinases is 69% with a median identity of 75%, suggesting the identity of the kinome corresponds well to the total genomic identity [22].

Unique zebrafish CDs consist of many alternate isoforms of homologous kinases

There are 78 unique zebrafish CDs, defined as those with no human CD of greater than 40% identity (Table 2). Several of these CDs match as isoforms of zebrafish CDs which have

percent identities greater than 40% (notably the PIM family and Aurora family); however, some of these CDs are isoforms of similar CDs with no strong human protein homolog (such as the TSSK family). This 'soform variation' may be a function of the diverged habitat of human and zebrafish (land vs. water) or some may be remnants of genome duplication events. These kinases are categorized as isoforms in this study based solely on maximum percent identity; therefore further functional characterization of these kinases is needed if they are to be investigated as drug target models.

Only 5% of human CDs are unique

The unique human CDs form a group of only 24 (Table 1). Interestingly, several of these CDs are related to each other, such as the FAM family, the NME family, and the SgK family. The FAM20 kinases are involved in phosphorylation of secreted substrates [32]. Given that humans and zebrafish live in fundamentally different environments, it is not surprising that pathways which interact with the environment (i.e. secretion) would be divergent. The NME family of kinases has been implicated in the function of tumor metastasis; however, the function of these kinases has been scarcely studied outside of humans [33]. Some pseudokinases such as sgk495 are represented, but unlike the FAM20 and NME kinase families, the SGK subfamily has several members with homologs between human and zebrafish (Table S1). The SGKs are generally involved in stress responses and cellular channel activation [34]. It is possible that some of these channel signaling processes have evolved to be highly divergent in humans, but several of these kinases are still under-explored.

There are seven typical human kinases that have no ortholog in mice [35]. Interestingly, while for many of these 7 there is likely not a strict zebrafish homolog, there is a CD with high identity to the human kinase 'missing' from the mouse kinome. For example neither the mouse kinome nor the zebrafish kinome has a CDK3 ortholog, but the CD of human CDK3 is 80% identical to that of zebrafish CDK2 (Table S1). Another example is that for human CK1alpha2: no ortholog is present in mice, but a high identity (98%) CD is present in zebrafish. The weakest zebrafish match among the 7 human CDs not present in mice is for human PKSH2 at only 74%. The differences in mouse and zebrafish kinomes suggest that organism selection for a particular drug-target model is critically dependent on information gleaned by the type of analysis presented in this study.

Pseudokinases are generally less conserved between zebrafish and humans

Forty-eight human kinases have been classified as 'pseudokinases' on the basis of variation in one of the three Hank's domains (VAIK, HRD, or DFG) necessary for catalysis, and 28 have conserved homologues in yeast, worms, mice, and humans [17]. Although these 48 have atypical variation in key residues based on primary sequence, several have been shown to be catalytically active [17, 18]. For example, WNK1 does not have the classical VIAK motif in β-strand 3, but instead has a lysine residue (K233) that enters the active site from β-strand 2 and confers catalytic activity [36]. This suggests that structural and functional analysis is key to pseudokinase classification [18]. The CDs of twenty-two of these pseudokinases are also conserved between zebrafish and humans (Table 3). Of these 22, differences between zebrafish and human pseudokinases owing to the loss of critical Hanks

domains (DFG, HRD, and VIAK) were only seen in three: Trib1, Trib2, and Sgk494A (Table 3). The zebrafish Trib kinases deviate more from the classical Hank's domains than do their human equivalents; however, Sgk494A has retained all of these domains. This suggests that there may be potential differences in structure and function between some of these pseudokinases which evolved after the human/zebrafish divergence. In addition to unique human pseudokinases, several groups which are represented in zebrafish have additional isoforms in humans (e.g. ANPa/b and RSKL1/2) further suggesting additional roles in humans that may be accomplished by fewer isoforms in zebrafish or not at all.

Structural identity is conserved in active sites of varying sequence identity

Although considerable attention is given to sequence identity in the literature, structural identity may be a more accurate predictor of functional conservation [8]. Fortunately, several of the FDA approved kinase inhibitors have been crystallized with their targets, and structures are available [4, 8]. The identity of the targets of these FDA approved drugs is variable, but analysis of these structures can provide information as to whether the most critical of contacts is preserved.

The FDA approved kinase inhibitor vemurafinib is a BRAF kinase inhibitor [4, 29]. Human BRAF kinase has 99% identity to zebrafish BRAF kinase in the catalytic active site (Table S1). Vemurafinib is coordinated in the BRAF active site by hydrogen bonding to the hinge region backbone and the activation loop backbone as well as through several stacking interactions and hydrophobic interactions (Figure 3). The BRAF kinase CDs are essentially identical between human and zebrafish; therefore the drug is almost certain to bind to zebrafish BRAF as it would in human BRAF.

Tofacitinib was FDA approved for arthritis and it inhibits human Jak3 kinase [4, 30]. Jak3 has one canonical CD and one pseudokinase domain. The canonical kinase domain has a 68% identity to zebrafish Jak2A (Table S1). Despite the lower sequence identity in the CDs, they still maintain high structural similarity (Figure 4). Tofacitinib binds to the active site of Jak3's canonical CD similar to most ATP competitors by hydrogen bonding to the hinge region backbone (Figure 4). It is also stabilized by hydrogen bonding to the P-loop and through hydrophobic interactions at the base and back of the active site. Although the CD residues appeared unchanged from the primary sequence, the higher overall divergence may have created conformational changes not evident from primary sequence alone. The I-Tasser-predicted structure shows the residues pointing toward the catalytic site in the CD and overall conformation are mostly unchanged (Figure 4). There is a small kink introduced in human Jak3 by substituting an alanine for glycine at aa966 (152), and there is a cysteine to leucine conversion at aa909 (142). Neither of these changes appears to affect the coordination of the drug in the active site, suggesting that despite a larger amount of evolutionary divergence compared to BRAF, the Jak3 active sites are structurally conserved and would be accurate models for testing new drugs. The structures of many human kinases are known, and several have been co-crystallized with drugs [8]. The sequence analysis presented here, combined with structural information from published and modeled drug-kinase interactions, provide potentially insightful tools for drug development.

Limitations

There are several limitations to both our analysis and the general use of zebrafish as an alternative to more costly mammalian models for kinase inhibitor drug development. Testing drugs in zebrafish requires a certain amount of water solubility. Solvents such as DMSO, used to deliver insoluble drugs have toxicity effects on their own, reducing the effectiveness of a freshwater organism model. Fortunately, most kinase inhibitors are less than 1000 Daltons in size and have some polar moieties, allowing them to have sufficient water solubility [8], and all 25 of the currently FDA approved kinase inhibitors are orally available [4].

The zebrafish genome is rapidly improving in annotation but is certainly not as comprehensively annotated as the human genome. For example, twice the zebrafish kinases from zfin were recovered as were in ensemble, but there could still be unannotated kinases that were not included in this study. Very few human kinases were not matched to a zebrafish kinase in this analysis, which could indicate the coverage is sufficient and additional zebrafish kinases would likely be used for biological processes specific to the zebrafish.

Although this work has presented divergent or unique zebrafish CDs based on a histogram of identity and an alignment, the specific utility of zebrafish for drug development for a specific human kinase depends on a number of factors beyond the cut-off presented here. The zebrafish kinome consists of several isoforms of kinases with moderate-to-high identity to their human counterpart. These isoforms are mostly uncharacterized and may be unique to zebrafish and have unique functions. Attempts at drug development targeting kinases which have several other isoforms may be clouded by off target effects. Careful scrutiny should be made in using zebrafish as a model for a particular kinase for which zebrafish possesses multiple isoforms.

This analysis focused only on primary sequences. Although the structures of the zebrafish kinases are predicted to be similar to their human counterparts based on primary sequences and known similar structures, subtle nuances may still create situations where certain inhibitor interactions are disrupted or strengthened. The quaternary interactions of human kinases have been elucidated thoroughly in some cases such as cyclins and CDKs [37], but this is generally lacking in zebrafish. It is possible that species-specific accessory proteins could cause appreciable conformational changes in the structures of these kinases and alter potential binding patterns. As discussed above, the role of pseudokinases in zebrafish is not well known, and allosteric activation by or of pseudokinases is generally underexplored [17, 18]. Additional structural information will be needed in the future to address these concerns.

CONCLUSIONS

Overall, this work should facilitate decisions on when zebrafish represent a useful choice to study human kinase inhibitors. The utility clearly varies from high for inhibitors targeting highly homologous kinase CDs (BRAF) and subtly divergent CDs (Jak3), to low for more highly divergent kinases (FAM20 kinases) or if several divergent isoforms (PIM kinases) exist.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors gratefully acknowledge Anna Huttenlocher for editorial advice on the manuscript, and Kelly Pittman for advice on selecting programs to create the BLAST matrix. We would also like to thank Caitlin Pepperell for early advice on the manuscript.

Funding for N. Wlodarchak and R. Striker is from the Hartwell Foundation and the Institute for Clinical and Translational Research (ICTR). The project described was supported by the Clinical and Translational Science Award (CTSA) program, through the NIH National Center for Advancing Translational Sciences (NCATS), grant UL1TR000427. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

REFERENCES

1. Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S. Science. 2002; 298(5600):1912–1934. [PubMed: 12471243]
2. Cohen P. Nat. Rev. Drug Discov. 2002; 1(4):309–315. [PubMed: 12120282]
3. Hunter T. Cell. 1995; 80(2):225–236. [PubMed: 7834742]
4. Hantschel O. ACS Chem. Biol. 2015; 10(1):234–245. [PubMed: 25531586]
5. Zhang C, Bollag G. Curr. Opin. Genet. Dev. 2010; 20(1):79–86. [PubMed: 19932020]
6. Patterson H, Nibbs R, McInnes I, Siebert S. Clin. Exp. Immunol. 2014; 176(1):1–10. [PubMed: 24313320]
7. Fisk M, Gajendragadkar PR, Maki-Petaja KM, Wilkinson IB, Cheriyan J. Am. J. Cardiovasc. Drugs. 2014; 14(3):155–165. [PubMed: 24504769]
8. Wang Q, Zorn JA, Kuriyan J. Methods Enzymol. 2014; 548:23–67. [PubMed: 25399641]
9. Mulder NJ, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Buillard V, Cerutti L, Copley R, Courcelle E, Das U, Daugherty L, Dibley M, Finn R, Fleischmann W, Gough J, Haft D, Hulo N, Hunter S, Kahn D, Kanapin A, Kejariwal A, Labarga A, Langendijk-Genevaux PS, Lonsdale D, Lopez R, Letunic I, Madera M, Maslen J, McAnulla C, McDowall J, Mistry J, Mitchell A, Nikolskaya AN, Orchard S, Orengo C, Petryszak R, Selengut JD, Sigrist CJ, Thomas PD, Valentin F, Wilson D, Wu CH, Yeats C. Nucleic Acids Res. 2007; 35:D224–D228. [PubMed: 17202162]
10. Hanks SK, Hunter T. FASEB J. 1995; 9(8):576–596. [PubMed: 7768349]
11. Stout TJ, Foster PG, Matthews DJ. Curr. Pharm. Des. 2004; 10(10):1069–1082. [PubMed: 15078142]
12. Hanks SK, Quinn AM. Methods Enzymol. 1991; 200:38–62. [PubMed: 1956325]
13. Hunter T. Cold Spring Harb. Perspect. Biol. 2014; 6(5):a020644. [PubMed: 24789824]
14. Schomburg I, Chang A, Hofmann O, Ebeling C, Ehrentreich F, Schomburg D. Trends Biochem. Sci. 2002; 27(1):54–56. [PubMed: 11796225]
15. Manning G, Plowman GD, Hunter T, Sudarsanam S. Trends Biochem. Sci. 2002; 27(10):514–520. [PubMed: 12368087]
16. Miranda-Saavedra D, Barton GJ. Proteins. 2007; 68(4):893–914. [PubMed: 17557329]
17. Boudeau J, Miranda-Saavedra D, Barton GJ, Alessi DR. Trends Cell Biol. 2006; 16(9):443–452. [PubMed: 16879967]
18. Zeqiraj E, van Aalten DM. Curr. Opin. Struct. Biol. 2010; 20(6):772–781. [PubMed: 21074407]
19. Paksa A, Raz E. Curr. Opin. Cell Biol. 2015; 36:80–85. [PubMed: 26232877]
20. Phillips JB, Westerfield M. Dis. Model Mech. 2014; 7(7):739–743. [PubMed: 24973743]
21. Gibert Y, Trengove MC, Ward AC. Curr. Med. Chem. 2013; 20(19):2458–2466. [PubMed: 23521675]

22. Howe K, Clark MD, Torroja CF, Torrance J, Berthelot C, Muffato M, Collins JE, Humphray S, McLaren K, Matthews L, McLaren S, Sealy I, Caccamo M, Churcher C, Scott C, Barrett JC, Koch R, Rauch GJ, White S, Chow W, Kilian B, Quintais LT, Guerra-Assuncao JA, Zhou Y, Gu Y, Yen J, Vogel JH, Eyre T, Redmond S, Banerjee R, Chi J, Fu B, Langley E, Maguire SF, Laird GK, Lloyd D, Kenyon E, Donaldson S, Sehra H, Almeida-King J, Loveland J, Trevanion S, Jones M, Quail M, Willey D, Hunt A, Burton J, Sims S, McLay K, Plumb B, Davis J, Clee C, Oliver K, Clark R, Riddle C, Elliot D, Threadgold G, Harden G, Ware D, Begum S, Mortimore B, Kerry G, Heath P, Phillimore B, Tracey A, Corby N, Dunn M, Johnson C, Wood J, Clark S, Pelan S, Griffiths G, Smith M, Glithero R, Howden P, Barker N, Lloyd C, Stevens C, Harley J, Holt K, Panagiotidis G, Lovell J, Beasley H, Henderson C, Gordon D, Auger K, Wright D, Collins J, Raisen C, Dyer L, Leung K, Robertson L, Ambridge K, Leongamornlert D, McGuire S, Gilderthorp R, Griffiths C, Manthravadi D, Nichol S, Barker G, Whitehead S, Kay M, Brown J, Murnane C, Gray E, Humphries M, Sycamore N, Barker D, Saunders D, Wallis J, Babbage A, Hammond S, Mashreghi-Mohammadi M, Barr L, Martin S, Wray P, Ellington A, Matthews N, Ellwood M, Woodmansey R, Clark G, Cooper J, Tromans A, Grafham D, Skuce C, Pandian R, Andrews R, Harrison E, Kimberley A, Garnett J, Fosker N, Hall R, Garner P, Kelly D, Bird C, Palmer S, Gehring I, Berger A, Dooley CM, Ersan-Ürün Z, Eser C, Geiger H, Geisler M, Karotki L, Kirn A, Konantz J, Konantz M, Oberländer M, Rudolph-Geiger S, Teucke M, Lanz C, Raddatz G, Osoegawa K, Zhu B, Rapp A, Widaa S, Langford C, Yang F, Schuster SC, Carter NP, Harrow J, Ning Z, Herrero J, Searle SM, Enright A, Geisler R, Plasterk RH, Lee C, Westerfield M, de Jong PJ, Zon LI, Postlethwait JH, Nüsslein-Volhard C, Hubbard TJ, Roest Crollius H, Rogers J, Stemple DL. Nature. 2013; 496(7446):498–503. [PubMed: 23594743]

23. Glasauer SM, Neuhauss SC. Mol. Genet. Genomics. 2014; 289(6):1045–1060. [PubMed: 25092473]

24. Kok FO, Shin M, Ni CW, Gupta A, Grosse AS, van Impel A, Kirchmaier BC, Peterson-Maduro J, Kourkoulis G, Male I, DeSantis DF, Sheppard-Tindell S, Ebarasi L, Betsholtz C, Schulte-Merker S, Wolfe SA, Lawson ND. Dev. Cell. 2015; 32(1):97–108. [PubMed: 25533206]

25. Cunningham F, Amode MR, Barrell D, Beal K, Billis K, Brent S, Carvalho-Silva D, Clapham P, Coates G, Fitzgerald S, Gil L, Girón CG, Gordon L, Hourlier T, Hunt SE, Janacek SH, Johnson N, Juettemann T, Kähäri AK, Keenan S, Martin FJ, Maurel T, McLaren W, Murphy DN, Nag R, Overduin B, Parker A, Patricio M, Perry E, Pignatelli M, Riat HS, Sheppard D, Taylor K, Thormann A, Vullo A, Wilder SP, Zadissa A, Aken BL, Birney E, Harrow J, Kinsella R, Muffato M, Ruffier M, Searle SMJ, Spudich G, Trevanion SJ, Yates A, Zerbino DR, Flicek P. Nucleic Acids Res. 2015; 43:D662–D669. [PubMed: 25352552]

26. Howe DG, Bradford YM, Conlin T, Eagle AE, Fashena D, Frazer K, Knight J, Mani P, Martin R, Moxon SA, Paddock H, Pich C, Ramachandran S, Ruef BJ, Ruzicka L, Schaper K, Shao X, Singer A, Sprunger B, Van Slyke CE, Westerfield M. Nucleic Acids Res. 2013; 41:D854–D860. [PubMed: 23074187]

27. Corpet F. Nucleic Acids Res. 1988; 16(22):10881–10890. [PubMed: 2849754]

28. McWilliam H, Li W, Uludag M, Squizzato S, Park YM, Buso N, Cowley AP, Lopez R. Nucleic Acids Res. 2013; 41:W597–W600. [PubMed: 23671338]

29. Bollag G, Hirth P, Tsai J, Zhang J, Ibrahim PN, Cho H, Spevak W, Zhang C, Zhang Y, Habets G, Burton EA, Wong B, Tsang G, West BL, Powell B, Shellooe R, Marimuthu A, Nguyen H, Zhang KY, Artis DR, Schlessinger J, Su F, Higgins B, Iyer R, D'Andrea K, Koehler A, Stumm M, Lin PS, Lee RJ, Grippo J, Puzanov I, Kim KB, Ribas A, McArthur GA, Sosman JA, Chapman PB, Flaherty KT, Xu X, Nathanson KL, Nolop K. Nature. 2010; 467(7315):596–599. [PubMed: 20823850]

30. Chrencik JE, Patny A, Leung IK, Korniski B, Emmons TL, Hall T, Weinberg RA, Gormley JA, Williams JM, Day JE, Hirsch JL, Kiefer JR, Leone JW, Fischer HD, Sommers CD, Huang HC, Jacobsen EJ, Tenbrink RE, Tomasselli AG, Benson TE. J. Mol. Biol. 2010; 400(3):413–433. [PubMed: 20478313]

31. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. Nat. Methods. 2015; 12(1):7–8. [PubMed: 25549265]

32. Tagliabracci VS, Pinna LA, Dixon JE. Trends Biochem. Sci. 2013; 38(3):121–130. [PubMed: 23276407]

33. Desvignes T, Pontarotti P, Fauvel C, Bobe J. BMC Evol. Biol. 2009; 9:256. [PubMed: 19852809]
34. Pearce D. Cell Physiol. Biochem. 2003; 13(1):13–20. [PubMed: 12649598]
35. Caenepeel S, Charydczak G, Sudarsanam S, Hunter T, Manning G. Proc. Natl. Acad. Sci. 2004; 101(32):11707–11712. [PubMed: 15289607]
36. Min X, Lee BH, Cobb MH, Goldsmith EJ. Structure. 2004; 12(7):1303–1311. [PubMed: 15242606]
37. Endicott JA, Noble ME. Biochem. Soc. Trans. 2013; 41(4):1008–1016. [PubMed: 23863171]

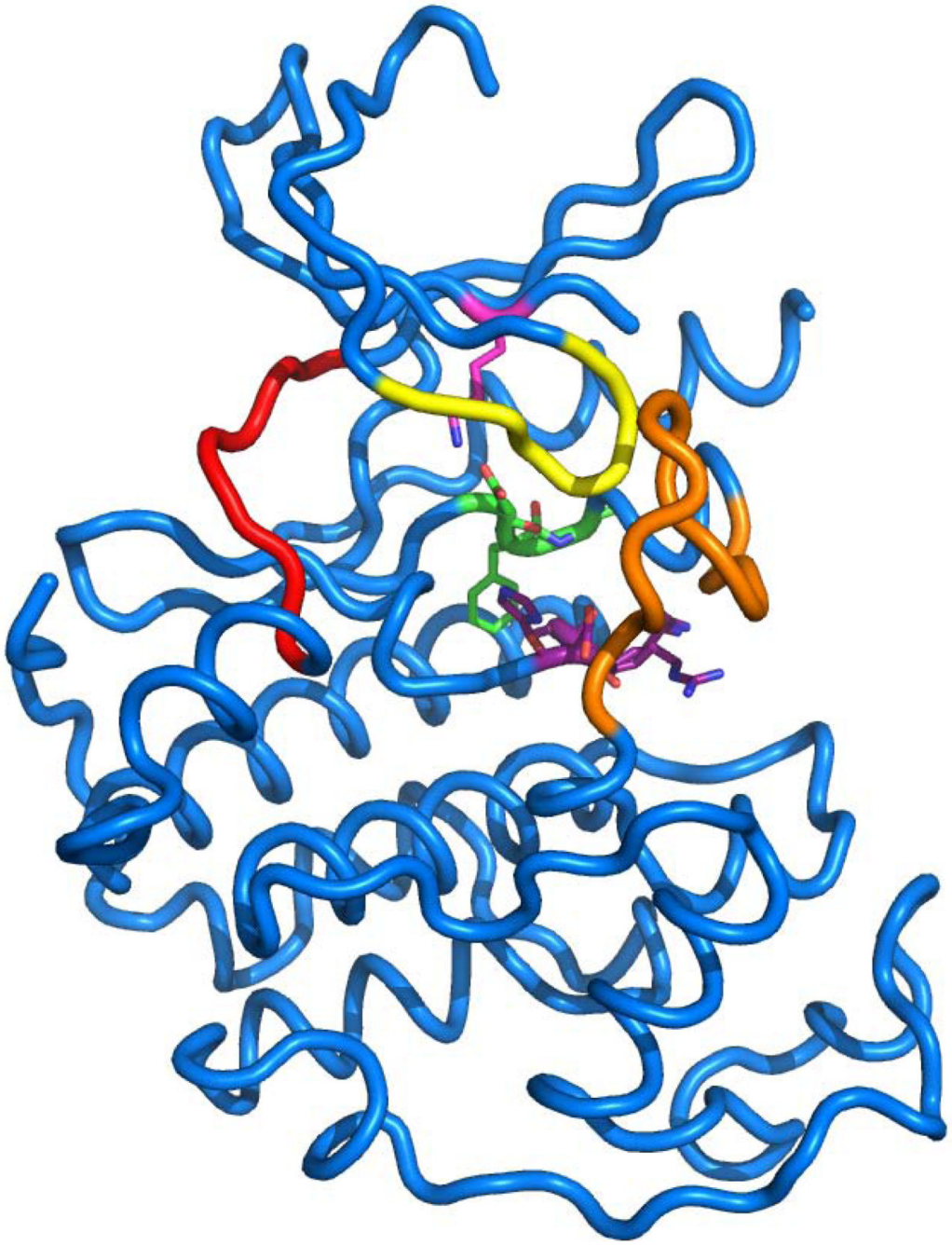

Figure 1.
CDK2 (4EK3) as an example of a classical Hank's kinase with critical Hanks motifs highlighted. Hinge region in red, P-loop in yellow, activation loop in orange, invariant lysine in magenta sticks, HRD in purple sticks, and DFG in green sticks.

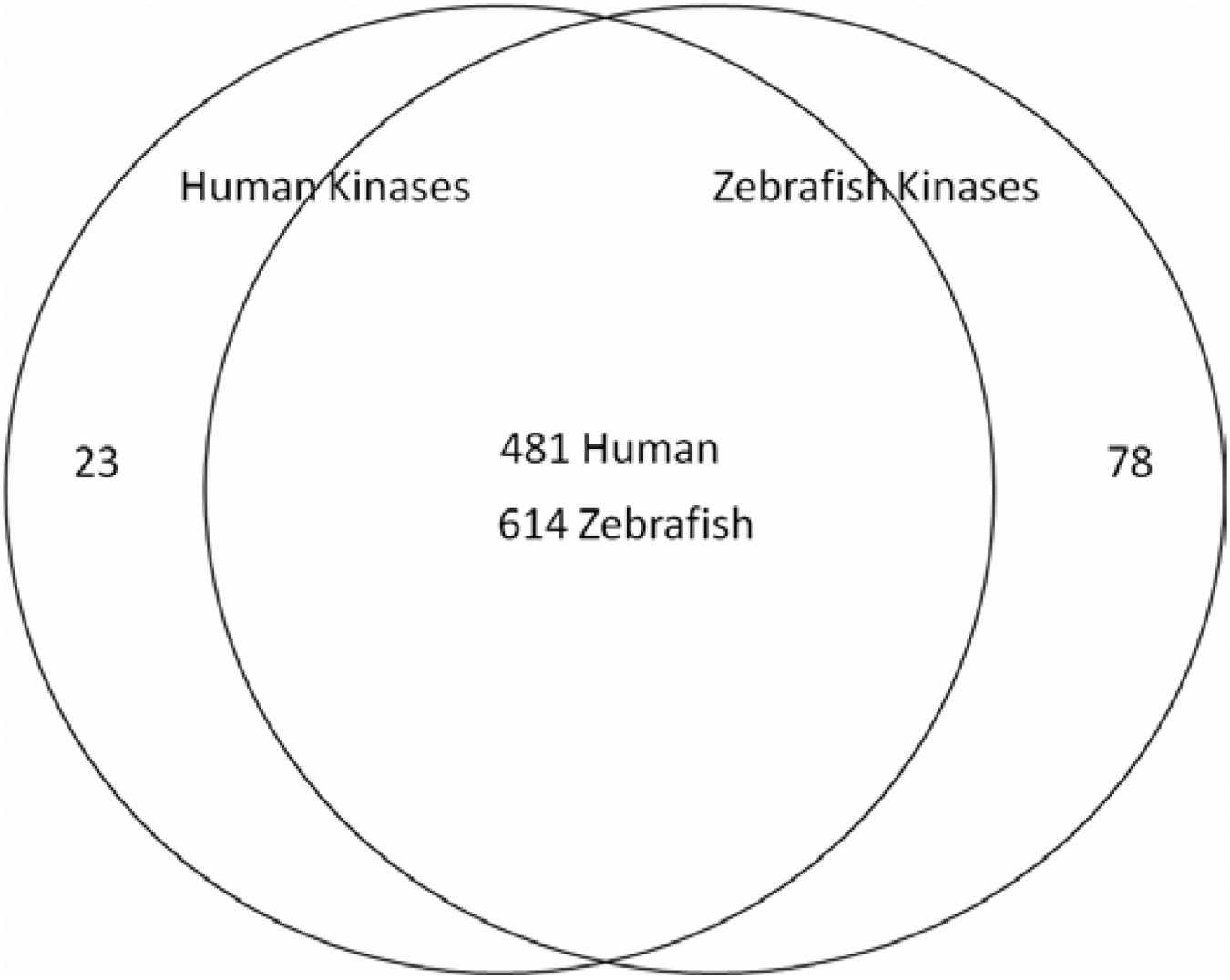

Figure 2.

Unique vs. shared kinase catalytic domains. Of the total 692 non-redundant zebrafish sequences analyzed, 614 had greater than 40% identity to 481 human kinase catalytic domain sequences. Only 23 human sequences were less than 40% identity to any zebrafish sequence and 78 zebrafish sequences were less than 40% identity to any human sequence.

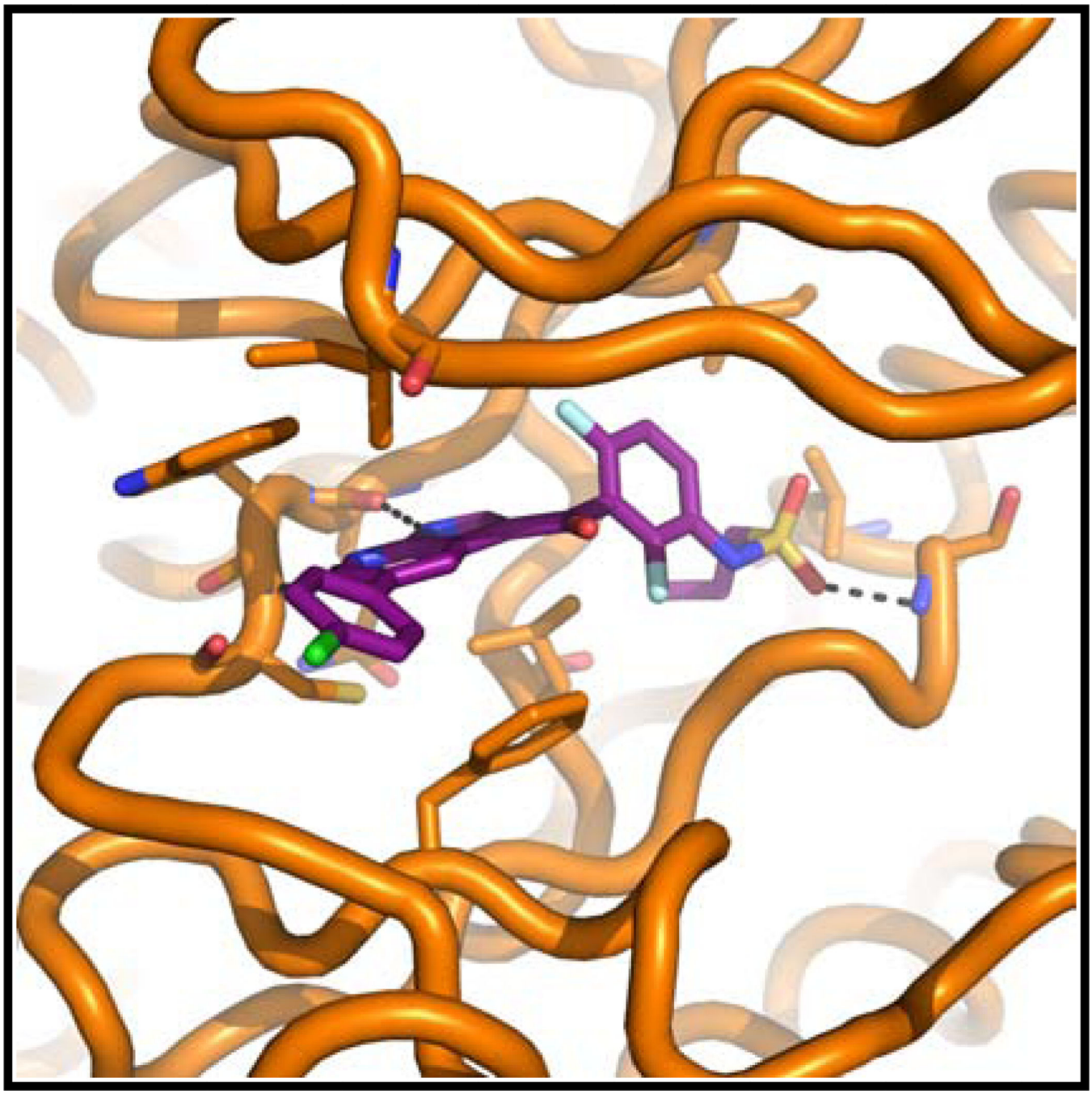

Figure 3.

Vemurafenib in human BRAF kinase. The structure of human BRAF with bound Vemurafenib (3OG7) is shown in orange cartoon tube with critical coordinating residues shown in orange sticks. Vemurafenib is shown in purple sticks and hydrogen bonds to the drug are shown as black dashes. Vemurafenib is an FDA approved kinase inhibitor which selectively blocks human BRAF kinase activity and is indicated to treat some melanomas [4, 29]. The zebrafish BRAF kinase catalytic domain has 92% identity to human BRAF.

Alignment of the zebrafish sequence to the human structure (3OG7) indicates that all drug coordinating residues in human BRAF are identical to zebrafish BRAF.

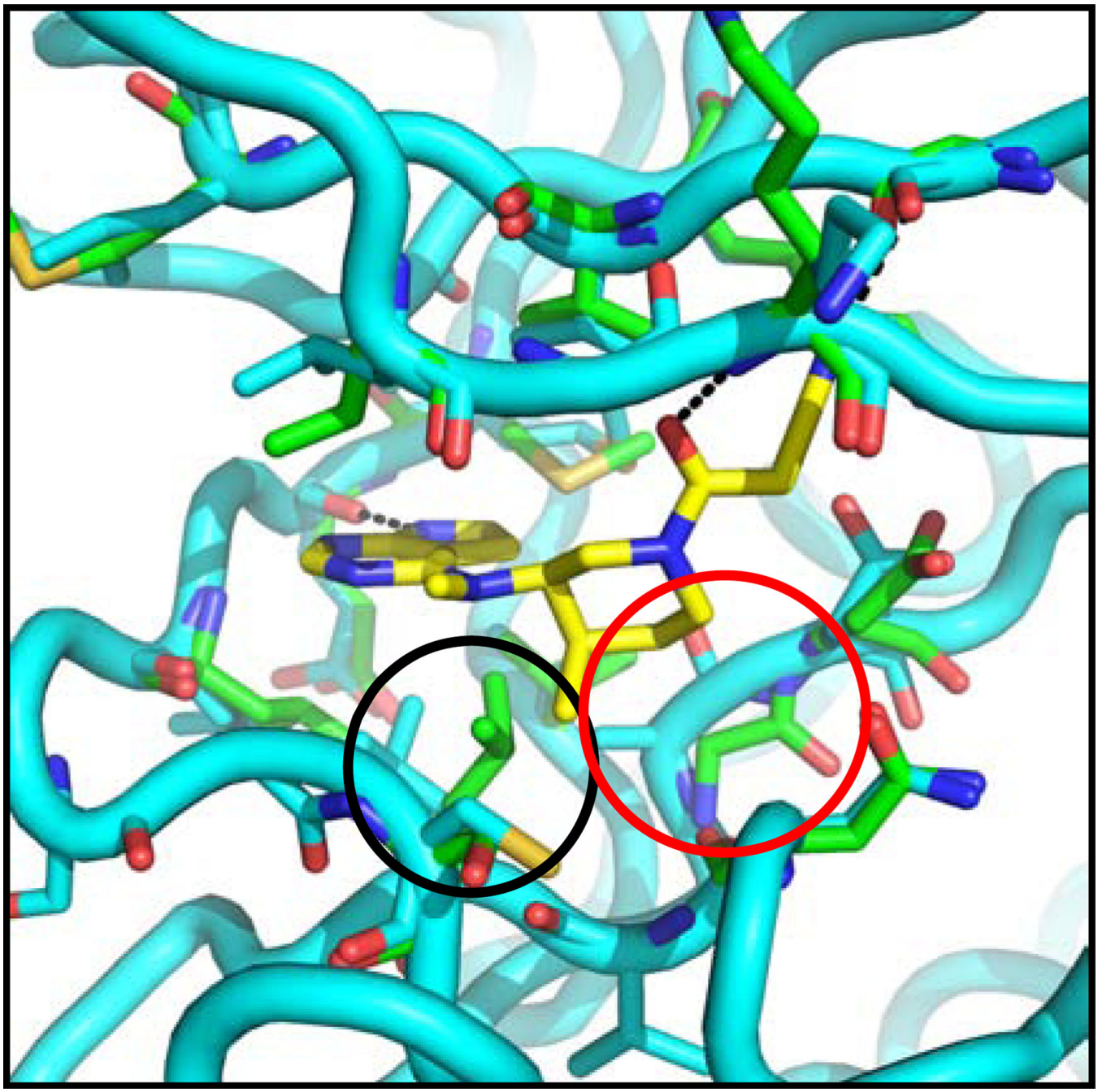

Figure 4.

Tofactinib in human JAK3 kinase. The structure of human JAK3 with bound Tofactinib is shown in cyan cartoon tube with critical coordinating residues shown in cyan sticks (3LXK). Tofactinib is shown in yellow sticks. Zebrafish JAK2B residues are shown as green sticks. Tofactinib hydrogen bonds are shown as black dashes. Tofactinib is an FDA approved kinase inhibitor which selectively blocks human JAK3 kinase activity and is indicated to treat rheumatoid arthritis [4, 30]. The zebrafish JAK2A kinase catalytic domain has 68% identity to human JAK3. Threading the zebrafish sequence to the human JAK3 structure

(3LXK) indicates that all but two residues in the human catalytic pocket JAK3 are identical to zebrafish JAK2A: C909-Hs is equivalent to L142-Dr (circled in black) and A966-Hs is equivalent to G152 (circled in red).

Table 1

Divergent or unique human kinase catalytic domains. Human catalytic domain sequences which do not match any zebrafish sequence with greater than 40% identity are shown below. Several of the unique sequences are members of distinct kinase families.

Unique human kinases	Percent identity	Zebrafish kinase match
COT	39	CHUK
FAM20A	24	MINK1
FAM20B	18	OXSR1B
FAM20C	26	PIM1
FJB1	27	PINK1
GCN2-P	31	MARK1
MLKL	37	TAOK1A
NME1A	21	SIK2A
NME1B	21	JAK1
NME3	20	PRKD3
NME4	21	JAK1
NME5	21	GUCY2F
NME6	21	POMK
NME7a	20	ZGC66101
PAN3	26	DCKL1B
SgK223	20	PRKD3
SgK269	25	PRKD3
SgK396	33	MAP3K12
SGK495	30	PTK6B
Slob	31	CDK13
TXNDC3	23	MYLKA
TXNDC6	19	CAMKVL
ULK4	39	STK36

Table 2

Catalytic domains of divergent or unique zebrafish kinases. Zebrafish catalytic domain sequences which do not match any human sequence with greater than 40% identity are shown below. Most (72/78) unmatched kinases are mutants or splice variants of otherwise well matched partners. Six of the catalytic domains are likely truly unique.

Unique zebrafish kinases	Percent identity	Human kinase match
SICH211-138G9.2	39	AurB
SICH211-138G9.3	39	AurB
SICH211-272B8.7	38	AurB
SIDKEY-22I16.10	39	AurB
SIDKEY-34D22.3	38	AurB
SIDKEY-80C24.1	39	AurB
SIDKEY-80C24.5	39	AurB
SIDKEYP-67E1.2	39	AurB
SIDKEY-183G16.2	38	HUNK
SIDKEY-206F10.6	39	MARK1
SIDKEY-83M22.16	38	MARK1
PAK7	26	MLK1
SIDKEY-155D18.1	37	MSK2A
SIDKEYP-67E1.4	39	MSK2A
PXK	24	NuAK1
SIDKEY-197D18.1	39	NuAK1
SIDKEY-197D18.2	39	NuAK1
G6PD	19	PAK4
SICH211-249H16.8	38	PASK
SICH211-57M13.8	38	PASK
SIDKEY-211E20.1	38	PASK
SIDKEY-37M8.14	39	PASK
SICH211-119D14.4	38	PIM1
SICH211-168M18.3	33	PIM1
SICH211-196G2.7	34	PIM1
SICH211-196G2.8	38	PIM1
SICH211-198E20.11	38	PIM1
SIDKEY-236A14.6	38	PIM1
SIDKEYP-100A1.5	37	PIM1
SICH211-176G13.9	39	PIM2
SICH211-215M21.11	39	PIM2
SICH211-237D5.4	39	PIM2
SICH211-57M13.10	36	PIM2

Unique zebrafish kinases	Percent identity	Human kinase match
SIDKEYP-110E4.3	39	PIM2
SICH211-147H1.4	39	PIM3
SICH211-196N4.5	32	PIM3
SICH211-196N4.7	35	PIM3
SICH211-196N4.8	32	PIM3
SICH211-214C11.2	35	PIM3
SICH211-214C11.4	32	PIM3
SICH211-214C11.7	31	PIM3
SICH211-214C11.8	32	PIM3
SICH211-57M13.3	39	PIM3
SICH211-57M13.5	39	PIM3
SICH211-57M13.6	39	PIM3
SICH211-57M13.7	39	PIM3
SIDKEY-222B8.5	34	PIM3
SIDKEYP-67E1.6	39	PIM3
SIDKEY-211E20.8	39	PKD1
SIDKEY-211E20.9	39	PKD1
SICH211-57M13.9	36	SIK
SICH211-10J20.2	38	TSSK1
SICH211-10J20.4	37	TSSK1
SICH211-215M21.13	37	TSSK1
SIDKEY-206F10.7	39	TSSK1
SIDKEY-217M5.9	38	TSSK1
SIDKEY-197D18.3	39	TSSK2
SICH211-160D20.5	38	TSSK3
SICH211-202N12.2	37	TSSK3
SICH211-207C6.4	37	TSSK3
SICH211-238G23.1	38	TSSK3
SICH211-57M13.1	36	TSSK3
SIDKEY-248F6.3	39	TSSK3
SIZFOS-754C12.2	38	TSSK3
SICH211-12E13.5	38	TSSK4
SICH211-12E13.6	38	TSSK4
PRKD3	27	TTBK2
PKN1B	23	TXNDC3

Table 3

Pseudokinase domain divergence. Of the 48 known human pseudokinase catalytic domains, 22 have various degrees of identity to zebrafish proteins. Pseudokinases with no clear homologous non-redundant partner are left unmatched. Modified critical Hanks domains (DFG, HRD, or VIAK) are indicated based on Boudeau and colleagues [17] and alignment of the matching pairs.

Human pseudokinase	Modified domains	% identity	Zebrafish pseudokinase	Modified domains
ANPA	HRD	84	NPR1A	HRD
ANPB	HRD	--	--	--
CASK	DFG	99	CASKA	DFG
CCK4	DFG	--	--	--
CYGD	HRD	60	GC3	HRD
CYGF	HRD	77	GC2	HRD
EPHA10	DFG, HRD, VIAK	--	--	--
EPHB6	DFG, HRD, VIAK	--	--	--
GCN2-P	HRD, VIAK	--	--	--
HER3 (ERBB3)	HRD	75	ERBB3A	HRD
HSER	HRD	--	--	--
ILK	DFG, HRD	86	ILK	DFG, HRD
IRAK2	DFG, HRD	--	--	--
JAK1-P	HRD	70	JAK1-P	HRD
JAK2-P	HRD	83	JAK2B-P	HRD
JAK3-P	HRD	60	JAK3-P	HRD
KSR1	VIAK	--	--	--
KSR2	VIAK	88	KSR2	VIAK
MLKL	DFG, HRD	--	--	--
NRBP1	DFG, HRD, VIAK	95	NRBP1	DFG, HRD, VIAK
NRBP2	DFG, HRD, VIAK	--	--	--
PSKH2	HRD	--	--	--
RSKL1	DFG, VIAK	--	--	--
RSKL2	DFG, VIAK	--	--	--
SCYL1	DFG, HRD, VIAK	--	--	--
SCYL2	DFG, HRD, VIAK	89	SICH211-244B2.1	DFG, HRD, VIAK
SCYL3	DFG, HRD, VIAK	73	SCYL3	DFG, HRD, VIAK
SGK071	HRD	--	--	--
SGK196	DFG, HRD, VIAK	60	POMK	DFG, HRD, VIAK
SGK223	DFG	--	--	--
SGK269	DFG	--	--	--
SGK307	DFG, HRD	--	--	--
SGK396	HRD	--	--	--

Human pseudokinase	Modified domains	% identity	Zebrafish pseudokinase	Modified domains
SGK494	DFG, HRD, VIAK	59	SGK494A	NONE
SGK495	DFG	--	--	--
SLOB	DFG, HRD, VIAK	--	--	--
STLK5 (STRADA)	DFG, HRD, VIAK	72	STRADA	DFG, HRD, VIAK
STKL6 (STRADB)	DFG, HRD	--	--	--
SURTK106	DFG	--	--	--
TBCK	DFG, HRD, VIAK	62	TBCK	DFG, HRD, VIAK
TRB1	DFG	55	TRIB1	HRD, VIAK
TRB2	DFG	58	TRIB3	DFG, HRD, VIAK
TRB3	DFG	--	--	--
TYK2-P	HRD	64	TYK2-P	HRD
ULK4	DFG, VIAK	--	--	--
VACAMKL	HRD	91	CAMKVA	HRD
VRK3	DFG, HRD, VIAK	46	VRK3	DFG, HRD, VIAK