

(19)



**Евразийское  
патентное  
ведомство**

(21) **201401199** (13) **A1**

(12) **ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОЙ ЗАЯВКЕ**

(43) Дата публикации заявки  
**2015.08.31**

(51) Int. Cl. **C12N 9/42 (2006.01)**

(22) Дата подачи заявки  
**2013.04.30**

---

(54) **ЭНДОГЛЮКАНАЗЫ С УЛУЧШЕННЫМИ СВОЙСТВАМИ**

---

(31) **12166458.5**

(32) **2012.05.02**

(33) **EP**

(86) **PCT/EP2013/058985**

(87) **WO 2013/164340 2013.11.07**

(71) Заявитель:

**КЛАРИАНТ ПРОДУКТЕ  
(ДОЙЧЛАНД) ГМБХ (DE)**

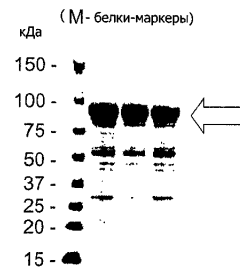
(72) Изобретатель:

**Райзингер Кристоф, Кларен Йорг,  
Унтерштрассер Изабель (DE),  
Митрович Александра (AT), Фликер  
Карлхайнц, Гебхард Габи (DE)**

(74) Представитель:

**Саломатина И.С. (RU)**

(57) Настоящее изобретение имеет отношение к термостабильным эндоглюканазам, конкретно к белкам, имеющим эндоглюканазную активность, которые содержат аминокислотную последовательность, имеющую по меньшей мере 96%-ную идентичность с SEQ. ID NO: 2, и белкам, имеющим эндоглюканазную активность, которые принадлежат к GH7-классу и которые демонстрируют активную термостабилизацию.



**201401199**

**A1**

**A1**

**201401199**

## ЭНДОГЛЮКАНАЗЫ С УЛУЧШЕННЫМИ СВОЙСТВАМИ

### Область изобретения и предшествующий уровень техники

Целлюлоза является основным компонентом растительного материала. Она является основой структурной целостности растений и часто встречается в лигноцеллюлозном матриксе, состоящем из целлюлозы, гемицеллюлоз и лигнина. Области применения с использованием целлюлозы используют преимущества или ее структурных свойств (волокна, текстиль, бумага и т.п.), или ее углеводную природу для получения D-глюкозы, целлобиозы и/или олигомеров целлюлозы.

Лигноцеллюлозы легко доступны из (продуктов) сельского и лесного хозяйства, включающие потоки побочных продуктов (переработки) зерновых, кукурузы, сахарного тростника, сахарной свеклы, древесины и т.п. В ближайшем будущем растения с оптимизированным содержанием лигноцеллюлозы и урожайностью ("энергетические культуры"), по-видимому, будут вносить вклад в качестве важного ресурса.

Целлюлазы содержат структурно и функционально разнообразный класс гликогидролаз, действующих на целлюлозу. Целлюлазы обнаружены в бактериях, археях, грибах и растениях. В целом, обладая активностью гидролитического расщепления гликозидных связей, присутствующих в полимерах целлюлозы или олигомерах, они (целлюлазы) различаются по субстратной специфичности, механизму действия и ферментативным параметрам, включающим процессивность, температурный и pH оптимум. Большинство целлюлаз действуют на  $\beta$ -1,4-связи между двумя остатками глюкозы. Однако другие связи, обнаруженные в лигноцеллюлозах, также могут быть гидролизованы. По механизму их действия целлюлазы могут быть подразделены на эндо- и экзоферменты. Эндоглиюканазы расщепляют полимер целлюлозы в случайных местах, посредством чего снижая степень полимеризации. Экзоферменты, такие как целлобиогидролазы, работают в последовательном механизме действия, высвобождая целлобиозу (D-глюкозо- $\beta$ -1,4-D-глюкопиранозид) из редуцирующего или нередуцирующего конца полимера.

База данных CAZY [Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert, resource for Glycogenomics. *Nucleic Acids Res* 37:D233-238 PMID: 18838391] содержит, среди прочих, коллекцию известных гликогидролаз, включающих ферменты, разрушающие целлюлозу, (т.е. целлюлазы). В этой базе данных ферменты классифицированы в различные GH-классы (гликозилгидролаз (GH)) согласно структурным элементам. Некоторые GH-классы

включают эндоглюканызы, в частности, классы GH5, GH7, GH9, GH12, GH16, GH45, GH48, GH61 и GH74. Несмотря на большое разнообразие внутри некоторых GH-классов, члены одного GH-класса часто имеют сходные физические и ферментативные параметры. Это позволяет сделать общие утверждения для членов определенного GH-класса, такие как субстратная специфичность, диапазон pH, стабильность или каталитическая эффективность.

Микроорганизмы, разрушающие целлюлозу, часто производят и секретируют сложную смесь целлюлаз. Например, в секретоме *Trichoderma reesei* (во всех белках, секретируемых *Trichoderma reesei*) было идентифицировано 7 эндоглюканаз, принадлежащим к 6 различным GH-классам (Cel5A, Cel7B, Cel12A, Cel45A, Cel61A, Cel61B, Cel74A). Различные эндоглюканызы демонстрируют спектр свойств (Karlsson J, Siika-aho M, Tenkanen M, Tjerneld F. Enzymatic properties of the low molecular mass endoglucanases Cel12A (EG III) and Cel45A (EG V) of *Trichoderma reesei*. J Biotechnol. 2002 Oct 9;99(1):63-78. PubMed PMID: 12; Karlsson J, Momcilovic D, Wittgren B, Schülein M, Tjerneld F, Brinkmalm G. Enzymatic degradation of carboxymethyl cellulose hydrolyzed by the endoglucanases Cel5A, Cel7B, and Cel45A from *Hemicella insolens* and Cel7B, Cel12A and Cel45A core from *Trichoderma reesei*. Biopolymers. 2002 Jan;63(1):32-40. PubMed PMID: 11754346.). Считается, что две преобладающих эндоглюканызы, EGI (Cel7B, GH7) и EGII (Cel5A) являются наиболее активными их ферментами.

Синергическая активность целлюлолитических ферментов обеспечивает эффективное разрушение сложных субстратов (B. Henrissat, H. Driguez, C. Viet & M. Schülein: Synergism of Cellulases from *Trichoderma reesei* in the Degradation of Cellulose; Nature Biotechnology 3, 722 - 726 (1985) doi:10.1038/nbt0885-722) и исключает замену компонента одного структурного класса, ферментом из другого класса, тогда как, в то же самое время, гидролитическую эффективность необходимо поддерживать на максимальном уровне (неэквивалентность различных эндоглюканаз (EGs)). Простая замена другим классом ферментов GH не всегда возможна. В общем, члены эндоглюканаз семейства GH5 (включающие EGs термофильных бактерий) демонстрируют более высокую термостабильность по сравнению с эндоглюканазми семейства GH7; тем не менее, применение термостабильного белка семейства GH7 часто является выгодным для высоких скоростей гидролиза.

Сообщалось о многих применениях эндоглюканаз, как в виде части сложных композиций ферментов, так и отдельных ферментативных активностей. Целлюлазы являются важными для производства биологических видов топлива, полученных из целлюлозы. После резки и, возможно, химической и/или физической предварительной

обработки, лигноцеллюлозы инкубируют с целлюлазами для высвобождения мономеров сахаров, которые дополнительно подвергают обработке. Режимы обработки должны быть адаптированы для оптимизации скоростей гидролиза, выходов продукта и/или стабильности. Более высокие температуры часто являются предпочтительными в этих способах, но требуются более термостабильные ферменты. Способы одновременного осахаривания и гидролиза (ферментации) (SSF) требуют целлюлолитических ферментов, которые являются активными в условиях ферментации. Консолидированная биообработка сырья (CBP) дополнительно требует объединения свойств ферментов, для выполнения в одну стадию получения фермента, осахаривания и ферментации.

Другие применения эндоглюканаз нацелены только на частичный гидролиз или модификацию целлюлозных волокон (модификацию волокна, биополировку, биологическое состаривание и т.д.). Следовательно, применяемые эндоглюканазы должны работать и/или быть стабильными при повышенных температурах, экстремальных значениях pH (например, щелочных, кислых) и химических условиях (например, прачечной, (в виде) детергентов, протеаз, растворителей и т.д.). Повреждение волокон должно быть сведено к минимуму для таких применений. Эндоглюканазы также могут помогать в отделении нецеллюлозных фракций от волоконного материала в способах варки (производство пульпы и бумаги) или улучшать реологические свойства технологических потоков. Устойчивость к детергентам и устойчивость к протеазам может быть рассмотрена в виде продукта с повышенной стабильностью структуры фермента, свойство, которое также является связанным с повышенной термостабильностью. Эндоглюканазы также находят применения в производстве пищевых продуктов и кормов (в пивоварнях, виноделии, извлечении масла из жмыха, выпечке, получении теста). Часто стерилизация или пастеризация требует более высоких температур. Для сокращения времени обработки может быть выгодной технологическая стабильность эндоглюканаз.

Белки эндоглюканазы I (Cel7B), полученные из грибов рода *Trichoderma* (анаморфа *Hyphocrea*), демонстрируют высокую степень идентичности и считаются мезофильными. Сообщается, что наиболее стабильные эндоглюканазы члены семейства GH7 являются нативными ферментами из *Humicola insolens* (Cel7B) и *Fusarium oxysporum* (egl) (US5912157). Согласно указанному сообщению, EGI не обнаруживает активность (при температуре) выше 60°C. Таким образом, существует необходимость в области обеспечения более термостабильных эндоглюканаз из семейства GH7.

Сообщалось, что некоторые эндоглюканазы могут быть термоинактивированы при более высоких температурах (Dominguez JM, Acebal C, Jimenez J, de la Mata I, Macarron R, Castillon MP. Mechanisms of thermoinactivation of endoglucanase I from *Trichoderma reesei*

QM 9414. Biochem J. 1992 Oct 15;287 ( Pt 2):583-8). Авторы указанного исследования также пытались повторно активировать термоинактивированную эндоглюканазу, но это требовало жестких условий, с использованием 8 М мочевины и дополнительных агентов. Эффекты, описанные в виде продуктивной повторной укладки (рефолдинга), были показаны на других белках, кроме эндоглюканаз, [Zhang N, Suen WC, Windsor W, Xiao L, Madison V, Zaks A. Improving tolerance of *Candida antarctica* lipase B towards irreversible thermal inactivation through directed evolution. Protein Eng. 2003 Aug; 6(8):599-605], но по знаниям авторов изобретения, не для эндоглюканаз, в частности, эндоглюканаз GH7. В данной области считается, что термоинактивированные эндоглюканы являются малопригодными в промышленном разложении целлюлозы. С другой стороны, повышенная термостабильность часто является необходимой для эндоглюканаз, в частности, для ферментов грибкового происхождения. До настоящего времени сообщалось только о некоторых улучшениях для эндоглюканаз GH12 и GH45. Сообщалось о термостабильных эндоглюканазах со структурными укладками цепей из GH5 и GH48. Указанные эндоглюканы в отношении их кинетических свойств и предпочтении к субстрату по существу отличаются от эндоглюканаз GH7-класса.

Таким образом, существует необходимость в процессивных эндоглюканазах, конкретно семейства GH7, с более высокими температурными профилями. Более того, было бы желательным достижение хорошей продуктивности от их экспрессирующего хозяина. Необходимость дополнительно подкрепляется фактом, что многие процессы промышленного значения протекают в жестких условиях и при повышенных температурах. Проблема, подлежащая решению настоящим изобретением, состоит в обеспечении улучшенных эндоглюканаз, конкретно эндоглюканаз с улучшенными термическими свойствами. Дополнительные проблемы, рассматриваемые и решаемые настоящим изобретением, станут очевидными из приведенных ниже разделов

### **Изложение сущности изобретения**

Настоящее изобретение имеет отношение к термостабильным белкам эндоглюканазам (полипептидам). Представленными решениями являются:

1. Белок, имеющий эндоглюканазную активность, который принадлежит к GH7-классу, и который демонстрирует активную термостабилизацию.
2. Белок, имеющий эндоглюканазную активность, который содержит аминокислотную последовательность, имеющую по меньшей мере 96%, предпочтительно по меньшей мере 97%, более предпочтительно по меньшей мере 98%, даже более

предпочтительно по меньшей мере 99%, такую как по меньшей мере 99,5% идентичность с SEQ. ID NO.: 2.

Предпочтительно, белки эндоглюканаз по изобретению демонстрируют более чем 95% остаточной активности при 60°C.

Дополнительным аспектом изобретения являются нуклеиновые кислоты, кодирующие указанные полипептиды, и экспрессирующие конструкции, содержащие эти полинуклеотиды в остове вектора, содержащегося в организме. Другим аспектом изобретения является применение белков по изобретению для переработки материалов лигноцеллюлозы и целлюлозы, в особенности, осахаривании лигноцеллюлозного сырья в консолидированных, частично консолидированных или неконсолидированных способах, или в производстве пищевых продуктов, кормов, целлюлозных волокон или в применениях для очистки.

Изобретение также имеет отношение к продуцированию/экспрессии организмов для получения белков по изобретению, и к способам культивирования таких организмов с целью получения белка. Организмы выбраны из организмов, включающих микроорганизмы (грибковые, бактериальные или архейные) или растения.

#### **Краткое описание чертежей**

Фиг. 1 — SDS-гель демонстрирует экспрессию последовательности Seq. ID NO 8 вариантного белка Seq. ID NO 2, секретируемого в супернатант. Полоса экспрессируемого белка видна между 75 и 100 кДа.

Фиг. 2 — экспрессирующая плаزمида *Trichoderma reesei*. Последовательность ДНК, кодирующая зрелый ген эндоглюканазы, клонируют в слиянии с сигнальной пептидной последовательностью *TrCBN1* под контролем промотора *TrCBN1*. Подлежащая вырезанию (эксцизии) экспрессирующая кассета *SwaI/SbfI*, содержит кассету устойчивости к гигромицину для селекции трансформантов.

Фиг. 3 — варианты эндоглюканаз, демонстрирующие повышенную температурную стабильность ([6] и [3]), и варианты с повышенной температурной стабильностью и активной термостабилизацией [1], [4], [5], [7], [8], [9] и [10] в сравнении с нативным белком GH7 [2].

Фиг. 4 — калибровка частей по 200 мкл раствора щелочного 4-метилумбеллиферона для считывания флуоресценции в планшетном ридере Tecan Infinite 200. Раствор 10 мМ готовили растворением 440 мг 4-метилумбеллиферона (Sigma Aldrich, № в каталоге 69580) в 250 мл 0,5 М раствора углекислого натрия. Серийные разведения

готовили в 0,5 М углекислом натрии. Интенсивность флуоресценции измеряли при 360нм/454нм с усилением 50.

Фиг. 5 — термостабилизация Seq. ID NO: 13 по сравнению с Seq. ID NO: 4

Фиг. 6 — определение периода полураспада при 70°C (Пример7) для Seq. ID NO: 14 по сравнению с Seq. ID NO: 4.

### **Определения**

"**Термостабильность**" является термином, используемым для описания внутреннего свойства конкретного белка с эндолюкканазной активностью по изобретению.

"**Активная термостабилизация**" является термином, используемым для описания внутреннего свойства конкретного белка с эндолюкканазной активностью по изобретению.

**Определение термостабильности и/или активной термостабилизации:** Термостабильность и активную термостабилизацию определяют следующим образом.

1. Фермент экспрессируется в *Pichia pastoris* как описано в примере 2. Если требуется, фермент очищают.

2. Корректировка концентрации фермента

Раствор фермента подходящей концентрации делают путем разведения очищенного фермента или культурального супернатанта *Pichia pastoris* в натрий-ацетатном буфере (50 мМ, рН 5) для применимой рабочей концентрации. Для определения применимой рабочей концентрации серийное разведение фермента, полученного на стадии 1 выше, готовят в натрий-ацетатном буфере, и аликвоты 10 мкл тестируют в температурном градиенте, как описано в примере 4. Применимую рабочую концентрацию определяют в виде концентрации, которая в результате дает сигнал флуоресценции между 5 000 и 15 000 в планшетном ридере Tecan Infinite M200 с усилением 50, или эквивалентной концентрации от 5,4 мкМ до 19 мкМ 4-метилумбеллиферона после инкубации, как описано в примере 4.

3. Определение способности к превращению субстрата, как описано в примере 4, за исключением того, что аликвоту 10 мкл культурального супернатанта заменяют на аликвоту 10 мкл раствора фермента в применимой рабочей концентрации, как определено на стадии 2.

4. Нормализация измерения путем деления всех считываний относительных единиц флуоресценции (rfu) на максимальное считывание (значение) rfu в пределах градиента

температуры для получения относительного превращения субстрата для каждого тестируемого белка при каждой протестированной температуре.

5. Графическое изображение относительного превращения субстрата в зависимости от протестированных температур реакции.

б. Определение температурной стабильности, как описано в (а), или определение активной термостабилизации, как описано в (b), являются следующими.

**а. Определение температурной стабильности:** белок отличается температурной стабильностью, если относительное превращение субстрата при 60°C составляет 0,5 или более, предпочтительно 0,7 или более, и более предпочтительно 0,9 или более, например, 0,95 или более.

**Определение активной термостабилизации:** анализ графика, полученного на стадии 5) на присутствие плато при относительном превращении субстрата, которое находится ниже максимального уровня (который составляет 1), но которое находится по меньшей мере вплоть до 0,15.

Плато определяют в виде уровня относительного превращения субстрата, который является по существу неизменным в диапазоне температур по меньшей мере 5°C, предпочтительно от 70 до 75°C (т.е. в пределах +/- 0,1 вокруг среднего значения в пределах указанного диапазона температур).

**б.** Варианты, демонстрирующие отсутствие активной термостабилизации, имеют относительное превращение субстрата от 0 и ниже 0,15, обычно около 0,1. Не желая быть связанными любой конкретной теорией, можно думать, что измеренное относительное превращение субстрата обычно около 0,1 (а не 0,0, как следовало ожидать для неактивного фермента при данной температуре) является следствием неконтролируемых изменений предельных в амплификаторе и/или в ходе манипулирования пробами смесей.

«**Термические свойства**» является термином, обычно используемым для обозначения свойств фермента при более высоких температурах (например, 60°C или более). Термин может включать одну или обе "температурную стабильность", как определено выше, и "активную термостабилизацию", как описано выше.

«**Эндоглюканазная активность**» в контексте настоящего изобретения, определена как каталитическое ускорение белком разрыва  $\beta$ -1,4-гликозидных связей посредством нуклеофильной атаки полярной молекулой, такой как вода или органические молекулы с их гидроксил-, или меркапто-, или аминofункциями. Определение также включает расщепление синтетических молекул, имеющих неуглеводную молекулу, связанную с глюкозой, целлобиозой или лактозой, посредством  $\beta$ -1,4-гликозидной связи. Примеры реакций, катализируемых эндоглюканазами, перечислены в базе данных Brenda



(<http://www.brenda-enzymes.info> (Release 2012.1 (January 2012))); Enzyme data and metabolic information: BRENDA. a resource for research in biology, biochemistry, and medicine Schomburg, I., Hofmann, O., Baensch, C. Chang, A., Schomburg, D. *Gene Funct*, Dis. 3-4, 109-18 (2000)).

**«Остаточная активность»** определена как ферментативная активность, которая восстанавливается после инкубации фермента в течение определенного времени при определенной (повышенной) температуре по сравнению с активностью без стадии инкубации. Протокол для определения остаточной активности приводят в примере 4.

**Выравнивание последовательностей с SEQ ID NO: 2:** Парное выравнивание любой второй последовательности эндоглюканазы GH7 с родительской последовательностью (SEQ ID NO 2) делают с использованием алгоритма ClustalW Algorithm (Larkin M.A., Blackshields G., Brown N.P., Chenna R., McGettigan P.A., McWilliam H., Valentin F., Wallace I.M., Wilm A., Lopez R., Thompson J.D., Gibson T.J. and Higgins D.G. (2007) ClustalW and ClustalX version 2. *Bioinformatics* 2007 23(21): 2947-2948). Парное выравнивание покажет номера положений для SEQ ID NO: 2. Указанные номера могут быть использованы для ссылки, например, когда говорят, что, например, остаток, соответствующий положению номер 2 (последовательности) SEQ ID NO: 2, является мутированным во второй эндоглюканазе GH7. В международных правилах нумерации аминокислот и обозначения вариантов белка для описания вариантов белка, аминокислоту в родительской белковой последовательности SEQ ID NO: 2 называют положением номер 1, или S1, или серин 1. Нумерация всех аминокислот будет согласно их положению в родительской последовательности, данной в SEQ ID NO: 2 относительно этого положения номер 1.

**Идентичность последовательностей:** для определения идентичности последовательностей используют программное обеспечение AlignX VectorNTI Package, которое продается Life Technology Corporation, с использованием стандартных установочных параметров (штраф (penalty) открывания гэпа 10, штраф (penalty) удлинения гэпа 0,1).

**Варианты белков** являются полипептидами, аминокислотная последовательность которых отличается в одном или нескольких положениях от данного родительского белка, в результате чего различия могут быть замещениями одного аминокислотного остатка (остатков) другими, делециями единственного аминокислотного остатка или нескольких аминокислотных остатков, или инсерцией (вставкой) дополнительного аминокислотного остатка (остатков), или фрагментами аминокислотного остатка (остатков) в родительскую последовательность. Белки могут быть модифицированы в определенных положениях

введением точечных мутаций в кодирующие нуклеиновые кислоты. Термин модифицированная белковая последовательность в настоящем документе всегда относится к белкам, полученным в результате транскрипции и трансляции, а также возможных процессов посттрансляционной модификации и транслокации из модифицированных соответствующим образом нуклеиновых кислот, или *in vitro* или путем подходящего экспрессирующего хозяина. Способы создания таких вариантов белков хорошо известны в данной области техники и таким образом, не ограничиваясь, примеры включают случайный или сайт-направленный мутагенез, сайт-насыщающий мутагенез, сборку синтезированных ПЦР-фрагментов, перетасовку ДНК, гомологичную рекомбинацию *in vitro* или *in vivo*, и методы синтеза генов на основе химического синтеза ДНК.

**Номенклатуру аминокислот**, пептидов, нуклеотидов и нуклеиновых кислот выполняют согласно IUPAC (Международный союз теоретической и прикладной химии, ИЮПАК). В общем, аминокислоты названы в настоящем изобретении согласно однобуквенному коду.

**Замены** единичных аминокислот описывают путем именованного однобуквенного кода исходной аминокислоты, за которым следует номер ее положения и однобуквенный код замещающей аминокислоты, т.е. замену глутамина в положении 1 на лейцин в этом положении описывают в виде "Q1L". Для делеций одиночных положений последовательности, символ замещающей аминокислоты заменяют тремя буквами аббревиатурой "del", таким образом, делеция аланина в положении 3 может быть названа "A3del". Вставленные (inserted) дополнительные аминокислоты получают номер предшествующего положения, продолженный строчной буквой в алфавитном порядке относительно их (аминокислот) расстояния до их точки вставки. Таким образом, вставка двух триптофанов после положения 3 называется "3aW, 3bW". Введение нетранслируемых кодонов TAA, TGA и TAG в последовательность нуклеиновой кислоты обозначается как "\*" в аминокислотной последовательности, таким образом, введение терминирующего кодона в положение 4 аминокислотной последовательности называется "G4\*". Множественные мутации разделяют знаком плюс, или косой чертой, или запятой. Например, две мутации в положениях 20 и 21, замещающие аланин и глутаминовую кислоту на глицин и серин, соответственно, обозначаются в виде "A20G+E21S" или "A20G/E21S" "A20G,E21S". Когда аминокислотный остаток в данном положении замещают двумя или несколькими альтернативными аминокислотными остатками, то эти остатки разделяют запятой или косой чертой. Например, замещение аланина в положении 30 или глицином, или глутаминовой кислотой обозначают "A20G,E", или "A20G/E", или

"A20G,A20E". Когда положение, подходящее для модификации, идентифицируют здесь без какой-либо предлагаемой специфической модификации, должно быть понятно, что любой аминокислотный остаток может быть замещен аминокислотным остатком, присутствующим в данном положении. Таким образом, например, при модификации аланина в положении 20, который упоминается, но не указан, должно быть понятно, что аланин может быть делегирован (удален) или замещен любым другим аминокислотным остатком (т.е., любым из R, N, D, C, Q, E, G, H, I, L, K, M, F, P, S, T, W, Y и V).

Термины "сходная мутация" или "сходная замена" относятся к аминокислотной мутации, где аминокислотный остаток в первой мутации (по отношению к родительской последовательности, такой как, например, SEQ ID NO: 2) снова заменен второй мутацией, и в результате чего аминокислотный остаток, привнесенный второй мутацией, имеет сходные свойства с аминокислотным остатком, который был привнесен первой мутацией. Сходный, в контексте данного изобретения, означает аминокислоту, которая имеет сходные химические свойства. Например, если первая мутация в конкретном положении приводит к замене неалифатического аминокислотного остатка (например, Ser) алифатическим аминокислотным остатком (например, Leu), тогда замена в одном и том же положении другой алифатической аминокислотой посредством второй мутации (например, Ile или Val) называется сходной мутацией. Дополнительные химические свойства включают размер остатка, гидрофобность, полярность, заряд, значение pK и т.п.. Таким образом, сходная мутация может включать замены, такие как основные на основные, кислотные на кислотные, полярные на полярные и т.д.. Произведенные таким образом наборы аминокислот, по-видимому, будут стабильными по структурным причинам. Эти наборы могут быть описаны в форме диаграммы Венна (Livingstone CD. and Barton GJ. (1993) "Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation" *Comput. Appl Biosci.* 9: 745-756; Taylor W. R. (1986) "The classification of amino acid conservation" *J.Theor.Biol.* 119; 205-218). Сходные замены могут быть произведены, например, согласно последующим группам аминокислот: гидрофобные: F W Y H K M I L V A G; ароматические: F W Y H; алифатические: I L V; полярные: W Y H K R E D C S T; заряженные H K R E D; положительно заряженные: H K R; отрицательно заряженные: E D.

**Экспрессирующая конструкция** здесь определена как последовательность ДНК, содержащая все необходимые элементы последовательности для установления экспрессии в клетке-хозяине, содержащая открытую рамку считывания (ORF), включающая последовательности для инициации транскрипции (промоторы), терминации и регуляции, сайты для инициации трансляции, области стабильной репликации или интеграции в

геном хозяина и селективные генетические маркеры. Если требуется, открытая рамка считывания состоит из слияния нуклеиновой кислоты, кодирующей белок-мишень, с дополнительными элементами, особенно секреторными сигналами, целлюлозосвязывающим доменом, TAG-метками, для усиления уровня экспрессии или облегчения очистки или выделения из ферментативного бульона. Таким образом, функциональная структура может быть создана или достигнута уже путем упорядочивания (объединения и т.д.) события в клетке-хозяине. В предпочтительном воплощении экспрессирующая конструкция содержит промотор, функционально связанный с открытой рамкой считывания, за которым следует, если требуется, терминирующая последовательность (стоп-кодон). Предпочтительными промоторами являются промоторы от средних до сильных, функциональные в выбранных клетках-хозяевах в условиях ферментации. Для иллюстрации, примеры предпочтительных промоторов являются следующими:

- бактерии (например, *Escherichia coli*): lac, tac, trp, tet, T3 T7, CP7, CP21, araBAD;
- дрожжи (например, *Pichia. Saccharomyces*): AOXI, AOXII, FMDH, GAP, TEF, PFK1, FBA1, PGK1, ADH1, ADH2, TDH3;
- грибы (например, *Trichoderma*): CBHI, CBHII, EGI, PGK, BGL, XYL1, XYL2.

Дополнительные примеры подходящих промоторов для гетерологичной экспрессии описаны в литературе. К другим частям экспрессирующей конструкции предъявляются требования к генетическим элементам для стабильно наследуемых признаков введенных нуклеиновых кислот и селективных маркеров, включающих генетические элементы, относящиеся к (маркерам) устойчивости к антибиотикам, или комплектации определенной ауксотрофии штамма-хозяина.

Последовательность всех нуклеиновых кислот настоящего изобретения, или нуклеиновых кислот, кодирующих полипептиды/белки данного изобретения, может быть скорректирована для использования оптимальных кодонов в выбранном экспрессирующем хозяине. Нуклеиновые кислоты, имеющие такое оптимизированное/оптимальное использование кодона для конкретного экспрессирующего хозяина, также являются частью настоящего изобретения. «**Хозяин для получения**» используется в данном документе как синоним к «экспрессирующий хозяин» и означает организм, который, при культивировании, продуцирует белок настоящего изобретения. В одном воплощении, белок настоящего изобретения не секретируется хозяином для получения; однако, в предпочтительном варианте, белок секретируется в окружающую среду. Такие организмы предпочтительно выбраны из

царства бактерий, архей, дрожжей, грибов и/или растений. Одним предпочтительным экспрессирующим хозяином является *Pichia pastoris*.

"**Бактерии**" в данном документе будет относиться к прокариотическим организмам. В предпочтительном воплощении бактериями являются эубактерии, и даже более предпочтительно они выбраны среди рода *Escherichia*, *Bacillus*, *Klebsiella*, *Streptomyces*, *Lactococcus* и *Lactobacillus*, в частности *Escherichia coli*, *Bacillus subtilis*, *Bacillus licheniformis*, *Bacillus amyloliquefaciens*, *Bacillus megaterium*, *Klebsiella planticola*, *Streptomyces lividans*, *Lactococcus lactis*, *Lactobacillus brevis*.

"**Дрожжи**" в данном документе будет относиться ко всем низшим эукариотическим организмам, демонстрирующим одноклеточное вегетативное состояние в их жизненном цикле. Дрожжи особенно включают организмы класса *Saccharomycetes*, в частности рода *Saccharomyces*, *Pachysolen*, *Pichia*, *Candida*, *Yarrowina*, *Debaromyces*, *Klyveromyces*, *Zygosaccharomyces*.

"**Мицелиальные грибы**" или "грибы" в данном документе будет относиться ко всем низшим эукариотическим организмам, демонстрирующим рост гифов на протяжении по меньшей мере одной стадии их жизненного цикла. Грибы особенно включают организмы типа *Ascomycota* и *Basidiomycota*, в частности, рода *Trichoderma*, *Taiaromyces*, *Aspergillus*, *Penicillium*, *Chrysosporium*, *Phanerochaete*, *Thermoascus*, *Agaricus*, *Pleutrus*, *Irpex*.

"**Растение**" в данном документе будет относиться ко всем эукариотическим организмам, принадлежащим к царству растений. В предпочтительном воплощении, экспрессирующий хозяин выбран из растений рода *Zea*, *Triticum*, *Hordeum*, *Secale*, *Miscanthus*, *Saccharum*, *Solanum*, *Ipomea*, *Manihot*, *Helianthus*, *Camellia*, *Aspalathus*, *Eucalyptus*, *Beta*, *Fagus*, членов семейства *Pinaceae*, *Betulaceae*, *Malvaceae*, *Cupressaceae*, *Rosaceae*, *Arecaceae*.

**Композиция фермента** предназначена быть в виде любой жидкости или композиции твердого вещества, содержащей фермент в качестве фракции. Дополнительные компоненты предпочтительно содержат воду, полиолы, сахара, детергенты, буферные агенты, восстановители, неорганические соли, твердые носители, консерванты, особенно с антибактериальной или противогрибковой активностью, красители, ароматизирующие вещества и/или отдушки.

**Применение эндоглюканаз**, в частности, таких как эндоглюканазы настоящего изобретения (неограничивающие примеры): гидролиз лигноцеллюлозного сырья для производства мономерных, димерных или олигомерных сахаров; производство пульпы и бумаги; применения в текстильной промышленности для улучшения или общего

производства волокон, нитей или денима; применения для очистки для применений в промышленности или в бытовой химии; выпуск питательных веществ, повышение выхода продукции или улучшение свойств теста в области пищевых продуктов и кормов.

### Подробное описание изобретения

Данное изобретение имеет отношение к эндоглюканазам GH7 с улучшенными свойствами. Более конкретно, изобретение имеет отношение к термостабильным белкам эндоглюканазам (полипептидам). Представленными решениями являются:

1. Белок, имеющий эндоглюканазную активность, который принадлежит к GH7-классу, и который демонстрирует активную термостабилизацию.

2. Белок, имеющий эндоглюканазную активность, который содержит аминокислотную последовательность, имеющую по меньшей мере 96%, предпочтительно по меньшей мере 97%, более предпочтительно по меньшей мере 98%, даже более предпочтительно по меньшей мере 99%, такую как по меньшей мере 99,5% идентичность с SEQ. ID NO.: 2.

Эти два воплощения подробно описаны ниже.

Температурная стабильность определена выше. Пример определения температурной стабильности дан в примере 4. Эндоглюканазы GH7-класса приведены в таблице 1 (ЕС 3.2.1.4). За исключением конкретных ограничений идентичностей последовательностей в конкретной формуле изобретения, изобретение имеет отношение к вариантам всех эндоглюканаз GH7-класса, содержащиеся в данном документе варианты эндоглюканаз показаны в таблице 1.

Таблица 1. Известные эндоглюканазы GH7-класса

	Название белка	Организм	Банк генов	Банк данных 3-D структур белков
1	целлюлаза III-A (пептидный фрагмент)	<i>Acremonium cellulolyticus</i>		
2	эндо- $\beta$ -1,4-глюканаза (EglB; AN3418.2)	<i>Aspergillus nidulans</i> FGSC A4	EAA63386.1	
3	эндо- $\beta$ -1,4-глюканаза (CelB)	<i>Aspergillus oryzae</i> KBN616	BAA22589.1	
4	эндо- $\beta$ -1,4-глюканаза (CelB; AO090010000314)	<i>Aspergillus oryzae</i> RIB40	AEB00821.1	
5	эндо- $\beta$ -1,4-глюканаза I	<i>Aspergillus terreus</i> MS-31	ADR78837.1	
6	эндо- $\beta$ -1,3-1,4-глюканаза (Bgl7A)	<i>Bispora</i> sp. MEY-1 / CGMCC 2500	ACT53749.1	
7	эндоглюканаза I (EG I) (пептидный фрагмент) (Cel7C)	<i>Chrysosporium lucknowense</i>		
8	эндо- $\beta$ -1,4-глюканаза (CLhgEG1)	<i>Coptotermes lacteus</i> symbiont WH2002	BAC07551.1	
9	эндо- $\beta$ -1,4-глюканаза (CLhgEG2)	<i>Coptotermes lacteus</i> symbiont WH2002	BAC07552.1	

	Название белка	Организм	Банк генов	Банк данных 3-D структур белков
10	эндо- $\beta$ -1,4-глюканаза (EglB)	<i>Emericella nidulans</i>	AAM54071.1	
11	эндо- $\beta$ -1,4-глюканаза I (EG I;Egl;Fof7) (Cel7B)	<i>Fusarium oxysporum</i>	AAA65586.1	1OYW[A,B,C,D]
12	эндоглюканаза 3 (HmEG3) (фрагмент)	<i>Holomastigotoides mirabile</i>	BAB64565.1	
13	эндоглюканаза 2 (HmEG2)	<i>Holomastigotoides mirabile</i>	BAB64564.1	
14	эндоглюканаза 1 (HmEG1)	<i>Holomastigotoides mirabile</i>	BAB64563.1	
15	эндо- $\beta$ -1,4-глюканаза I (Egl1;EG-I)	<i>Humicola grisea</i> var. <i>thermoidea</i>	BAA09786.1	
16	эндоглюканаза I (EG I;EG1) (Cel7B)	<i>Humicola insolens</i>	AAE25068.1	1A39[A]
17	эндо- $\beta$ -1,4-глюканаза I (EGI;Egl1;EG-I) (Cel7B)	<i>Hypocrea jecorina</i>	AAA34212.1	1EG1[A,C]
18	эндоглюканаза I (Egl)	<i>Hypocrea jecorina</i> M5	ADM08177.1	
19	эндо- $\beta$ -1,4-глюканаза (Egl1)	<i>Hypocrea jecorina</i> PTCC 5142	AAX28897.1	
20	эндоглюканаза I	<i>Hypocrea pseudokoningii</i>	ABM90986.1	
21	эндоглюканаза I (Egl)	<i>Hypocrea pseudokoningii</i> 3.3002	AEQ29501.1	
22	эндоглюканаза I (Egl)	<i>Hypocrea rufa</i>	AEO17039.1	
23	эндо- $\beta$ -1,4-глюканаза I (EGI;Bgl1)	<i>Hypocrea rufa</i> AS 3.3711	AAQ21382.1	
24	эндоглюканаза I	<i>Hypocrea rufa</i> HK-75		
25	эндо- $\beta$ -1,4-глюканаза (Egl1;MG02532.4)	<i>Magnaporthe grisea</i> 70-15	XP_366456.1	
26	эндоглюканаза	<i>Myceliophthora thermophila</i> CBS 117.65	AAE25067.1	
27	эндоглюканаза I (Egl1) (Cel7B)	<i>Penicillium decumbens</i> 114-2	ABY56790.1	
28	эндоглюканаза I (Egl1)	<i>Penicillium decumbens</i> L-06	ACJ15337.1	
29	эндоглюканаза I (Egl1;Egl)	<i>Penicillium oxalicum</i>	ACS32299.1	
30	эндоглюканаза (Cel7B)	<i>Penicillium purpurogenum</i>	AEL78899.1	
31	эндоглюканаза (Bgl7C7)	<i>Penicillium</i> sp. C7	AEG74551.1	
32	эндо- $\beta$ -1,4-глюканаза (EGI) (пептидные фрагменты) (Cel7B)	<i>Penicillium verruculosum</i>		
33	эндоглюканаза 3 (PgEG3)	<i>Pseudotriconympha grassii</i>	BAB64562.1	
34	эндоглюканаза 2 (PgEG2)	<i>Pseudotriconympha grassii</i>	BAB64561.1	
35	эндоглюканаза 1 (PgEG1h)	<i>Pseudotriconympha grassii</i>	BAB64553.1	
36	Egl1 (фрагмент)	<i>Trichoderma asperellum</i> T203	AAS37698.1	
37	эндоглюканаза I	<i>Trichoderma longibrachiatum</i> 3.1029	AEI71804.1	
38	эндоглюканаза I (Egl1)	<i>Trichoderma longibrachiatum</i> 36MS	AEC03714.1	
39	эндо- $\beta$ -1,4-глюканаза (Egl1;Egl1;T1Cel7A) (Cel7A)	I <i>Trichoderma longibrachiatum</i> CECT 2606	1920181A	
40	эндоглюканаза I	<i>Trichoderma longibrachiatum</i> FU05	ACZ34302.1	
41	эндоглюканаза I (Egl;EGI)	<i>Trichoderma</i> sp. SSL	ACH68455.1	

	Название белка	Организм	Банк генов	Банк данных 3-D структур белков
42	эндо- $\beta$ -1,4-глюканаза (RsSymEG1;SM2038B11)	некультивированные симбиотические простейшие (термитов) <i>Reticulitermes speratus</i>	BAF57296.1	

Первый и второй аспекты ниже будут описаны подробно.

**Первый аспект. Белок, имеющий эндоглюканазную активность, который принадлежит к GH7-классу и который демонстрирует активную термостабилизацию.**

В первом аспекте данного изобретения белки имеют эндоглюканазную активность и улучшенные термические свойства. Улучшенные термические свойства определены в виде температурной стабильности, которая проявляется в относительной активности превращения субстрата выше, чем 90% (например, выше, чем 95 %) при инкубации при температурах 60°C или выше, и (в виде) активной термостабилизации. Активная термостабилизация описана ниже.

Авторами настоящего изобретения неожиданно было обнаружено, что белки, демонстрирующие активную термостабилизацию, также демонстрируют температурную стабильность.

Это было показано последующим примером. Авторы получили эндоглюканазу GH7, которая является конкретным вариантом SEQ ID NO: 4 (т.е., вариантом, заданным SEQ ID NO: 2), следующим образом. Нуклеиновую кислоту, кодирующую полипептид с SEQ ID NO: 2, получали случайным мутагенезом (допускающей ошибки ПЦР, как описано в примере 1). Способы случайного мутагенеза хорошо известны в данной области. Кроме того, теперь, когда авторы в данном изобретении раскрыли пригодность полипептида, кодируемого SEQ ID NO: 2, соответствующая нуклеиновая кислота, кодирующая этот белок, может быть получена непосредственно специалистом в данной области техники. Способы ее получения включают, например, синтез гена или сайт-направленный мутагенез, начиная с нуклеиновой кислоты с высокой степенью идентичности последовательностей (например, более чем 90 %) с SEQ ID NO: 4 и введение мутаций сайт-направленным мутагенезом (в одну или несколько стадий) для получения нуклеиновой кислоты, кодирующей белок SEQ ID NO: 2. Стартовой последовательностью, из которой нуклеиновая кислота, кодирующая SEQ ID NO: 2, может быть получена путем мутагенеза, является Cel7B из *Hypocrea pseudokonigii*, представленная в данном документе в виде SEQ ID NO: 4 (номер доступа в банке генов ABM90986).



Авторы настоящего изобретения исследовали температурную стабильность белка, имеющего SEQ ID NO: 2. Как видно на фиг. 3, этот белок решает техническую проблему, лежащую в основе настоящего изобретения, т.е. (белок) имеет более высокую температурную стабильность, чем его родительский белок (SEQ ID NO: 4). Это является очевидным, например, из того факта, что относительное превращение субстрата все еще находится вблизи его максимума, например, при 60°C, в то время как относительное превращение субстрата белка, имеющего SEQ ID NO: 4, находится на очень низком уровне при указанной температуре (см., фиг. 3).

К удивлению авторов настоящего изобретения оказалось, что даже при высоких температурах, например, в диапазоне от 68 до 76°C (включая от 70 до 74°C), относительное превращение субстрата существенно не падает с повышением температуры. Это находится в резком контрасте со свойствами родительского белка, имеющего SEQ ID NO: 4, который, на графике в зависимости от повышающихся температур, демонстрирует уменьшение относительного превращения субстрата, уменьшение падает до фоновых уровней без какого-либо промежуточного плато. Считается, что белок, имеющий SEQ ID NO: 4, под действием высоких температур, например, 60°C или более, таких как 70°C или более, не находится в его активном состоянии. Не желая быть связанными любой конкретной теорией, считается, что этот эффект обусловлен тепловым разворачиванием (или сворачиванием (фолдингом) в неактивные конформации) белка. Не желая быть связанными любой конкретной теорией, эффект потери активности при высоких температурах в последующем будет называться тепловым разворачиванием. Тепловое разворачивание является хорошо известным феноменом для белков практически любого типа, конкретно ферментов, при высоких температурах. Таким образом, тепловое разворачивание, наблюдаемое для белка, имеющего SEQ ID NO: 4, соответствует ожиданиям специалистов в данной области техники. Белок SEQ ID NO: 4 не является частью данного изобретения.

В резком контрасте, белок этого аспекта изобретения демонстрирует фазу плато при более высоких температурах, например, в диапазоне от 68 до 76°C (включая от 70 до 74°C). Это плато находится ниже максимума относительного превращения субстрата, но выше, чем фон относительного превращения субстрата. Не желая быть связанными любой конкретной теорией, авторы настоящего изобретения делают заключение, что белок данного изобретения находится при этих высоких температурах в состоянии, которое отличается от свернутого состояния при низких температурах (например, 46°C), но этот белок еще является ферментативно активным. Таким образом, можно предположить, что при высоких температурах белок данного изобретения интенсивно свернут повторно, т.е.

повторно свернут для получения дополнительного активного состояния (и таким образом, предоставляя возможность наблюдать относительное превращение субстрата при более высоких температурах). Следовательно, авторы изобретения назвали это свойство, которое также определено выше в разделе определений, как “активная термостабилизация”.

Таким образом, белок изобретения решает технические проблемы, лежащие в основе настоящего изобретения, будучи термостабильным. Кроме того, на основе раскрытия настоящего изобретения, квалифицированному специалисту в данной области дается методическое руководство для идентификации других белков, согласно этому первому аспекту изобретения.

Такие дополнительные белки могут быть найдены следующим образом. Сначала любой тип мутаций (включающих, делецию, вставку или замену одного или нескольких аминокислотных остатков, и причем случайно или направленно) может быть введен в любую эндоглюканазу семейства GN7, конкретно в любую из названных в таблице 1, для получения мутантного белка или его библиотеки. Полученный таким образом мутантный белок или его библиотека могут быть подвергнуты скринингу на активную термостабилизацию как определено выше. Известные белки, представленные в таблице 1 выше, не являются частью изобретения, но любые их мутанты, демонстрирующие активную термостабилизацию, включены в данное изобретение.

Важно, что все ферменты согласно первому аспекту настоящего изобретения, т.е. ферменты, которые демонстрируют активную термостабилизацию, как определено выше, демонстрируют температурную стабильность, как определено выше. Таким образом, активная термостабилизация является, по первому аспекту, решением проблемы, лежащей в основе настоящего изобретения. Так или иначе, любой данный белок соответствует первому аспекту данного изобретения может быть надежно протестирован путем анализа на активную термостабилизацию, предоставленную выше.

**Второй аспект. Белок, имеющий эндоглюканазную активность, который содержит аминокислотную последовательность, имеющую по меньшей мере 96%, предпочтительно по меньшей мере 97%, более предпочтительно по меньшей мере 98%, даже более предпочтительно по меньшей мере 99%, такую как по меньшей мере 99,5% идентичность с SEQ. ID NO.: 2.**

В поиске второго решения проблемы, лежащей в основе настоящего изобретения, авторы изобретения приступили к проекту по мутагенезу, начиная с белка SEQ ID NO: 2. Таким образом, авторы изобретения ввели мутации, такие как точечные мутации, в белок, имеющий SEQ ID NO: 2 (т.е. путем модификации, лежащей в основе нуклеиновой

кислоты, как описано ниже). Авторы изобретения обнаружили, что многие такие мутанты также демонстрируют температурную стабильность и таким образом решают, лежащую в основе проблему, во втором аспекте. Примеры решений приведены на фиг. 3.

Таким образом, *во втором аспекте* изобретение имеет отношение к белку, имеющему эндоглюканазную активность, который содержит аминокислотную последовательность, имеющую по меньшей мере 96%, предпочтительно по меньшей мере 97%, более предпочтительно по меньшей мере 98%, даже более предпочтительно по меньшей мере 99%, такую как по меньшей мере 99,5% идентичность с SEQ. ID NO.: 2. Этот белок обычно может принадлежать к GH7-классу.

Конкретно, настоящее изобретение также обеспечивает конкретные мутанты белка с последовательностью SEQ ID NO: 2. Таким образом, последовательность, предоставленную в SEQ ID NO: 2, модифицируют в одном или нескольких (предпочтительно 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18 или 19) положениях. Такая модификация может состоять из замены, делеции, вставки и т.п.. В конкретном ее воплощении, модификация заключается в замещении. Даже в более конкретном воплощении, модификация заключается в замещении одного или нескольких конкретных положений SEQ ID NO: 2, которые выделены в крайней левой колонке таблиц 2, 3, 4.

Наряду с такой модификацией в любом из этих предоставленных положений, в принципе может быть замещение на любой аминокислотный остаток, предпочтительно, что замещение является заменой на аминокислотный остаток, предоставленный в линии номер 4 любой одной или нескольких таблиц 2, 3 или 4, или на аминокислотный остаток, сходный с ним (сходная мутация, как определено выше). Таким образом, сходные мутации, как определено выше, могут быть введены вместо перечисленных мутаций. Пример 5 демонстрирует некоторые из таких мутантов. Способы введения мутаций известны в данной области. Приводимое в качестве примера руководство может быть взято из примера 1. Другими словами, предпочтительное воплощение данного изобретения имеет отношение к предпочтительным положениям для мутагенеза эндоглюканазы GH7-класса. Перечень предпочтительных замен предоставляют в таблице 2, линия 2. В другом предпочтительном воплощении предпочтительные мутации выбраны из перечня в таблице 3, линия 2. В другом предпочтительном воплощении изобретения предпочтительные мутации выбраны из таблицы 4, линия 2. Также возможно объединить два или три из этих предпочтительных воплощений, например, одна или несколько предпочтительных замен, предоставленных в таблице 2, могут быть объединены с одной

или несколькими предпочтительными заменами, предоставленными в таблице 3 и/или таблице 4.

Таблица 2. Предпочтительные замены аминокислот в отношении Seq. ID NO. 2

Альтернативные замены аминокислот для Seq. ID NO. 2	Наиболее предпочтительные аминокислоты в эндоглюканазах GH7 путем выравнивания ClustalW с Seq. ID NO. 2			
	линия 1	линия 2	линия 3	линия 4
2	L	L,Q	L,Q	Q
8	T	T,C	T	
16	T	T,C	T	
19	K	K,E	K	
23	S	S,H	S,H	H
30	N	N,D	N	
32	Y	Y,S	Y,S	S
41	W	W,R	W	
42	I	I,M	I,M	M
48	N	N,Y	N	
55	G	G,C	G,C	C
64	E	E,H,K	E,H,K	H,K
65	A	A,D	A,D	D
67	G	G,C	G,C	C
68	S	S,C,G	S,C,G	C,G
75	G	G,C	G	
86	N	N,S	N,S	S
88	S	S,C,D,F,T	S,D	D
93	N	N,H,R,Y	N,H,R	H,R
104	T	T,S	T,S	S
107	S	S,T	S	
118	K	K,E	K	
137	A	A,D	A,D	D
144	A	A,G	A	
145	S	S,A	S	
150	Q	Q,E	Q,E	E
153	E	E,K	E	
164	G	G,S	G	
179	Q	Q,L	Q	
185	T	T,D,E	T,D,E	D,E
191	Q	Q,K	Q,K	K
193	F	F,S	F	
201	L	L,F	L,F	F
210	L	L,M,Y	L,M	M
212	P	P,L,S	P	
216	N	N,T	N	

Альтернативные замены аминокислот для Seq. ID NO. 2	Наиболее предпочтительные аминокислоты в эндоглюканазах GH7 путем выравнивания ClustalW с Seq. ID NO. 2			
	линия 1	линия 2	линия 3	линия 4
217	A	A,Y	A	
231	R	R,G,H,K	R,H,K	H,K
233	G	G,N	G	
235	P	P,S	P	
242	G	G,D	G	
249	P	P,R	P,R	R
261	G	G,C	G	
263	P	P,T	P,T	T
271	T	T,K	T,K	K
277	N	N,D,E	N,D,E	D,E
290	T	T,E	T,E	E
293	S	S,T	S	
299	T	T,A	T,E	E
312	E	E,D,S	E,D,S	D,S
317	I	I,V	I	
322	N	N,W	N	
323	D	D,N	D	
325	S	S,T	S	
327	Y	Y,F	Y	
328	M	M,K	M	
335	D	D,E,S	D,E,S	E,S
352	N	N,V,W	N	
357	H	H,E	H,E	E
360	Y	Y,F	Y	
379	P	P,L	P,L	L
382	P	P,del	P	
383	P	P,del	P	
390	S	S,L	S	
391	T	T,I	T,I	I
392	A	A,T	A,T	T
398	S	S,T	S,T	T
405	I	I,T	I,T	T
431	Y	Y,H	H	
432	S	S,G	S,G	G
434	D	D,Y	D,Y	Y
448	H	H,Y	H	

Таблица 3. Предпочтительные замены аминокислот в отношении Seq. ID NO. 2

Положение Seq. ID NO. 2	Положение аминокислоты в Seq. ID NO. 2	Предпочтительные аминокислоты в эндотоксинах GH7 путем выравнивания ClustalW с Seq. ID NO. 2	Наиболее предпочтительные аминокислоты в эндотоксинах GH7 путем выравнивания ClustalW с Seq. ID NO. 2	Альтернативные замены аминокислот для Seq. ID NO. 2
				линия 4
	линия 1	линия 2	линия 3	линия 4
23	S	S,H	S,H	H
30	N	N,D	N	
41	W	W,R	W	
55	G	G,C	G,C	C
64	E	E,H,K	E,H,K	H,K
65	A	A,D	A,D	D
67	G	G,C	G,C	C
118	K	K,E	K	
137	A	A,D	A,D	D
144	A	A,G	A	
150	Q	Q,E	Q,E	E
164	G	G,S	G	
179	Q	Q,L	Q	
185	T	T,D,E	T,D,E	D,E
191	Q	Q,K	Q,K	K
201	L	L,F	L,F	F
212	P	P,L,S	P	
216	N	N,T	N	
231	R	R,G,H,K	R,H,K	H,K
242	G	G,D	G	
249	P	P,R	P,R	R
261	G	G,C	G	
263	P	P,T	P,T	T
271	T	T,K	T,K	K
277	N	N,D,E	N,D,E	D,E
290	T	T,E	T,E	E
299	T	T,A	T,E	E
312	E	E,D,S	E,D,S	D,S
323	D	D,N	D	
325	S	S,T	S	
328	M	M,K	M	
335	D	D,E,S	D,E,S	E,S
357	H	H,E	H,E	E
379	P	P,L	P,L	L
390	S	S,L	S	
391	T	T,I	T,I	I
405	I	I,T	I,T	T
432	S	S,G	S,G	G
434	D	D,Y	D,Y	Y

Таблица 4. Предпочтительные замены аминокислот в отношении Seq. ID NO. 2

Положение Seq. ID NO: 2	Альтернативные замены аминокислот для Seq. ID NO: 2			
	Положение аминокислоты в Seq. ID NO: 2	Предпочтительные аминокислоты в эндоглюканазах GH7 путем выравнивания ClustalW с Seq. ID NO: 2	Наиболее предпочтительные аминокислоты в эндоглюканазах GH7 путем выравнивания ClustalW с Seq. ID NO: 2	Альтернативные замены аминокислот для Seq. ID NO: 2
	линия 1	линия 2	линия 3	линия 4
2	L	L,Q	L,Q	Q
8	T	T,C	T	
16	T	T,C	T,C	C
19	K	K,E	K	
32	Y	Y,S	Y,S	S
42	I	I,M	I,M	M
48	N	N,Y	N	
68	S	S,G	S,G	G
75	G	G,C	G	
86	N	N,S	N,S	S
88	S	S,C,D,F,T	S,C	C
93	N	N,H,R,Y	N	
104	T	T,S	T,S	S
107	S	S,T	S	
145	S	S,A	S	
153	E	E,K	E	
193	F	F,S	F	
210	L	L,M,Y	L	
217	A	A,Y	A	
233	G	G,N	G	
235	P	P,S	P	
293	S	S,T	S	
317	I	I,V	I	
322	N	N,W	N	
327	Y	Y,F	Y	
352	N	N,V,W	N,V	V
360	Y	Y,F	Y	
382	P	P,del	P	
383	P	P,del	P	
392	A	A,T	A,T	T
398	S	S,T	S,T	T
431	Y	Y,H	H	
448	H	H,Y	H	

Первый и второй аспекты изобретения, хотя и являются различными решениями одной и той же проблемы, не обязательно являются взаимоисключающими. Таким образом, данное изобретение имеет отношение к белкам, выполняющим условия как

первого аспекта, так и второго аспекта, описанных выше. Важно отметить, что первый аспект и второй аспект являются двумя альтернативными решениями проблемы обеспечения ферментов GH7 с улучшенной температурной стабильностью. Эти решения являются независимыми (хотя для некоторых примеров частично перекрывающимися) и таким образом, не обязательно должны быть объединены. Например, белок, идентифицированный в виде [6] на фиг. 3, демонстрирует температурную стабильность по сравнению с белком, имеющим SEQ ID NO: 4 ([2] на фиг. 3), но еще не демонстрирует активную термостабилизацию.

Таким образом, изобретение обеспечивает много различных вариантов фермента согласно первому аспекту, выше и/или согласно второму аспекту, выше. Демонстрирует ли любой предоставленный фермент необходимые термические свойства (температурную стабильность и/или активную термостабилизацию) может быть легко протестировано тестом, озаглавленным “Определение термостабильности и/или активной термостабилизации”, выше.

Как подробно изложено выше в разделе определений, а также детально определено в примерах ниже, здесь кратко изложено как необходимые мутации могут быть получены:

- попарным выравниванием любой последовательности эндоглюканазы GH7 с Seq ID NO 2 с использованием алгоритма ClustalW;
- идентификацией соответствующих положений (линия 1) в последовательности-мишени эндоглюканазы GH7;
- модификацией соответствующих положений последовательности-мишени эндоглюканазы GH7 согласно предполагаемым предпочтительным заменам, предоставленным в линии 2, или предпочтительно в линии 3;
- экспрессией модифицированной последовательности и тестированием экспрессируемого белка на улучшенные термические свойства.

Считается, что термостабильные ферменты по изобретению также получают с пониженным образованием агрегатов при более высоких температурах, и таким образом с пониженной преципитацией. Предотвращение таких преципитатов особенно предпочтительно при наличии в разработанных лоскутах, дениме или тканевых материалах, а также для применений в мембранных реакторах, для снижения характеристик засорения мембраны.

Слитые белки, содержащие любой белок по изобретению, также являются частью данного изобретения.

Другой аспект настоящего изобретения имеет отношение к получению белков по изобретению путем гетерологичной экспрессии в хозяине для получения, также

называемым экспрессирующим хозяином. Способы гетерологичной экспрессии содержат перенос нуклеиновой кислоты, кодирующей белок по изобретению (экспрессирующая конструкция), в хозяин для получения путем трансформации, трансфекции, скрещивания или эквивалентными способами переноса в отношении нуклеиновой кислоты (ДНК или РНК). Способы трансформации в смысле данного изобретения конкретно не ограничиваются. Сообщалось о примерах для различных видов, и (способы трансформации) включают электропорацию, трансформацию протопластов, химическую трансформацию и перенос с использованием частиц для баллистической трансформации, микроинъекцию, инфицирование вирусом, перекрестное скрещивание или применение природных компетентных штаммов или клеточных линий. Предпочтительный хозяин для получения совместно секретирует эндоглюканазу по изобретению с другими целлюлазами, гемицеллюлазами или пектиназами в культуральную жидкость. Таким образом, предпочтительным является то, что кодирующая последовательность в экспрессирующей конструкции кодирует эндоглюканазу данного изобретения, с предшествующим сигналом для секреции из конкретного штамма-хозяина. Такие сигналы хорошо известны специалистам в данной области; например, в *Eubacteria*, они называются сигнальными пептидами. Не желая быть связанными конкретной теорией, эти сигналы имеют общую способность направлять секрецию белка, типично котрансляционным способом. Предпочтительным экспрессирующим хозяином является *Trichoderma reesei*.

Дополнительным аспектом данного изобретения является применение описанных выше белков эндоглюканаз. Дополнительный аспект включает применения очищенных, частично очищенных или неочищенных препаратов белка как таковых или в композиции фермента, а также применение цельных клеток или организмов, экспрессирующих белок-мишень. Области применений эндоглюканаз могут быть найдены в главе «Область изобретения». Как указано там, применение термостабильных белков крайне необходимо. Предпочтительное применение эндоглюканазы лежит в области ферментативного превращения лигноцеллюлозы.



## Обзор описанных здесь последовательностей

Seq. ID NO	Тип	Функция	Источник
1	ДНК	Последовательность ДНК, кодирующая Seq. ID NO:2 (эндоглюканаза GH7 по изобретению) –использование кодона, адаптированного для <i>Pichia pastoris</i>	Искусственный
2	Белок	Эндоглюканаза GH7 по изобретению – последовательность зрелого белка	Искусственный
3	ДНК	Последовательность ДНК, кодирующая Seq. ID NO: 4 (эндоглюканаза GH7 Cel7B DNA) – <i>использование адаптированного кодона</i>	Искусственный
4	Белок	Эндоглюканаза I (Cel7B) – последовательность зрелого белка	<i>Hypocrea pseudokonigii</i> - AVM90986
5	Белок	Эндоглюканаза I (Cel7B) – последовательность зрелого белка	Искусственный
6	Белок	Эндоглюканаза GH7 по изобретению – последовательность зрелого белка	Искусственный
7	Белок	Эндоглюканаза GH7 по изобретению– последовательность зрелого белка	Искусственный
8	Белок	Эндоглюканаза I (Cel7B) – вариант последовательности зрелого белка	Искусственный
9	Белок	Эндоглюканаза GH7 по изобретению– последовательность зрелого белка	Искусственный
10	Белок	Эндоглюканаза GH7 по изобретению – последовательность зрелого белка	Искусственный
11	Белок	Эндоглюканаза GH7 по изобретению – последовательность зрелого белка	Искусственный
12	Белок	Эндоглюканаза GH7 по изобретению – последовательность зрелого белка	Искусственный
13	Белок	Эндоглюканаза I (Cel7B) – вариант последовательности зрелого белка	Искусственный
14	Белок	Эндоглюканаза I (Cel7B) – вариант последовательности зрелого белка	Искусственный
15	ДНК	Последовательность ДНК, кодирующая Seq. ID NO. 2 с N-концевой 6 х гистидиновой меткой Histidine TAG ( <i>italic</i> ) с сигнальным пептидом SP <sub>mfa</sub> ( <u>подчеркнуто</u> ) в <i>pichia pastoris</i>	Искусственный
16	ДНК	Последовательность ДНК, кодирующая Seq. ID NO. 2 в слиянии с сигнальным пептидом СВН1 ( <u>подчеркнуто</u> ) для экспрессии в <i>Trichoderma reesei</i>	Искусственный

## Последовательности, описанные здесь (NO: 1-16)

SEQ ID NO: 1

TCTCTGCAGCCAGGAACCTTCTACTCCAGAGGTGCACCCAAAGCTGACCACCTACAAGTGTACCACCTCTGGT  
 GGTGTGTGTTGCTCAGAACACCTATGTTGTTCTGGACTGGAACACAGATGGATCCACGACGCCAACTACAACCTCTTG  
 TACCSTGAACGGTGGTGTCAACACTACTCTGTGTCCAGACGAGGCTACTGGTAGCAAGAACTGCTTCATCGAGGGTG  
 TTGACTACGCTGCTTCTGGTGTACTGCCAATGGTTCTACSTTGACCCTGAACCAGTACATGCCATCTTCCCTCTGCC

GGTTACACTTCTGTGTCGCCAAGACTGTACTTGGTTGGGTCCAGACGGTAAGTACGTTATGCTGAAGCTGAACGGACA  
 GGAGCTGTCTTTTGACGTTGACCTGTCTGCTTTGCCATGTGGAGAGAACGCTTCTCTGTACCTGTCTCAGATGGACG  
 AGAACGGTGGAGCTAACCAGTACAACACCCGCCGGTGTAACTACGGTTCTGGTTACTGTGACGCCCAGTGTCCAGTT  
 CAGACTTGGAGAAACGGAACCCTGAACACTTCTGGCCAGGGATTCTGCTGTAACGAGATGGACATCTTGAGGGGAAA  
 CTCTAGAGCTAACGCTCTGACCCACACTCTTGTAATGCTACCGCTTGTGACTCTGCTGGTTGCGGTTTTAACCCAT  
 ACCGCTCGGGTTACCCAACTACTTTGGCCCAGGTGGCACTGTTGACACCTCGAAGCCATTACCATCATCACCAG  
 TTCAACACCGACAACGGTTCTCCATCTGGTAACCTGGTGTGATCACCAGAAAGTACAGACAGAACGGCGTTGACAT  
 CCCATCTGCTAAACCAGGTGGCGACACCATTTTCGTCTTGTCCATCTGCCTCTACTTACGGTGGATTGGCTACCATGG  
 GAAAGGCTCTGTCCGAGGGAATGGTGTGATCTTCTCGATCTGGAACGACAACCTCGCAGTACATGAACTGGCTGGAC  
 TCTGGTGATGCTGGTCCATGTTCTTCTACCGAGGGCAACCCATCTAACATCCTGGCTAACAAACCCTGGTACTCACGT  
 GGTGTACTCGAACATTAGATGGGGCGACATTGGTTCTACCACCAACTCTACCGGTGGTAACCCACCACCACCCTG  
 CATCTTCTACCACCTTCTCGACCGCCAGAAGATCGTCTACCTCCTCTTCTTCTCCATCTTGTATCCAGACTCACTGG  
 GGTGAGTGTGGTGGTATTGGCTACACCGGCTGTAAGACCTGTACCTCTGGAACCACTTGCCAGTACAGCAACGACTA  
 CTACTCTCAGTGCCTGTGA

**SEQ ID NO: 2**

SLQPGTSTPEVHPKLTITYKCTTSGGCVAQNTYVVLWDNYRWIHDANYNSCTVNGGVNNTLCPDEATGSKNCF  
 IEGVDYAASGVTANGSTLTLNQYMPSSSGGYTSVSPRLYLLGPDGKYVMLKLNQELSFDVDLSALPCGENASLYLS  
 QMDENGGANQYNTAGANYGSGYCDQCPVQTRNGLTNTSGQGFCCNEMDILEGNSRANALTPHSCNATAACDSAGCG  
 FNPYRSGYPNYFGPGGTVDTSKPFTIITQFNTDNGSPSGNLVSI TRKYRQNGVDI PSAKPGGDTISSCPSASTYGG  
 ATMGKALSEGVLIFSIWNDNSQYMNWLDSDAGPCSSTEGNPSNILANNPGTHVVYSNIRWGDIGSTTNSTGGNPP  
 PPPASSTTFSTARRSSTSSSSPSCIQTHWGQCGGIGYTGCKTCTSGTTCQYSNDYYSQCL\*

**SEQ ID NO: 3**

TCTCAGCAGCCAGGAACCTTCTACTCCAGAGGTGCACCCAAAGCTGACCACCTACAAGTGTACCACCTCTGGT  
 GGTGTGTTGCTCAGGACACCTCTGTTGTTCTGGACTGGAACCTACAGATGGATGCACGACGCCAACTACAACCTCTTG  
 TACCGTGAACGGTGGTGTCAACACTACTCTGTGTCCAGACGAGGCTACTTGTGGCAAGAACTGCTTCATCGAGGGTG  
 TTGACTACGCTGCTTCTGGTGTACTGCCTCTGGTTCTACCTTGACCCTGAACCAGTACATGCCATCTTCTCTGGC  
 GGTTACTCTTCTGTGTCGCCAAGACTGTACTTGGTTGGGTCCAGACGGTGTAGTACGTTATGCTGAAGCTGAACGGACA  
 GGAGCTGTCTTTTGACGTTGACCTGTCTGCTTTGCCATGTGGAGAGAACGGTTCTCTGTACCTGTCTCAGATGGACG  
 AGAACGGTGGAGCTAACCAGTACAACACCCGCCGGTGTAACTACGGTTCTGGTTACTGTGACGCCCAGTGTCCAGTT  
 CAGACTTGGAGAAACGGAACCCTGAACACTTCTGGCCAGGGATTCTGCTGTAACGAGATGGACATCTTGAGGGGAAA  
 CTCTAGAGCTAACGCTCTGACCCACACTCTTGACTGCTACCGCTTGTGACTCTGCTGGTTGCGGTTTTAACCCAT  
 ACCGCTCGGGTTACCCAACTACTTTGGCCCAGGTGACACTGTTGACACCTCGAAGCCATTACCATCATCACCAG  
 TTCAACACCGACAACGGTTCTCCATCTGGTAACCTGGTGTGATCACCAGAAAGTACAGACAGAACGGCGTTGACAT  
 CCCATCTGCTAAACCAGGTGGCGACACCATTTTCGTCTTGTCCATCTGCCTCTGCTTACGGTGGATTGGCTACCATGG  
 GAAAGGCTCTGTCTCTGGAATGGTGTGATCTTCTCGATCTGGAACGACAACCTCGCAGTACATGAACTGGCTGGAC  
 TCTGGTTCTGCTGGTCCATGTTCTTCTACCGAGGGCAACCCATCTAACATCCTGGCTAACAAACCCTGGTACTCACGT  
 GGTGTACTCGAACATTAGATGGGGCGACATTGGTTCTACCACCAACTCTACCGGTGGTAACCCACCACCACCCTG  
 CATCTTCTACCACCTTCTCGACACCAGAAGATCGTCTACCACCTCTTCTTCTCCATCTTGTATCCAGACTCACTGG  
 GGTGAGTGTGGTGGTATTGGCTACACCGGCTGTAAGACCTGTACCTCTGGAACCACTTGCCAGTACGGCAACGACTA  
 CTACTCTCAGTGCCTGTGA

## SEQ ID NO: 4

SQQPGTSTPEVHPKLTITYKCTTSGGCVAQDTSVVLWDWNYRWMHDANYNSCTVNGGVNTTLCPEATCGKNCF  
 IEGVDYAASGVTASGSTLTLNQYMPSSSGGYSSVSPRLLYLLGPDGEYVMLKLNQELSFVDVLSALPCGENGLYLS  
 QMDENGGANQYNTAGANYGSGYCAQCPVQVQTRNGTLNTSGQGFCCNEMDILEGNSRANALTPHSCATACDSAGCG  
 FNPYSGGYPNYFGPGDVTDTSKPFTIITQFNTDNGSPSGNLVSI TRKYRQNGVDI P SAKPGGDTISSCPSASAYGGL  
 ATMKGALSSGMVLI FSIWWDNSQYMNWLD S G S A G P C S S T E G N P S N I L A N N P G T H V V Y S N I R W G D I G S T T N S T G G N P P  
 PPPASSTTFSTTRRSSTSSSPSCTQTHWGQCGGIGYTGCKTCTSGTTCQYGN DYYSQCL\*

## SEQ ID NO: 5

SQQPGTSTPEVHPKLTITYKCTTSGGCVAQDTSVVLWDWNYRWMHDANYNSCTVNGGVNTTLCPEATCGKNCF  
 IEGVDYAASGVTASGSTLTLNQYMPSSSGGYSSVSPRLLYLLGPDGEYVMLKLNQELSFVDVLSALPCGENGLYLS  
 QMDKNGGANQYNTAGANYGSGYCAQCPVQVQTRNGTLNTSGQGFCCNEMDILEGNSRANALTPHSCATACDSAGCG  
 FNPYSGGYPNYFGPGDVTDTSKPFTIITQFNTDNGSPSGNLVSI TRKYRQNGVDI P SAKPGGDTISSCPSASAYGGL  
 ATMKGALSSGMVLI FSIWWDNSQYMNWLD S G S A G P C S S T E G N P S N I L A N N P G T H V V Y S N I R W G D I G S T T N S T G G N P P  
 PPPASSTTFSTTRRSSTSSSPSCTQTHWGQCGGIGYTGCKTCTSGTTCQYGN DYYSQCL\*

## SEQ ID NO: 6

SLQPGTSTPEVHPKLTITYKCTTSGGCVAQNTSVVLWDWNYRWMHDANYNSCTVNGGVNTTLCPEATGGKNCF  
 IEGVDYAASGVTASGSTLTLNQYMPSSSGGYSSVSPRLLYLLGPDGKYVMLKLNQELSFVDVLSALPCGENASLYLS  
 QMDENGGANQYNTAGANYGSGYCAQCPVQVQTRNGTLNTSGQGFCCNEMDILEGNSRANALTPHSCNATAACDSAGCG  
 FNPYSGGYPNYFGPGGTVDTSKPFTIITQFNTDNGSPSGNLVSI TRKYRQNGVDI P SAKPGGDTISSCPSASTYGGGL  
 ATMKGALSSGMVLI FSIWWDNSQYMNWLD S G S A G P C S S T E G N P S N I L A N N P G T H V V Y S N I R W G D I G S T T N S T G G N P P  
 PPPASSTTFSTTRRSSTSSSPSCIQTHWGQCGGIGYTGCKTCTSGTTCQYSNDYYSQCL\*

## SEQ ID NO: 7

SLQPGTSTPEVHPKLTITYKCTTSGGCVAQNTYVVLWDWNYRWIHDANYNSCTVNGGVNTTLCPEATGSKNCF  
 IEGVDYAASGVTANGSTLTLNQYMPSSSGGYTSVSPRLLYLLGPDGKYVMLKLNQELSFVDVLSALPCGENASLYLS  
 QMDENGGANQYNTAGANYGSGYCAQCPVQVQTRNGTLNTSGQGFCCNEMDILEGNSRANALTPHSCNATAACDSAGCG  
 FNPYSGGYPNYFGPGGTVDTSKPFTIITQFNTDNGSPSGNLVSI TRKYRQNGVDI P SAKPGGDTISSCPSASTYGGGL  
 ATMKGALSSGMVLI FSIWWDNSQYMNWLD S G S A G P C S S T E G N P S N I L A N N P G T H V V Y S N I R W G D I G S T T N S T G G N P P  
 PPPASSTTFSTTRRSSTSSSPSCIQTHWGQCGGIGYTGCKTCTSGTTCQYSNDYYSQCL\*

## SEQ ID NO: 8

SQQPGTSTPEVHPKLTITYKCTTSGGCVAQNTSVVLWDWNYRWMHDANYNSCTVNGGVNTTLCPEATCGKNCF  
 IEGVDYAASGVTASGSTLTLNQYMPSSSGGYSSVSPRLLYLLGPDGKYVMLKLNQELSFVDVLSALPCGENASLYLS  
 QMDENGGANQYNTAGANYGSGYCAQCPVQVQTRNGTLNTSGQGFCCNEMDILEGNSRANALTPHSCNATAACDSAGCG  
 FNPYSGGYPNYFGPGGTVDTSKPFTIITQFNTDNGSPSGNLVSI TRKYRQNGVDI P SAKPGGDTISSCPSASAYGGL  
 ATMKGALSSGMVLI FSIWWDNSQYMNWLD S G S A G P C S S T E G N P S N I L A N N P G T H V V Y S N I R W G D I G S T T N S T G G N P P  
 PPPASSTTFSTTRRSSTSSSPSCTQTHWGQCGGIGYTGCKTCTSGTTCQYGN DYYSQCL\*

## SEQ ID NO: 9

SLQPGTSTPEVHPKLTITYKCTTSGGCVAQNTSVVLWDWNYRWMHDANYNSCTVNGGVNTTLCPEATCCKNCF  
 IEGVDYAASGVTASGSTLTLNQYMPSSSGGYSSVSPRLLYLLGPDGKYVMLKLNQELSFVDVLSALPCGENASLYLS  
 QMDENGGANQYNTAGANYGSGYCAQCPVQVQTRNGTLNTSGKGFCCNEMDILEGNSRANALTPHSCNATAACDSAGCG  
 FNPYSGGYPNYFGPGGTVDTSKPFTIITQFNTDNGSPSGNLVSI TRKYRQNGVDI P SAKPGGDTISSCPSASAYGGL

ATMGKALSSGMVLI FSIWNDNSQYMNWLD SSGSAGPCSSTEGNPSNILANNPGTHVVYSNIRWGDIGSTTNSTGGNPP  
PPPASSTTFSTTRRSSTTSSSPSCTQTHWQC GCGIGYTGCKTCTSGTTCQYSNDYYSQCL\*

SEQ ID NO: 10

SQQPGTSTPEVHPKLT TYKCTTSGGCVAQNTSVVLDWNYRWMHDANYNSCTVNGGVNTTLC PDEATCGKNCF  
IEGVDYAASGVTASGSTLT LNQYMPSSSGGYSSVSPRLLYLLGPDGKYVMLKLNQELSF DVDLSALPCGENASLYLS  
QMDENGGANQYNTAGANYGSGY CDAQCPVQ TWRNGLNTSGQGFCCNEMDILEGNSRANALTPHSCNATA CDSAGCG  
FNPYKSGYPNYFGPGGTVD TSKPFTIITQFNTDNGSPSGNLVSI TRKYRQNGVDI PSAKPGGDTISSCPSASAYGGL  
ATMGKALSEG MVLIFSIWNDNSQYMNWLD SSGSAGPCSSTEGNPSNILANNPGTHVVYSNIRWGDIGSTTNSTGGNPP  
PPPASSTTFSTTRRSSTTSSSPSCTQTHWQC GCGIGYTGCKTCTSGTTCQY GNDYYSQCL\*

SEQ ID NO: 11

SQQPGTSTPEVHPKLT TYKCTTSGGCVAQNTSVVLDWNYRWMHDANYNSCTVNGGVNTTLC PDEATCGKNCF  
IEGVDYAASGVTASGSTLT LNQYMPSSSGGYSSVSPRLLYLLGPDGKYVMLKLNQELSF DVDLSALPCGENASLYLS  
QMDENGGANQYNTAGANYGSGY CDAQCPVQ TWRNGLNTSGQGFCCNEMDILEGNSRANALTPHSCNATA CDSAGCG  
FNPYKSGYPNYFGPGGTVD TSKPFTIITQFNTDNGSPSGNLVSI TRKYRQNGVDI PSAKPGGDTISSCPSASAYGGL  
ATMGKALSDGMVLIFSIWNDNSQYMNWLD SGEAGPCSSTEGNPSNILANNPGTHVVYSNIRWGDIGSTTNSTGGNPP  
PPPASSTTFSTTRRSSTTSSSPSCTQTHWQC GCGIGYTGCKTCTSGTTCQY GNDYYSQCL\*

SEQ ID NO: 12

SQQPGTSTPEVHPKLT TYKCTTSGGCVAQNTSVVLDWNYRWMHDANYNSCTVNGGVNTTLC PDEATCGKNCF  
IEGVDYAASGVTASGSTLT LNQYMPSSSGGYSSVSPRLLYLLGPDGKYVMLKLNQELSF DVDLSALPCGENASLYLS  
QMDENGGANQYNTAGANYGSGY CDAQCPVQ TWRNGLNTSGQGFCCNEMDILEGNSRANALTPHSCNATA CDSAGCG  
FNPYKSGYPNYFGPGGTVD TSKPFTIITQFNTDNGSPSGNLVSI TRKYRQNGVDI PSAKPGGDTISSCPSASAYGGL  
ATMGKALSDGMVLIFSIWNDNSQYMNWLD SSGSAGPCSSTEGNPSNILANNPGTHVVYSNIRWGDIGSTTNSTGGNPP  
PPPASSTTFSTTRRSSTTSSSPSCTQTHWQC GCGIGYTGCKTCTSGTTCQY GNDYYSQCL\*

SEQ ID NO: 13

SQQPGTSTPEVHPKLT TYKCTTSGGCVAQNTSVVLDWNYRWIHDANYNSCTVNGGVNTTLC PDEATCSKNCF  
IEGVDYAASGVTANGSTLT LNQYMPSSSGGYSSVSPRLLYLLGPDGEYVMLKLNQELSF DVDLSALPCGENGSLYLS  
QMDENGGANQYNTAGANYGSGY CDAQCPVQ TWRNGLNTSGQGFCCNEMDILEGNSRANALTPHSC TATA CDSAGCG  
FNPYKSGYPNYFGPGDTVD TSKPFTIITQFNTDNGSPSGNLVSI TRKYRQNGVDI PSAKPGGDTISSCPSASAYGGL  
ATMGKALSSGMVLIFSIWNDNSQYMNWLD SSGSAGPCSSTEGNPSNILANNPGTHVVYSNIRWGDIGSTTNSTGGNPP  
PPPASSTTFSTTRRSSTTSSSPSCTQTHWQC GCGIGYTGCKTCTSGTTCQY GNDYYSQCL\*

SEQ ID NO: 14

SQQPGTSTPEVHPKLT TYKCTTSGGCVAQNTSVVLDWNYRWIHDANYNSCTVNGGVNTTLC PDEATCGKNCF  
IEGVDYAASGVTANGSTLT LNQYMPSSSGGYSSVSPRLLYLLGPDGEYVMLKLNQELSF DVDLSALPCGENGSLYLS  
QMDENGGANQYNTAGANYGSGY CDAQCPVQ TWRNGLNTSGQGFCCNEMDILEGNSRANALTPHSC TATA CDSAGCG  
FNPYKSGYPNYFGPGDTVD TSKPFTIITQFNTDNGSPSGNLVSI TRKYRQNGVDI PSAKPGGDTISSCPSASAYGGL  
ATMGKALSSGMVLIFSIWNDNSQYMNWLD SSGSAGPCSSTEGNPSNILANNPGTHVVYSNIRWGDIGSTTNSTGGNPP  
PPPASSTTFSTTRRSSTTSSSPSCTQTHWQC GCGIGYTGCKTCTSGTTCQY GNDYYSQCL\*

SEQ ID NO: 15

atgagatttccttcaat ttttactgcag ttttattcgcagcatcctccgcattagctgctccagtcaacact  
acaacagaagatgaaacggcaca aattccggctgaagctgtcatcggttacttagatttagaaggggatttcgatgt  
tgctgttttgccat tttccaacagcaca aataacgggttattgtttataaatactactattgccagcattgctgcta

aagaagaaggggtatctttggataaacgtgaggcggaagca *tgccaccaccaccaccactcctccggctctctg*  
*cagccaggaacttctactccagaggtgcacccaaagctgaccacctacaagtgtaccacctctggtggttgtgttgc*  
*tcagaacacctatggttctggactggaactacagatggatccacgacgccaactacaactcttgtaccgtgaacg*  
*gtggtgtcaacactactctgtgtccagacgaggctactggtagcaagaactgcttcatcgagggtgttgactacgt*  
*gcttctggtgttactgccaatggttctaccttgaccctgaaccagtacatgccatcttctctggcggttacacttc*  
*tgtgtcgccaagactgtacttgttgggtccagacggtaagtagctttagctgaagctgaacggacaggagctgtctt*  
*ttgacgttgacctgtctgctttgccatgtggagagaacgcttctctgtacctgtctcagatggacgagaacggtgga*  
*gctaaccagtagaacaccgcgggtgctaactacggttctgggtactgtgacgcccagtgtccagttcagacttgag*  
*aaacggaaacctgaacacttctggccaggattctgctgtaacgagatggacatcttggagggaaactctagagcta*  
*acgctctgacccccacactcttghtaatgctaccgcttgtgactctgctggttgcggttttaaccataccgctcgggt*  
*tacccaaactactttggcccagggtggcactgttgacacctcgaagccattcaccatcatcaccagttcaacaccga*  
*caacggttctccatctggtaacctgggtgtcgatcaccagaaagtagacagaaacggcggttgacatcccatctgcta*  
*aaccaggtggcgacaccatttctgtctgtccatctgcctctacttacggtggattggctaccatgggaaaggctctg*  
*tccgaggggaatggtgctgatcttctcgatctggaacgacaactcgcagtacatgaaactggctggactctggtgatgc*  
*tggtccatgttcttctaccgagggcaaccatctaacatcctggctaaacaacctggtaactcagtggtgtactcga*  
*acattagatggggcgacattgggttctaccaccaactctaccggtggtaaccaccaccaccacctgcatcttctacc*  
*accttctcgaccgccagaagatcgtctacctcctcttcttctccatcttgtatccagactcactggggtcagtggtg*  
*tggtataggctacaccggctgtaagacctgtacctctggaaccacttgccagtagcaacgactactactctcag*  
*gcctgtga*

#### SEQ ID NO: 16:

atgtatcggaagttggccgtcatctcggccttcttggccacagcacgggcttctctgcaaccgggtaccag  
*acccccgaggtccatcccaagttgacaacctacaagtgtagaacctccgggggggtgctggtggccagaacacctatgt*  
*ggtccttgactggaactaccgctggatccacgacgcaactacaactcgtgcaaccgtcaacggcggtcaacacca*  
*cgctctgccctgacgaggcgaccggttagcaagaactgcttcatcgaggcgctcgaactacgccgctcggggcgtcacg*  
*gccaatggcagcaccctcaacctgaaccagtagatgccagcagctctggcggttaactagcgtctctcctcggct*  
*gtatctcctgggtccagacggtaagtagctgatgctgaagctcaacggccaggagctgagcttcgacgtcgacctct*  
*ctgctctgccgtgtggagagaacgctcgtctacctgtctcagatggacgagaacggggggcgccaaccagtataac*  
*acggccgggtgccaactacgggagcggctactgcgatgctcagtgccccgtccagacatggaggaacggcaccctcaa*  
*cactagcggccagggttctgctgcaacgagatggatatacctggagggcaactcgagggcgaatgccttgaccctc*  
*actcttgcaatgccacggcctggaactctgcccgttgcggttcaaccctatcgacgcggtaccctaaactacttc*  
*ggccccggaggcaccggtgacacctccaagccattcaccatcatcaccagttcaacacggacaacggctcgcctc*  
*gggcaaccttgtgagcatcaccgcgaagtagacaaaaacggcgctcgacatcccagcgccaaacccggcgggcaca*  
*ccatctcgtcctgccgctccgctcaactacggcggcctcgccaccatgggcaaggccctgagcaggggcatggtg*  
*ctcatcttcagcatttggaaacgacaacagccagtagatgaaactggctcgacagcggcgatgccggccctgcagcag*  
*caccgagggcaaccatccaacatcctggccaacaacccgggtacgcacgtcgtctactccaacatccgctggggag*  
*acattgggtctactacgaactcgaactgggtgggtccgcccccgctcgtccagcagcagcttttcgactgcccgagg*  
*agctcgacgtcctcgagcagcccagctgcattccagactcactgggggcagtgcggtggcattgggtacaccgggtg*  
*caagacgtgcacgtcgggcactacgtgccagtagcaacgactactactcgcaatgcctttaa*

#### Примеры

##### Пример 1. Создание библиотек и конкретных вариантов

Библиотеку на основе Seq. ID NO: 3 (библиотека “N7”) получали с использованием SEQ ID NO: 3 в качестве матрицы путем ошибочно-направленной ПЦР с использованием

полимеразы Taq, исходя из опубликованного протокола (Joyce et al) с использованием следующих условий ПЦР: 2 мин. при 95°C, 30 циклов (1 мин. при 95°C, 1 мин. при 56°C, 1 мин. при 72°C), 5 мин. при 72°C). Все продукты, полученные в результате ПЦР, очищали с помощью набора QIAquick PCR Purification Kit (Qiagen, Hilden, Germany).

Конкретные варианты Seq. ID NO: 3 получали путем модифицированного протокола ПЦР с использованием праймеров, содержащих мутировавшую нуклеотидную последовательность (Ho, S. N. et al. Gene; 1989; 77; 51-9).

### **Пример 2. Экспрессия в *Pichia pastoris***

Конструирование линейной экспрессирующей кассеты (LEC). LECs (Liu Z, et al. Chembiochem. 2008 Jan 4; 9 (1):58-61) с маркером Zeocin и GAP-промотором конструировали путем модифицированного протокола ПЦР.

Трансформация и культивирование *Pichia pastoris*. Компетентные клетки получали и трансформировали как описано (Lin-Cereghino, J., et al. BioTechniques. 2005, 38, 44-48). Трансформанты отбирали на чашках с YPD на агаровой основе, содержащей Zeocin (зеоцин) 100мг/л, и собирали в чашки с глубокими лунками (DWP) (BMD5% 250мл/лунку) с использованием автоматического устройства для сбора (колоний) (QPix2, Genetix). Инокулированные DWPs культивировали в течение 60 часов при 28°C, влажности 80% и встряхивании 280 об./мин.

### **Пример 3. Экспрессия в *Trichoderma reesei***

#### **Конструкция экспрессирующего вектора *Trichoderma reesei***

Расщепленную SbfI/SwaI, линейаризованную ДНК плазмиды pV7 (фиг. 2) трансформировали в *Trichoderma reesei* SCF41 по существу как описано Penttilä et al 1997. Селекцию трансформантов делали на чашках со средой Mandel's Andreotti, содержащей гигромицин в качестве селективного агента (100 мг/л). Трансформанты проверяли ПЦР.

### **Пример 4. Определение способности к превращению субстрата при различных температурах для указания термостабильности вариантов Seq ID NO. 2 с использованием 4-метилумбеллиферрил-β-D-целлобиозида (4-MUC)**

Для точного сравнения термостабильности 10 мкл культуральных супернатантов *Pichia pastoris*, содержащих секретрируемые варианты эндоглоканазы, инкубировали с 90 мкл 100 мкМ 4-MUC (растворенного в натрий-ацетатном буфере (50мМ, pH 5,0)) в градиенте температуры термоциклера Eppendorff Gradient Thermocycler. 24 реакционных смеси инкубировали в градиенте температуры, простирающимся от 45°C до 65°C и от 55°C до 75°C (каждую реакционную смесь выдерживали при уникальном постоянном уровне температуры) в течение одного часа. Ферментативная активность при соответствующей температуре может быть определена после добавления 100 мкл 1М

раствора углекислого натрия в каждую реакционную смесь и (после) измерения интенсивности флуоресценции при 360нм/454нм в планшет-ридере Tecan Infinite M200. Для сравнения термостабильности значения (сигналов) флуоресценции в каждой температурной точке, относительную ферментативную активность определяли путем деления на максимальное значение серий (нормализация к 1). Температурный профиль для любого данного фермента получали построением графика относительной ферментативной активности в диапазоне измеряемых температур.

#### **Пример 5. Активная термостабилизация некоторых вариантов эндоглюканазы**

Этот пример описывает примеры удивительного эффекта *активной термостабилизации*. В этом примере, использовали белки (культуральный супернатант) (таблица 5 ниже), экспрессируемые в *Pichia pastoris*.

Фигура 3 демонстрирует установленные свойства белков по изобретению: белки, названные как [1], [4], [5], [7], [8], [9] и [10], демонстрируют активную термостабилизацию и температурную стабильность, тогда как белки, названные как [3] и [6], демонстрируют температурную стабильность, но не активную термостабилизацию.

Таблица 5. Белки, протестированные в примерах 5, 6 и 7

Обозначение	SEQ ID NO:	Фигура, №	Мутации по отношению к Seq. ID NO: 2
EGI <i>Hypocrea pseudoconingii</i>	Seq. ID NO: 4	[2]	L2Q, N30D, Y32S, I42M, G67C, S68G, N86S, T104S, K118E, A144G, N216T, R231G, G242D, T299A, E312S, D335S, A392T, S398T, I405T, S432G
Вариант EGI <i>Hypocrea pseudokoningii</i> (E153K по отношению к Seq. ID NO: 4)	Seq. ID NO: 5	[3]	L2Q, N30D, Y32S, I42M, G67C, S68G, N86S, T104S, K118E, A144G, E153K, N216T, R231G, G242D, T299A, E312S, D335S, A392T, S398T, I405T, S432G
EG/вариант по изобретению; полученный мутагенезом из последовательности EGI <i>Hypocrea pseudokoningii</i>	Seq. ID NO: 2	[1]	
Пример варианта 1 Seq. ID NO: 2	Seq. ID NO: 6	[4]	Y32S, I42M, S68G, N86S, T104S, R231G, E312S, D335S, A392T
Пример варианта 2 Seq. ID NO: 2	Seq. ID NO: 7	[5]	R231G, E312S, D335S, A392T

Вариант EGI <i>Hypocrea pseudokoningii</i> (по отношению к Seq. ID NO: 4)	<b>Seq. ID NO: 8</b>	[6]	,L2Q,Y32S,I42M,G67C,S68G,N86S,T104S,R231 G,T299A, E312S,D335S,A392T,S398T,I405T,S432G
Пример варианта 3 Seq. ID NO:2	<b>Seq. ID No: 9</b>	[7]	,Y32S,I42M,G67C,S68C,N86S,T104S,Q191K,R23 1G,T299A,E312S,D335S,A392T,S398T,I405T
Пример варианта 4 Seq. ID NO:2	<b>Seq. ID No: 10</b>	[8]	,L2Q,Y32S,I42M,G67C,S68G,N86S,T104S,R231 K,T299A, D335S,A392T,S398T,I405T,S432G
Пример варианта 5 Seq. ID NO:2	<b>Seq. ID No: 11</b>	[9]	,L2Q,Y32S,I42M,G67C,S68G,N86S,T104S,R231 G,T299A, E312D,D335E,A392T,S398T,I405T,S432G
Пример варианта 6 Seq. ID NO:2	<b>Seq. ID NO:12</b>	[10]	,L2Q,Y32S,I42M,G67C,S68G,N86S,T104S,R231 K,T299A, E312D,D335S,A392T,S398T,I405T,S432G
Пример варианта 7 Seq. ID No: 2	<b>Seq. ID NO:13</b>		,L2Q,N30D,Y32S,G67C,T104S,K118E,A144G,N2 16T,R231G,G242D,T299A,E312S,D335S,A392T, S398T,I405T,S432G
Пример варианта 8 Seq. ID No: 2	<b>Seq. ID NO:14</b>		,L2Q,N30D,Y32S,G67C,S68G,T104S,K118E,A14 4G,N216T, R231G,G242D,T299A,E312S,D335S,A392T,S398 T,I405T, S432G

#### **Пример 6. Определение высвобождения редуцирующих сахаров на соломе**

Высвобождение редуцирующих сахаров на соломе определяли путем нанесения предварительно обработанной кислотой пшеничной соломы с массой сухого вещества 2,5%. Последующие ферменты добавляли в реакционную смесь: целлобиогидролазу I (12,5 мг/л), бета-глюкозидазу (40 СБУ/мг (*пер.*, СБУ - единиц карбоксипептидазы В) целлобиогидролазы I) и тестировали вариант эндоглюканазы GH7 (12,5 мг/л). Гидролиз соломы инкубировали при 60°C путем постоянного встряхивания в течение 48 часов.

#### **Пример 7. Определение температурного профиля вариантов Seq ID. No2**

Для анализа активности с использованием MUL (4-метилумбеллиферрил-β-D-лактопиранозид), 10 мкл супернатанта после культивирования смешивали с 90 мкл 100 мкМ MUL в 25 мМ Na-ацетатном буфере, pH 4,8. Планшеты заклеивали и инкубировали в течение 2 ч, со встряхиванием при 300 об./мин., при 45°C и 59°C каждый (для повторного скрининга также при 65°C). Реакцию гасили добавлением 100 мкл Na<sub>2</sub>CO<sub>3</sub> на лунку.



Возбуждение осуществляли при 365 нм, и флуоресценцию измеряли при 450 нм. Результаты показаны на фигуре 5.

#### **Пример 8. Определение температурного профиля вариантов Seq ID. No2**

Период полураспада ферментов определяли измерением остаточной активности с использованием анализа MUL, описанного в примере 7 после инкубации супернатантов культур *Pichia pastoris* после экспрессии при 70°C в течение от 0 до 7 минут на водяной бане. Пробы помещали в лед после точного периода инкубации до постановки анализа на активность.

Sequence listing.txt  
SEQUENCE LISTING

<110> Clariant Produkte (Deutschland) GmbH  
<120> Endoglucanases with improved properties  
<130> 164 411 n17  
<140>  
<141>  
<150> EP12166458  
<151> 2012-05-02  
<160> 16  
<170> PatentIn version 3.5  
<210> 1  
<211> 1323  
<212> DNA  
<213> artificial  
<220>  
<223> artificial construct

<400> 1  
tctctgcagc caggaacttc tactccagag gtgcacccaa agctgaccac ctacaagtgt 60  
accacctctg gtggttgtgt tgctcagaac acctatgttg ttctggactg gaactacaga 120  
tggatccacg acgccaacta caactcttgt accgtgaacg gtggtgtcaa cactactctg 180  
tgtccagacg aggctactgg tagcaagaac tgcttcatcg aggggtgttga ctacgctgct 240  
tctggtgtta ctgccaatgg ttctaccttg accctgaacc agtacatgcc atcttctctt 300  
ggcggttaca cttctgtgtc gccaagactg tacttgttgg gtccagacgg taagtacggt 360  
atgctgaagc tgaacggaca ggagctgtct tttgacgttg acctgtctgc tttgccatgt 420  
ggagagaacg cttctctgta cctgtctcag atggacgaga acggtggagc taaccagtac 480  
aacaccgccg gtgctaacta cggttctggt tactgtgacg cccagtgctc agttcagact 540  
tggagaaacg gaaccctgaa cacttctggc cagggattct gctgtaacga gatggacatc 600  
ttggagggaa actctagagc taacgctctg accccacct cttgtaatgc taccgcttgt 660  
gactctgctg gttgcggttt taaccatac cgctcgggtt acccaaacta ctttgccca 720  
ggtggcactg ttgacacctc gaagccattc accatcatca cccagttcaa caccgacaac 780  
ggttctccat ctggtaacct ggtgtcgatc accagaaagt acagacagaa cggcgttgac 840  
atcccatctg ctaaacaggg tggcgacacc atttcgtctt gtccatctgc ctctacttac 900  
ggtggattgg ctaccatggg aaaggctctg tccgagggaa tgggtgctgat cttctcgatc 960  
tggaacgaca actcgcagta catgaactgg ctggactctg gtgatgctgg tccatgttct 1020  
tctaccgagg gcaaccatc taacatctcg gctaacaacc ctggtactca cgtggtgtac 1080  
tcgaacatta gatggggcga cattggttct accaccaact ctaccggtgg taaccacca 1140  
ccaccacctg catcttctac caccttctcg accgccagaa gatcgtctac ctctcttct 1200  
tctccatctt gtatccagac tcaactggggt cagtgtgggtg gtattggcta caccggctgt 1260

sequence listing.txt

aagacctgta cctctggaac cacttgccag tacagcaacg actactactc tcagtgacctg 1320  
tga 1323

<210> 2  
<211> 440  
<212> PRT  
<213> artificial

<220>  
<223> artificial construct

<400> 2

Ser Leu Gln Pro Gly Thr Ser Thr Pro Glu Val His Pro Lys Leu Thr  
1 5 10 15

Thr Tyr Lys Cys Thr Thr Ser Gly Gly Cys Val Ala Gln Asn Thr Tyr  
20 25 30

Val Val Leu Asp Trp Asn Tyr Arg Trp Ile His Asp Ala Asn Tyr Asn  
35 40 45

Ser Cys Thr Val Asn Gly Gly Val Asn Thr Thr Leu Cys Pro Asp Glu  
50 55 60

Ala Thr Gly Ser Lys Asn Cys Phe Ile Glu Gly Val Asp Tyr Ala Ala  
65 70 75 80

Ser Gly Val Thr Ala Asn Gly Ser Thr Leu Thr Leu Asn Gln Tyr Met  
85 90 95

Pro Ser Ser Ser Gly Gly Tyr Thr Ser Val Ser Pro Arg Leu Tyr Leu  
100 105 110

Leu Gly Pro Asp Gly Lys Tyr Val Met Leu Lys Leu Asn Gly Gln Glu  
115 120 125

Leu Ser Phe Asp Val Asp Leu Ser Ala Leu Pro Cys Gly Glu Asn Ala  
130 135 140

Ser Leu Tyr Leu Ser Gln Met Asp Glu Asn Gly Gly Ala Asn Gln Tyr  
145 150 155 160

Asn Thr Ala Gly Ala Asn Tyr Gly Ser Gly Tyr Cys Asp Ala Gln Cys  
165 170 175

Pro Val Gln Thr Trp Arg Asn Gly Thr Leu Asn Thr Ser Gly Gln Gly  
180 185 190

Phe Cys Cys Asn Glu Met Asp Ile Leu Glu Gly Asn Ser Arg Ala Asn  
195 200 205

Sequence listing.txt

Ala Leu Thr Pro His Ser Cys Asn Ala Thr Ala Cys Asp Ser Ala Gly  
 210 215 220

Cys Gly Phe Asn Pro Tyr Arg Ser Gly Tyr Pro Asn Tyr Phe Gly Pro  
 225 230 235 240

Gly Gly Thr Val Asp Thr Ser Lys Pro Phe Thr Ile Ile Thr Gln Phe  
 245 250 255

Asn Thr Asp Asn Gly Ser Pro Ser Gly Asn Leu Val Ser Ile Thr Arg  
 260 265 270

Lys Tyr Arg Gln Asn Gly Val Asp Ile Pro Ser Ala Lys Pro Gly Gly  
 275 280 285

Asp Thr Ile Ser Ser Cys Pro Ser Ala Ser Thr Tyr Gly Gly Leu Ala  
 290 295 300

Thr Met Gly Lys Ala Leu Ser Glu Gly Met Val Leu Ile Phe Ser Ile  
 305 310 315 320

Trp Asn Asp Asn Ser Gln Tyr Met Asn Trp Leu Asp Ser Gly Asp Ala  
 325 330 335

Gly Pro Cys Ser Ser Thr Glu Gly Asn Pro Ser Asn Ile Leu Ala Asn  
 340 345 350

Asn Pro Gly Thr His Val Val Tyr Ser Asn Ile Arg Trp Gly Asp Ile  
 355 360 365

Gly Ser Thr Thr Asn Ser Thr Gly Gly Asn Pro Pro Pro Pro Ala  
 370 375 380

Ser Ser Thr Thr Phe Ser Thr Ala Arg Arg Ser Ser Thr Ser Ser Ser  
 385 390 395 400

Ser Pro Ser Cys Ile Gln Thr His Trp Gly Gln Cys Gly Gly Ile Gly  
 405 410 415

Tyr Thr Gly Cys Lys Thr Cys Thr Ser Gly Thr Thr Cys Gln Tyr Ser  
 420 425 430

Asn Asp Tyr Tyr Ser Gln Cys Leu  
 435 440

<210> 3  
 <211> 1323  
 <212> DNA  
 <213> Hypocrea pseudokonigii

<400> 3  
 tctcagcagc caggaacttc tactccagag gtgcaccca agctgaccac ctacaagtgt

Sequence listing.txt

accacctctg gtggttgtgt tgctcaggac acctctgttg ttctggactg gaactacaga 120  
 tggatgcacg acgccaacta caactcttgt accgtgaacg gtggtgtcaa cactactctg 180  
 tgtccagacg aggctacttg tggcaagaac tgcttcatcg aggggtgttga ctacgctgct 240  
 tctggtggtta ctgcctctgg ttctaccttg accctgaacc agtacatgcc atcttctctt 300  
 ggcggttact cttctgtgtc gccaagactg tacttgttgg gtccagacgg tgagtacggt 360  
 atgctgaagc tgaacggaca ggagctgtct tttgacgttg acctgtctgc tttgccatgt 420  
 ggagagaacg gttctctgta cctgtctcag atggacgaga acggtggagc taaccagtac 480  
 aacaccgccg gtgctaacta cggttctggt tactgtgacg cccagtgtcc agttcagact 540  
 tggagaaacg gaaccctgaa cacttctggc cagggattct gctgtaacga gatggacatc 600  
 ttggagggaa actctagagc taacgctctg accccacact cttgtactgc taccgcttgt 660  
 gactctgctg gttgcggttt taaccatac ggctcggggt acccaaacta ctttgcccca 720  
 ggtgacactg ttgacacctc gaagccattc accatcatca cccagttcaa caccgacaac 780  
 ggttctccat ctggtaacct ggtgtcgate accagaaagt acagacagaa cggcgttgac 840  
 atcccatctg ctaaaccagg tggcgacacc atttcgtctt gtccatctgc ctctgcttac 900  
 ggtggattgg ctaccatggg aaaggctctg tcctctggaa tgggtgctgat cttctcgatc 960  
 tggaaacgaca actcgcagta catgaactgg ctggactctg gttctgctgg tccatgttct 1020  
 tctaccgagg gcaaccatc taacatcctg gtaacaacc ctggactca cgtgggtgac 1080  
 tcgaacatta gatggggcga cattggttct accaccaact ctaccggtgg taaccacca 1140  
 ccaccacctg catcttctac caccttctcg accaccagaa gatcgtctac caccttctt 1200  
 tctccatctt gtaccagac tctactggggt cagtgtgggtg gtattggcta caccggctgt 1260  
 aagacctgta cctctggaac cacttgccag tacggcaacg actactactc tcagtgcctg 1320  
 tga 1323

<210> 4  
 <211> 440  
 <212> PRT  
 <213> Hypocrea pseudokonigii

<400> 4

Ser Gln Gln Pro Gly Thr Ser Thr Pro Glu Val His Pro Lys Leu Thr  
 1 5 10 15

Thr Tyr Lys Cys Thr Thr Ser Gly Gly Cys Val Ala Gln Asp Thr Ser  
 20 25 30

Val Val Leu Asp Trp Asn Tyr Arg Trp Met His Asp Ala Asn Tyr Asn  
 35 40 45

Ser Cys Thr Val Asn Gly Gly Val Asn Thr Thr Leu Cys Pro Asp Glu  
 50 55 60

Sequence listing.txt

Ala Thr Cys Gly Lys Asn Cys Phe Ile Glu Gly Val Asp Tyr Ala Ala  
65 70 75 80

Ser Gly Val Thr Ala Ser Gly Ser Thr Leu Thr Leu Asn Gln Tyr Met  
85 90 95

Pro Ser Ser Ser Gly Gly Tyr Ser Ser Val Ser Pro Arg Leu Tyr Leu  
100 105 110

Leu Gly Pro Asp Gly Glu Tyr Val Met Leu Lys Leu Asn Gly Gln Glu  
115 120 125

Leu Ser Phe Asp Val Asp Leu Ser Ala Leu Pro Cys Gly Glu Asn Gly  
130 135 140

Ser Leu Tyr Leu Ser Gln Met Asp Glu Asn Gly Gly Ala Asn Gln Tyr  
145 150 155 160

Asn Thr Ala Gly Ala Asn Tyr Gly Ser Gly Tyr Cys Asp Ala Gln Cys  
165 170 175

Pro Val Gln Thr Trp Arg Asn Gly Thr Leu Asn Thr Ser Gly Gln Gly  
180 185 190

Phe Cys Cys Asn Glu Met Asp Ile Leu Glu Gly Asn Ser Arg Ala Asn  
195 200 205

Ala Leu Thr Pro His Ser Cys Thr Ala Thr Ala Cys Asp Ser Ala Gly  
210 215 220

Cys Gly Phe Asn Pro Tyr Gly Ser Gly Tyr Pro Asn Tyr Phe Gly Pro  
225 230 235 240

Gly Asp Thr Val Asp Thr Ser Lys Pro Phe Thr Ile Ile Thr Gln Phe  
245 250 255

Asn Thr Asp Asn Gly Ser Pro Ser Gly Asn Leu Val Ser Ile Thr Arg  
260 265 270

Lys Tyr Arg Gln Asn Gly Val Asp Ile Pro Ser Ala Lys Pro Gly Gly  
275 280 285

Asp Thr Ile Ser Ser Cys Pro Ser Ala Ser Ala Tyr Gly Gly Leu Ala  
290 295 300

Thr Met Gly Lys Ala Leu Ser Ser Gly Met Val Leu Ile Phe Ser Ile  
305 310 315 320

Trp Asn Asp Asn Ser Gln Tyr Met Asn Trp Leu Asp Ser Gly Ser Ala  
325 330 335

Sequence listing.txt

Gly Pro Cys Ser Ser Thr Glu Gly Asn Pro Ser Asn Ile Leu Ala Asn  
 340 345 350

Asn Pro Gly Thr His Val Val Tyr Ser Asn Ile Arg Trp Gly Asp Ile  
 355 360 365

Gly Ser Thr Thr Asn Ser Thr Gly Gly Asn Pro Pro Pro Pro Ala  
 370 375 380

Ser Ser Thr Thr Phe Ser Thr Thr Arg Arg Ser Ser Thr Thr Ser Ser  
 385 390 395 400

Ser Pro Ser Cys Thr Gln Thr His Trp Gly Gln Cys Gly Gly Ile Gly  
 405 410 415

Tyr Thr Gly Cys Lys Thr Cys Thr Ser Gly Thr Thr Cys Gln Tyr Gly  
 420 425 430

Asn Asp Tyr Tyr Ser Gln Cys Leu  
 435 440

<210> 5  
 <211> 440  
 <212> PRT  
 <213> artificial

<220>  
 <223> artificial construct

<400> 5

Ser Gln Gln Pro Gly Thr Ser Thr Pro Glu Val His Pro Lys Leu Thr  
 1 5 10 15

Thr Tyr Lys Cys Thr Thr Ser Gly Gly Cys Val Ala Gln Asp Thr Ser  
 20 25 30

Val Val Leu Asp Trp Asn Tyr Arg Trp Met His Asp Ala Asn Tyr Asn  
 35 40 45

Ser Cys Thr Val Asn Gly Gly Val Asn Thr Thr Leu Cys Pro Asp Glu  
 50 55 60

Ala Thr Cys Gly Lys Asn Cys Phe Ile Glu Gly Val Asp Tyr Ala Ala  
 65 70 75 80

Ser Gly Val Thr Ala Ser Gly Ser Thr Leu Thr Leu Asn Gln Tyr Met  
 85 90 95

Pro Ser Ser Ser Gly Gly Tyr Ser Ser Val Ser Pro Arg Leu Tyr Leu  
 100 105 110

Leu Gly Pro Asp Gly Glu Tyr Val Met Leu Lys Leu Asn Gly Gln Glu  
 115 120 125

Sequence listing.txt

Leu Ser Phe Asp Val Asp Leu Ser Ala Leu Pro Cys Gly Glu Asn Gly  
 130 135 140  
 Ser Leu Tyr Leu Ser Gln Met Asp Lys Asn Gly Gly Ala Asn Gln Tyr  
 145 150 155 160  
 Asn Thr Ala Gly Ala Asn Tyr Gly Ser Gly Tyr Cys Asp Ala Gln Cys  
 165 170 175  
 Pro Val Gln Thr Trp Arg Asn Gly Thr Leu Asn Thr Ser Gly Gln Gly  
 180 185 190  
 Phe Cys Cys Asn Glu Met Asp Ile Leu Glu Gly Asn Ser Arg Ala Asn  
 195 200 205  
 Ala Leu Thr Pro His Ser Cys Thr Ala Thr Ala Cys Asp Ser Ala Gly  
 210 215 220  
 Cys Gly Phe Asn Pro Tyr Gly Ser Gly Tyr Pro Asn Tyr Phe Gly Pro  
 225 230 235 240  
 Gly Asp Thr Val Asp Thr Ser Lys Pro Phe Thr Ile Ile Thr Gln Phe  
 245 250 255  
 Asn Thr Asp Asn Gly Ser Pro Ser Gly Asn Leu Val Ser Ile Thr Arg  
 260 265 270  
 Lys Tyr Arg Gln Asn Gly Val Asp Ile Pro Ser Ala Lys Pro Gly Gly  
 275 280 285  
 Asp Thr Ile Ser Ser Cys Pro Ser Ala Ser Ala Tyr Gly Gly Leu Ala  
 290 295 300  
 Thr Met Gly Lys Ala Leu Ser Ser Gly Met Val Leu Ile Phe Ser Ile  
 305 310 315 320  
 Trp Asn Asp Asn Ser Gln Tyr Met Asn Trp Leu Asp Ser Gly Ser Ala  
 325 330 335  
 Gly Pro Cys Ser Ser Thr Glu Gly Asn Pro Ser Asn Ile Leu Ala Asn  
 340 345 350  
 Asn Pro Gly Thr His Val Val Tyr Ser Asn Ile Arg Trp Gly Asp Ile  
 355 360 365  
 Gly Ser Thr Thr Asn Ser Thr Gly Gly Asn Pro Pro Pro Pro Ala  
 370 375 380  
 Ser Ser Thr Thr Phe Ser Thr Thr Arg Arg Ser Ser Thr Thr Ser Ser  
 385 390 395 400



Sequence listing.txt

Ser Pro Ser Cys Thr Gln Thr His Trp Gly Gln Cys Gly Gly Ile Gly  
 405 410 415

Tyr Thr Gly Cys Lys Thr Cys Thr Ser Gly Thr Thr Cys Gln Tyr Gly  
 420 425 430

Asn Asp Tyr Tyr Ser Gln Cys Leu  
 435 440

<210> 6  
 <211> 440  
 <212> PRT  
 <213> artificial

<220>  
 <223> artificial construct

<400> 6

Ser Leu Gln Pro Gly Thr Ser Thr Pro Glu Val His Pro Lys Leu Thr  
 1 5 10 15

Thr Tyr Lys Cys Thr Thr Ser Gly Gly Cys Val Ala Gln Asn Thr Ser  
 20 25 30

Val Val Leu Asp Trp Asn Tyr Arg Trp Met His Asp Ala Asn Tyr Asn  
 35 40 45

Ser Cys Thr Val Asn Gly Gly Val Asn Thr Thr Leu Cys Pro Asp Glu  
 50 55 60

Ala Thr Gly Gly Lys Asn Cys Phe Ile Glu Gly Val Asp Tyr Ala Ala  
 65 70 75 80

Ser Gly Val Thr Ala Ser Gly Ser Thr Leu Thr Leu Asn Gln Tyr Met  
 85 90 95

Pro Ser Ser Ser Gly Gly Tyr Ser Ser Val Ser Pro Arg Leu Tyr Leu  
 100 105 110

Leu Gly Pro Asp Gly Lys Tyr Val Met Leu Lys Leu Asn Gly Gln Glu  
 115 120 125

Leu Ser Phe Asp Val Asp Leu Ser Ala Leu Pro Cys Gly Glu Asn Ala  
 130 135 140

Ser Leu Tyr Leu Ser Gln Met Asp Glu Asn Gly Gly Ala Asn Gln Tyr  
 145 150 155 160

Asn Thr Ala Gly Ala Asn Tyr Gly Ser Gly Tyr Cys Asp Ala Gln Cys  
 165 170 175

Sequence listing.txt

Pro Val Gln Thr Trp Arg Asn Gly Thr Leu Asn Thr Ser Gly Gln Gly  
 180 185 190

Phe Cys Cys Asn Glu Met Asp Ile Leu Glu Gly Asn Ser Arg Ala Asn  
 195 200 205

Ala Leu Thr Pro His Ser Cys Asn Ala Thr Ala Cys Asp Ser Ala Gly  
 210 215 220

Cys Gly Phe Asn Pro Tyr Gly Ser Gly Tyr Pro Asn Tyr Phe Gly Pro  
 225 230 235 240

Gly Gly Thr Val Asp Thr Ser Lys Pro Phe Thr Ile Ile Thr Gln Phe  
 245 250 255

Asn Thr Asp Asn Gly Ser Pro Ser Gly Asn Leu Val Ser Ile Thr Arg  
 260 265 270

Lys Tyr Arg Gln Asn Gly Val Asp Ile Pro Ser Ala Lys Pro Gly Gly  
 275 280 285

Asp Thr Ile Ser Ser Cys Pro Ser Ala Ser Thr Tyr Gly Gly Leu Ala  
 290 295 300

Thr Met Gly Lys Ala Leu Ser Ser Gly Met Val Leu Ile Phe Ser Ile  
 305 310 315 320

Trp Asn Asp Asn Ser Gln Tyr Met Asn Trp Leu Asp Ser Gly Ser Ala  
 325 330 335

Gly Pro Cys Ser Ser Thr Glu Gly Asn Pro Ser Asn Ile Leu Ala Asn  
 340 345 350

Asn Pro Gly Thr His Val Val Tyr Ser Asn Ile Arg Trp Gly Asp Ile  
 355 360 365

Gly Ser Thr Thr Asn Ser Thr Gly Gly Asn Pro Pro Pro Pro Ala  
 370 375 380

Ser Ser Thr Thr Phe Ser Thr Thr Arg Arg Ser Ser Thr Ser Ser Ser  
 385 390 395 400

Ser Pro Ser Cys Ile Gln Thr His Trp Gly Gln Cys Gly Gly Ile Gly  
 405 410 415

Tyr Thr Gly Cys Lys Thr Cys Thr Ser Gly Thr Thr Cys Gln Tyr Ser  
 420 425 430

Asn Asp Tyr Tyr Ser Gln Cys Leu  
 435 440

Sequence listing.txt

<210> 7  
 <211> 440  
 <212> PRT  
 <213> artificial

<220>  
 <223> artificial construct

<400> 7

Ser Leu Gln Pro Gly Thr Ser Thr Pro Glu Val His Pro Lys Leu Thr  
 1 5 10 15

Thr Tyr Lys Cys Thr Thr Ser Gly Gly Cys Val Ala Gln Asn Thr Tyr  
 20 25 30

Val Val Leu Asp Trp Asn Tyr Arg Trp Ile His Asp Ala Asn Tyr Asn  
 35 40 45

Ser Cys Thr Val Asn Gly Gly Val Asn Thr Thr Leu Cys Pro Asp Glu  
 50 55 60

Ala Thr Gly Ser Lys Asn Cys Phe Ile Glu Gly Val Asp Tyr Ala Ala  
 65 70 75 80

Ser Gly Val Thr Ala Asn Gly Ser Thr Leu Thr Leu Asn Gln Tyr Met  
 85 90 95

Pro Ser Ser Ser Gly Gly Tyr Thr Ser Val Ser Pro Arg Leu Tyr Leu  
 100 105 110

Leu Gly Pro Asp Gly Lys Tyr Val Met Leu Lys Leu Asn Gly Gln Glu  
 115 120 125

Leu Ser Phe Asp Val Asp Leu Ser Ala Leu Pro Cys Gly Glu Asn Ala  
 130 135 140

Ser Leu Tyr Leu Ser Gln Met Asp Glu Asn Gly Gly Ala Asn Gln Tyr  
 145 150 155 160

Asn Thr Ala Gly Ala Asn Tyr Gly Ser Gly Tyr Cys Asp Ala Gln Cys  
 165 170 175

Pro Val Gln Thr Trp Arg Asn Gly Thr Leu Asn Thr Ser Gly Gln Gly  
 180 185 190

Phe Cys Cys Asn Glu Met Asp Ile Leu Glu Gly Asn Ser Arg Ala Asn  
 195 200 205

Ala Leu Thr Pro His Ser Cys Asn Ala Thr Ala Cys Asp Ser Ala Gly  
 210 215 220

Cys Gly Phe Asn Pro Tyr Gly Ser Gly Tyr Pro Asn Tyr Phe Gly Pro  
 225 230 235 240

Sequence listing.txt

Gly Gly Thr Val Asp Thr Ser Lys Pro Phe Thr Ile Ile Thr Gln Phe  
245 250 255

Asn Thr Asp Asn Gly Ser Pro Ser Gly Asn Leu Val Ser Ile Thr Arg  
260 265 270

Lys Tyr Arg Gln Asn Gly Val Asp Ile Pro Ser Ala Lys Pro Gly Gly  
275 280 285

Asp Thr Ile Ser Ser Cys Pro Ser Ala Ser Thr Tyr Gly Gly Leu Ala  
290 295 300

Thr Met Gly Lys Ala Leu Ser Ser Gly Met Val Leu Ile Phe Ser Ile  
305 310 315 320

Trp Asn Asp Asn Ser Gln Tyr Met Asn Trp Leu Asp Ser Gly Ser Ala  
325 330 335

Gly Pro Cys Ser Ser Thr Glu Gly Asn Pro Ser Asn Ile Leu Ala Asn  
340 345 350

Asn Pro Gly Thr His Val Val Tyr Ser Asn Ile Arg Trp Gly Asp Ile  
355 360 365

Gly Ser Thr Thr Asn Ser Thr Gly Gly Asn Pro Pro Pro Pro Ala  
370 375 380

Ser Ser Thr Thr Phe Ser Thr Thr Arg Arg Ser Ser Thr Ser Ser Ser  
385 390 395 400

Ser Pro Ser Cys Ile Gln Thr His Trp Gly Gln Cys Gly Gly Ile Gly  
405 410 415

Tyr Thr Gly Cys Lys Thr Cys Thr Ser Gly Thr Thr Cys Gln Tyr Ser  
420 425 430

Asn Asp Tyr Tyr Ser Gln Cys Leu  
435 440

<210> 8

<211> 440

<212> PRT

<213> artificial

<220>

<223> artificial construct

<400> 8

Ser Gln Gln Pro Gly Thr Ser Thr Pro Glu Val His Pro Lys Leu Thr  
1 5 10 15

Sequence listing.txt

Thr Tyr Lys Cys Thr Thr Ser Gly Gly Cys Val Ala Gln Asn Thr Ser  
 20 25 30

Val Val Leu Asp Trp Asn Tyr Arg Trp Met His Asp Ala Asn Tyr Asn  
 35 40 45

Ser Cys Thr Val Asn Gly Gly Val Asn Thr Thr Leu Cys Pro Asp Glu  
 50 55 60

Ala Thr Cys Gly Lys Asn Cys Phe Ile Glu Gly Val Asp Tyr Ala Ala  
 65 70 75 80

Ser Gly Val Thr Ala Ser Gly Ser Thr Leu Thr Leu Asn Gln Tyr Met  
 85 90 95

Pro Ser Ser Ser Gly Gly Tyr Ser Ser Val Ser Pro Arg Leu Tyr Leu  
 100 105 110

Leu Gly Pro Asp Gly Lys Tyr Val Met Leu Lys Leu Asn Gly Gln Glu  
 115 120 125

Leu Ser Phe Asp Val Asp Leu Ser Ala Leu Pro Cys Gly Glu Asn Ala  
 130 135 140

Ser Leu Tyr Leu Ser Gln Met Asp Glu Asn Gly Gly Ala Asn Gln Tyr  
 145 150 155 160

Asn Thr Ala Gly Ala Asn Tyr Gly Ser Gly Tyr Cys Asp Ala Gln Cys  
 165 170 175

Pro Val Gln Thr Trp Arg Asn Gly Thr Leu Asn Thr Ser Gly Gln Gly  
 180 185 190

Phe Cys Cys Asn Glu Met Asp Ile Leu Glu Gly Asn Ser Arg Ala Asn  
 195 200 205

Ala Leu Thr Pro His Ser Cys Asn Ala Thr Ala Cys Asp Ser Ala Gly  
 210 215 220

Cys Gly Phe Asn Pro Tyr Gly Ser Gly Tyr Pro Asn Tyr Phe Gly Pro  
 225 230 235 240

Gly Gly Thr Val Asp Thr Ser Lys Pro Phe Thr Ile Ile Thr Gln Phe  
 245 250 255

Asn Thr Asp Asn Gly Ser Pro Ser Gly Asn Leu Val Ser Ile Thr Arg  
 260 265 270

Lys Tyr Arg Gln Asn Gly Val Asp Ile Pro Ser Ala Lys Pro Gly Gly  
 275 280 285

Sequence listing.txt

Asp Thr Ile Ser Ser Cys Pro Ser Ala Ser Ala Tyr Gly Gly Leu Ala  
 290 295 300

Thr Met Gly Lys Ala Leu Ser Ser Gly Met Val Leu Ile Phe Ser Ile  
 305 310 315 320

Trp Asn Asp Asn Ser Gln Tyr Met Asn Trp Leu Asp Ser Gly Ser Ala  
 325 330 335

Gly Pro Cys Ser Ser Thr Glu Gly Asn Pro Ser Asn Ile Leu Ala Asn  
 340 345 350

Asn Pro Gly Thr His Val Val Tyr Ser Asn Ile Arg Trp Gly Asp Ile  
 355 360 365

Gly Ser Thr Thr Asn Ser Thr Gly Gly Asn Pro Pro Pro Pro Ala  
 370 375 380

Ser Ser Thr Thr Phe Ser Thr Thr Arg Arg Ser Ser Thr Thr Ser Ser  
 385 390 395 400

Ser Pro Ser Cys Thr Gln Thr His Trp Gly Gln Cys Gly Gly Ile Gly  
 405 410 415

Tyr Thr Gly Cys Lys Thr Cys Thr Ser Gly Thr Thr Cys Gln Tyr Gly  
 420 425 430

Asn Asp Tyr Tyr Ser Gln Cys Leu  
 435 440

<210> 9  
 <211> 440  
 <212> PRT  
 <213> artificial

<220>  
 <223> artificial construct

<400> 9

Ser Leu Gln Pro Gly Thr Ser Thr Pro Glu Val His Pro Lys Leu Thr  
 1 5 10 15

Thr Tyr Lys Cys Thr Thr Ser Gly Gly Cys Val Ala Gln Asn Thr Ser  
 20 25 30

Val Val Leu Asp Trp Asn Tyr Arg Trp Met His Asp Ala Asn Tyr Asn  
 35 40 45

Ser Cys Thr Val Asn Gly Gly Val Asn Thr Thr Leu Cys Pro Asp Glu  
 50 55 60

Ala Thr Cys Cys Lys Asn Cys Phe Ile Glu Gly Val Asp Tyr Ala Ala  
 65 70 75 80

Sequence listing.txt

Ser Gly Val Thr Ala Ser Gly Ser Thr Leu Thr Leu Asn Gln Tyr Met  
85 90 95

Pro Ser Ser Ser Gly Gly Tyr Ser Ser Val Ser Pro Arg Leu Tyr Leu  
100 105 110

Leu Gly Pro Asp Gly Lys Tyr Val Met Leu Lys Leu Asn Gly Gln Glu  
115 120 125

Leu Ser Phe Asp Val Asp Leu Ser Ala Leu Pro Cys Gly Glu Asn Ala  
130 135 140

Ser Leu Tyr Leu Ser Gln Met Asp Glu Asn Gly Gly Ala Asn Gln Tyr  
145 150 155 160

Asn Thr Ala Gly Ala Asn Tyr Gly Ser Gly Tyr Cys Asp Ala Gln Cys  
165 170 175

Pro Val Gln Thr Trp Arg Asn Gly Thr Leu Asn Thr Ser Gly Lys Gly  
180 185 190

Phe Cys Cys Asn Glu Met Asp Ile Leu Glu Gly Asn Ser Arg Ala Asn  
195 200 205

Ala Leu Thr Pro His Ser Cys Asn Ala Thr Ala Cys Asp Ser Ala Gly  
210 215 220

Cys Gly Phe Asn Pro Tyr Gly Ser Gly Tyr Pro Asn Tyr Phe Gly Pro  
225 230 235 240

Gly Gly Thr Val Asp Thr Ser Lys Pro Phe Thr Ile Ile Thr Gln Phe  
245 250 255

Asn Thr Asp Asn Gly Ser Pro Ser Gly Asn Leu Val Ser Ile Thr Arg  
260 265 270

Lys Tyr Arg Gln Asn Gly Val Asp Ile Pro Ser Ala Lys Pro Gly Gly  
275 280 285

Asp Thr Ile Ser Ser Cys Pro Ser Ala Ser Ala Tyr Gly Gly Leu Ala  
290 295 300

Thr Met Gly Lys Ala Leu Ser Ser Gly Met Val Leu Ile Phe Ser Ile  
305 310 315 320

Trp Asn Asp Asn Ser Gln Tyr Met Asn Trp Leu Asp Ser Gly Ser Ala  
325 330 335

Gly Pro Cys Ser Ser Thr Glu Gly Asn Pro Ser Asn Ile Leu Ala Asn  
340 345 350

Sequence listing.txt

Asn Pro Gly Thr His Val Val Tyr Ser Asn Ile Arg Trp Gly Asp Ile  
 355 360 365

Gly Ser Thr Thr Asn Ser Thr Gly Gly Asn Pro Pro Pro Pro Ala  
 370 375 380

Ser Ser Thr Thr Phe Ser Thr Thr Arg Arg Ser Ser Thr Thr Ser Ser  
 385 390 395 400

Ser Pro Ser Cys Thr Gln Thr His Trp Gly Gln Cys Gly Gly Ile Gly  
 405 410 415

Tyr Thr Gly Cys Lys Thr Cys Thr Ser Gly Thr Thr Cys Gln Tyr Ser  
 420 425 430

Asn Asp Tyr Tyr Ser Gln Cys Leu  
 435 440

<210> 10  
 <211> 440  
 <212> PRT  
 <213> artificial

<220>  
 <223> artificial construct

<400> 10

Ser Gln Gln Pro Gly Thr Ser Thr Pro Glu Val His Pro Lys Leu Thr  
 1 5 10 15

Thr Tyr Lys Cys Thr Thr Ser Gly Gly Cys Val Ala Gln Asn Thr Ser  
 20 25 30

Val Val Leu Asp Trp Asn Tyr Arg Trp Met His Asp Ala Asn Tyr Asn  
 35 40 45

Ser Cys Thr Val Asn Gly Gly Val Asn Thr Thr Leu Cys Pro Asp Glu  
 50 55 60

Ala Thr Cys Gly Lys Asn Cys Phe Ile Glu Gly Val Asp Tyr Ala Ala  
 65 70 75 80

Ser Gly Val Thr Ala Ser Gly Ser Thr Leu Thr Leu Asn Gln Tyr Met  
 85 90 95

Pro Ser Ser Ser Gly Gly Tyr Ser Ser Val Ser Pro Arg Leu Tyr Leu  
 100 105 110

Leu Gly Pro Asp Gly Lys Tyr Val Met Leu Lys Leu Asn Gly Gln Glu  
 115 120 125



Sequence listing.txt

Leu Ser Phe Asp Val Asp Leu Ser Ala Leu Pro Cys Gly Glu Asn Ala  
 130 135 140

Ser Leu Tyr Leu Ser Gln Met Asp Glu Asn Gly Gly Ala Asn Gln Tyr  
 145 150 155 160

Asn Thr Ala Gly Ala Asn Tyr Gly Ser Gly Tyr Cys Asp Ala Gln Cys  
 165 170 175

Pro Val Gln Thr Trp Arg Asn Gly Thr Leu Asn Thr Ser Gly Gln Gly  
 180 185 190

Phe Cys Cys Asn Glu Met Asp Ile Leu Glu Gly Asn Ser Arg Ala Asn  
 195 200 205

Ala Leu Thr Pro His Ser Cys Asn Ala Thr Ala Cys Asp Ser Ala Gly  
 210 215 220

Cys Gly Phe Asn Pro Tyr Lys Ser Gly Tyr Pro Asn Tyr Phe Gly Pro  
 225 230 235 240

Gly Gly Thr Val Asp Thr Ser Lys Pro Phe Thr Ile Ile Thr Gln Phe  
 245 250 255

Asn Thr Asp Asn Gly Ser Pro Ser Gly Asn Leu Val Ser Ile Thr Arg  
 260 265 270

Lys Tyr Arg Gln Asn Gly Val Asp Ile Pro Ser Ala Lys Pro Gly Gly  
 275 280 285

Asp Thr Ile Ser Ser Cys Pro Ser Ala Ser Ala Tyr Gly Gly Leu Ala  
 290 295 300

Thr Met Gly Lys Ala Leu Ser Glu Gly Met Val Leu Ile Phe Ser Ile  
 305 310 315 320

Trp Asn Asp Asn Ser Gln Tyr Met Asn Trp Leu Asp Ser Gly Ser Ala  
 325 330 335

Gly Pro Cys Ser Ser Thr Glu Gly Asn Pro Ser Asn Ile Leu Ala Asn  
 340 345 350

Asn Pro Gly Thr His Val Val Tyr Ser Asn Ile Arg Trp Gly Asp Ile  
 355 360 365

Gly Ser Thr Thr Asn Ser Thr Gly Gly Asn Pro Pro Pro Pro Ala  
 370 375 380

Ser Ser Thr Thr Phe Ser Thr Thr Arg Arg Ser Ser Thr Thr Ser Ser  
 385 390 395 400

Sequence listing.txt

Ser Pro Ser Cys Thr Gln Thr His Trp Gly Gln Cys Gly Gly Ile Gly  
 405 410 415

Tyr Thr Gly Cys Lys Thr Cys Thr Ser Gly Thr Thr Cys Gln Tyr Gly  
 420 425 430

Asn Asp Tyr Tyr Ser Gln Cys Leu  
 435 440

<210> 11  
 <211> 440  
 <212> PRT  
 <213> artificial

<220>  
 <223> artificial construct

<400> 11

Ser Gln Gln Pro Gly Thr Ser Thr Pro Glu Val His Pro Lys Leu Thr  
 1 5 10 15

Thr Tyr Lys Cys Thr Thr Ser Gly Gly Cys Val Ala Gln Asn Thr Ser  
 20 25 30

Val Val Leu Asp Trp Asn Tyr Arg Trp Met His Asp Ala Asn Tyr Asn  
 35 40 45

Ser Cys Thr Val Asn Gly Gly Val Asn Thr Thr Leu Cys Pro Asp Glu  
 50 55 60

Ala Thr Cys Gly Lys Asn Cys Phe Ile Glu Gly Val Asp Tyr Ala Ala  
 65 70 75 80

Ser Gly Val Thr Ala Ser Gly Ser Thr Leu Thr Leu Asn Gln Tyr Met  
 85 90 95

Pro Ser Ser Ser Gly Gly Tyr Ser Ser Val Ser Pro Arg Leu Tyr Leu  
 100 105 110

Leu Gly Pro Asp Gly Lys Tyr Val Met Leu Lys Leu Asn Gly Gln Glu  
 115 120 125

Leu Ser Phe Asp Val Asp Leu Ser Ala Leu Pro Cys Gly Glu Asn Ala  
 130 135 140

Ser Leu Tyr Leu Ser Gln Met Asp Glu Asn Gly Gly Ala Asn Gln Tyr  
 145 150 155 160

Asn Thr Ala Gly Ala Asn Tyr Gly Ser Gly Tyr Cys Asp Ala Gln Cys  
 165 170 175

Pro Val Gln Thr Trp Arg Asn Gly Thr Leu Asn Thr Ser Gly Gln Gly  
 180 185 190

Sequence listing.txt

Phe Cys Cys Asn Glu Met Asp Ile Leu Glu Gly Asn Ser Arg Ala Asn  
 195 200 205

Ala Leu Thr Pro His Ser Cys Asn Ala Thr Ala Cys Asp Ser Ala Gly  
 210 215 220

Cys Gly Phe Asn Pro Tyr Gly Ser Gly Tyr Pro Asn Tyr Phe Gly Pro  
 225 230 235 240

Gly Gly Thr Val Asp Thr Ser Lys Pro Phe Thr Ile Ile Thr Gln Phe  
 245 250 255

Asn Thr Asp Asn Gly Ser Pro Ser Gly Asn Leu Val Ser Ile Thr Arg  
 260 265 270

Lys Tyr Arg Gln Asn Gly Val Asp Ile Pro Ser Ala Lys Pro Gly Gly  
 275 280 285

Asp Thr Ile Ser Ser Cys Pro Ser Ala Ser Ala Tyr Gly Gly Leu Ala  
 290 295 300

Thr Met Gly Lys Ala Leu Ser Asp Gly Met Val Leu Ile Phe Ser Ile  
 305 310 315 320

Trp Asn Asp Asn Ser Gln Tyr Met Asn Trp Leu Asp Ser Gly Glu Ala  
 325 330 335

Gly Pro Cys Ser Ser Thr Glu Gly Asn Pro Ser Asn Ile Leu Ala Asn  
 340 345 350

Asn Pro Gly Thr His Val Val Tyr Ser Asn Ile Arg Trp Gly Asp Ile  
 355 360 365

Gly Ser Thr Thr Asn Ser Thr Gly Gly Asn Pro Pro Pro Pro Ala  
 370 375 380

Ser Ser Thr Thr Phe Ser Thr Thr Arg Arg Ser Ser Thr Thr Ser Ser  
 385 390 395 400

Ser Pro Ser Cys Thr Gln Thr His Trp Gly Gln Cys Gly Gly Ile Gly  
 405 410 415

Tyr Thr Gly Cys Lys Thr Cys Thr Ser Gly Thr Thr Cys Gln Tyr Gly  
 420 425 430

Asn Asp Tyr Tyr Ser Gln Cys Leu  
 435 440

<210> 12  
 <211> 440

Sequence listing.txt

<212> PRT  
 <213> artificial

<220>  
 <223> artificial construct

<400> 12

Ser Gln Gln Pro Gly Thr Ser Thr Pro Glu Val His Pro Lys Leu Thr  
 1 5 10 15

Thr Tyr Lys Cys Thr Thr Ser Gly Gly Cys Val Ala Gln Asn Thr Ser  
 20 25 30

Val Val Leu Asp Trp Asn Tyr Arg Trp Met His Asp Ala Asn Tyr Asn  
 35 40 45

Ser Cys Thr Val Asn Gly Gly Val Asn Thr Thr Leu Cys Pro Asp Glu  
 50 55 60

Ala Thr Cys Gly Lys Asn Cys Phe Ile Glu Gly Val Asp Tyr Ala Ala  
 65 70 75 80

Ser Gly Val Thr Ala Ser Gly Ser Thr Leu Thr Leu Asn Gln Tyr Met  
 85 90 95

Pro Ser Ser Ser Gly Gly Tyr Ser Ser Val Ser Pro Arg Leu Tyr Leu  
 100 105 110

Leu Gly Pro Asp Gly Lys Tyr Val Met Leu Lys Leu Asn Gly Gln Glu  
 115 120 125

Leu Ser Phe Asp Val Asp Leu Ser Ala Leu Pro Cys Gly Glu Asn Ala  
 130 135 140

Ser Leu Tyr Leu Ser Gln Met Asp Glu Asn Gly Gly Ala Asn Gln Tyr  
 145 150 155 160

Asn Thr Ala Gly Ala Asn Tyr Gly Ser Gly Tyr Cys Asp Ala Gln Cys  
 165 170 175

Pro Val Gln Thr Trp Arg Asn Gly Thr Leu Asn Thr Ser Gly Gln Gly  
 180 185 190

Phe Cys Cys Asn Glu Met Asp Ile Leu Glu Gly Asn Ser Arg Ala Asn  
 195 200 205

Ala Leu Thr Pro His Ser Cys Asn Ala Thr Ala Cys Asp Ser Ala Gly  
 210 215 220

Cys Gly Phe Asn Pro Tyr Lys Ser Gly Tyr Pro Asn Tyr Phe Gly Pro  
 225 230 235 240

sequence listing.txt

Gly Gly Thr Val Asp Thr Ser Lys Pro Phe Thr Ile Ile Thr Gln Phe  
 245 250 255

Asn Thr Asp Asn Gly Ser Pro Ser Gly Asn Leu Val Ser Ile Thr Arg  
 260 265 270

Lys Tyr Arg Gln Asn Gly Val Asp Ile Pro Ser Ala Lys Pro Gly Gly  
 275 280 285

Asp Thr Ile Ser Ser Cys Pro Ser Ala Ser Ala Tyr Gly Gly Leu Ala  
 290 295 300

Thr Met Gly Lys Ala Leu Ser Asp Gly Met Val Leu Ile Phe Ser Ile  
 305 310 315 320

Trp Asn Asp Asn Ser Gln Tyr Met Asn Trp Leu Asp Ser Gly Ser Ala  
 325 330 335

Gly Pro Cys Ser Ser Thr Glu Gly Asn Pro Ser Asn Ile Leu Ala Asn  
 340 345 350

Asn Pro Gly Thr His Val Val Tyr Ser Asn Ile Arg Trp Gly Asp Ile  
 355 360 365

Gly Ser Thr Thr Asn Ser Thr Gly Gly Asn Pro Pro Pro Pro Ala  
 370 375 380

Ser Ser Thr Thr Phe Ser Thr Thr Arg Arg Ser Ser Thr Thr Ser Ser  
 385 390 395 400

Ser Pro Ser Cys Thr Gln Thr His Trp Gly Gln Cys Gly Gly Ile Gly  
 405 410 415

Tyr Thr Gly Cys Lys Thr Cys Thr Ser Gly Thr Thr Cys Gln Tyr Gly  
 420 425 430

Asn Asp Tyr Tyr Ser Gln Cys Leu  
 435 440

<210> 13  
 <211> 440  
 <212> PRT  
 <213> artificial

<220>  
 <223> artificial construct

<400> 13

Ser Gln Gln Pro Gly Thr Ser Thr Pro Glu Val His Pro Lys Leu Thr  
 1 5 10 15

Thr Tyr Lys Cys Thr Thr Ser Gly Gly Cys Val Ala Gln Asp Thr Ser  
 20 25 30

Sequence listing.txt

Val Val Leu Asp Trp Asn Tyr Arg Trp Ile His Asp Ala Asn Tyr Asn  
 35 40 45  
 Ser Cys Thr Val Asn Gly Gly Val Asn Thr Thr Leu Cys Pro Asp Glu  
 50 55 60  
 Ala Thr Cys Ser Lys Asn Cys Phe Ile Glu Gly Val Asp Tyr Ala Ala  
 65 70 75 80  
 Ser Gly Val Thr Ala Asn Gly Ser Thr Leu Thr Leu Asn Gln Tyr Met  
 85 90 95  
 Pro Ser Ser Ser Gly Gly Tyr Ser Ser Val Ser Pro Arg Leu Tyr Leu  
 100 105 110  
 Leu Gly Pro Asp Gly Glu Tyr Val Met Leu Lys Leu Asn Gly Gln Glu  
 115 120 125  
 Leu Ser Phe Asp Val Asp Leu Ser Ala Leu Pro Cys Gly Glu Asn Gly  
 130 135 140  
 Ser Leu Tyr Leu Ser Gln Met Asp Glu Asn Gly Gly Ala Asn Gln Tyr  
 145 150 155 160  
 Asn Thr Ala Gly Ala Asn Tyr Gly Ser Gly Tyr Cys Asp Ala Gln Cys  
 165 170 175  
 Pro Val Gln Thr Trp Arg Asn Gly Thr Leu Asn Thr Ser Gly Gln Gly  
 180 185 190  
 Phe Cys Cys Asn Glu Met Asp Ile Leu Glu Gly Asn Ser Arg Ala Asn  
 195 200 205  
 Ala Leu Thr Pro His Ser Cys Thr Ala Thr Ala Cys Asp Ser Ala Gly  
 210 215 220  
 Cys Gly Phe Asn Pro Tyr Gly Ser Gly Tyr Pro Asn Tyr Phe Gly Pro  
 225 230 235 240  
 Gly Asp Thr Val Asp Thr Ser Lys Pro Phe Thr Ile Ile Thr Gln Phe  
 245 250 255  
 Asn Thr Asp Asn Gly Ser Pro Ser Gly Asn Leu Val Ser Ile Thr Arg  
 260 265 270  
 Lys Tyr Arg Gln Asn Gly Val Asp Ile Pro Ser Ala Lys Pro Gly Gly  
 275 280 285  
 Asp Thr Ile Ser Ser Cys Pro Ser Ala Ser Ala Tyr Gly Gly Leu Ala  
 290 295 300

Sequence listing.txt

Thr Met Gly Lys Ala Leu Ser Ser Gly Met Val Leu Ile Phe Ser Ile  
305 310 315 320

Trp Asn Asp Asn Ser Gln Tyr Met Asn Trp Leu Asp Ser Gly Ser Ala  
325 330 335

Gly Pro Cys Ser Ser Thr Glu Gly Asn Pro Ser Asn Ile Leu Ala Asn  
340 345 350

Asn Pro Gly Thr His Val Val Tyr Ser Asn Ile Arg Trp Gly Asp Ile  
355 360 365

Gly Ser Thr Thr Asn Ser Thr Gly Gly Asn Pro Pro Pro Pro Ala  
370 375 380

Ser Ser Thr Thr Phe Ser Thr Thr Arg Arg Ser Ser Thr Thr Ser Ser  
385 390 395 400

Ser Pro Ser Cys Thr Gln Thr His Trp Gly Gln Cys Gly Gly Ile Gly  
405 410 415

Tyr Thr Gly Cys Lys Thr Cys Thr Ser Gly Thr Thr Cys Gln Tyr Gly  
420 425 430

Asn Asp Tyr Tyr Ser Gln Cys Leu  
435 440

<210> 14  
<211> 440  
<212> PRT  
<213> artificial

<220>  
<223> artificial construct

<400> 14

Ser Gln Gln Pro Gly Thr Ser Thr Pro Glu Val His Pro Lys Leu Thr  
1 5 10 15

Thr Tyr Lys Cys Thr Thr Ser Gly Gly Cys Val Ala Gln Asp Thr Ser  
20 25 30

Val Val Leu Asp Trp Asn Tyr Arg Trp Ile His Asp Ala Asn Tyr Asn  
35 40 45

Ser Cys Thr Val Asn Gly Gly Val Asn Thr Thr Leu Cys Pro Asp Glu  
50 55 60

Ala Thr Cys Gly Lys Asn Cys Phe Ile Glu Gly Val Asp Tyr Ala Ala  
65 70 75 80

Sequence listing.txt

Ser Gly Val Thr Ala Asn Gly Ser Thr Leu Thr Leu Asn Gln Tyr Met  
 85 90 95  
 Pro Ser Ser Ser Gly Gly Tyr Ser Ser Val Ser Pro Arg Leu Tyr Leu  
 100 105 110  
 Leu Gly Pro Asp Gly Glu Tyr Val Met Leu Lys Leu Asn Gly Gln Glu  
 115 120 125  
 Leu Ser Phe Asp Val Asp Leu Ser Ala Leu Pro Cys Gly Glu Asn Gly  
 130 135 140  
 Ser Leu Tyr Leu Ser Gln Met Asp Glu Asn Gly Gly Ala Asn Gln Tyr  
 145 150 155 160  
 Asn Thr Ala Gly Ala Asn Tyr Gly Ser Gly Tyr Cys Asp Ala Gln Cys  
 165 170 175  
 Pro Val Gln Thr Trp Arg Asn Gly Thr Leu Asn Thr Ser Gly Gln Gly  
 180 185 190  
 Phe Cys Cys Asn Glu Met Asp Ile Leu Glu Gly Asn Ser Arg Ala Asn  
 195 200 205  
 Ala Leu Thr Pro His Ser Cys Thr Ala Thr Ala Cys Asp Ser Ala Gly  
 210 215 220  
 Cys Gly Phe Asn Pro Tyr Gly Ser Gly Tyr Pro Asn Tyr Phe Gly Pro  
 225 230 235 240  
 Gly Asp Thr Val Asp Thr Ser Lys Pro Phe Thr Ile Ile Thr Gln Phe  
 245 250 255  
 Asn Thr Asp Asn Gly Ser Pro Ser Gly Asn Leu Val Ser Ile Thr Arg  
 260 265 270  
 Lys Tyr Arg Gln Asn Gly Val Asp Ile Pro Ser Ala Lys Pro Gly Gly  
 275 280 285  
 Asp Thr Ile Ser Ser Cys Pro Ser Ala Ser Ala Tyr Gly Gly Leu Ala  
 290 295 300  
 Thr Met Gly Lys Ala Leu Ser Ser Gly Met Val Leu Ile Phe Ser Ile  
 305 310 315 320  
 Trp Asn Asp Asn Ser Gln Tyr Met Asn Trp Leu Asp Ser Gly Ser Ala  
 325 330 335  
 Gly Pro Cys Ser Ser Thr Glu Gly Asn Pro Ser Asn Ile Leu Ala Asn  
 340 345 350



Sequence listing.txt

Asn Pro Gly Thr His Val Val Tyr Ser Asn Ile Arg Trp Gly Asp Ile  
 355 360 365

Gly Ser Thr Thr Asn Ser Thr Gly Gly Asn Pro Pro Pro Pro Ala  
 370 375 380

Ser Ser Thr Thr Phe Ser Thr Thr Arg Arg Ser Ser Thr Thr Ser Ser  
 385 390 395 400

Ser Pro Ser Cys Thr Gln Thr His Trp Gly Gln Cys Gly Gly Ile Gly  
 405 410 415

Tyr Thr Gly Cys Lys Thr Cys Thr Ser Gly Thr Thr Cys Gln Tyr Gly  
 420 425 430

Asn Asp Tyr Tyr Ser Gln Cys Leu  
 435 440

<210> 15  
 <211> 1620  
 <212> DNA  
 <213> artificial

<220>  
 <223> artificial construct

<400> 15  
 atgagatttc cttcaatttt tactgcagtt ttattcgcag catcctccgc attagctgct 60  
 ccagtcaaca ctacaacaga agatgaaacg gcacaaattc cggctgaagc tgtcatcggc 120  
 tacttagatt tagaagggga tttcgatggt gctgttttgc cattttccaa cagcacaaat 180  
 aacgggttat tgtttataaa tactactatt gccagcattg ctgctaaaga agaaggggta 240  
 tctttggata aacgtgaggc ggaagcatgc caccaccacc accaccactc ctccggctct 300  
 ctgcagccag gaacttctac tccagaggtg cacccaaagc tgaccaccta caagtgtacc 360  
 acctctggcg gttgtgttgc tcagaacacc tatgttgttc tggactggaa ctacagatgg 420  
 atccacgacg ccaactacaa ctcttgtagc gtgaacggcg gtgtcaacac tactctgtgt 480  
 ccagacgagg ctactggtag caagaactgc ttcacgagg gtgttgacta cgctgcttct 540  
 ggtgttactg ccaatggttc taccttgacc ctgaaccagt acatgccatc ttcctctggc 600  
 ggttacactt ctgtgtcggc aagactgtac ttgttgggtc cagacggtaa gtacgttatg 660  
 ctgaagctga acggacagga gctgtctttt gacgttgacc tgtctgcttt gccatgtgga 720  
 gagaacgctt ctctgtacct gtctcagatg gacgagaacg gtggagctaa ccagtacaac 780  
 accgccggcg ctaactacgg ttctggttac tgtgacgccc agtgtccagt tcagacttgg 840  
 agaaacggaa ccctgaacac ttctggccag ggattctgct gtaacgagat ggacatcttg 900  
 gagggaaact ctagagctaa cgctctgacc ccacactctt gtaatgctac cgcttgtgac 960  
 tctgctgggt gcggttttaa cccataccgc tcgggttacc caaactactt tggcccaggc 1020  
 ggcactgttg acacctcga gccattcacc atcatcacc agttcaacac cgacaacggc 1080

Sequence listing.txt

tctccatctg gtaacctggt gtcgatcacc agaaagtaca gacagaacgg cgttgacatc 1140  
 ccattctgcta aaccaggtgg cgacaccatt tcgtcttgtc catctgcctc tacttacggt 1200  
 ggattggcta ccatgggaaa ggctctgtcc gagggaatgg tgctgatctt ctcgatctgg 1260  
 aacgacaact cgcagtacat gaactggctg gactctgggt atgctgggtcc atgttcttct 1320  
 accgagggca acccatctaa catcctggct aacaaccctg gtactcacgt ggtgtactcg 1380  
 aacattagat ggggcgacat tggttctacc accaactcta ccggtggtaa cccaccacca 1440  
 ccacctgcat cttctaccac cttctcgacc gccagaagat cgtctacctc ctcttcttct 1500  
 ccattctgta tccagactca ctggggtcag tgtgggtggt ttggctacac cggctgtaag 1560  
 acctgtacct ctggaaccac ttgccagtac agcaacgact actacttctca gtgcctgtga 1620

<210> 16  
 <211> 1368  
 <212> DNA  
 <213> artificial

<220>  
 <223> artificial construct

<400> 16  
 atgtatcgga agttggccgt catctcggcc ttcttggcca cagcacgggc ttctctgcaa 60  
 ccgggtacca gcacccccga ggtccatccc aagttgacaa cctacaagtg tacaacctcc 120  
 ggggggtgcg tggcccagaa cacctatgtg gtccttgact ggaactaccg ctggatccac 180  
 gacgcaaact acaactcgtg caccgtcaac ggcggcgtca acaccacgct ctgccctgac 240  
 gaggcgaccg gtagcaagaa ctgcttcacg gagggcgctc actacgccgc ctcgggcgctc 300  
 acggccaatg gcagcacctt caccctgaac cagtacatgc ccagcagctc tggcggctac 360  
 actagcgtct ctctcggct gtatctcctg ggtccagacg gtaagtacgt gatgctgaag 420  
 ctcaacggcc aggagctgag cttcgacgtc gacctctctg ctctgccgtg tggagagaac 480  
 gcctcgtctt acctgtctca gatggacgag aacgggggcg ccaaccagta taacacggcc 540  
 ggtgccaaact acgggagcgg ctactgcgat gctcagtgcc ccgtccagac atggaggaac 600  
 ggcacctca aactagcgg ccagggcttc tgctgcaacg agatggatat cctggagggc 660  
 aactcgaggg cgaatgcctt gacccctcac tcttgcaatg ccacggcctg cgactctgcc 720  
 ggttgccgct tcaaccctta tcgcagcggc taccctaaact acttcggccc cggaggcacc 780  
 gttgacacct ccaagccatt caccatcatc acccagttca acacggacaa cggctcggcc 840  
 tcgggcaacc ttgtgagcat cccccgaaag tacagacaaa acggcgtcga catccccagc 900  
 gccaaacctg gcggcgacac catctcgtcc tgcccgtccg cctcaactta cggcggcctc 960  
 gccaccatgg gcaaggccct gagcgagggc atggtgctca tcttcagcat ttggaacgac 1020  
 aacagccagt acatgaactg gctcgacagc ggcgatgccg gccctgagc cagcaccgag 1080  
 ggcaaccat ccaacatcct ggccaacaac cccggtacgc acgtcgtcta ctccaacatc 1140  
 cgctggggag acattgggtc tactacgaac tcgactgggt gtccgcccc gcctgcgtcc 1200

Sequence listing.txt

agcacgacgt ttcgactgc ccggaggagc tcgacgtcct cgagcagccc gagctgcatc	1260
cagactcact gggggcagtg cggtagcatt gggtacaccg ggtgcaagac gtgcacgtcg	1320
ggcactacgt gccagtatag caacgactac tactcgcaat gcctttaa	1368

**ФОРМУЛА ИЗОБРЕТЕНИЯ**

1. Белок, имеющий эндоглюканазную активность, который содержит аминокислотную последовательность, имеющую по меньшей мере 96%, предпочтительно по меньшей мере 97%, более предпочтительно по меньшей мере 98%, даже более предпочтительно по меньшей мере 99%, такую как по меньшей мере 99,5% идентичность с SEQ. ID NO.: 2.

2. Белок по п. 1, который демонстрирует по меньшей мере 90%-ную остаточную способность к превращению субстрата при температуре 60°C при проведении инкубации в течение одного часа.

3. Белок, имеющий эндоглюканазную активность, который принадлежит к GH7-классу, и который демонстрирует активную термостабилизацию.

4. Белок по п. 3, и имеющий по меньшей мере 70%, предпочтительно по меньшей мере 90%, более предпочтительно по меньшей мере 95%, более предпочтительно по меньшей мере 96%, более предпочтительно по меньшей мере 97%, более предпочтительно по меньшей мере 98%, даже более предпочтительно по меньшей мере 99%, такую как по меньшей мере 99,5% идентичность с SEQ. ID NO.: 2.

5. Белок по любому из п.п. 1, 3 или 4, который является идентичным любой из SEQ ID NO: 5; SEQ ID NO: 6; SEQ ID NO: 7; SEQ ID NO: 8; SEQ ID NO: 9; SEQ ID NO: 10; SEQ ID NO: 11; SEQ ID NO: 12; SEQ ID NO: 13; SEQ ID NO: 14; SEQ ID NO: 2; или который имеет по меньшей мере одну мутацию относительно SEQ ID NO: 2.

6. Полипептид по п. 5, в котором по меньшей мере одна мутация является мутацией, выбранной из предпочтительных, более предпочтительных или альтернативных замен, показанных в таблице 2.

7. Полипептид по п. 5, в котором по меньшей мере одна мутация является мутацией, выбранной из предпочтительных, более предпочтительных или альтернативных замен, показанных в таблице 3.

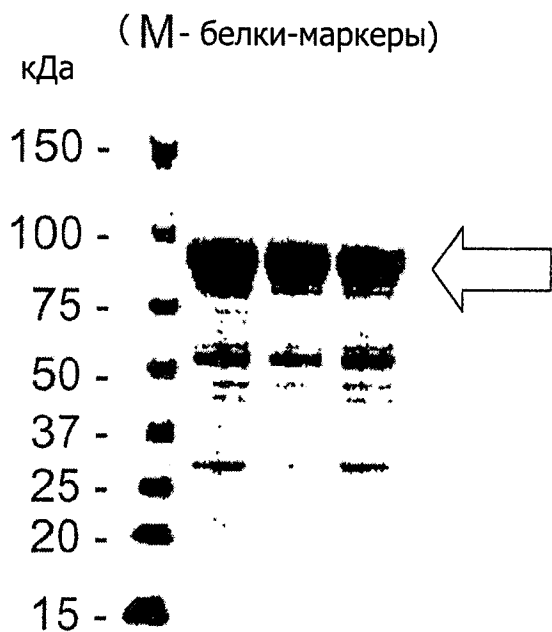
8. Полипептид по п. 5, в котором по меньшей мере одна мутация является мутацией, выбранной из предпочтительных, более предпочтительных или альтернативных замен, показанных в таблице 4.

9. Нуклеиновая кислота, кодирующая белок по любому из предшествующих пунктов.

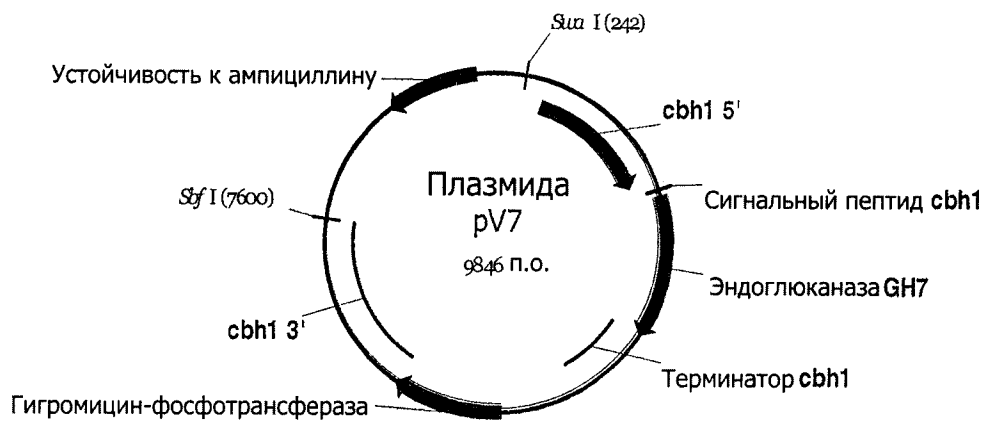
10. Экспрессирующий вектор, содержащий нуклеиновую кислоту по п. 9.

11. Микроорганизм, содержащий векторную конструкцию по п. 10.

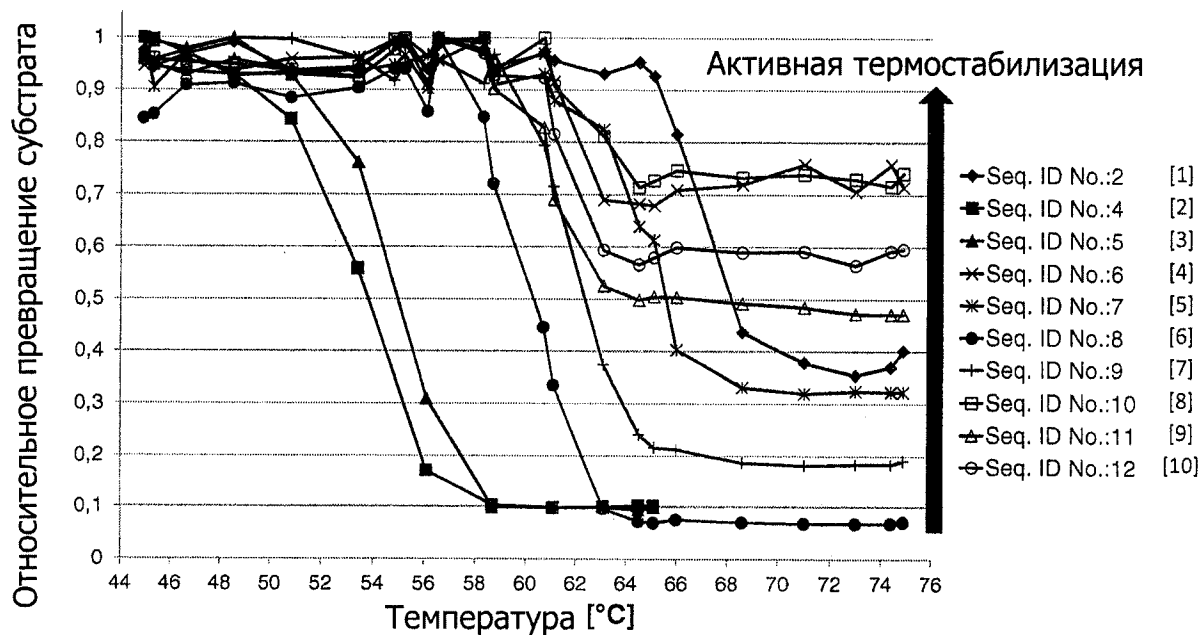
12. Композиция, содержащая белок по любому из пп. 1-8, и один дополнительный фермент или несколько дополнительных ферментов, предпочтительно выбранных из одной или нескольких целлюлаз, гемицеллюлаз и пектиназ.
13. Применение белка по любому из пп. 1-8 или композиции по п. 12.
14. Применение п. 13 для осахаривания лигноцеллюлозы.



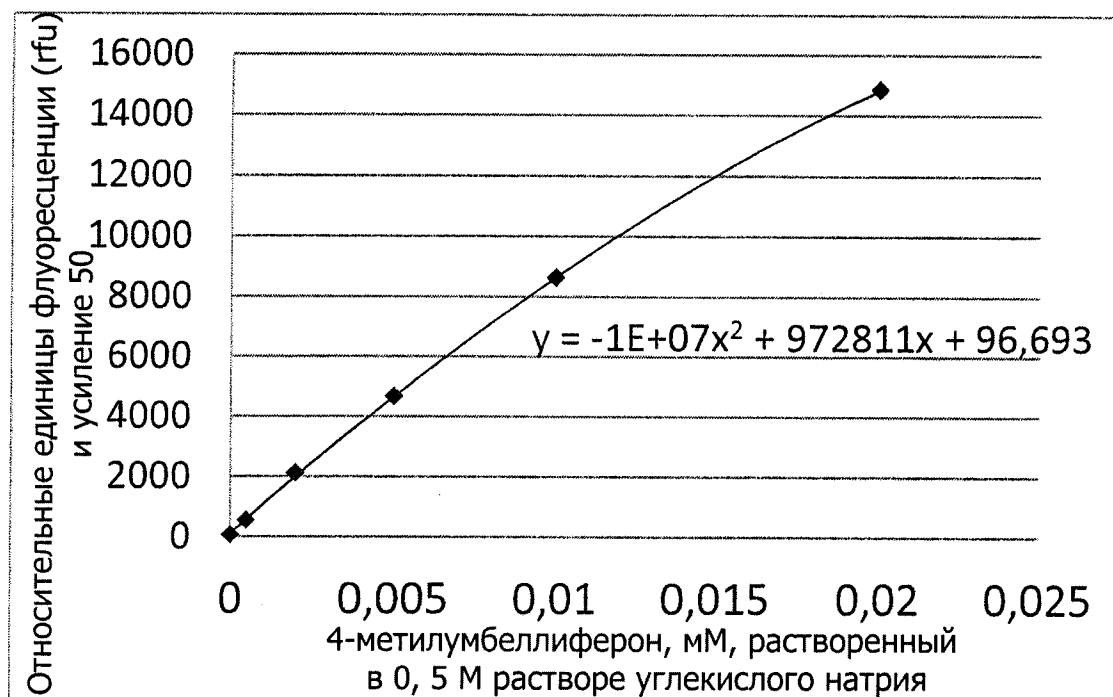
Фиг. 1



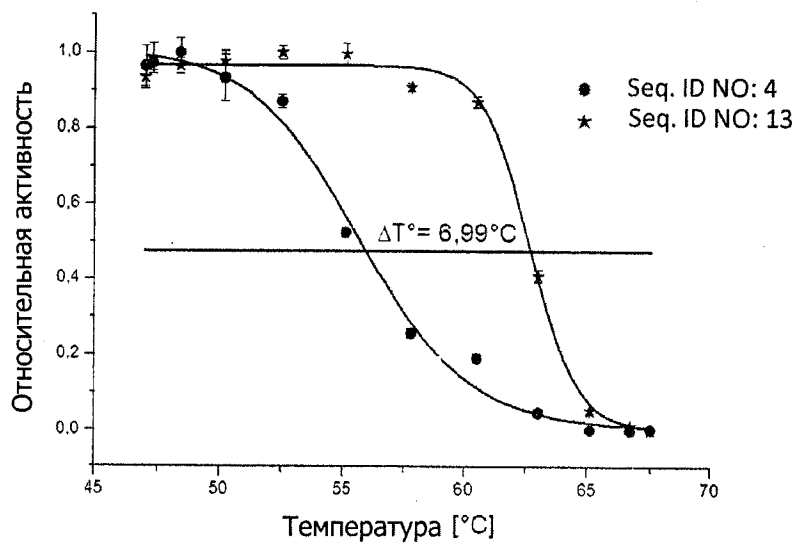
Фиг. 2



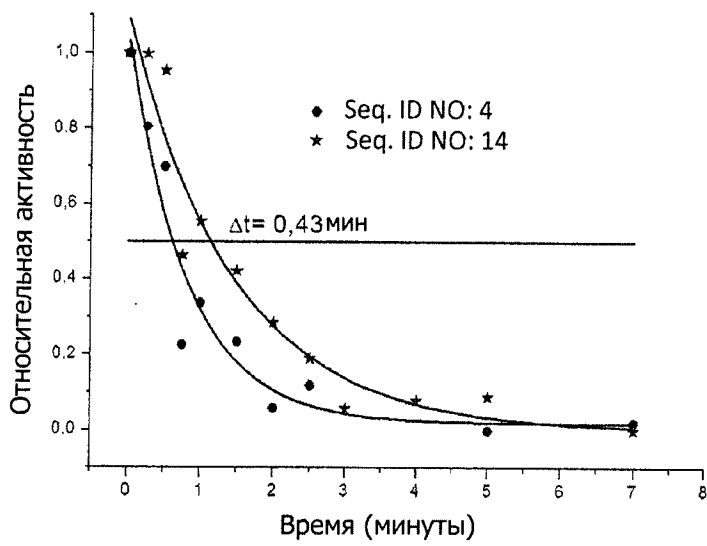
Фиг. 3



Фиг. 4



Фиг. 5



Фиг. 6