Endocrine, *vol. 19, no. 3, 333–337, December 2002* 0969–711X/02/19:333–337/$20.00

Prediction of Tyrosine Sulfation in Seven-Transmembrane Peptide Receptors

Kristine M. Yu,[1] Justin Liu,[1] Ryan Moy,[1] Henry C. Lin,[1] Hugh B. Nicholas, Jr.,[2] and Grace L. Rosenquist[1]

[1]Section of Neurobiology, Physiology and Behavior, University of California, Davis, CA; and [2]Pittsburgh Supercomputing Center, Mellon Institute, 4400 Fifth Avenue, Pittsburgh, PA

Posttranslational modification by tyrosine sulfation regulates many important protein–protein interactions and modulates the binding affinity and specificity of seven-transmembrane peptide receptors. We developed a log-odds position-specific-scoring-matrix (PSSM) to accurately predict tyrosine sulfation using 62 tyrosine sites known to be sulfated and 421 tyrosine sites known not to be sulfated. We predict that 49 tyrosines of 32 seven-transmembrane peptide receptors are sulfated. Although we did not incorporate characteristics of confirmed sulfation sites such as clustering and conservation across species into our PSSM, our predicted sites nevertheless exhibited these characteristics. The observed conservation suggests that there are strong evolutionary pressures to preserve selected biological activity of seven-transmembrane receptors. The predicted tyrosine sulfation sites predominantly occur in the extracellular tail and extracellular loop 2, regions consistent with their association with binding pockets of the receptor.

Key Words: Tyrosine sulfation; posttranslational processing; seven-transmembrane receptors; G-coupled protein receptors; peptides.

Introduction

Tyrosine sulfation, catalyzed by tyrosylprotein sulfotransferases (TPSTs), is the most common posttranslational modification of tyrosine residues transported through the Golgi system in eukaryotes *(1,2)*. The importance of tyrosine sulfation became widely recognized in 1999 after the demonstration of sulfation in the seven-transmembrane (7TM) receptor CCR5 *(3,4)*. Recent studies have strongly supported that tyrosine sulfation is also required for optimal protein–protein interactions and specific functions of many other 7TM receptors, such as CCR2 *(5)*, CX3CR1 *(6)*, C5aR *(7)*, and CXCR4 *(8)*.

Received September 9, 2002; Revised October 17 and November 15, 2002; Accepted November 15, 2002.

Author to whom all correspondence and reprint requests should be addressed: Grace L. Rosenquist, Section of Neurobiology, Physiology and Behavior, University of California, 1 Shields Avenue, Davis, CA 95616. E-mail: rosenqui@psc.edu

While information on tyrosine sulfation of chemokine and chemotactic factor receptors is well established, knowledge about tyrosine sulfation in 7TM peptide receptors is more preliminary. Currently, only the thyrotropin receptor (TSHr) has been directly shown to be tyrosine sulfated, although experimental evidence suggests that the luteinizing hormone receptor (LHr) and the follicle-stimulating receptor (FSHr) are likely to be tyrosine sulfated as well *(9)*. However, tyrosine sulfation of other 7TM peptide receptors has not been reported. While the impact of tyrosine sulfation on 7TM peptide receptors remains to be described, this posttranslational modification may be involved with many of the functions associated with regulatory peptide receptors, such as anti-inflammation, glucose metabolism, food intake, and the response to stress. Thus, the identification of tyrosine sulfation in 7TM peptide receptors may expedite the study of 7TM peptide receptors and elucidate the nature of their interactions with other ligands and proteins.

In our study, a position-specific-scoring-matrix (PSSM) was used to predict the sulfation of 49 tyrosines of 32 7TM peptide receptors. These predictions are consistent with what is known about tyrosine sulfation. Most notably, the locations of the predicted sulfated tyrosines in the receptors correspond to those of experimentally confirmed sites.

Results

Forty-nine tyrosines of 32 receptors were predicted to be sulfated out of a total of 756 tyrosine sites (Table 1). Out of these 49 predicted tyrosines, 36 (73.5%) are located in the N-terminal extracellular tail of the receptors or in the second extracellular loop between transmembrane helices 4 and 5 (Fig. 1). In addition, a few sites in the first and third extracellular loops were also predicted to be sulfated. Unexpectedly, five predicted tyrosine sulfation sites are located in the cytoplasmic tail.

When multiple tyrosines of a receptor were predicted to be sulfated, they were often located in proximity to each other. For instance, the corticotropin-releasing-factor (CRF) receptor 1, FMLP-related, FSH, glucagon-like peptide (GLP), and LSH receptors possess clusters of predicted sulfated tyrosines located within five residues of each other (Table 1, highlighted). This clustering has previously been noted in other experimentally determined tyrosine sulfation sites.

Table 1
Scores of Tyrosine Sites in Peptide Receptors Predicted to be Sulfated

Common Name	Swiss-Prot Name	Tyr #	Sulfation Site	Score
Calcitonin gene-related peptide receptor	CGRR_HUMAN	365	iaeevYdyimh	4.252
Corticotropin releasing factor receptor 1 precursor	CRF1_HUMAN	282	igklyYdnekc	3.920
		299	gvytdYiyqgp	2.816
		301	ytdyiYqgpmi	2.771
Corticotropin releasing factor receptor 2 precursor	CRF2_HUMAN	249	igklyYeneqc	2.900
Diuretic hormone receptor	DIHR_ACHDO	116	rnysdYvhcre	2.583
		291	wmaedYfdwih	7.892
Endothelin-1 receptor	ET1R_HUMAN	251	kfmefYqdvkd	2.371
FMLP-related receptor 1	FML1_HUMAN	12	tplneYeevsy	3.867
		17	yeevsYesagy	2.407
Follicle-stimulating hormone receptor	FSHR_HUMAN	330	gfdmtYtefdy	2.544
		335[b]	ytefdYdlcne	6.733
Glucagon-like peptide receptor	GLP1_HUMAN	69	rtfdeYacwpd	7.195
		289	wgivkYlyede	2.249
		291	ivkylYedegc	2.321
Glucagon receptor	GLR_HUMAN	65	rtfdkYscwpd	5.003
Isotocin receptor	ITR_CATCO	189	igngvYdcwgd	3.790
Leucine-rich repeat-containing G protein-coupled receptor 4 precursor	LGR4_HUMAN	843	eqdfyYdcgmy	5.572
		908	sahsdYadeed	6.676
Leucine-rich repeat-containing G protein-coupled receptor 5 precursor	LGR5_HUMAN	387	rhneiYeikvd	3.636
Leucine-rich repeat-containing G protein-coupled receptor 6	LGR6_HUMAN	384	qaenhYdqdld	7.698
Relaxin receptor	LGR8_HUMAN	578	cfplyYdqted	2.846
Luteinizing hormone receptor	LSHR_HUMAN	331[b]	lsgwdYeygfc	1.714
		333[b]	lsgwdYeygfc	3.053
Neuropeptide FF receptor 1	NFF1_HUMAN	201	rsyplYscwea	2.120
Neuropeptide FF receptor 2	NFF2_HUMAN	401	mmlsdYadlsp	4.441
Neuromedin-B receptor	NMBR_HUMAN	293	yrsfnYneidp	2.651
Neuropeptide Y receptor type 1	NY1R_HUMAN	347	srdddYetiam	5.399
Neuropeptide Y receptor type 5	NY5R_HUMAN	9	yskqdYnmdle	2.358
Neuropeptide Y receptor	NYR_DROME	18	neddnYqegyf	7.388
		76	fscddYdllse	4.700
Orexin receptor 1	OX1R_HUMAN	27	pvppdYedefl	4.737
		41	wrdylYpkqye	3.936
		211	waddlYpkiyh	2.362
Orexin receptor 2	OX2R_HUMAN	34	lnptdYddeef	9.047
		219	wggeiYpkmyh	2.813
Somatostatin receptor 1	SSR1_HUMAN	353	eepvdYyatal	2.529
Somatostatin receptor 2	SSR2_HUMAN	38	qtepyYdltsn	2.941
Somatostatin receptor 4	SSR4_HUMAN	346	eepldYyatal	3.932
Tachykinin-like receptor 1	TLR1_DROME	245	wpdgrYptsma	3.684
Tackykinin-like receptor 2	TLR2_DROME	12	feaddYgdisw	7.186
		159	nvtfnYyymld	3.976
		255	nrtvcYpewpd	4.372
Thyrotropin releasing hormone receptor	TRFR_HUMAN	88	itdsiYgswvy	2.464
Thyrotropin receptor	TSHR_HUMAN	385[a]	afdshYdytic	4.499
		387[a]	dshydYticgd	5.319
		482	thseyYnhaid	2.976
Vasopressin receptor	V1AR_HUMAN	216	wgsraYvtwmt	3.611
Vasoactive intestinal polypeptide receptor 1	VIPR_HUMAN	39	qeecdYvqmie	5.594
		283	ihfedYgcwdt	8.807

[a]Experimentally proven tyrosine sulfation sites. These sites were incorporated into the PSSM and their scores are presented for the purpose of discussion.

[b]Tyrosine sites with some experimental evidence for sulfation.

The predicted tyrosine sulfation sites that exhibit clustering are highlighted.

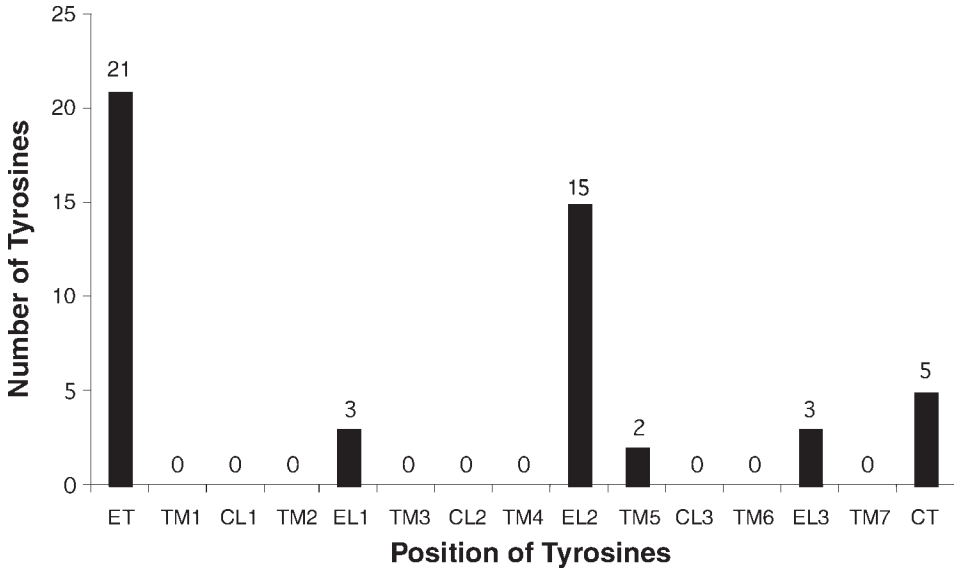

Fig. 1. Overall distribution of tyrosines predicted to be sulfated. The predictions are highly consistent and the majority of the predicted sulfation sites occur at the extracellular regions of the receptors, especially on the extracellular tail (ET) and the second extracellular loop (EL2) between helices 4 and 5. TM, transmembrane helix; CL, cytoplasmic loop; EL, extracellular loop 1, 3; CT, cytoplasmic tail.

Table 2
Conservation of Tyrosine Sulfation Sites

Species	Tyr #	Sulfation Site
Follicle stimulating hormone receptor		
Human	335	y t e f d Y d l c n e
Bovine	335	y s e f d Y d l c n e
Chicken	335	e n e f d Y g l c n e
Donkey	327	y s e f d Y d l c n e
Horse	334	y s e f e Y d l c n e
Cynomologus Monkey	335	y a e f d Y d l c n e
Mouse	334	y s e f d Y d l c n e
Pig	335	y s e f d Y d l c n e
Rat	334	y n e f d Y d l c n e
Sheep	335	y s e f d Y d l c s e
Corticotropin releasing factor receptor 1		
Human	301	y t d y i Y q g p m i
Chicken	277	y t d y i Y q g p m i
Mouse	272	y t d y i Y q g p m i
Rat	272	y t d y i Y q g p m i
Sheep	272	y t d y i Y q g p m i
Xenopus	272	y t d f i Y q g p v i

Conservation Percent
100%
80% or more
60% or more
Less than 60%

Tyrosine sites predicted to be sulfated are conserved in many species. The high degree of conservation may reflect the importance of tyrosine sulfation and its contribution toward evolutionary fitness.

Predicted tyrosine sulfation sites are conserved across numerous species. Both FSH receptor and CRF receptor 1 exhibit a high degree of conservation surrounding predicted tyrosines across several species (Table 2).

Discussion

Posttranslational modification by tyrosine sulfation is important for modulating protein–protein interactions in many receptors *(3)*. While it has been estimated that up to 1% of all tyrosines in eukaryotes are sulfated *(10)*, it is likely that sulfation is much more common than previously thought. In particular, this study has predicted that of the 70 7TM peptide receptors *(11)*, 32 are tyrosine sulfated. These predictions are supported by observations consistent with previous studies: the clustering of the predicted sulfation sites *(12)*, the preponderance of predicted sites near the N-terminal and

other extracellular regions of the receptors, and the conservation of predicted sites across species. The high degree of conservation observed in the predicted sulfation sites, as well as their abundance in ligand binding domains of 7TM peptide receptors, shows the importance of tyrosine sulfation in 7TM peptide receptors. Previously, studies have established the importance of tyrosine sulfation in chemokine and glycoprotein receptors *(9)*.

The most striking aspect of the predicted sites is that most of the sites are located in the extracellular regions of the receptor. For tyrosines to be accessible to TPST in the Golgi lumen, they must be located close to the N-terminal or in the extracellular loops. This is because the active site of TPST is lumenally oriented *(13)*, as are the hydrophilic extracellular N terminal tail and loops *(14)*. Moreover, it is well known that the binding sites for ligands of the majority of 7TM peptide receptors involve the extracellular domains *(15)*, and, in particular, the second extracellular loop. Of the 49 tyrosines predicted to be sulfated, 42 (85.7%) are located in the extracellular region of the receptors as expected, and all but 6 of these (73.5%) are located in the extracellular tail and the second extracellular loop between transmembrane helices 4 and 5. Interestingly, the predicted tyrosine sulfation sites of the glucagon and glucagon-like receptors fall in a putative hormone-binding domain, indicating a possible role for tyrosine sulfation in hormone binding. The predicted sites, tyr65 of the glucagon receptor and tyr69 of the glucagon-like peptide receptor 1, are two residues from an aspartate in the conserved region that may be critical for ligand binding in the family B hormone receptors *(16)*. In addition, tyr103 of the glucagon-like peptide receptor 2, also two residues downstream from the conserved aspartate, was the only site to have a positive score for that receptor (0.700).

The large percentage of predicted tyrosine sulfation sites in the extracellular terminal and second extracellular loop points to the importance of exposing the putative sulfation site to TPST. However, five of the tyrosines predicted to be sulfated are located on the cytoplasmic tail of the 7TM receptors and are thus not able to be sulfated by TPST. Experiments have shown that tyrosine phosphorylation rather than tyrosine sulfation may occur in the cytoplasmic tail; indeed, two of the cytoplasmic sites predicted to be sulfated are in the leucine-rich repeat-containing GPCR 4 precursor, which is known to be tyrosine phosphorylated in nematodes *(17)*. The PSSM predicts that these cytoplasmic tyrosines are sulfated due to the similarity between tyrosine sulfation and phosphorylation sites *(18)*. The only tyrosines predicted to be sulfated in the transmembrane region are tyr299 and tyr301 in TM5 of the CRF receptor 1. These two tyrosines flank the boundary between the second extracellular loop and the head of TM5, which begins at residue 298. While it may be possible for TPST to access TM5, boundaries between TMs and loops may be dynamic, so that the tyrosines predicted to be sulfated in TM5 are actually in the second extracellular loop *(19)*.

Aligning homologous sites revealed a high degree of conservation across animal species. Conservation of the FSH receptor sequences supports the evidence for sulfation at tyr335, as does the evidence that tyrosine at this location is necessary for signaling of this receptor *(9)*. Although CRF1_XENLA (*Xenopus*) had a negative score due to the substitution of val (−0.978) for met (2.001) at position Y+4, it contains only two amino acid differences from the otherwise completely conserved set of mammalian CRF1 sequences. This contradictory negative score is likely an artifact of the small size of the positive sites training set and such discrepancies will be minimized as more positive sites are experimentally confirmed and added to the PSSM. In any case, the extensive conservation of the tyrosine sequences may be indicative of strong evolutionary pressure to preserve vital biological functions encoded by the conserved amino acids.

The experimental work *(9)* on the glycoprotein hormone receptors LSHr, FSHr, and TSHr corroborates the accuracy of the PSSM. In these experiments, sulfation of tyr385 and tyr387 was shown to be necessary for high-affinity binding in TSHr, and it was suggested that sulfation of the homologous motif in LSHr and in FSHr played a similar role *(9)*. As expected, because they were included in the training set of experimentally confirmed sulfation sites for the PSSM, tyr385 and tyr387 of TSHr had high scores. Moreover, tyr335 of FSHR and tyr333 and tyr331 of LSHr are probably sulfated, according to experimental evidence, and had PSSM scores of 6.733, 3.053, and 1.714, respectively, with the score for tyr331 of LSHr just below the threshold score of 2.

The PSSM can reliably direct the discovery of new 7TM peptide receptor tyrosine sulfation sites as demonstrated by the agreement between the predicted tyrosine sulfation sites and those ascertained by experiment in the glycoprotein hormone receptors. The advantage of using the PSSM over the strict consensus sequence derived from the highest-scoring residue in every position, WDDDDYDDMMD, is that the much less informative consensus sequence does not convey all of the properties of the tyrosine sulfation site. That is, it conveys no information about the relative importance of or the degree and nature of the variation allowed at different positions in the site. This information, which is explicit in the log-odds scores of the PSSM, is important, because, except for the tyrosine to be sulfated, no position in the site is completely invariant. Most positions within the site have more than one amino acid that result in a positive, favorable score. For example, one of the putative tyrosine sulfation sites, the vasopressin receptor (V1AR_HUMAN), contains no acidic residues and even has a basic arginine and some bulky residues. Our interpretation of this observation is that rather than conforming to a consensus sequence, tyrosine sulfation sites occur where the tyrosine is accessible to TPST, and secondary and tertiary structure of the

site are factors for tyrosine sulfation in addition to primary structure *(2)*. Knowledge of tyrosine sulfation sites in 7TM receptors and the roles they play in mediating the interactions between the receptors and their various ligands will undoubtedly improve our understanding of signaling pathways and processes, of which these 7TM peptide receptors play an integral part.

Materials and Methods

Acquisition of Hormone and Regulatory Peptide Receptors Protein Sequences

The 7TM peptide receptor sequences were obtained from the SWISS-PROT Database Release 40.22 of June 24, 2002 *(20)*. Tyrosine sulfation sites, which were comprised of the target tyrosine and the five flanking amino acids on either side, were extracted from the sequences using the Pick sites Program accessed from the Pittsburgh Supercomputing Center (PSC). Each known tyrosine sulfation site must have direct evidence of tyrosine O-sulfate for it to be considered sulfated.

The Position-Specific-Scoring-Matrix

The position-specific-scoring-matrix (PSSM) used 62 tyrosine sites known to be sulfated and 421 sites known not to be sulfated. All known tyrosine sulfation sites, including sites from 7TM receptors, were incorporated into the positive training set of 62 tyrosine sites. The known sulfation sites included in the PSSM from 7TM peptide receptors were tyr385 and tyr387 of TSHr. The PSSM formally organizes experimental observations into an information theory framework. This framework allows us to ask, in an objective and repeatable manner, whether the amino acid sequence around any tyrosine is more similar to those around the sulfated tyrosines or more similar to those around the nonsulfated tyrosines. Any score above zero indicates that the amino acids in the site are more similar to those in sites that are known to be sulfated than to those in sites known not to be sulfated, and the magnitude of the score indicates the degree of similarity. A detailed description of this method was published previously *(2)*. Henikoff pseudocounts *(21)*, calculated with a multiplier of 5 ($m = 5$), were added to compensate for zeros in the observed counts of amino acids *(2)*. Sixty-two jackknifed PSSMs were constructed by using only 61 of the 62 tyrosine sites from the known sequences. Then, each PSSM was used to score the one missing site and the receiver operating characteristic (ROC) was calculated. The ROC score for the jackknifed PSSM contrasting known positive and negative sites was 0.979 *(22)*. This score indicates that the PSSM has a prediction accuracy of 97.9% for unknown sites *(2)*. The minimum score for a predicted tyrosine sulfation site was set at 2 to equalize the probability of false positive and false negative predictions.

Acknowledgments

We would like to thank Michelle Louie and Timothy Yu for editorial assistance. This research was supported by grant P41 RR06009 from the National Center for Research Resources.

References

1. Niehrs, C., Beisswanger, R., and Huttner, W. B. (1994). *Chem. Biol. Interact.* **92,** 257–271.
2. Nicholas, H. B., Chan, S., and Rosenquist, G. L. (1999). *Endocrine* **11,** 285–292.
3. Kehoe, J. W. and Bertozzi, C. R. (2000). *Chem. Biol.* **7,** R57–R61.
4. Farzan, M., Babcock, G. J., Vasilieva, N., et al. (1999). *Cell* **96,** 667–676.
5. Preobrazhensky, A., Dragan, S., Kawano, T., et al. (2000). *J. Immunol.* **165,** 5295–5303.
6. Fong, A., Alam, S. M., Imai, T., Bodduluri, H., and Patel, D. (2002). *J. Biol. Chem*. **277,** 19418–19423.
7. Farzan, M., Schnitzler, C. E., Vasilieva, N., et al. (2001). *J. Exp. Med.* **193,** 1095–1066.
8. Farzan, M., Mirzabekov, T., Kolchinsky, P., et al. (2002). *J. Biol. Chem.* **277,** 29484–29489.
9. Costagliola, S., Panneels, V., Bonomi, M., et al. (2002). *EMBO J.* **21,** 504–513.
10. Baeuerle, P. A. and Huttner, W. B. (1985). *J. Biol. Chem.* **260,** 6434–6439.
11. GPCRDB Sequence-Derived Data. *GPCRDB* Online. Internet. 8 Mar 2002. Available at http://www.gpcr.org/7tm/mutali/mutali.html.
12. Hortin, G. L., Farries, T. C., Graham, J. P., and Atkinson, J. P. (1989). *Proc. Natl. Acad. Sci. USA* **86,** 1338–1342.
13. Ouyang, Y., Lane, W. S., and Moore, K. L. (1998). *Proc. Natl. Acad. Sci. USA* **95,** 2896–2901.
14. Gether, U. and Kobilka, B. K. (1998). *J. Biol. Chem.* **273,** 17979–17982.
15. Gether, U. (2000). *Endocr. Rev*. **21,** 90–113.
16. Harmar, A. J. (2001). *Genome Biol.* **2,** 1–10.
17. Kudo, M., Chen, T., Nakabayashi, K., Hsu, S. Y., and Hsueh, A. J. (2000). *Mol. Endocrinol.* **14,** 272–284.
18. Niehrs, C., Kraft, M. Kee, R. W. H., and Huttner, W. B. (1990). *J. Biol. Chem*. **265,** 8525–8532.
19. Ji, T. H., Grossman, M., and Ji, I. (1998). *J. Biol. Chem.* **273,** 17299–17302.
20. Swiss Prot. *Swiss Prot Protein Knowledgebase* Online. Internet. 29 June 2002. Available at http://www.expasy.ch/cgi-bin/lists?7tmrlist.txt.
21. Henikoff, J. G. and Henikoff, S. (1996). *CABIOS* **12,** 135–143.
22. Metz, C. E. (1978). *Sem. Nucl. Med.* **8,** 283–298.